UNIVERSITÄT HOHENHEIM

Faculty of Agricultural Sciences

Institute of Agricultural Economics and Social Sciences
in the Tropics and Subtropics (490)

Land Use Economics in the Tropics and Subtropics
(Josef G. Knoll Professorship)

Dissertation

Agent-based modeling of climate change adaptation in agriculture:

A case study in the Central Swabian Jura

submitted by

Christian Troost

in partial fulfillment of the requirements for the degree
'Doktor der Agrarwissenschaften' (Dr. sc. agr.)

2014

This thesis was accepted as a doctoral dissertation in fulfillment of the requirements for the degree "Doktor der Agrarwissenschaften" (Dr. sc. agr.) by the Faculty of Agricultural Sciences of the University of Hohenheim.

Dean:	Prof. Dr. Martina Brockmeier
Date of oral examination:	6th August, 2014
Examination committee:	
Supervisor and reviewer:	Prof. Dr. Thomas Berger
Co-reviewer:	Prof. Dr. Thilo Streck
Third examiner:	Prof. Dr. Harald Grethe
Head of committee:	Prof. Dr.-Ing. Stefan Böttinger

Summary

Agent-based models allow for an in-depth analysis of processes determining the extent and speed of agricultural adaptation to climate change that may escape other agricultural economic simulation approaches. Using the MPMAS multi-agent software, the present thesis implements an agro-economic agent-based model to analyze climate change adaptation of agricultural production in the Central Swabian Jura. It contributes to the DFG PAK 346/ FOR 1695 research projects dedicated to improve the understanding of processes that shape structure and functions of agricultural landscapes in the context of climate change at regional scale. In the context of this example, this thesis discusses, develops and tests novel approaches to deal with four notorious challenges that have so far hampered the empirical use of agent-based models for applied economic analysis: data availability, process uncertainty, model validity and computational requirements. The model is used to examine climatic effects on agriculture, changes in agricultural price responses and biogas support and agri-environmental policies illustrating the applicability of the model to adaptation analysis.

The work is divided into two parts: The first part is dedicated to a methodological discussion of the use of mathematical programming-based multi-agent systems, such as MPMAS, for the analysis of agricultural adaptation to climate change. It synthesizes knowledge about the potential impacts of climate change and processes of farmer adaptation and reviews existing agent-based models for their potential contribution to adaptation analysis.

The major focus of the first part is a discussion of available approaches to model validation, calibration and uncertainty analysis and their suitability for the use with mathematical programming-based agent-based models. This discussion is based on four principles required to ensure the validity of conclusions drawn from modeling studies: (i) a transparent model documentation, (ii) that the invariant elements of the model can really be expected to be invariant between scenarios assessed, (iii) that empirical calibration of the model is limited to the extent warranted by available observation and knowledge about the expected error distribution, and (iv) that the effect of process uncertainty on the conclusions is evaluated and communicated.

For the case of agricultural agent-based models that employ mathematical programming there is little knowledge of error distributions and suitably structured observations are often scarce. In this situation, modelers are well advised to rely on robust calibration criteria and refrain from identifying only one single best-fitting parameter combination. Rather, scenarios should be run for a suitably designed sample of the remaining parameter space and results be presented as ranges or distributions over all repetitions of the sample communicating the influence of uncertainty on outcomes. Based on these conclusions, generic extensions of the MPMAS toolbox are developed to allow the application of suitable approaches for validation and uncertainty analysis.

The second part of the thesis describes the application of the newly developed methodology in the construction and use of the Central Swabian Jura model. The model focuses on an endogenous representation of heterogeneity in agent behavior, an empirical parameterization of the model, and an incorporation of climate effects on possible crop rotations and suitable days for field work besides the expected effects on yields. It extends the demographic, investment and land market components of MPMAS to improve the simulation of structural change over time.

Data to parameterize the model was gathered from statistical offices, extension services, expert interviews and a farm survey. The Monte-Carlo approach to create synthetic, but representative agent populations usually used for MPMAS models was enhanced to better

reflect joint distributions and allow for the use of privacy-restricted agricultural census and farm survey data. Process uncertainty could be partly reduced in a conservative calibration approach designed to avoid overfitting the model: The uncertain parameter space was reduced only when goodness-of-fit improved across the structural breaks incorporated in all three observation years of a calibration dataset. The remaining uncertainty is explicitly communicated by presenting simulation results as distributions over the space of potential parameter combinations instead of point estimates. The simulation results show that despite a rather modest reduction of parameter uncertainty the conclusions drawn with respect to climate change effects are robust. The model predictions held up well to expert scrutiny in a Turing test and the calibrated model reproduced observed land use data with a satisfactory level of accuracy. The observed biases were largely consistent with model simplifications. The computational requirements of the model are significant and were addressed by the use of grid-computing facilities, the use of efficient experimental designs to keep the necessary model repetitions low, and the development of efficient pre- and postprocessing tools that allow a flexible creation of scenarios and repetitions based on a basic complete model version.

The model was used to analyze potential effects of climate change adaptation on agricultural production and land use in the study area. The results show that besides effects on yields also other climate change-induced effects on the conditions of agricultural production may have important impacts on land use decisions of farmers and deserve more attention in climate change impact analysis. Potential impacts of changes in the time slots suitable for field work and an additional rotation option are predicted to be comparable to the impact of the changes in yields predicted by a crop growth model. Results point to an expansion of wheat and silage maize areas at the expense of barley areas. The partial crowding out of summer barley by wheat area held for current price relations and is less strong at higher relative prices for summer barley. Price response analysis indicated that winter wheat production enters into a substitutive relationship with summer barley production under climate change conditions, while competition with winter barley area diminishes. This leads also to a higher elasticity of the wheat area with respect to relative summer barley prices.

The model was then used to analyze biogas support through the Renewable Energy Act (EEG) and the support for grassland extensification and crop rotation diversification through the MEKA scheme. Especially simulated participation in crop rotation diversification is strongly reduced in the climate change scenarios, while the investments in biogas plants are slightly increased. The conditions established by the latest EEG revision imply that further development of biogas capacity will crucially depend on the existence of demand for excess process heat, because the alternative option of using high manure shares seems to be rather unattractive for farmers in the area according to the simulation results. The simulations suggest a noteworthy interplay between the two policy schemes: Increased investments in biogas plants increase the demand for silage maize production and intensive grassland areas and consequently decrease the attractiveness of participation in MEKA crop diversification and grassland extensification measures. The MEKA scheme may have a minor dampening and delaying effect on biogas investments.

Zusammenfassung

Agentenbasierte Modelle ermöglichen eine vertiefte Analyse des Ausmaßes und der Geschwindigkeit landwirtschaftlicher Anpassung an den Klimawandel, da sie Prozesse einbeziehen, die in anderen agrarökonomischen Modellansätzen gewöhnlich nicht berücksichtigt werden können. In der vorliegenden Arbeit wird mithilfe der Modellierungssoftware MPMAS ein agrarökonomisches Multiagentenmodell entwickelt, um die Anpassung der Landwirtschaft auf der Mittleren Schwäbischen Alb zu untersuchen. Die Arbeit ist Teil der DFG Forschungsprojekte PAK346 und FOR1695, deren Ziel es ist das Verständnis jener Prozesse zu verbessern, die auf der regionalen Ebene Struktur und Funktionen von Agrarlandschaften unter Bedingungen des Klimawandels bestimmen. Vor dem Hintergrund dieser Anwendung werden neue Ansätze diskutiert, entwickelt und getestet, um vier typischen Problemen zu begegnen, die sich bei der empirischen Anwendung agentenbasierter Modelle für ökonomische Analysen ergeben: Datenverfügbarkeit, Prozessunsicherheit, Modellvalidierung und benötigte Rechenkapazität. Mithilfe des erstellten Modells untersucht die Arbeit Klimaeffekte auf die Landwirtschaft, Veränderungen landwirtschaftlicher Angebotsfunktionen sowie Auswirkungen von Fördermaßnahmen für erneuerbare Energieproduktion und Agrarumweltmaßnahmen und demonstriert auf diese Weise seine Anwendbarkeit in der Anpassungsforschung.

Die Arbeit besteht aus zwei Teilen: Der erste Teil diskutiert methodische Aspekte der Nutzung agentbasierter Modelle wie MPMAS, die Entscheidungen als mathematischer Optimierungprobleme darstellen, in der landwirtschaftlichen Anpassungsforschung: Bisherige Erkenntnisse zu den Auswirkungen des Klimawandels auf die Landwirtschaft und den sich daraus ergebenden Anpassungsprozessen werden zusammengefasst und bestehende agentenbasierte Modelle hinsichtlich ihres potentiellen Beitrags zur Anpassungsforschung untersucht. Der Hauptfokus des ersten Teils liegt dann auf der Diskussion bestehender Ansätze zur Modellvalidierung, -kalibrierung und Unsicherheitsanalyse und ihrer Anwendbarkeit auf optimierungsorientierte, agentenbasierte Modelle. Die Diskussion orientiert sich an vier Prinzipien, die die Validität der aus der Modellierung gezogenen Schlussfolgerungen sicherstellen sollen: (i) eine transparente Modelldokumentation, (ii) dass als konstant angenommene Modellelemente tatsächlich nicht zwischen den untersuchten Szenarien variieren, (iii) dass das Modell nicht stärker kalibriert wird als es die verfügbaren Beobachtungen und die erwartete Fehlerverteilung erlauben, und (iv) dass etwaige Auswirkungen der Prozessunsicherheit auf Ergebnisse und Schlussfolgerungen untersucht und kommuniziert werden.

Bei agrarökonomischen agentenbasierten Modelle, die auf mathematischer Optimierung beruhen, ist die zu erwartende Fehlerverteilung oft unklar und es stehen in der Regel nur wenige Beobachtungen zur Verfügung. In diesem Fall ist es ratsam mit robusten Kalibrierungskriterien zu arbeiten und auf die Identifikation einer einzigen, bestangepassten Parameterkombination zu verzichten. Stattdessen sollte der Parameterraum mithilfe robuster Kriterien nur eingegrenzt und Szenarien für ein geeignetes den eingegrenzten Parameterraum abbildendes experimentelles Design wiederholt werden. Simulationsergebnisse sollten dann als Spanne oder Verteilung über alle Wiederholungen des Designs angegeben werden, so dass die Unsicherheit der Ergebnisse deutlich wird. Für die Umsetzung dieses Ansatzes der Validierung und Umsicherheitsanalyse waren generische Erweiterungen des MPMAS Softwarepakets notwendig, die in dieser Arbeit entwickelt wurden.

Der zweite Teil der Arbeit beschreibt die Anwendung der neuentwickelten Verfahren bei der Erstellung und Nutzung eines Multiagentenmodells für die Mittlere Schwäbische Alb. Der Schwerpunkt der Modellentwicklung lag hierbei auf der Abbildung der Heterogenität

des Agentenverhaltens, der empirischen Parametrisierung, und der Berücksichtigung klimatischer Effekte auf mögliche Fruchtfolgen und zur Feldarbeit geeignete Arbeitstage – neben den klimatischen Auswirkungen auf Ernteerträge. Darüberhinaus wurde die Modellierung von Demographie, Investitionsentscheidungen und Pachtmärkten in MPMAS ergänzt, um die Simulation des landwirtschaftlichen Strukturwandels über die Zeit zu verbessern.

Die Daten zur Modellparametrisierung stammen von Statistischen Ämtern, Agrarberatungen, Expertenbefragungen und aus einer Erhebung unter Landwirten der Region. Der traditionell in MPMAS-Modellen benutze Ansatz künstliche, aber repräsentative Agentenpopulationen mithilfe eines Monte-Carlo-Verfahrens zu erstellen wurde ergänzt, um multivariate Verteilungen besser abbilden zu können und die Datenschutzbestimmungen bei der Nutzung von Daten aus der Landwirtschaftszählung und der Agrarstrukturerhebung einhalten zu können. Prozessunsicherheiten konnte teilweise mithilfe eines konservativen Kalibrierungsansatzes reduziert werden, der darauf angelegt war eine Überkalibrierung des Modells zu vermeiden: Der Parameterraum wurde nur in solchen Fällen verkleinert, in denen eine Verbesserung der Anpassungsgüte in allen drei Beobachtungsjahren, d.h. über strukturelle Brüche hinweg, zu beobachten war. Die verbleibende Unsicherheit wird explizit dargestellt, indem alle Simulationsergebnisee nicht als Punktschätzungen, sondern als Verteilungen über den Parameterraum angegeben werden. Die Simulationsergebnisse zeigen robuste Klimaeffekte trotz einer großen, verbleibenden Prozessunsicherheit. Die simulierten Produktionsentscheidungen überzeugten Experten in einem Turing-Test und das kalibrierte Modell reproduziert die beobachtete Landnutzung mit zufriedenstellender Genauigkeit. Beobachtete Abweichungen entsprachen den aufgrund der bei der Modellerstellung angenommenen Vereinfachungen zu erwartenden Fehlern. Die Modellsimulationen benötigen eine hohe Rechenkapazität und wurde nur möglich durch die Nutzung von Grid-Computing Clustern, die Entwicklung von Pre/Postprocessing-Skripten für flexible Erstellung von Szenarien und Wiederholungen, und die Nutzung effizienter experimenteller Designs.

Mithilfe des Modells wurden potentielle Anpassungsreaktionen der Landwirte auf den Klimawandel hinsichtlich ihrer Auswirkungen auf landwirtschaftliche Produktion und Landnutzung in der Untersuchungsregion analysiert. Die Ergebnisse zeigen, dass neben Ertragsveränderungen auch andere klimainduzierte Veränderungen der landwirtschaftlichen Produktionsbedingungen bedeutende Auswirkungen auf die Landnutzungsentscheidungen der Landwirte haben können und mehr Aufmerksamkeit verdienen: Potentielle Klimaeffekte auf Feldarbeitstage und zusätzliche Fruchtfolgeoptionen zeigten ähnliche Auswirkungen wie die von einem Pflanzenwachstumsmodell vorhergesagten Ertragsveränderungen. Die Ergebnisse deuten auf eine Ausweitung der Weizen- und Silomaisanbaufläche auf Kosten des Gersteanbaus hin. Die Verdrängung von Sommergerstefläche durch Weizenfläche gilt allerdings für momentane Preisrelationen und ist bei höheren Relativpreisen für Sommergerste weniger stark ausgeprägt. Eine Analyse der Angebotsreaktionen zeigte, dass die Winterweizenfläche unter Klimawandelbedingungen in ein Substitutionsverhältnis mit der Sommergersteproduktion tritt, während die Konkurrenz mit Wintergerste abnimmt. Das bedeutet auch, dass die Kreuzpreiselastizität der Winterweizenfläche hinsichtlich des Braugerstepreises zunimmt.

Das Modell wurde außerdem genutzt, um die Förderung der Biogaserzeugung durch das Erneuerbare-Energien-Gesetz (EEG) und die Förderung der Grünlandextensivierung und Fruchtfolgediversifizierung durch das MEKA-Programm zu untersuchen. Speziell die Beteiligung an der Fruchtfolgediversifizierung zeigte einen starken Rückgang in den Klimawandelszenarien, während die Investition in Biogasanlagen leicht stieg. Nach der letzten Änderung des EEG, die die Nutzung von Prozessabwärme zur Voraussetzung für eine

Förderung macht, muss davon ausgegangen werden, dass weitere Investitionen in Biogasanlagen stark von der lokalen Vermarktbarkeit von Überschusswärme abhängen werden, da die Alternativoption erhöhter Güllenutzung nach den Simulationsergebnissen für die Landwirte eher unattraktiv erscheint. Die Simulationen zeigen auch den Zielkonflikt zwischen den beiden Politikmaßnahmen auf: Erhöhte Investitionen in Biogasanlagen steigern die Silomaisnachfrage und Grünlandintensivierung und verringern die Attraktivität der Beteiligung an MEKA-Fruchtfolge- und Grünlanddiversifizierungsmaßnahmen. Die MEKA-Maßnahmen haben auf der anderen Seite eine leicht verringernde und verzögernde Wirkung auf Biogasinvestitionen.

Contents

List of Figures

List of Tables

List of acronyms

AA	agricultural area
ABM	agent-based model
AIC	Akaike information criterion
AMS	automatic milking system
AOGCM	Atmosphere-Ocean General Circulation Model
AR4	the Fourth Assessment Report of the IPCC
BIC	Bayesian information criterion
BW	Baden-Württemberg
CAP	EU Common Agricultural Policy
CCM	corn cob mix
CGE	Computable General Equilibrium Model
EEG	*Erneuerbare Energien Gesetz* 'Renewable Energy Act'
EU	European Union
ESU	European Size Unit (EU farm typology)
EMIC	Earth System Models of Intermediate Complexity
FADN	Farm Accounting Data Network
GCM	General Circulation Model
GHG	greenhouse gas
GTOF	general type of farm (EU farm typology, level 1)
ha	hectare
IPCC	Intergovernmental Panel on Climate Change
LAI	leaf area index
LHS	Latin-hypercube sample
LU	livestock unit
KWK	*Kraft-Wärme-Kopplung*, 'combined heat-power generation'
MAS	multi-agent system
MEKA	*Marktentlastungs- und Kulturlandschaftsausgleich*, 'Compensation Scheme for Market Easing and Landscape Protection' (an agri-environmental support scheme in BW)
ME	metabolizable energy (animal nutrition)
MF	main forage area
MIP	mixed integer programming
MP	mathematical programming
MPMAS	Mathematical Programming-based Multi-Agent System
MTR	Mid-term Review

NaWaRo	*Nachwachsende Rohstoffe* (Renewables)
NEL	net energy lactation (animal nutrition)
NPV	net present value
OA	orthogonal array
nXP	usable raw protein (animal nutrition)
PMP	positive mathematical programming
PTOF	principal type of farm (EU farm typology, level 2)
py	person-year: yearly workload of a full-time employee
RCM	Regional Climate Model
rLHS	replicated Latin hypercube sample
RLU	roughage-consuming livestock unit
SCM	Simple Climate Model
SRC	standardized regression coefficient
SRES	Special Report on Emissions Scenarios
UAA	utilized agricultural area
UPCS	unbiased permuted column sample (cf. Morris et al. 2008)
VBA	Visual Basic for Applications
XP	raw protein (animal nutrition)

Acknowledgements

I would like to express my gratitude to Prof. Thomas Berger for encouraging and supervising the work presented in the current thesis. His guidance and openness to my ideas have been essential in the development of this work. I want to thank Prof. Thilo Streck and Prof. Herald Grethe for taking the time to evaluate my doctoral dissertation.

I thank the Ministry of Rural Areas and Consumer Protection of Baden-Württemberg (MLR) for data on MEKA participation, and the Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) for meteorological data. I am grateful to the Forschungsdatenzentrum der Statistischen Landesämter (FDZ) for providing the AFiD-Panel data on agricultural and forestry and especially to Alexander Vogel for his patience and willingness to cope with my unusual requests. I thank Norbert Sauer at KTBL for his helpful comments regarding the calculation of standard gross margins. Without the computational resources provided by the bwGRiD project [bwGRiD, 2013] my simulations would not have been possible.

This work was funded as part of the DFG Integrated Project (PAK) 346 and the DFG Research Unit 1695 'Regional Climate Change' and conducted at the Department of Land Use Economics in the Tropics and Subtropics (490d) of the University of Hohenheim. It would not have been possible without the contributions of the project team and department members. I am especially indebted to Germán Calberto, Joachim Ingwersen, Kirsten Warrach-Sagi, Petra Högy and Eckhart Priesack (and many others behind the scenes) for their work in providing data, parameterizing and running the crop simulations with Expert-N. Marius Eisele, Joachim Aurbacher, Matthias Siebold, Viktoriya Latysheva, Evelyn Reinmuth and Jennifer Steinbach collaborated in the collection of socioeconomic data and stakeholder information. Teresa Walter redacted and processed most of the information related to biogas electricity production. Apart from these, the discussions with Thorsten Arnold, Lutz Göhring, Evgeny Latynskiy, Christian Grovermann, Tesfamichael Assfaw Wossen and Marcelo Precoppe provided many inspirations for this thesis.

I want to thank my parents and family for keeping me anchored and grounded and always being there when I need them, and most of all, I thank Ines for tolerating my physical and mental absence and always encouraging me in my work.

Introduction

Climate change has the potential to profoundly affect agriculture in many regions of the world. Farmers may have to deal with weather conditions that have not been observed in their area before. As a consequence, scientists and policy makers alike are increasingly looking for tools that enable them to make useful and well-founded predictions of the consequences of climate change for farming.

In most cases, such analysis will directly or indirectly build on climate models, which provide projections of future climatic development at increasingly detailed levels. These can then, for example, be fed into hydrological, land-surface and plant growth models in order to analyze the changes in water availability, soil moisture or crop yields triggered by changing weather conditions.

Farmers will not remain passive, merely observing a changing natural and socioeconomic environment, but they will actively react to changes and adapt their production: They will try to mitigate potentially adverse effects and, if possible, strive to benefit from newly arising opportunities. It is paramount to take farm adaptation into account when analyzing the potential effects of climate change on agricultural production, food security, the economic well-being of farms, or the environment. Moreover, adaptation of agricultural production may feed back into the geophysical system: Agricultural land cover and irrigation practices affect heat and water fluxes in the lower atmosphere. Crop management influences carbon fixation in soils, and energy and resource consumption of agricultural production contribute to greenhouse gas emissions. Hence, besides models from the natural science domain, also agricultural economic models constitute an important pillar of the analysis of climate change effects [Reidsma et al., 2010].

Since many of the conditions expected for the future have not yet been observed in the past, researchers cannot simply rely on a statistically estimated simple parametric relationship of the outcome of interest (e.g. agricultural supply, land use) to exogenous inputs (e.g. prices): Many of the statistical relationships that can be estimated from past observations will likely be altered by climate change, but also by other factors like inherent structural change within the agricultural sector or changes in policy. Instead, models have to build on the understanding of processes and regularities that are expected to remain unaffected by climate change [Antle and Capalbo, 2001].

It is for this reason that climatologists employ General Circulation Models (GCMs) to analyze future climatic developments. They rely on the fundamental laws of physics (especially fluid dynamics and thermodynamics), and – where this is prevented by coarse resolution – on parametric equations of well-observed basic meteorological processes (e.g. cloud formation). Models are rigorously evaluated by trying to reproduce meteorological observations of the past [Randall et al., 2007].

It would be desirable if also agricultural economics could provide such models for the analysis of agricultural adaptation, albeit the difficulties to predict human behavior, the complexity of natural processes involved in agricultural production, and the limits to the observation of local circumstances make it hard to imagine that agricultural economic models could get anywhere near the predictive capacity of weather forecast models or at least climate models.

Nevertheless, there are generalizable patterns of human behavior and agricultural processes that can be represented in models. For example, agricultural economists would expect that, also in the future, farmers will weigh expected benefits and costs of investment and production activities intending to choose the combination they deem best suited for their farm and family. If, however, agricultural economic models are employed in the analysis of adaptation, then researchers should also strive to narrow the gap to advanced biophysical models by preferring theoretical representations of invariant patterns of behavior over mere statistical relationships and by emphasizing an empircial parameterization, conscious validation and uncertainty analysis of their models.

Several authors [e.g. Balbi and Giupponi, 2009; Patt and Siebenhüner, 2005; Moss et al., 2001] suggest the use of agent-based models (ABM) to analyze climate change adaptation in agriculture. Agent-based models represent every relevant real-world decision-maker, e.g. every farm manager, by an individual agent in the model. Aggregate outcomes, e.g. regional agricultural production, result from the individual actions of all agents and their interactions. This setup allows reflecting the individually specific decision problems and capture the heterogeneity of utility functions, resource constraints, knowledge and perception, which tend to differ i.a. with farm size, location, or household composition. Explicitly capturing interactions enables the modeler to examine fragmented markets, innovation diffusion processes, social learning, or farm cooperation problems, all of which are not easily incorporated into other modeling frameworks.

For the problem of climate change analysis, agent-based models enhance the modelers' options to formulate cost and benefit functions based on generalizable patterns and include climatic conditions and other factors varying with time or between farms as explanatory variables. As an example, consider the cost for harvesting barley: The amount of manpower, fuel and maintenance cost required to harvest a hectare of barley with a certain type of machinery combination can be assumed to be unaffected by changes in local climate.[1] The same is true for the maximum water content in the grain allowed at harvest. What will probably change is the timing of harvest and the probability of having suitable weather conditions allowing the grain to reach the required water content around that time. As a consequence, the harvest period may shrink or widen, or move to overlap with e.g. the harvest period of another crop. The farmer could then have to employ more manpower or more powerful machinery with higher cost. Or he could decide to produce less of either crop, which would then give rise to opportunity costs. The cost incurred by using more powerful machinery, using more manpower, reducing the production of another crop, or reducing cropping area – and whether these are viable options at all – depends very much on the specific circumstances of each farm. An agent-based model (ABM) not only allows to represent all the farm-specific circumstances faced by farm managers in the area, it can also predict the consequences of such a diverse farming population economically and socially interacting.

On the other hand, the use of ABM in agricultural economics is associated to a number of challenges that arise from the disaggregate nature of these models:

Data requirements Making use of the full power of ABM requires the availability of disaggregate datasets in order to reflect the actual heterogeneity of the farming population. Though detailed datasets like agricultural censuses, farm accounting databases or nation-wide household surveys are increasingly available, access to these datasets is often associated to privacy restrictions and they do not necessarily contain the complete scope and detail of information required.

Process uncertainty Modeling regional agricultural supply and land use by explicitly rep-

[1] Abstracting from changes in fuel prices and changes in yields, i.e. amount to be harvested, for a moment.

resenting individual farm decisions involves the necessity to model processes whose best theoretic representation is inherently uncertain, debatable or has not even been developed. Observation data can be scarce and modelers will often have to resort to *ad hoc* implementations and parameterizations.

Model validity and uncertainty Data gaps, process uncertainty and *ad hoc* parameterization entail considerable epistemic uncertainty. Moreover, the complexity of model approaches hampers the accessibility to stakeholders and scientific review. This raises doubts about the validity of agent-based modeling approaches, especially, since a shared understanding of suitable validation and calibration procedures for agent-based models in agriculture has not yet been established.

Computational requirements Depending on the number of individuals to be represented and the complexity of their decision problems and interactions, the time and resources required for simulation can be high limiting the number of scenarios and repetitions to represent model uncertainty that can be evaluated.

The present thesis examines whether these challenges can be overcome in the context of a case study of agricultural adaptation to climate change in the Central Swabian Jura, a mountainous area in Southern Germany, using the MPMAS multi-agent framework [Schreinemachers and Berger, 2011].

It discusses different approaches to model validation, calibration and uncertainty analysis and their suitability for the use with mathematical programming-based ABMs in order to deal with the uncertainty involved in using *ad hoc* parameterizations. As these approaches require large numbers of simulation runs, it presents tools to automatize the pre- and post-processing of MPMAS inputs and outputs.

Selected methodologies and tools are then employed for the development of a model of climate change adaptation for the Central Swabian Jura. The model illustrates the disaggregation of cost and benefits of different production options as a function of climatic conditions, farm size and resources and household composition in order to be able to reflect changes in profitability of different production options and analyze the capacity of farms to adapt.

Available disaggregated secondary farm datasets are used to initialize and validate the ABM using an approach that observes the associated privacy restrictions. The epistemic model uncertainty is partly reduced in a conservative calibration approach and the remaining uncertainty is explicitly communicated by presenting simulation results as distributions over the space of potential parameter combinations instead of point estimates.

This thesis is divided into two parts. The first part is dedicated to a methodological discussion of the use of mathematical programming based multi agent systems, such as MPMAS, for the analysis of agricultural adaptation of climate change, with a special emphasis on uncertainty analysis and validation. It starts by recapitulating the different pathways through which climate change will affect agricultural production, and the processes, options and decisions through which adaptations may occur (chapter 1). This summary serves as the basis to formulate requirements that an ABM of climate change adaptation should fulfill in chapter 2, which are then used to review the literature on ABMs in agricultural economics and land use sciences with these requirements in mind. On the one hand, ABMs that draw on the agricultural economics tradition of using mathematical programming to model farm decision-making (e.g MPMAS) stand out in their ability to reproduce the complex interlinkages between production options and resource constraints. On the other hand, other ABMs albeit using undercomplex representations of farming conditions offer interesting approaches to capture interaction and learning processes. What becomes especially apparent, however, is that those models which strive to provide empirical representations of specific agricultural regions, often struggle with data requirements and hardly reflect on the suitability of calibration and validation procedures they employ.

Chapter 3 summarizes basic principles and approaches of model validation, calibration and uncertainty analysis, recalling that model validity should not be equated to a high goodness-of-fit score, but rather involves a sound theoretical foundation, a thorough uncertainty analysis, and a prudent use of calibration. This lays the foundation for a discussion of the validation and calibration of mathematical programming-based ABMs in chapter 4.

Favored approaches for uncertainty analysis and calibration require repeating scenario simulations with many different parameter combinations and model setups. As the traditional MPMAS interface presents severe limitations to such an approach, chapter 5 introduces new pre- and postprocessing tools allowing to process model input and output for large numbers of model runs. It also shortly discusses the use of MPMAS on grid computing facilities. The presented tools are designed to be generic and to enhance the potential of MPMAS for many future applications.

The second part of the thesis describes the application of the MPMAS model in connection with the discussed methodologies and the newly developed tools in the context of a joint research project of the University of Hohenheim and the Helmholtz Center Munich that is dedicated to the analysis of climate change effects on agricultural landscapes on a regional scale. It starts with an introduction of the study area in the Central Swabian Jura (chapter 6), followed by a detailed description of the MPMAS model developed for the analysis of agricultural adaptation to climate change in the area (chapter 7). Chapter 8 discusses the model parameters which have to be considered uncertain and describes the calibration procedure used to reduce the uncertainty, which was developed based on the foundations laid out in the first part of the thesis. The model is then used for one period assessments of climate change adaptation focusing on the importance of different pathways of climate change impacts, structural effects of climate change on price response functions and the effects associated to agri-environmental and biogas support policies (chapter 9). The following chapter 10 tests the long-run stability of short-term results in first recursive-dynamic simulations. The model, methodology and technical infrastructure employed and the simulation results obtained in the practical application in the Central Swabian Jura are discussed in chapter 11.

Finally, the conclusions that can be drawn from both the methodological and the practical part of this thesis for the use of agent-based modeling in climate change adaptation analysis as well as other empirical use in agricultural economics are summarized.

Part I

Agent-based modeling of agriculture under conditions of climate change: a discussion of methodology

Chapter 1

Climate change and agriculture

In their "anatomy of adaptation", Smit et al. [2000] identify four essential questions regarding the analysis of adaptation to climate change:

(1) "who or what adapts?"
(2) "adaptation to what?"
(3) "how does adaptation occur?"
(4) "how good is the adaptation?"

The fourth question highlights an important motivation for conducting simulation studies assessing the adaptation of agricultural systems to climate change: How well will farmers adapt to climate change? How will food production and farm income be affected? Will the adaptation have adverse effects on the environment, maybe even reinforce climate change? Will it be necessary to design policies to influence the course of adaptation, and what effects would suggested policies really have?

Building a simulation model to address these issues requires answering the first three questions. This first chapter will discuss them in a general way in order to lay the conceptual basis for discussing how agent-based models can best be employed to increase our understanding of adaptation processes and their consequences. Case studies will most of the time consider only a certain part of the potential effects of climate change and only part of the potential ways to react to them, focusing on what is considered relevant for the specific research question and constrained by the scope and resources available for the research. For the discussion of methodology, however, it is important to have an overview of what could be considered providing a frame of reference against which to judge the completeness of an analysis and compare it with other research efforts.

The chapter starts by addressing the question "adaptation to what?", giving a short summary of current knowledge and expectations on the future climatic development (section 1.1) and exploring the different ways in which climate change will affect agricultural systems (section 1.2). Section 1.3 then turns to the questions "who or what adapts?" and "how does adaptation occur?" looking at the actors and mechanisms of adaptation in agriculture. The chapter ends with a summary and typology of motivating questions for climate change adaptation studies in section 1.4.

1.1 Projecting future climatic development

Any estimation of the effects of climate change upon agriculture requires an understanding and potentially a projection of the future development of the global climate system. Depending on the scale of analysis such a projection needs to be downscaled to the region of interest.

1.1.1 Global projections of climate change

Global climate projections are usually created using Atmosphere-Ocean General Circulation Models (AOGCMs). These models are based on a discretized representation of the Earth at a rather coarse resolution (about 1-3°, about 200 km) and simulate the atmospheric and oceanic processes determining the state of meteorological variables in each subdivision over time. Processes are based on fundamental physical laws, although in many cases parametric approximations are necessary e.g. in order to reflect processes that occur at lower scales than the chosen resolution [Randall et al., 2007].

A number of different AOGCMs have been developed and used in research institutions around the world. These models differ in resolution, upper limit of the atmosphere and scope, formulation and parameterization of processes included. An overview of 23 different AOGCMs can be found e.g. in Randall et al. [2007]. The differences in formulation and parameterization lead to different sensitivities of the models to exogenous shocks and consequently to different climate projections. Experience has shown that averaging over the results of climate projections of different models gives more reliable results than looking at the results of a single AOGCM [Phillips and Gleckler, 2006; Meehl et al., 2007].

The uncertainty in climate change projections stems from three different sources: natural variability of the climate system, the incomplete knowledge about the climate system, respectively its incomplete representation in climate models, and uncertainty about the future development of exogenous influences – especially anthropogenic emissions [Tebaldi and Knutti, 2010]. Natural variability is accounted for by running the same climate simulation with different initial conditions, and averaging over the results [Tebaldi and Knutti, 2010]. Computational constraints still prevent increasing the resolution and thus the number of explicitly modeled processes, and also limit the number of simulations that can be conducted with AOGCMs. Simple Climate Models (SCMs) and Earth System Models of Intermediate Complexity (EMIC) are simplified model architectures which have been developed to reflect the behavior of AOGCMs with lower computational costs at the cost of a less process-based architecture. These are used for repeated calculations at higher resolutions and a higher number of scenario evaluations [Randall et al., 2007].

The major exogenous shock applied in AOGCM simulations is the change in anthropogenic emissions of greenhouse gases. The Special Report on Emissions Scenarios (SRES) [Nakićenović and Swart, 2000] created 40 different scenarios following four different narratives on the future development of the world, which reflect expected emissions associated with different potential future trajectories of the global economic system. All of these scenarios are baseline scenarios that do not consider any mitigation measures specifically directed at climate change.

Scenario family A1 assumes an evermore globalized world with strong economic growth and convergence of economic prosperity as well as social and cultural relations between the different countries of the world. The world population is expected to grow until the middle of the 21st century and to decline slowly afterwards. Scenario family A2 assumes a more heterogeneous world with slow convergence of living conditions and regionally oriented economic growth. Population growth is highest in this scenario family and continues until the end of the century. Scenario family B1 is similar to the A1 family with respect to economic and cultural convergence and population growth, but assumes that the world economy will become more resource efficient and sustainable by a transition from an industrial to a service and information economy. Scenario family B2 combines the transition towards a more environmentally friendly economy of B1 with the globally fragmented development of A2. World population will continue to increase until the end of the century, but more slowly than in scenario family A2.

Scenario family A1 has been further subdivided into three subgroups representing different trajectories of technological development (A1FI - fossil-intensive, A1T - nonfossil , A1B - balanced between fossil and nonfossil energy sources), leading to six scenario groups in total. Out of the 40 scenarios, the SRES authors marked one scenario in each scenario group which they considered to provide the best illustration for the respective scenario group [Nakićenović and Swart, 2000].

The coordinated global climate simulation experiments conducted for AR4 focused resources on three of the SRES illustrative scenarios: B1, A1B and A2, representing low, medium and high future emission levels and spanning the full range of uncertainty of future greenhouse gas (GHG) emissions [Meehl et al., 2007]. Results of model means for changes in global temperature and precipitation are shown in table 1.1. There is little influence of the chosen SRES scenario on projected warming in the early 21st century (2011-2030), also because roughly half of the projected temperature increase of 0.64-0.69°C would occur even if greenhouse gas concentrations could be fixed at the levels observed in the year 2000. Advancing through the 21st century, the influence of the emission scenario increases and the share of temperature increase already 'committed' decreases [Meehl et al., 2007].

Table 1.1: *Projections of global annual mean surface air temperature and precipitation change compared to 1980-1999 from AR4 [Meehl et al., 2007, tabs. 10.5, S10.4]*

	Global mean warming (°C)			Global mean precipitation change		
Scenario	2011-2030	2046-2065	2080-2099	2011-2030	2046-2065	2080-2099
A2	0.64	1.65	3.13	1.38	1.33	1.45
A1B	0.69	1.75	2.65	1.45	1.51	1.63
B1	0.66	1.29	1.79	1.62	1.65	1.88
const. 2000	0.37	0.47	0.56	2.27	2.32	2.29

Geographically, the AR4 results suggest an above-average warming over land except for the Southern mid-latitudes and a lower than average warming over the oceans except for the high Northern latitudes. Warm weather extremes are expected to increase, while cold weather extremes are expected to decrease in frequency. Daily minimum temperatures increase faster than daily maximum temperatures leading i.a. to longer growing seasons in mid to high latitudes [Meehl et al., 2007]. Precipitation and evaporation are expected to increase on a global average, with the distribution becoming more unequal both geographically and over time: Precipitation is expected to increase around the equator and in the high latitudes and to decrease in the subtropics, i.e. generally speaking wet zones become wetter and dry zones become drier. In tropical, mid- and high latitude areas, rainfall intensity during extreme events increases overproportionally. In the subtropics, rainfall intensity increases, but time between rainfall events increases, too [Meehl et al., 2007].

1.1.2 Regional projections

Due to the coarse resolution of AOGCMs and their inability to model smaller-scale processes explicitly, results from AOGCMs are commonly judged to be meaningful at a horizontal resolution of at least 1000 km only [Christensen et al., 2007]. Assessing impacts on agriculture for specific locations or regions requires considerably higher precision, especially the projection of precipitation patterns needs to incorporate smaller scale processes explicitly [Ines and Hansen, 2006; Ehret et al., 2012]

Climate projections for individual locations as required for impact assessments are created either by statistical or by dynamic downscaling of the results of a GCM [Tebaldi and Knutti, 2010]. In statistical downscaling, statistical relationships between local weather patterns and large scale patterns are estimated from past weather records and then applied to derive local projections from large scale projections produced by the GCM. This is a purely statistical approach, not necessarily physically consistent with the GCM itself and the stability of statistical relationships between local weather and large scale circulation under climate change is questionable [Wilby, 1997; Wilby and Wigley, 1997; Ehret et al., 2012]. In dynamic downscaling, a Regional Climate Model (RCM) with a higher resolution is driven by boundary conditions taken from the results of a GCM. RCM resolutions of 10 to 50 km are common [Tebaldi and Knutti, 2010], ongoing efforts strive to increase the resolution towards cloud-resolving scales around 2 km [Hohenegger et al., 2008]. While using a RCM has the potential to improve the physical representation of regional scale processes, biases in projection are often still too large to allow direct use in impact assessments. Bias correction is then often applied to projections based on observed deviations. Like statistical downscaling methods, bias correction procedures are not necessarily physically consistent with the RCM and the stationarity of observed differences remains debated [Ehret et al., 2012].

1.2 Pathways of climate change effects on agriculture

Climate change affects regional agricultural systems along a large variety of different pathways. On the one hand, it directly alters the biophysical conditions under which agricultural production takes place in a region. On the other hand, it will also induce economic, social, political and technological developments that change the socioeconomic environment in which farms operate. The following two subsections provide an overview of the biophysical and socioeconomic pathways that have to be considered when analyzing the effects of climate change upon agriculture.

1.2.1 Biophysical pathways

The overview of biophysical effects largely follows the reviews redacted by Olesen and Bindi [2002] and Lotze-Campen and Schellnhuber [2009].

Apart from its effects on global climate, the *increased atmospheric concentration of* CO_2 will have a direct effect on plant growth. Photosynthesis, transpiration (due to effects on stomatal aperture and density) and maintenance respiration react to elevated CO_2 concentrations. As a combined effect, yield increases and increased resource use efficiencies of water, nitrogen and radiation can be observed [Olesen and Bindi, 2002]. Besides the effect on yield quantity, elevated CO_2 concentrations may also affect yield quality. Especially, reductions in protein contents are expected [Taub et al., 2008; Högy and Fangmeier, 2008]. Depletion of the stratospheric ozone layer leading to *higher UV-B radiation* as well as increases in *tropospheric ozone concentration* may adversely affect yields [Olesen and Bindi, 2002].

As an indirect effect of rising greenhouse gas concentrations, global mean temperature is expected to rise, albeit with strong regional differences. A *temperature increase* should, generally, accelerate plant growth and increase biomass production up to a certain plant specific limit, when the effect turns negative due to *heat stress*. At the same time, *evapotranspiration* increases and leads to higher water use. Apart from the mean temperature, also changes in extreme temperatures are relevant in this respect as they change the *length of growing seasons*. Accelerated development may have detrimental effects on yields of determinant crops, leading to early maturity [Olesen and Bindi, 2002]. Whether positive or negative yield effects dominate, will vary from region to region. In general, higher latitudes are expected to experience an extension of crop areas due to extended growing seasons and milder winters, allowing the growth of new crops and more productive cultivars, while in lower latitudes increased heat stress and evapotranspiration might reduce yields and potential cropping areas [Olesen and Bindi, 2002; Lotze-Campen and Schellnhuber, 2009].

Changes in *precipitation* will affect water availability for both rainfed and irrigated agriculture. Rainfall is expected to increase in high latitudes and the wet tropics, and to decrease in mid-latitudes and the dry tropics, i.e. in tendency already wetter regions get even wetter and dry regions get drier. Net irrigation requirements are expect to rise by 3-5% until 2020 and by 5-8% until 2070 on global average [Döll, 2002]. In general, *weather variability* is expected to increase making agricultural production more risky. This is true for 'normal' weather variations affecting yield quantities, especially for crops grown close to their limits of climatic tolerance, but also includes more frequent occurrences of extreme weather events like droughts, heat waves, extreme precipitation events, or floods that are likely to destroy or severely damage harvests [Olesen and Bindi, 2002; Lotze-Campen and Schellnhuber, 2009]. Higher temperatures increase the decomposition of organic matter and the chance of nutrient leaching, while more intensive rainfall accelerates soil erosion, drier soils speed up wind erosion and increased evapotranspiration fosters salinisation, leading overall to more severe *soil degradation* [Olesen and Bindi, 2002].

Knowledge on the effects of climate change on *pests* and *plant diseases* is still very fragmented. Like crops, also pests and pathogens are affected by the atmospheric and climatic changes described above, however, effects on yield quantity and quality are very much dependent on the specific crop-pest/crop-pathogen interaction and hard to generalize [Gregory et al., 2009; Luck et al., 2011]. This holds especially true for the potential damage that may be caused by invasive species, whose importance can rather be expected to increase with moving climatic zones and increasing global trade [Ziska et al., 2011]. Climate change might aggravate anthropogenic threats to *pollinator* populations [Kerr, 2001] and affect production of crops that depend on biotic pollination. Weather conditions constrain *crop management*: Cereal harvests require dry weather, moisture levels restrict the workability of soils and their capacity to be grazed by animals. Changing climates will thus also alter time slots for crop and herd management [Olesen and Bindi, 2002].

Animal husbandry is, of course, affected by effects on the production of fodder crops and pastures, as described above. Further, weather conditions directly affect *animal growth, health, reproduction* and product quantity and quality as animals only have limited ranges of heat and cold tolerance. The geographical range of animal diseases and pests will probably change as well [Rötter and van de Geijn, 1999; Olesen and Bindi, 2002]. Global fish production is likely negatively affected, though also here considerable uncertainty remains in scientific assessments [Brander, 2007]. Finally, what is true for animals also holds for humans: Extreme weather conditions and changed disease spreading patterns (e.g. malaria) may directly affect the well-being and workforce of farm households and workers [Morton, 2007].

1.2.2 Estimating effects of climate change on crop yields

The effect of climate change on crop growth can be estimated using statistical estimation based on observation of crop growth under different climatic conditions. Time-series models use historical crop yield and weather data for one location and determine the effect of year-to-year weather variability on crop yields. While this may be useful for estimations of climate change effects in the near future, it is not guaranteed to give good predictions for

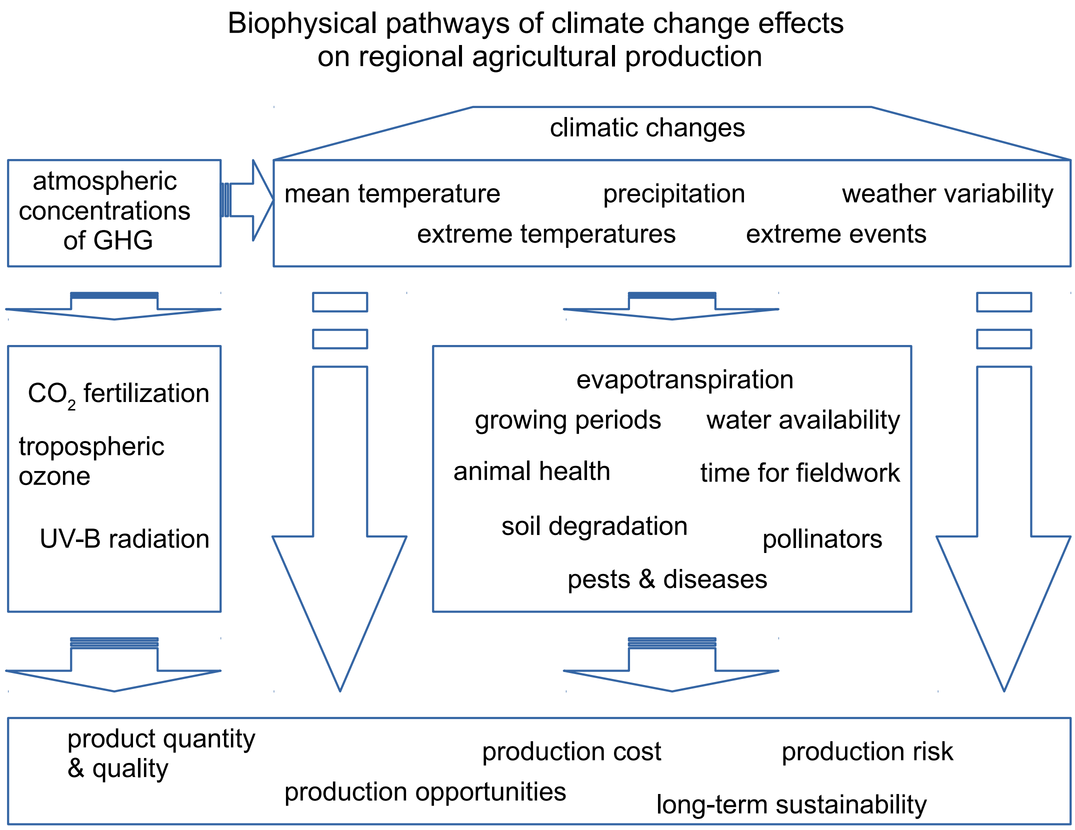

Figure 1.1: *Biophysical pathways of climate change effects on regional agricultural production*

previously unobserved situations: It assumes stationarity of crop-weather relationships that might, however, change in the future and it neglects that farmers may adapt to gradually proceeding climate change in a way they cannot to year-to-year weather variabilities [Lobell, 2010].

Cross-sectional analysis uses yield observations from different climatic regions to establish climate-yield relationships. This does capture the locally specific adaptations of farmers to the long-term climate, however, regional differences may be due to more factors than climatic differences, which can often not sufficiently be controlled for. Fixed effect models using panel data can overcome this omitted variable problem, however, as time series models they are based on year-to-year variability rather than regional climate and thus do again not capture adaptation [Schlenker, 2010].

As an alternative to statistical estimation, crop yields under different meteorological conditions can be simulated using crop growth models. Crop growth models incorporate the scientific understanding of biophysical processes within the plant and between the plant and its environment. Process representations range from simple parametric approaches to more detailed, complex approaches. While their parameterization requires much more extensive datasets than statistical analysis, well-parameterized crop growth models can be used to predict crop behavior under unobserved conditions, including testing adaptations like fertilization, the timing of sowing and harvesting or even the use of different cultivars. Still, to the extent the model has been calibrated on current conditions, a caveat on extrapolation also applies. Moreover, the potential of cultivars developed in the future remains inherently hard to assess [White and Hoogenboom, 2010].

In a series of studies, research teams around M. Parry and C. Rosenzweig estimated the effect of climate change on crop yields around the world and its consequences for global food supply [Rosenzweig and Parry, 1994; Parry et al., 1999, 2004]. In the 2004 edition of their study, Parry et al. use climate projections generated by the HadCM3 AOGCM for four different SRES scenarios in order to estimate changes in wheat, rice, maize, and soybean yields in the 2020s, 2050s and 2080s for every country. Their estimation is based on regional yield transfer functions that have been statistically estimated from crop growth model simulations for experimental sites around the world. The authors report results with and without CO_2 fertilization effect and consider plot level adaptation of crop management (fertilization, variety, timing). Scenario results show major differences only from the 2050s onwards. The estimations for the 2020s show little differences between scenarios, and results are dominated by multidecadal climate variations. Consideration of the effect of CO_2 fertilization is crucial for predicted outcomes. CO_2 fertilization overcompensates for expected losses due to temperature and precipitation changes in most of the developed world, and at least partly alleviates climatic pressure in the developing world. Without CO_2 fertilization by the 2080s losses are predicted for the whole world reaching up to 30% in Africa and surpassing 2.5% even in Canada and Europe. In general, a rather strong difference between developed and developing countries can be observed. With CO_2 fertilization yields are predicted to increase by more than 10% by the 2080s in Europe under A1FI, for example. Taken together the global yield effect is negative (0-5%) for nearly all scenarios, except A2, and strongly negative if the direct CO_2 effect cannot be realized as assumed (-9 to -22%).

1.2.3 Socioeconomic pathways

Biophysical effects of climate change will affect agricultural systems all over the world in different ways. When adaptation is analyzed in a specific region, it needs to be taken into account that also the climate change-induced adaptations in other agricultural systems have an indirect effect on the agricultural system under study via price effects in both product and input markets. Competition will increase or decrease if product supply or input demand change elsewhere [Juliá and Duchin, 2007; Lotze-Campen and Schellnhuber, 2009].

Furthermore, such interdependencies exist also with other economic sectors that compete with agriculture for ever scarcer resources (e.g. energy, labor, land) or adapt their demand for agricultural products due to climate change. The demand for energy production from biomass is certainly the most prominent example. All these kinds of market changes affect farming households not only as producers, but also consumers [Easterling et al., 2007]. At the same time, awareness of climate change among policy makers and consumers creates demand for new products or products that fulfill certain climate-related requirements. Payments for environmental services (e.g. carbon storage), environmental certification schemes, demand for regional products, reduced demand for meat are some examples. Apart from demand on markets, preference for climate-friendly production shapes new regulations constraining production and subsidy schemes setting incentives for certain products or production processes [Janssen and de Vries, 1998; Rosenzweig and Tubiello, 2007; Hall et al., 2004]. Climate change induces technological innovations that find their way into farm enterprises. For example, photovoltaic systems have nowadays become an additional income source for many German farms [Linder, 2013].

1.3 Agricultural adaptation to climate change

As Rosenzweig and Tubiello [2007, p.860] note: "Adaptation in agriculture is the norm rather than the exception". Farmers are used to react to constant changes in market prices, technological progress and changes in consumer preferences, and they are used to deal with weather variability. Adaptation of farming systems is therefore expected to occur autonomously to an important extent and consequently farmers are the main actors of adaptation.

Certainly, governments, associations and extension services can try to increase the adaptive capacity of farmers and try to avoid or mitigate adverse consequences of adaptation. The first thing they can do is foster the generation of and disseminate information about climate change impacts and adaptation options [Easterling et al., 2007]. Improved information may turn adaptation from a 'reactive' process, i.e. adaptation after agents have observed changes to their environment, into a partly 'anticipatory' process, i.e. adaptation based on the expectation that a certain change will occur, although it has not occurred yet [Smit et al., 2000]. They can support the research and development of new technologies including the breeding of new varieties, the establishment of weather and climate information systems and the development of water, nutrient and pest management innovations. Governments can support adaptation by financial and material support introducing or modifying subsidy, incentive and income stabilization programs, disaster aid, and public insurance. Of course it is also their task to engage in the adaptation of infrastructure e.g. for transport or irrigation water supply [Smit and Skinner, 2002; Olesen and Bindi, 2002; Easterling et al., 2007]. Besides producers and public sector organizations, also other private actors in the agricultural system have to react to climate impacts and resulting adaptations of producers. Insurance companies can offer private crop insurance schemes [Smit and Skinner, 2002], financial institutions crop futures and forwards and adapted financing. Processors, traders and input suppliers may change their contractual policies.

Before policy-makers engage in this kind of 'planned adaptation' [Easterling et al., 2007, p.294], however, they will want to anticipate which adaptations will occur autonomously and what consequences they will have in order to develop targeted policies to assist the adaptation process and avoid potentially harmful side-effects. Modeling farmers' reaction to climatic changes and climate-related policies will therefore be the focus of analysis in the following.

From an economic perspective, the climate change-induced alterations to the biophysical and socioeconomic environment of agricultural producers will affect all dimensions of agricultural production:

(1) the *range* of economic activities that can potentially be pursued
(2) the *revenue* that can be expected from these economic activities
(3) the *cost* that has to be incurred for these economic activities
(4) the *risk* associated with both, revenue and cost of economic activities

If – as suspected – effects differ between different economic activities, climate change will alter the relative preferability of agricultural activities. Agricultural producers will adapt their production choices such that the negative effects on their well-being are minimized or opportunities offered by changing conditions can be exploited. The simplest reactions are changes in crop management of already cultivated crops: This may include the choice of adapted varieties and shifts in the timing of sowing, harvesting, fertilization, or pest management, but also an intensification or reduction of fertilizer and pesticide application. Similarly the choice of breeds, grazing times and intensity, stocking rates and pasture rotation can be adapted for livestock management. Increased water scarcity, soil degradation and changes in pest and disease distributions may require the introduction, intensification or improvement of special field management practices like water 'harvesting' and irrigation management, conservation tillage and other measures of soil conservation, integrated pest management and changes in crop rotation [Smit and Skinner, 2002; Olesen and Bindi, 2002; Rosenzweig and Tubiello, 2007; Easterling et al., 2007]. Smit and Skinner [2002] even consider deliberate changes to land topography.

In other cases, farms may change the choice of crops or animals potentially even leading to a complete change in the farming system or at least to a specialization or diversification of production activities. Farmers may try to reduce production risk by informing decisions based on seasonal weather forecasts and relying on crop insurances, forwards and futures. Climate change may also change the relative preferability of agricultural activities compared to nonagricultural activities: Farmers may change their involvement in non-agricultural economic activities on and off their farm trying to diversify the sources of household income, and in some cases give up farming completely [Smit and Skinner, 2002; Olesen and Bindi, 2002; Rosenzweig and Tubiello, 2007; Easterling et al., 2007].

1.4 Consequences of adaptation: outlining research questions

The consequences of adaptation may be multiple and not only pertain to the agricultural and food sector alone. Assessing the effects on food production and farm incomes is an important motivation for the analysis of climate change adaptation, but also climatologists and ecologists are interested in the future development of agriculture: Agriculture contributes to climate change by greenhouse gas emissions and land cover change, but it also offers a huge potential for mitigation [Smith et al., 2008]. Climate change adaptation therefore has the potential to positively or negatively feed back on climatic development. Intensity of agricultural production and land cover changes affect climate, biodiversity, hydrological processes and even public health on local and global scales [Foley et al., 2005].

The research interests guiding simulation analysis of climate change adaptation in agriculture can be subsumed under three headlines:[1]

Predicting production, resource use and structure of the agricultural sector

For the evaluation of food security, food trade and economic development, it is important to know how agricultural supply as well as input and factor demand will develop in the future. Future agricultural land use, greenhouse gas emissions and production intensity may have significant impacts on climate change projections and other environmental assessments. Anticipating the future structure of the agricultural sector is the basis for designing farm support and agricultural development policies. In all of these cases, the research question is a variant of: 'What will it be like?' 'What will be different in the future?'

Assessing the adaptive capacity of farms and the resilience of adaptation strategies

In light of the uncertainty involved in climate change projections, it is important to assess the livelihoods and business strategies of farmers under a wide range of potential futures [Pielke et al., 2007]. As an example, one might ask whether the current coping strategies

[1] This classification was developed for Berger and Troost [2014].

of farms to survive a drought will also be feasible at an increased frequency of dry years. Authorities, donors, banks or insurance companies could then better prepare for such cases. These questions are of the type: 'What happens to ..., if ...?'

Ex ante policy analysis

Policy makers will want to know whether a proposed policy can improve the adaptive capacity of farms to adapt, whether it can mitigate adverse effects, or whether it would even aggravate problems. This includes policies specifically created for climate change adaptation or mitigation, but also other policies affecting the agricultural sector. Their questions admit to the form: 'What effect will ... have on ...?'

Chapter 2

A survey of the use of agent-based models in agricultural economics

Balbi and Giupponi [2009], Patt and Siebenhüner [2005], and Moss et al. [2001] have suggested agent-based models (ABM) as important tools for the analysis of agricultural adaptation to climate change. They could complement other modeling tools available to agricultural economics like Ricardian analysis, computable general equilibrium (CGE) and partial equilibrium (PE) models, regional farm and representative farm models. All of these modeling approaches have their individual strengths and weaknesses:

Ricardian analysis allows a quick analysis of potential economic effects based on scarce data. It completely abstracts from effects on individual production activities and links the economic production potential of a region to its climatic characteristics based on the observation of currently existing variation in climate and agricultural production between regions. Combined with climate projections, it gives a first overview which regions might rather benefit and which might be adversely affected by climate change, without providing details as to how the structure of the agricultural sector might change [Lippert et al., 2009].

CGE and *PE* models can be used to simulate how demand and supply are going to be balanced and trade flows are going to adapt given changes in supply and demand functions for agricultural products. They depend, however, on previous analysis of the potential changes in agricultural productivity and supply behavior [Eboli et al., 2010]. MP-based *regional farm models* [Buysse et al., 2007] contribute by constraining supply behavior to regional land, feedstock and organic fertilizer balances, but are mostly based on empirically estimated supply functions that seldom capture the potential structural changes entailed by climate change or allow the inclusion of new crops, previously not grown in the area.

Apart from the aggregation bias entailed by lumping all farms in an area, these types of models are not suited to provide answers to questions which require the analysis of decision making and outcomes at the individual farm level, e.g. when it comes to assessing participation in agri-environmental measures, technology adoption and learning processes. These are prerequisites for the estimation of area-specific adjustment costs and the valuation of climate impact scenarios at more localized levels, an objective which has been identified as a top research priority by Wreford et al. [2010].

> "The real issue in the coming decades will be the rate and nature of climate change compared to the adaptation capacity of farmers. If future changes are relatively smooth, farmers may successfully adapt to changing climates in the coming decades by applying a variety of agronomic techniques that already work well under current climates." [Rosenzweig and Tubiello, 2007, p. 860]

Representative farm models are a suitable tool for this kind of analysis since they allow for an assessment of constraints and adaptation behavior at a very detailed level [Janssen and van Ittersum, 2007], but their results cannot easily be scaled up to the regional level and take account of the interactions between farms. *Agent-based models* try to bridge this gap (figure 2.1).

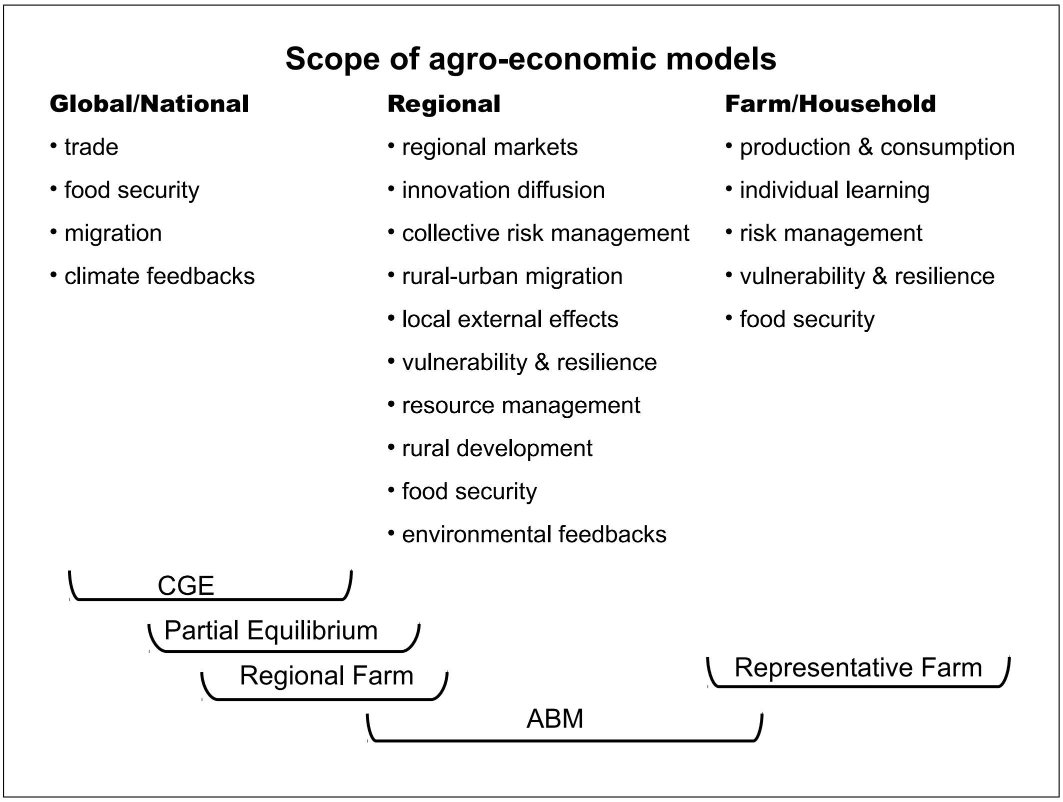

Figure 2.1: *Portofolio of agricultural economic models for climate change adaptation analysis (an adapted version has been published in Berger and Troost 2014)*

They can be as detailed as representative farm models in representing the opportunities and constraints shaping the decisions of individual farms. By incorporating a farm model for each farm of the region, they upscale individual results to the regional level and enhance the analysis by an explicit incorporation of local resource markets, cooperation, social learning and other agent-agent interactions [Berger et al., 2006]. ABM focus on the regional level and do usually not incorporate the reactions of world or even only national markets on climate change-induced shifts in production.

Agent-based models allow an in-depth analysis of processes determining the extent and speed of adaptation that may escape other simulation approaches. This includes, for example, micro-level constraints to adaptation like economies of scale, hysteresis and indivisibility of assets, the explicit simulation of individual and social learning processes to deal with new environmental conditions, direct interaction between agents and the environment as well as interactions among agents. Simulation outcomes can be assessed taking an aggregate and a disaggregate perspective at the same time: At the aggregate level, the modeler can observe changes in agricultural supply of goods, the demand of resources and changes of land use and land cover in the region. These aggregate outcomes are essential inputs in economic analysis of global or national food supply, but also important for the biophysical analysis of feedbacks in climate, ecosystems and resulting effects on the provision of environmental services. At the disaggregate level, the modeler can look at the individual farms and their existing adaptation strategies, and identify those households which are especially vulnerable. Both perspectives are necessary to evaluate different policies intended to assist adaptation or mitigate adverse outcomes of adaptation on ecosystems with respect to effectiveness, efficiency and equity.

While expectations are high, so far the use of agent-based models for climate change has been rather limited, which may be attributed to the fact that climate change research only recently shifted focus towards adaptation. On the other hand, there is a large diversity of agent-based models which emphasize very different aspects and usually focus only on a limited set of opportunities offered by the multi-agent structure. Some explore new paradigms of agent behavior or agent-agent interactions in rather theoretical frameworks, while others focus on empirical research that mainly differs in the representation of heterogeneity from more traditional economic modelling approaches. The following section identifies properties which are necessary or at least desirable in order to make an agent-based model fruitful for the analysis of climate change adaptation in agriculture. The main part of this chapter then surveys studies that use agent-based models in the field of land use science and agricultural economics, summarizes the experience that has already been gathered and identifies the gaps that have to be focused on, so that agent-based models can provide valuable insights into climate change adaptation.

2.1 Identifying desirable model properties for adaptation analysis

Prediction of future adaptation behavior cannot be based on mere extrapolation of behavior observed in the past. It requires a good theoretical understanding of the relevant processes that helps distinguishing their invariant structural parameters from those that could potentially change under future conditions. Models of adaptive farm behavior can therefore not rely on purely empirical-statistical parameterizations of agent behavior, but need to incorporate a theoretic understanding of economic decision making and its relation to the biophysical and socioeconomic environment of the farm [Lambin et al., 2000]. Such a sound theoretic foundation will also help to transfer results from locally confined case studies to more generalizable statements.

2.1.1 A conceptual model of the adaptation process

Before farmers take adaptation decisions, they will first have to notice that something has changed or will change in the future and they will want to understand these changes before taking action. Even after changes have been recognized, actions will not necessarily occur immediately, but often be incorporated into the regular decision schedule: Similar to other economic supply behavior, agricultural producer decisions need to take into account that the use of factors and resources cannot be changed freely from one moment to the next, and that factors have to be considered fixed for different time spans, giving rise to the distinction between short and long-run supply behavior [Nerlove, 1958].

In agriculture, short-term decisions are related to the management of production without major system changes, while long-term decisions determine the production system that is used and are usually associated with major investments into machinery and buildings [Antle and Capalbo, 2010; Olesen and Bindi, 2002]. As their name suggests, short-term decisions can be revised more frequently, e.g. planting decision at the start of each season, irrigation decisions every day, while long-term decisions usually bind farmers for a longer period and constrain the range of options they can choose from when making short-term decisions. Of course, in reality, the spectrum of frequency and transformative impact of decisions is continuous rather than dichotomous and the impact of certain types of decisions varies between different agricultural systems: While a change in crop choice – e.g. from corn to wheat production – constitutes a change of production system in the very specialized farms of the U.S. corn belt, giving up corn production on a diversified European farm in order to increase wheat area may not trigger significant changes in the production system and farm set-up at all.

For a first conceptualization, the process of adaptation can be divided into three phases: First, the new situation and potential adaptation options have to be recognized and understood by the farmer (*adaptation of knowledge*). Second, production decisions that are easily adjusted can be immediately adapted to a new optimum given the new situation and the old asset base (*short-term adaptation*). Third, the farm infrastructure and machinery is adapted as soon as it is possible and beneficial (*long-term adaptation*). Reasons for a delay of farm set-up adaptations are e.g. sunk costs, lack of liquidity, uncertainty about future conditions, expected decline of investment cost in the future and so on. In reality, climate change is not one discrete event and adaptation will not occur linearly, but iteratively, because farmers will learn over time and their actions may feed back into the system and trigger further changes in the biophysical and socioeconomic environment of the farm. As a consequence, the three phases will occur repeatedly and concurrently. Especially decisions with longer-term perspectives have to be taken subject to considerable uncertainty about the future development of natural and economic conditions. This involves the potential of mistake. Future decisions will have to be taken subject to boundaries set by previous actions, even if these turn out to be suboptimal in hindsight [Day, 2008]. As a consequence, producers will try to select "strategies that, over a range of likely future climate and socio-economic scenarios, minimize the potential negative impacts of climate change while maximizing opportunities for adjustment" [Rosenzweig and Tubiello, 2007, p.860].

2.1.2 Desirable model features for adaptation analysis

Based on this conceptualization of the adaptation process, the following desirable properties of adaptation models can be derived:

Time dimension When trying to simulate the adaptation of farming systems to climate change, one can distinguish two extreme perspectives of dealing with the time dimension of adaptation:

Comparative-static approaches ignore the time dimension of adaptation: Agents freely choose an optimal production strategy and farm setup for the respective situation, assuming full knowledge and no liquidity constraints and sunk costs. This amounts to a comparison of an assumed current equilibrium with a new predicted equilibrium in the future, once adaptation has terminated. This perspective facilitates predictions of aggregate agricultural land use and production in a more distant future by avoiding the need to explicitly model the trajectory of change and requiring a less precise observation of the current situation.

Dynamic approaches, on the other hand, explicitly simulate a gradual change in climate and a lagging gradual process of learning, short-term and long-term adaptation. This kind of approach allows an explicit evaluation of the speed and inertia in the process of adaptation. It is better suited to assess short and mid-term changes, to highlight the constraints to adaptation and evaluate the resilience of farms and farming systems. Furthermore, sunk costs, unexpected extreme events and the uncertainty involved in climate change may create path dependencies and lead to different long-run outcomes compared to the idealized equilibrium situation of the comparative-static perspective [Balmann et al., 1996]. While agent-based models can also be used in comparative-static approaches, being capable of fully dynamic simulations certainly adds to their attractiveness for adaptation analysis.

Learning, knowledge sharing and innovation Simulating the speed of autonomous adaptation and understanding path-dependent effects requires understanding the farmers' perception of a change in farming conditions and subsequent learning about the new situation. This includes individual observation, but also the exchange of knowledge among farmers and the information communicated by scientists and consultants [Grothmann and Patt, 2005; Berkhout et al., 2006].

Long-term structural change While the model is expected to adequately represent the current farming system, a dynamic simulation of adaptation processes requires a sound understanding of the processes that shape the structure of farms on the long run. These include e.g. investment decisions, financing and economic stability, land and resource markets, technological progress, demographical development, inheritance, off-farm labor, migration and exit from farming. Additionally, this may also include the interaction of agriculture with other types of land users in order to, for example, capture the effects of urbanization, industrialization, or de- and afforestation [Gaube et al., 2009].

Variability and risk management Antle and Capalbo [2010] reiterate that many adaptation decisions can be framed as investment decisions under uncertainty. Farmers are used to price and weather variability and take precautions to manage the associated risk, and it is likely that individual and collective risk management strategies will also play an important role in the future.

Scope of climate change effects While often climate change effects on crop yields are in the focus of debate, climate change may affect agriculture via several other biophysical and socioeconomic pathways (see section 1.2). Further, policies developed to address climate change effects create incentives or regulations that affect farming decisions. The model needs to contain a clear formulation of the cause-effect chain from exogenous effect to land use or production decisions and outcomes. While not every analysis will require covering the full spectrum of pathways, it is certainly beneficial if the model is capable to incorporate and assess the relevance of a broad spectrum of potential effects.

Feedbacks: integration and coupling The adaptation of farming systems will likely induce changes in land cover and intensity of input use that may feed back into the biophysical system and trigger further environmental changes or even affect the further trajectory of local climatic conditions. Depending on the purpose of analysis, a tight integration with biophysical process models may be necessary [Falloon and Betts, 2010].

Empirical parameterization and data availability Purely theoretical models or models relying only on *ad hoc* parameterizations can sometimes be helpful to discover new emergent phenomena and highlight the importance of certain processes. However, in the case of adaptation analysis, empirical parameterization ensuring an adequate relationship of magnitudes is crucial to examine the resilience and adaptive capacity of farming systems. It is not the mere existence of an effect, but the comparison of its magnitude compared to other relevant effects that the researcher is interested in [Frisch, 1933]. Further, farming systems are specifically adapted to local environmental conditions and very often the constraints and opportunities resulting from these conditions are decisive determinants in shaping the production and land use decisions of farmers. Models need to be able to account for and incorporate these locally specific conditions. Although the out-of-sample nature of climate change assessments sets limits to behavioral validation, confidence in models can usually be improved if they or their components are subjected to empirical testing of their predictive accuracy against suitable real world observations. Both, empirical parameterization and validation entail the requirement that the necessary data are readily available or can be gathered with a reasonable amount of effort.

Technical aspects The assessment of climate change adaptation is associated with a high degree of uncertainty. Not only the future climatic development itself, but also the understanding of relevant processes and empirical information used to construct models is often limited and imprecise. A thorough uncertainty and sensitivity analysis is indispensable in order to arrive at robust conclusions and considered a prerequisite for environmental simulation projects [Jakeman et al., 2006]. Such analysis is facilitated if the analyst can use established and well-tested procedures and draw on previous experience with similar models in similar problem situations. The same holds for empirical parameterization and validation. While this is mainly a question of methodology, conceptual and technical characteristics of a model can encourage or discourage sensitivity analysis. Here, especially data requirements, computation time and portability to high performance computing facilities play a role. Also the integration with other models (e.g. crop growth, hydrological, meteorological and soil process models) is considerably easier if the model already provides technical interfaces for data exchange.

2.2 Agent-based models of agricultural and land use economics

Agent-based modeling has been used in a variety of applications in agricultural economics and the neighboring field of land use sciences for about 15 years. This section will provide an overview of the existing studies and survey them with respect to the requirements formulated in the previous sections. The intention is not to provide a complete account of all studies that have been conducted in the field. Rather, the idea is to depict the experiences that have been gathered with agent-based models so far and highlight the gaps and challenges to be addressed in order to use ABM for assessing climate adaptation policies. This section mainly covers studies that have an applied empirical focus and only occasionally mentions pure "thought experiments" if these follow an interesting approach that might prove useful for empirical applications in the future. Other reviews of multi-agent models in agriculture and land use sciences can be found in Kaye-Blake et al. [2010], Matthews et al. [2007], Robinson et al. [2007].

2.2.1 Applications of ABM

Multi-agent models of agricultural systems have been applied to a wide range of research questions in both industrial and developing countries. The agricultural sectors that have been studied with ABM span from commercial and highly mechanized farming in Europe or North America [e.g. Happe et al., 2008, 2009; Freeman et al., 2009] to subsistence farming in Africa or South East Asia [e.g. Castella, Boissau, Trung and Quang, 2005; Castella, Trung and Boissau, 2005; Matthews, 2006; Bithell and Brasington, 2009; Schreinemachers et al., 2007] including very heterogeneous settings where large-scale, export oriented farms operate side-by-side with small-scale family farms, like in South Central Chile [Berger, 2001; Berger et al., 2007]. A recurring feature is the analysis of settings which are characterized by the dichotomy of several rather sharply distinguished groups like e.g. forest managers and pastoralists [Simon and Etienne, 2010]. The spatial extent of study areas is generally local to regional; it may range from a small village to a large watershed. Time spans covered vary widely, from a few years to the extreme case of several hundred years in the investigation of historical settlement patterns [Reynolds et al., 2003]. With respect to their field of analysis, the models discussed in the following can be subsumed under three headlines: the analysis of agricultural policies, natural resource management, and urbanization.

Agricultural policy The disaggregated nature of ABM allows differentiating benefits and costs of policy interventions by groups of farmers and estimating effects upon the structural set-up of the agricultural sector. For example, Berger [2001] studied the dynamic effects of Chile's association to the MERCOSUR and analyses the likely impacts on resource use, technology adoption, and structural change. Happe et al. [2008, 2009] used their AgriPoliS model to assess the effects of agricultural policies upon structural change in Europe. A similar assessment was done by Lobianco and Esposti [2010] for an Italian study region. Freeman et al. [2009] examined drivers of structural change in Saskatchewan, Canada, over 40 years.

Especially in the context of development policy, ABM can help identifying the constraints to agricultural development: Schreinemachers et al. [2007] examined the origin of the gap between potential and average yields in Uganda and showed that this gap results from the individual optimization behavior of small-holder farmers making the best out of their constrained situations. Milner-Gulland et al. [2006] identified winter forage availability as a key determinant of livestock numbers in Kazakhstan.

On the other hand, many policy measures are well-intended but not necessarily followed by individual farms. ABM can be useful in modeling innovation diffusion and adoption of policy and technology. Schreinemachers et al. [2009] examined the diffusion of greenhouses in Northern Thailand. Ziervogel et al. [2005] and Bharwani et al. [2005] assessed adoption of seasonal weather forecasts among farms in Lesotho and South Africa and the benefits and risks associated with relying upon fallible projections. Saqalli et al. [2011] analyzed participation in development schemes promoting fertilizer use in Niger.

Natural resource management Also in the context of environmental policy assessments, ABM have been employed to analyze adoption of agro-environmental policy measures [Weisbuch and Boudjema, 1999], organic farming [Deffuant et al., 2005], green manure [Schreinemachers et al., 2007] or other soil conservation incentives [Sengupta et al., 2005]. A more important focus, however, has been to (i) explore complex human-environment interactions in order to identify drivers for environmental degradation, (ii) explore the sustainability of individual and collective resource management strategies, or (iii) explain the emergence of collective management regimes [see Gotts et al., 2003, for an overview of the use of ABM in social dilemma situations].

Bithell and Brasington [2009], Le et al. [2008], An et al. [2005], Manson [2005], Castella, Boissau, Trung and Quang [2005], Castella, Trung and Boissau [2005] as well as Huigen [2004] analyzed population dynamics and other drivers that lead to forest degradation or deforestation. BenDor et al. [2009] analyzed the economic and ecological sustainability of individual fishing strategies, while Janssen and Ostrom [2006] explored the role of mutual trust in adopting and sustaining resource exploitation rules and related this to fishery. Gross et al. [2006] and Janssen et al. [2000] examined the role of individual learning and threshold-driven policy adaptations for rangeland management, respectively. Matthews [2006] analyzed the exchange of farm-yard manure in a subsistence community in Nepal. Jepsen et al. [2006] examined spatial patterns of shifting cultivation.

Collective resource management and interactions between agents via the natural resource system is also an important feature of irrigation agriculture: Lansing and Kremer [1993] explained the balanced rice planting schedule in Bali as an optimization in the trade-off between water allocation and pest prevention (see also the generalization by Janssen 2007). Becu et al. [2003] investigated relationships between upstream and downstream farmers testing different cooperative or conflict-laden scenarios. Van Oel et al. [2010] reproduced land-use patterns in a Brazilian watershed. Berger [2001] included watershed dynamics into his assessment of MERCOSUR policies, while Berger et al. [2007], Troost et al. [2010], and

Latynskiy et al. [2010] undertook policy analysis with ABM in multi-level water governance structures. Izquierdo et al. [2003] analyzed the spatial dimension of eutrophication caused by agricultural land users.

In companion modeling approaches, ABM have been employed as a communication tool in order to help stakeholders develop a systems perspective, realize the consequences of their actions and foster negotiations about collective resource use arrangements [Bousquet et al., 1999; Barreteau et al., 2001; Etienne, 2003; Campo et al., 2009].

Urbanization In the wider field of land use science, ABM have been used to assess processes of urbanization and urban sprawl [e.g. Brown et al., 2004; Loibl and Toetzer, 2003]. Many of these studies focused explicitly on the competition between agricultural and urban land uses [Gaube et al., 2009; Filatova et al., 2009; Guzy et al., 2008; Parker and Meretsky, 2004; Gotts and Polhill, 2009]. Some of these studies replicated the observed historical development of land use change [Hoffmann et al., 2002; Evans and Kelley, 2004; Deadman et al., 2004] including the settlement patterns of ancient cultures over several hundred years [Sanders et al., 1997; Dean et al., 2000; Kohler et al., 2000; Reynolds et al., 2003].

2.2.2 Model designs and theoretical foundations

To a certain degree, all agent-based models are process-based models: instead of simulating aggregates, they explain aggregate outcomes as the consequence of the interplay of actions of individuals, the agents. Each computational agent can be characterized by a number of properties and processes. Properties comprise variant and invariant attributes associated with an agent, for example, age, education, location of its farmstead, ownership of assets, knowledge and expectations about the environment. Models differ in number and kind of properties, and whether these are invariant, i.e. fixed at initialization, or whether they can change over time. Changes of properties over time are determined by various processes which can be subdivided into internal processes (e.g. the aging of household members) and the activities of the agent during the simulation (i.e. the agent behavior in a narrower sense). Furthermore, there may be influences on properties that are exogenous to the system, like rainfall or prices.

Besides the agents, the second major component of nearly all agent-based models of agricultural economics and land use sciences is a landscape, with which and within which the agents interact. The landscape fulfills several roles for agents: (i) it defines a topology between agents; (ii) it influences the available decision options of the agent; and (iii) it consists of plots, which agents may exchange with each other on land markets.

In most cases, the landscape is implemented as a raster grid consisting of equally sized cells or pixels. The state of a landscape unit can be represented by a broad land cover class (forest, grassland), or by a whole array of variables like slope, soil moisture, nitrogen content, leaf area index and so on. This state can either be constant throughout a simulation run, or it can change as a consequence of the agents' actions or biophysical processes. Some models dedicated to watershed management additionally include node-link networks of water flows that are topological in nature, without full spatial extension [van Oel et al., 2010; Becu et al., 2003; Berger et al., 2007]. Agents are usually linked to a part of a landscape by ownership or user rights that define the sphere of influence of one agent. Few models implement free-roaming agents that can abandon land and encroach upon unused land [Le et al., 2008], wander over the landscape to collect fuel wood [An et al., 2005] or find a place to settle [Huigen, 2004; Loibl and Toetzer, 2003] .

The distinction between process-based and purely empirical models discussed in section 2.1 mainly refers to the conceptualization of the behavior of individual agents, and the determination of outcomes generated by interactions among agents and between agents and the environment, and – in case the ABM includes the representation of biophysical processes – also extends to the conceptualization of processes in the landscape.

Agent decision-making

The crucial choice to be made by the modeler is the conceptualization of the decision-making process that agents use to choose between their available options. One of the prevailing characterizations of agent decision-making has been the distinction between rule-based decisions ("heuristics") and optimization decisions. As has been noted by Schreinemachers and Berger [2006] and Haas and Jaeger [2005], however, there is a continuum between heuristics and optimization. Often the same conceptualization of agent decision-making can be formulated as a decision tree or as a mathematical programming problem. Further, many heuristic decision strategies employ rules that are implicitly optimizing and many mathematical programming based models also employ heuristics for certain decisions. For our purpose, it is useful to distinguish decision algorithms according to whether they are primarily based on observation or theory, deterministic or stochastic, and, especially, whether they allow agents to depart from observed behavior under new circumstances, in extreme case even in ways that may not be foreseen by the modeler.

On the empirical side of the spectrum, there are deterministic decision trees, which are usually specified after observing or explicitly asking real-world actors about their decision-making procedures. These decision trees are often used in small-holder contexts where resource constraints or cultural preferences determine land uses. For example, Castella, Boissau, Trung and Quang [2005] and Castella, Trung and Boissau [2005] report that upland fields in their Vietnam study area are only used for cash crops if enough area has been allocated to rice, in order to cover the subsistence requirements of the household. Van Oel et al. [2010] observed that a group of farmers in Brazil plants 50% of their irrigated area with rice if they expect low water supply, and 60% if they expect normal conditions. Ziervogel et al. [2005] note that farmers tend to ignore bad forecasts, while taking good forecasts into account. Saqalli et al. [2011] include a rich set of social rules and economic interrelationships between village members that determine individual behavior. Empirical decision rules of this kind are then usually transformed into a decision tree with several branches. As an alternative, rules can be formulated as an empirical function estimated from statistical data. For example, An et al. [2005] estimate the yearly fuel wood demand of a household as a linear function of household size, the presence or absence of a senior person, and the area cultivated with corn and potato.

Instead of using deterministic rules, decisions may also be made stochastically, based on observed probabilities. In the study of of An et al. [2005], young unmarried female adults out-migrate from the study area because of marriage at a probability of 0.28%. In the simulations of Freeman et al. [2009], farmers above 50 years of age have a certain probability to retire. While in these cases the probability is independent of any other factor, it can of course also be calculated as a function of agent properties and environmental status. For example, An et al. [2005] estimate the probability to switch from fuel wood to electricity as a function of the reliability of electricity supply, distance to nearest fuel wood supply, age and education of household head. Bayesian belief networks like the ones used by Lei et al. [2005] also fall into this class of empirical stochastic approaches. The probability of a certain decision outcome is estimated conditional on agent and environmental characteristics using training datasets. Valbuena et al. [2010], for example, use empirical data to estimate a probability to buy, keep or sell land, which depends on farm type and past history of land acquisition.

Theory-based decision algorithms can usually be associated with the paradigm of recursive dynamic programming of Day [2008] starting from a theoretical assumption about the intentions or goals of agents, and using a choice algorithm that ensures some gradual achievement or optimization towards these goals.[1]

A natural choice for economists is to assume utility maximizing behavior. In the simplest case, this consists in (i) choosing the alternative that (expectedly) provides higher utility when comparing two alternatives (e.g. Parker and Meretsky 2004, or also the decision to participate in the fertilizer scheme in Saqalli et al. 2011); or (ii) maximizing a classical microeconomic one-input-one-output production function [BenDor et al., 2009].

Production decisions of farm households, however, are often more complex as land resources and capital assets have to be allocated among different products. AgriPoliS and MPMAS models, as well as the models of Lobianco and Esposti [2010] and Freeman et al. [2009] follow the whole-farm planning approach of agricultural economics. They use constrained optimization based on mathematical programming techniques to find optimal production and investment plans, while being able to consider a large number of constraints. These constraints may not only be resource or technical constraints, but may also include additional goals reflecting, for example, the priority of household food requirements through subsistence production before maximizing cash income. Freeman et al. [2009] use expected income as a utility function, including risk considerations into their model. Schreinemachers et al. [2007] include an empirically estimated consumption function into their three-stage agent decision-making, determining the allocation of income among savings, nonfood expenditure and a number of food categories. The optimization at agent level then simultaneously determines the agent's income, consumption pattern, production plan, as well as purchase and self-consumption of food.

Alternative algorithms to mathematical optimization are *prioritization* and *satisficing*. Le et al. [2008], Matthews [2006], and Deadman et al. [2004] use prioritization algorithms that rank production activities by utility and then select activities starting with the highest rank until a certain resource (land or labor) has been exhausted. Compared to full mathematical optimization, this neglects optimal combinations and opportunity costs between different activities.

Satisficing means that the search for a better alternative is stopped once a satisfactory level of the target variable has been reached or initiated once this aspiration level is not reached anymore by the current choice [Gigerenzer and Goldstein, 1996]. For example, agents may start searching for a new economic strategy if income falls below a minimum threshold [Gotts and Polhill, 2009; Gaube et al., 2009] or an economic performance measure evaluates inferior in comparison to a certain peer group [Gross et al., 2006]. Holtz and Pahl-Wostl [2012] use a utility function with weighted multiple objectives and implement a random variation and filtering algorithm that varies the agent's current land use and keeps the variation with the highest utility value.

As said before, there is a continuum between theory-based and empirical approaches, and one model may use more empirically based approaches for one decision and theory-based ones for others [e.g. Saqalli et al., 2011], or even mix them for a single decision. (Becu et al. 2003 use linear programming at certain nodes in their decision tree, namely selecting a cash crop, once branching along the decision tree has resulted in the decision that a cash crop and not rice shall be grown.) In general, theory-based models that are used in empirical contexts are parameterized with empirically observed parameters where available, and often they

[1]A more detailed summary of the concept of recursive dynamic programming will be given in section 4.1.2.

are also calibrated, i.e. uncertain parameters are chosen such that the resulting behavior resembles the one observed in reality [see Berger and Schreinemachers, 2006].

Theory-based approaches can quite easily be adapted to out-of-sample situations by adapting parameters and exogenous variables accordingly without actually changing the basic model structure. Purely empirical approaches would require some form of experimentation to extend the observation sample. Compared to aggregate models, agent-based models have the advantage that they facilitate such experimentation, because experiments do not have to be performed with the whole system, but the researcher can observe individual behavior under different scenarios in role-playing games and companion modeling approaches [see e.g. Simon and Etienne, 2010], leaving it to the model to determine system-wide outcomes.

Learning, innovation and knowledge sharing

Generally, agent-based models explicitly consider the limited knowledge of agents about their environment. Empirically observed behavioral rules implicitly incorporate the knowledge of real-world decision makers.

Theory-based approaches usually let agents take their decisions based on expectations rather than on actual values of key parameters, often because the realizations of these variables are not known in advance even to the modeler, e.g. when they result from agent interactions (e.g. prices in endogenous markets or the water supply in interconnected irrigation schemes). Expectations can be constant or can be formed based on indicators: Irrigator agents in van Oel et al. [2010] estimate dry season water availability based on rainfall during the wet season and reservoir filling. Ziervogel et al. [2005] let their agents base production decisions on weather forecasts once they have established trust in them. Trust is generated and lost by observing the performance of the forecast. Manson [2005] lets each agent perform a symbolic regression based on genetic programming in order to derive a crop suitability function for each land cell. A common indicator is the observation of a variable in previous time steps: For example, pastoralists of Gross et al. [2006] adapt their expectation of savannah carrying capacity based on last year's feed utilization. AgriPoliS and MPMAS agents use adaptive expectations based on weighted arithmetic or geometric means of previous expectation and actually obtained results. While in most cases agents adapt their expectations continuously, Janssen and de Vries [1998] require a certain threshold of deviation between previously expected and observed values for learning to be triggered.

Apart from this type of individual learning by updating expectations based on observed indicators, agents may also learn by imitating or observing strategies of other agents. Berger [2001], Schreinemachers et al. [2009] and Schreinemachers et al. [2010] use the threshold model of innovation diffusion [Granovetter, 1978], classifying agents by their innovativeness. An agent considers adopting an innovation only if the proportion of peers that already adopted is larger than a certain threshold that is determined by the agent's degree of innovativeness. Whether agents eventually adopt the innovation depends on their personal evaluation of its benefits by investment calculus. Deffuant et al. [2005] and Kaufmann et al. [2009] use a more detailed approach that explicitly represents opinion formation, information transmission in social networks and takes into account agent interest, knowledge and a priori convictions about the social value of an innovation. In all of these cases, models focus on discrete, predefined innovations that merely have to diffuse.

Models by Gotts and Polhill [2009] and Gross et al. [2006], in contrast, allow for the generation of innovations within the model and for its subsequent diffusion through the agent population. Both models work under the satisficing paradigm with relatively abstract, predefined decision strategies. Agents that are not satisfied with their current strategy either randomly select a new strategy, which may result in an innovation, or they imitate a successful strategy of their peers. While the peer group or information network is statically defined in most cases, Izquierdo et al. [2003] include social approval as a function of another agent's actions and allow for changes in the composition of peer groups.

Heterogeneity and long-term structural change

The strength of ABM is to allow for the representation of disaggregated and heterogeneous behavior within a population of land users. On the one hand, this can be reached by including different types of agents: Agricultural economic ABM usually feature computational agents that represent farm holdings, farm households or individual household members, but often also other kinds of land users, companies, governments, marketing organizations, irrigation water managers etc. have been represented by agents. Even institutions that are not actors in the narrower sense can be modeled as agents: e.g. a market can be represented by an auctioneer agent [e.g. Freeman et al., 2009].

On the other hand, sources of heterogeneity among agents of the same type include measurable agent properties like resource endowments constraining production options, or household size determining demand for food and other goods. More subtle differences rooted in personality or cultural traits can be represented in utility functions [Loibl and Toetzer, 2003; Holtz and Pahl-Wostl, 2012], risk aversion coefficients [Freeman et al., 2009], aspiration levels [Gotts and Polhill, 2009; Gross et al., 2006], personal opinions on social value [Deffuant et al., 2005], or expectations [Berger, 2001].

Other modelers classify their agents into groups and specify separate decision rules or use different parameters and probabilities [Huigen, 2004; van Oel et al., 2010; Janssen and de Vries, 1998; Valbuena et al., 2010].

The scope of actions an agent can pursue varies widely between different models. For example, in Parker and Meretsky's [2004] model agents have only two options: they can either opt for an agricultural or an urban land use for one single parcel they own. Evans and Kelley [2004] let their agents choose between four types of land use (farming, haying, timber harvesting, nonextractivist land use). Agents representing fishermen in BenDor et al. [2009] catch only fish, and agents representing pastoralists in Gross et al. [2006] just raise cattle, but both types of agents can determine the size of fishing or livestock-herding efforts. Agents in AgriPoliS [Happe et al., 2008, 2009] and MPMAS models [Schreinemachers and Berger, 2011], in contrast, have the choice between all kinds of different agricultural production activities, purchase of investment goods, giving-up farming, renting land and so on.

While a greater scope of actions does increase heterogeneity *between* agents, it does also increase the potential heterogeneity of the decisions of one agent over the course of its lifetime, the potential adaptation options that can be considered and the extent of structural change and land use change the model is able to produce. Simulation of long-term adaptation requires the reflection of processes that fundamentally change the structure of the farm or household. Many agent-based models include such dynamic processes. Depending on their purpose, they lay more emphasis on demographical developments, migration, investment or land markets.

Demography and migration An et al. [2005], Huigen [2004] and Saqalli et al. [2011] employ sophisticated models of household demography, where marrying, household formation, migration and reproduction are the result of explicit agent decisions and agent-agent interaction. MPMAS models, e.g. Schreinemachers et al. [2007], use statistical propensities to give birth or die to change household composition over time. In other cases, household composition is irrelevant: Only the age of the manager evolves over time and triggers bequeathing of the farm to the next generation or farm exit in case the heir does not find it profitable to take over the farm based on some economic indicator [Happe et al., 2008; Freeman et al., 2009]. In Freeman et al.'s model also the general behavior of the agent is related to the age of the farm manager. Farmers close to retirement lay more emphasis on securing equity and avoid risky transactions on the land market. Like other decisions, out-migration from farming can be either based on expected utility maximization, e.g. higher off-farm wages [Berger, 2001] or statistical probabilities [An et al., 2005]. Bankruptcy usually leads to forced exits.

Investments MP-based multi-agent systems usually integrate investment options into the set of potential farm actions, representing investment decisions according to the net present value rule (NPV). Recently modelers have started to incorporate real options theory into agricultural agent-based models in order to better capture the effect of uncertainty and waiting for improved information before investing [Feil et al., 2011].

Endogenous (land) markets Most ABM employ the small-region assumption and set market prices exogenously. However, for certain regionally traded factors and products this assumption cannot be upheld. In the case that agents are only suppliers of a certain good, the market price can be determined using a specified demand curve, once the supply of all agents has been computed internally [e.g. Parker and Meretsky, 2004]. In this case, agents do not interact explicitly, but implicitly. Exogenous prices have also been used with explicit interactions: In Matthews [2006], model agents sell farmyard-manure in equal proportions to all requesting agents at a fixed price. Agents of Lobianco and Esposti [2010] rent land to or from an anonymous land-owner agent at a fixed price.

Some models in agricultural economics [Balmann, 1997; Happe et al., 2008, 2009; Freeman et al., 2009] use an auctioneer agent to represent land rental markets. Potential buyers estimate a bid value for a certain plot based on the expected economic rent of a plot, while the seller estimates a minimum acceptable bid. The auctioneer then awards each plot to the highest bidder iteratively until no bid surpasses the minimum acceptable bid value for any plot that has not been rented yet. The price paid by the bidder is not his bid, but adjusted toward the average rental price paid in the region.

Tesfatsion [2006] illustrates the challenges and the gaps in economic theory encountered when trying to construct a working market without a fictitious auctioneer. Few multi-agent models devoted to agriculture or land-use change have started this endeavor. Filatova et al. [2009] implement a bilateral land market, albeit in a relatively abstract setting that does not require meeting all the challenges described by Tesfatsion. Again, buyers determine their willingness-to-pay according to their utility function and make a bid to the seller of the plot, which maximizes their utility. The seller determines both its willingness-to-accept (equal to its opportunity cost) and its asking price, which is estimated according to the agent's own utility function. The seller sells to the highest bid that exceeds the asking price. If the agent does not find a buyer after a certain number of iterations, the asking price will be reduced by say 3%, but never below the agent's willingness-to-accept. Berger [2001] implements a similar algorithm of bilateral trade between potential buyers and sellers and generates endogenous land-rental prices that match observed prices in Chile.

Risk management

Observed heuristics often implicitly incorporate risk management considerations. In theory-based approaches, most models use expected production outcomes, thus effectively modeling risk-neutral agents. Freeman et al. [2009] are an exception as they explicitly use expected utility theory and risk aversion coefficients to capture the heterogeneity of risk preferences among agents. They simplify their analysis by neglecting covariance between production options, as in their case these are highly positively correlated and thus provide no option for hedging. Holtz and Pahl-Wostl [2012] associate a certain risk level with each land use activity and include risk reduction as one objective into the utility function.

Scope of climate change effects and feedback

Implementing climate change scenarios in the described agent-based models requires changing exogenous input variables and empirical parameters of biophysical process and decision models. The scope of climate change effects that can be considered will depend on whether the associated variables and parameters are taken account of in the model. For some models, a pre-estimation of parameters/variables, e.g. crop yields, will be necessary, other models integrate the relevant biophysical processes and are able to directly work with climatic variables as input.

Integration of biophysical processes Biophysical processes can be implemented by simple rules, e.g. a look-up table defining the yield of a certain activity on a certain soil class or a fixed rate of forest growth if no logging takes place [An et al., 2005]. They can, however, also be simulated by more complex biophysical models taking into account environmental conditions like weather or water inflows and activities realized on neighboring plots. The later allows for the incorporation of indirect interactions between agents via the landscape, often the main type of agent-agent interaction considered in a land use model.

ABM can either use embedded coupling, i.e. the biophysical model is directly coded into the ABM, or external coupling, where both models are linked by appropriate software infrastructure [see discussion in Schreinemachers and Berger, 2011]. Some ABM also integrate biophysical processes using a simplified modeling approach (a kind of meta-model) that uses parameters derived from more sophisticated biophysical models. For example, Gaube et al. [2009] represent the C-N-Cycle using a stock-flow model that has been calibrated with the more complex ecosystem model Biome-BGC [see Petritsch et al., 2007]. The following studies are examples for integrations of ABM with biophysical models and shall give an overview on the range of biophysical processes considered:

Becu et al. [2003]'s CATCHSCAPE model and van Oel et al. [2010]'s ABSTRACT model both use the CATCHCROP approach of Perez et al. [2002] to simulate the water balance of a landscape cell at 10-days time resolution. The water balance is established using cell-specific crop and soil parameters and exogenously provided rainfall and potential evapotranspiration data. The hydraulic module consists of a network of river branches (including alluvial aquifers) from which water is diverted based on agent irrigation decisions. Van Oel et al. [2010] also include an explicit module for the reservoir water balance. All modules are embedded: Both the ABM and the CATCHCROP model were implemented using the CORMAS modeling platform.

Le et al. [2008]'s LUDAS model estimates crop yields using an empirical production function. Forest dynamics are implemented using a system dynamics model based on equations by Vanclay [1994] and Alder and Silva [2000]. Natural land cover transitions follow simple rules. Gross et al. [2006] and Janssen et al. [2000] embed system dynamics models of rangeland ecosystem and livestock herd dynamics, which are based on empirically estimated coefficients, into their models.

Bithell and Brasington [2009] model forest dynamics using an individual-based model (the ecologists' equivalent to a multi-agent model) that allows for explicit competition between trees according to their relative positions in the landscape and estimates tree growth using an enhanced version of SORTIE [Pacala et al., 1996] which is based on empirically derived allometric relations between tree growth parameters and shading. They calculate explicit water flows between landscape cells based on a finite volume model of saturated horizontal flows and evaporation, in- and exfiltration processes in four soil reservoirs.

Matthews [2006] directly codes biophysical processes into his PALM model, borrowing soil water and nitrogen dynamics from the DSSAT model [Ritchie, 1985; Godwin and Jones, 1991], organic matter decomposition routines from CENTURY [Parton et al., 1988], crop growth processes from CERES [Ritchie et al., 1998] and CROPGRO [Boote et al., 1998] and forest tree growth from the GenW model [Kirschbaum, 1999].

Berger [2001]'s MPMAS uses an empirically parameterized water balance model (EDIC) to represent return-flows between irrigation sectors. The yield response to water deficit is calculated using the FAO56 Cropwat model [Allen et al., 1998]. Both modules have been coded directly into the ABM model. The parameterization of the hydrological module has later been improved by calibration using simulation results of the process-based hydrological software package WaSiM-ETH [Schulla and Jasper, 2000], and finally MPMAS has been externally coupled to WaSiM-ETH [Arnold et al., 2010; Arnold, 2010].

In a different set-up, Schreinemachers et al. [2007] integrate the Tropical Soil Productivity Calculator (TSPC: Aune and Lal, 1995) into code of MPMAS, simulating soil fertility dynamics (including N, P, K, acidity, and soil organic matter) and resulting yield responses based on empirically calibrated relationships (see also Quang et al. 2014). In current studies, MPMAS is being coupled externally with two more process-based models of soil nutrient and moisture dynamics: (i) the Land Use Change Impact Assessment (LUCIA) model is a raster-based, spatially-explicit dynamic model that simulates watershed functions, soil fertility and plant growth for small catchment areas [Marohn et al., 2010, 2013] and is used for applications of MPMAS in tropical environments; (ii) In the second part of the present thesis, the Expert-N modeling package is used. Expert-N integrates state-of-the-art crop growth models such as CERES, SPASS, SUCROS, and GECROS into a modeling system that is complemented by modules for soil water flow, soil heat transfer, soil carbon and nitrogen turnover, and other soil-plant-atmosphere processes on a daily basis [Priesack et al., 2001, 2006]. Both LUCIA and EXPERT-N are coupled to MPMAS using a lightweight model coupling framework [Schreinemachers and Berger, 2011].

Collective action, institutions, and policies

In most ABM applications, policy interventions are implemented as settings of exogenous variables or parameters and are not endogenous to the model. A notable exception is the model of global policy response to climate change by Janssen and de Vries [1998], where the implemented policy is determined as a weighted average of the individual policy choices of agents, which determine their individual choices based on their worldview and the experience with previous implementations. Another example is Janssen et al. [2000], where the changes in policy regimes governing a rangeland area are triggered by the surpassing of economic or ecological thresholds. Both studies are examples for modeling institutions as abstract "institutional" agents whose actions are governed by predefined decision rules rather than arising from collective decision-making. The latter approach has been implemented by Becu et al. [2003] , where water allocation in the ABM is negotiated between villages (themselves specified as institutional agents). citetjanssenostrom2006 examine the adoption of fishery management rules through approval by fishermen, where approval depends on trust in other fellow fishermen and the success of the rule.

2.2.3 Model parameterization and empirical data collection

Empirically-oriented ABM draw on a large number of different data sources to develop and parameterize their agents. High data requirements have often been cited as an important drawback of ABM [e.g. Parker et al., 2003; Zimmermann et al., 2009]; on the other hand, ABM provide considerable flexibility in the representation of processes, which can be tailored according to the data that are available.

Robinson et al. [2007] review data collection approaches for empirical multi-agent models and classify them into five categories: (i) sample surveys; (ii) participant observation; (iii) field and laboratory experiments; (iv) companion modeling; and (v) GIS and remotely sensed spatial data. As an extension to this list, it may be useful to add (vi) secondary data sources such as agricultural censuses, farm accounting data, and agronomic data provided by extension services. Though similar to (i) in terms of structure, the researcher has much less influence on the data content and has to deal with the information as is. The information modelers have to obtain can be classified into (i) parameters for the behavioral modules (e.g. decision rules, or coefficients for an MP problem); (ii) data that represent fixed characteristics of agents (e.g. risk aversion coefficients); (iii) data on the initial states of dynamic variables (e.g. dairy cows owned).

Participant observation, experiments and companion modeling are mainly used to elicit rules or coefficients for the behavioral models. The strength of these data collection methods is the possibility to elucidate behavior that (i) is not easily captured by structured interviews; (ii) cannot be inferred from statistical data; or (iii) does not necessarily comply with standard economic assumptions. Sample surveys and secondary statistics are mainly used to parameterize theory-based agent decision modules, especially mathematical programming problems, as these require parameters that can usually be obtained by closed-ended questions.

Two major approaches have been used to generate agent populations in ABM: (i) Happe et al. [2008, 2009] and Lobianco and Esposti [2010] clone typical agents from available survey data – in their case the European FADN –, such that the difference of aggregate statistics of the real-world population and the model population is minimized; (ii) Berger and Schreinemachers [2006], in contrast, estimate probability distribution functions from the survey or census population and use Monte-Carlo sampling to generate statistically consistent agent populations.

2.2.4 Empirical validation, uncertainty and sensitivity analysis

Behavioral validation of agent-based models, if conducted at all, has so far often been restricted to an informal comparison of overall trends in observed and simulated datasets [e.g. Castella, Trung and Boissau, 2005; Lobianco and Esposti, 2010; Gross et al., 2006]. More formal measures of comparison that have been employed are regressions of simulated on observed values [Berger, 2001], R^2 [Evans and Kelley, 2004], Nash-Sutcliffe efficiencies [van Oel et al., 2010] and statistical tests for the comparison of distributions [An et al., 2005; Freeman et al., 2009]. Schreinemachers et al. [2009] and Schreinemachers et al. [2010] explicitly check the goodness-of-fit at different levels of aggregation of agents to groups showing that, in their case, the model overall achieves a statistically satisfactory fit overall, but may show considerable deviations for individual agents.

A specific feature offered by agent-based models is the possibility to conduct interactive validation, i.e. the behavior of agents can be discussed with their real world counterparts in role-playing games and group discussions in order to check the behavioral validity of agent decisions [Castella, Trung and Boissau, 2005; Berger et al., 2010]. In general, most empirically oriented agent-based models are a combination of several submodels, which often provide the opportunity for separate validation of processes [Troost et al., 2010].

While most ABM modelers perform scenario analysis, formal uncertainty and sensitivity analysis on parameters [like described e.g. by Helton et al., 2006] has rarely been used. Exceptions are the use of Monte-Carlo techniques in connection with stochastic submodels [Bithell and Brasington, 2009; Valbuena et al., 2010] or sampling of agent characteristics [Berger and Schreinemachers, 2006; Holtz and Pahl-Wostl, 2012]. As a consequence, it is not surprising that a perceived lack of established formal measures for validation and calibration is one of the frequently cited problems of socioeconomic agent-based models [e.g. Zimmermann et al., 2009]. In the field of socioeconomic agent-based modeling itself, there has been a fierce debate on the correct approaches to validate agent-based models, which mainly centered around the discussion whether ensuring realisticness of assumptions or correct predictions is more important [Brenner and Werker, 2007; Windrum et al., 2007; Moss, 2008; Deichsel and Pyka, 2009].

2.2.5 Examples for the use of agent-based models in climate change analysis

To date, there are only a few ABM studies published that focus explicitly on climate change. One reason is that adaptation to climate change has only recently gained relevance as a major research topic.

The study of Janssen and de Vries [1998] has already been mentioned above. It simulates adaptive policy responses to climate change on the global level, based on agent's convictions and learning about the earth system. Haas and Jaeger [2005] present the LAGOM model, a multi-agent modeling framework that is designed to assess risk management and learning of economic agents about climate change, though it is not specifically focused on agriculture.

Acosta-Michlik and Espaldon [2008] assess the vulnerability of Philippine farmers to changes in crop yields and prices induced by climate change and global economic development using an ABM that is based on a heuristic decision-making module and includes exchange of information, imitation and mutual credit relations as important agent-agent interactions. The work of Bharwani et al. [2005] and Ziervogel et al. [2005] are good examples of policy assessments aimed at facilitating farmers' adaptation to climate change (in this case, seasonal weather forecasts). Hailegiorgis et al. [2010] analyze the vulnerability of pastoralists to rainfall reductions in the Madera triangle in East Africa. They replicate the observed nonlinear relationship between rainfall and carrying capacity. However, their model still abstracts from important risk-coping mechanisms.

Currently, a number of projects are developing integrated model systems targeted at climate change analysis which include agent-based components for socioeconomic submodules: Angus et al. [2009] sketch an ABM for Bangladesh to examine the dynamic impacts of climate change (affecting river-flow, rainfall, temperature and extreme weather events) on poverty, migration, mortality and conflict. They cover the complete 21st century at weekly time resolution and the whole country, though having each computational agent represent not one, but 14,500 persons. Janmaat and Anputhas [2010] report on ongoing work to integrate an agent-based module of land use change into their hydrological model in order to assess climate impacts in the Deep Creek watershed in Canada.

2.3 Discussion and conclusions

The review shows that many of the features required for a dynamic modeling of the process of climate change adaptation have already been used and implemented in agent-based models: Many of the studies reviewed above include variability in meteorological conditions and several are linked to biophysical process models. Incorporating a different climate into ABM usually requires just changing weather-related input variables for the biophysical component. A wide range of agent-decision models has been employed including theory-based, empirically parameterized recursive-dynamic programming models which are able to create short- and long-term adaptation endogenously. Learning by experience, experimentation, imitation and communication has been a focus of several agent-based models, while others have focused on market interactions.

While many of these process representations can certainly be improved, the major challenge that remains at the moment is to combine these features into a single model. To my knowledge, there has not yet been a journal publication of an agricultural or land use ABM available that would combine realistic agent learning algorithms, process-based biophysical modules and a generic empirically parameterized, theory-based agent decision that considers risk management at the same time. Current models either have a sophisticated, comprehensive decision-making process, but simple learning mechanisms, or vice versa. Explicit risk management has so far been largely neglected in agent-based models, though mathematical programming based ABM may be able to borrow from risk management implementations of farm models [Janssen and van Ittersum, 2007].

A second challenge remains the validation and the use of sensitivity and uncertainty analysis with agent-based models. While agent-based models provide the opportunity of interactive validation with stakeholders, confidence into model outcomes of more generic models and simulations at larger scales and over longer time-scales can be improved by some form of behavioral validation. Where formal behavioral validation of agricultural ABM has been conducted so far, it has been based on statistical measures that seem to have been chosen rather *ad hoc*, without theoretical foundation or a discussion of the suitability of the measure in question. For example, both regressions of simulated on observed values and statistical tests of differences have been found questionable devices for model validation in general [Kleijnen et al., 1998; Forster, 2000; Moriasi et al., 2007]. The suitability of specific loss functions or goodness-of-fit measures depends on assumptions on the likelihood of different types and sizes of error [Kuczera et al., 2006; Schaeffli and Gupta, 2007]. This to my knowledge has not been explicitly considered or analyzed in the context of agricultural economic agent-based models so far. Likewise, uncertainty and sensitivity analysis has mostly been rudimentary, and can certainly be improved considerably relying on established techniques for other types of simulation models in neighboring fields [see for example Helton et al., 2006; Saltelli and Annoni, 2010].

Chapter 3

Validation, calibration, uncertainty analysis: principles and approaches

In the past and in the context of agricultural mathematical programming models sometimes still today [e.g. Buysse et al., 2007], the validity of a model has often been narrowly defined as the accuracy of the model in representing empirical facts. As Oreskes et al. [1994] have most prominently argued, this view of validation as verification cannot be upheld because an absolute proof of the correspondence of a model to all dimensions of reality is impossible. While Oreskes et al. argue for a replacement of the term validation by confirmation, most authors from agricultural economics as well as neighboring fields like economics, operations research or hydrological modeling stick to the terms validation and validity. They redefine the validity of a model as the adequacy of the model with respect to the purpose for which it is intended to be used [e.g. Gass, 1983; McCarl and Apland, 1986; Barlas, 1996; Kydland and Prescott, 1996; Rykiel, 1996; Beck et al., 1997; Jakeman et al., 2006].

Barlas [1996] links the adoption of the more general – but also more fuzzy – concept of adequacy instead of accuracy to the development of a more relativist, more holistic school of thought in science theory in general: Up to about the 1960s, science was dominated by what Hands [2001] calls the 'Received view'. The empirical testability of results was regarded as a strict criterion for the scientificness of a theory and the ability of a theory to correctly predict empirical facts as the only criterion to choose between different scientific theories. In economics, this school of thought is often associated with the methodological writings of Terrence Hutchison, Paul Samuelson and Milton Friedman[1] [Hands, 2001]. The works of Kuhn [1962, 1977] and Quine [1951], who emphasized the theory-ladenness of empirical observation and the underdetermination of theories, then started a lively debate on the foundations of the Scientific method that finally eroded most of the central tenets of the Received view: It illustrated how the refutation or confirmation of theories relies on ideal conditions that are never fulfilled in reality [Hands, 2001; McCloskey, 1983].

This debate has not led to the development of a new, revised 'Scientific method', rather, as McCloskey [1983] concludes, it has become clear that there is not the one and only 'Methodology' that ensures the scientificness of a theory and, by extension, the validity of a model. As Klappholz and Agassi formulated already earlier:

> "[T]here is only one generally applicable methodological rule, and that is the exhortation to be critical and always ready to submit one's hypothesis to critical scrutiny." [Klappholz and Agassi, 1959, p. 60]

This view, that the most important and characterizing element of scientific inquiry is criticism, is the central tenet of the epistemological theory of critical rationalism attributed to Karl Popper [Klappholz and Agassi, 1959; Hands, 2001]. According to critical rationalism, a theory or model has to be criticizable and bear the potential to be improved by criticism.

[1] Though Friedman has been reinterpreted and connected to other methodological currents by some authors [Hands, 2001; Deichsel and Pyka, 2009].

Empirical testability and testing of theories and models is one criterion of criticism, but it is not an exclusive one: A theory does not necessarily have to be empirically testable to be criticizable. Rather, a strict requirement on testability may even preclude valuable criticism to a theory [Klappholz and Agassi, 1959]. While this seems to be at odds with the theory of falsificationism, for which Popper is more renowned, Caldwell [1991] and Boland [1994] see critical rationalism as the broader theory, and falsificationism as a specific incarnation of criticism especially suited for many situations in the natural sciences [Hands, 2001].

Popper himself [1963, as cited by Caldwell 1991] as well as other authors like Longino [1992] point out that the essential requirement to ensure a critical attitude, scientific progress and the development of valid theories are institutions that allow, invite and accept criticism and participation by a wide range of diverse people. Scientists themselves should do everything to facilitate criticism.

> "It is preferable, we think, to help our critic to formulate his rival hypothesis rather than to discourage him." [Klappholz and Agassi, 1959, p. 68]

Both, criticism and defense against criticism have to be voiced by argumentation and theories are, as Klappholz and Agassi [1959] and McCloskey [1983] argue, ultimately accepted, refuted or improved by scientists, if the arguments for or against them are found to be convincing. It has to be emphasized that, despite denying the existence of a more specific scientific method, both, critical rationalists and McCloskey as a proponent of the rhetorics of economics, do value methodological advice and the establishment of methodological standards. Argumentation usually benefits from shared knowledge of established practice, i.e. in the case of validation the use of established validation methods. What they reiterate, however, is that standards cannot guarantee scientificness, require argumentation in their support and may be revised or abandoned if they are found not to be useful.

This understanding of validation as criticism is exemplified in Jakeman et al.'s summary of good practice in the development and evaluation of environmental models, which has been published as a position paper in the journal of Environmental Modelling & Software and can be regarded as setting a standard for the field. As the authors summarize their understanding:

> "[T]he modeling process is about constructing or discovering purposeful, credible models from data and prior knowledge, in consort with end-users, with every stage open to critical review and revision."[Jakeman et al., 2006, p.612]

The critical attitude is also reflected in Jakeman et al.'s intention to raise awareness of the components of the modeling process in order to enable stakeholders to fruitfully criticize simulation processes. To achieve this goal, all elements of the modeling process need to be exposed and made explicit.

A thorough model documentation is a necessary basis for any critical assessment of a model and therefore a prerequisite of a valid model. Documentation needs to cover the whole process of modeling from the inception to the interpretation of results. Jakeman et al. emphasize that model documentation needs to include a clear definition of the problem and the intended role of the model in solving it, and a clear delineation of the scope and resources of the simulation study. A clear conceptualization of the knowledge of the system is equally important as the design of a model can only be appraised and criticized with respect to these frameworks.

The following sections discuss three aspects of the modeling process that can be considered critical for model validity. These are:

(i) the invariance of the model over all situations to which it is to be applied,

(ii) how uncertainty is dealt with,

(iii) the behavior and predictive accuracy of the model.

3.1 The invariance of the model

Models are usually used to predict the state or behavior of a variable of interest in a situation that cannot be or at least has not been observed. In a very general way, a quantitative model can be written as $\vec{y} = f(\vec{x})$, i.e. as a function (f) that links exogenous variables (x) to a number of endogenous variables (y), some of which are the outcome variables the researcher is actually interested in. The endogenous variables are those which are determined by running the model. The values of the exogenous variables are observed or anticipated by the modeler and constitute the input of the model. The model f describes the association between exogenous and endogenous variables.

When comparing two situations A and B using a model, the modeler needs to ensure that all the anticipated (relevant) differences between A and B are reflected in $\vec{x}$, while f is invariant between the situations. The Lucas Critique [Lucas, 1976] is a famous example of fruitful criticism along this line of argument. It challenged the forecasting practice of macroeconomic modelling of the 1960s/70s by arguing that the parameters/functional relationships estimated from past data could not be assumed to remain the same for the forecasting situation. Validation, therefore, should always provide arguments that explain why the model is generalizable over the envisioned domain of prediction. For estimated parameters it needs to justify why parameters and model structure can be assumed to be the same in the new situation as they were in the context of their estimation. Model validity is never general, but always qualified by necessary conditions, and the use of a model for prediction in a new situation requires that these conditions apply [Reichert and Omlin, 1997; Rykiel, 1996].

3.2 Dealing with uncertainty

The modern view of validation recognizes that models are abstractions from reality and most of the times a specific model is only one of many possible ways of describing reality [Barlas, 1996]. On the one hand, the modeler will encounter many situations during the design process, where knowledge of the system is incomplete or ambiguous. Relationships between different variables may be largely uncertain, and very different functional relationships may be theoretically plausible. In other cases, relationships are well-defined and only the magnitude of a certain effect depends on unobserved local conditions. On the other hand, the modeler cannot build a complete, exact image of a real system and has to simplify some process representations and completely omit others and it may not be obvious, which degree of simplification is still adequate and which processes can safely be omitted without causing significant errors in model conclusions. Error and uncertainty are distinguished by the cause of inaccuracy and the methods for minimizing it differ. In case of error, the correct value or process representation is known; that is, error is potentially identifiable but not necessarily identified. Uncertainty, in contrast, stems from an unknown value or process representation and cannot be fully eliminated even with severe efforts [Oberkampf et al., 2002].

Oberkampf et al. [2002] distinguish (i) acknowledged error, (ii) unacknowledged error, (iii) aleatory uncertainty and (iv) epistemic uncertainty:

- *Acknowledged errors* are recognized by the modeler and may occur due to mathematical simplification or the chosen level of spatial and temporal resolution. These errors are usually tolerated due to time and resource constraints and, because the correct specification is known, the modeler usually has an idea about the magnitude of error.

- *Unacknowledged errors* are mistakes or program bugs that pass unrecognized and can only be avoided by repeated double-checking and application of good software engineering practices, for example, object-oriented programming and component-based software architectures.

- *Aleatory uncertainty* is inherent in the process modeled and therefore irreducible. It follows from a stochastic process that is well understood; still its outcome for a specific sample point is unknown and will differ from model run to model run. Examples are rainfall in a specific month in the future, the sex of newborn cattle, or the date of farm machinery breakdown. Aleatory uncertainty cannot be reduced in advance at all.

- *Epistemic uncertainty* results from incomplete knowledge, usually due to the vagueness of concepts, unobserved values or conflicting evidence, and can be associated to both model structure as well as model parameters.

Uncertainty and error may critically affect the conclusions that may be drawn from model outcomes and modelers can defend the validity of conclusions only by showing that they (i) reduced unacknowledged errors and epistemic uncertainty as far as possible and (ii) communicated the effects of acknowledged error, aleatory uncertainty and remaining epistemic uncertainty on model outcomes and conclusions. Assessment and communication of uncertainty is an essential element of a modeling study right from its beginning and cannot merely be attached to an existing analysis at the end [Refsgaard et al., 2007].

Epistemic uncertainty can be reduced by incorporating well-founded, established theory and reliable observations. Similarly, experts and stakeholders familiar with the processes modeled can provide valuable information. However, wherever this is not available, the modeler still has to make a choice and choose a specific functional relationship with specific parameters [Walker et al., 2003; Refsgaard et al., 2007]. Calibration may provide arguments for selecting a specific model or parameterization which performs best in reproducing observed behavior. Still, also calibration procedures are subject to considerable epistemic uncertainty and selecting a single, best-fitting model may actually only hide uncertainty instead of truly reducing it, as is discussed in more detail in subsection 3.2.1, which also includes a review of different calibration techniques. Rather it may be warranted to include all potential candidate models into further analysis and assess the robustness of conclusions with respect to the uncertainty involved in the modeling process [Jakeman et al., 2006]. If robustness cannot be established, it is important to clearly identify the conditions under which certain conclusions hold. As remaining uncertainty may be large, the number of potential candidate models is usually very high, and it is important to find an efficient representation of this uncertainty to reduce the necessary number of model runs. Suitable sampling strategies and approaches for sensitivity analysis are discussed in subsection 3.2.2.

3.2.1 Parameter estimation and model selection

Often a distinction is made between model selection, i.e. choice of a functional form, and parameter estimation, i.e. choice of values for function parameters. In the following, the term calibration is used to refer to both unless otherwise noted. The conditions necessary for calibration to be usefully applied for uncertainty reduction, can be subsumed under five points:

(1) The parameters to be estimated are invariant between the situation used for calibration/estimation and the situation for which the model is to be used [Lucas, 1976].

(2) A dataset of observations of exogenous and endogenous variables is available [Bellman and Åström, 1970].

(3) There is a suitable goodness-of-fit measure, loss function or likelihood functions to distinguish the predictive performance of different models [Hansen and Heckman, 1996].

(4) Calibration can be ensured to fit the systematic part of the relationship between exogenous and endogenous variables and not also the noise in the data: overfitting can be avoided [Jakeman et al., 2006].

(5) The number of model runs necessary for the algorithm or sampling strategy employed during calibration is feasible.

While the first two points are rather self-explanatory, the choice of a goodness-of-fit measure and the danger of overfitting deserve a more detailed discussion. A short reflection on calibration algorithms will conclude this subsection.

Goodness-of-fit: Loss functions and likelihoods

In a general formulation of the model, the vector $\vec{\theta}$ may represent epistemic uncertainty in the form of the different candidate implementations[2] among which the modeler wishes to select by calibration. Further, aleatory uncertainty can be represented by random realizations of a stochastic process ζ in the case of a stochastic model. Given that the model is not a perfect representation of reality and that reality cannot be measured without errors, ε accounts for deviations between model results and corresponding observations.

$$y = f(\theta, x, \zeta) + \varepsilon(x) \tag{3.1}$$

Deviations may be due to acknowledged or unacknowledged error, unknown systematic effects (making ε potentially dependent on x) and random noise, e.g. measurement error in the data, different outcomes of ζ between simulation and observation, and irreducible randomness not captured by f. Lack of accuracy in reproducing observation data can then be caused by any of these components and it is impossible to attribute it to the model or a specific parameter in the model without making further assumptions on ε. This is an instance of the Duhem-Quine problem of underdetermination [Quine, 1951], already mentioned above, which says that a model or theory always requires auxiliary hypotheses to be tested, and be they only that the model or theory is applicable to the situation in question [Fagiolo et al., 2007]. Consequently, they can never be confirmed or disconfirmed, rather only the whole construct of theory/model and auxiliary hypothesis together can be confirmed or disconfirmed [Oreskes et al., 1994].

Classical calibration approaches proceed by minimizing a loss function calculated from the observed deviations [Hunt et al., 2007]. Using a specific loss function implies making

[2] The choice between different functional forms can mathematically be represented by a parameter and forms part of $\vec{\theta}$.

assumptions on the source of the deviations between observed and simulated data. For example, a loss function that reaches its minimum at zero deviations assumes that the model can be expected to be unbiased. If, on the other hand, there is an acknowledged error, e.g. the omission of a certain process that in theory should lead to an underestimation of results, this bias should be reflected in the loss function. Otherwise, the parameter estimates will also be biased and compensate for the omitted process. The correspondence of prediction and observation for the calibration sample will then be better, however, the reliability of the parameters for prediction in other situations is compromised. Measures based on weighted squared deviations, as often used in classical calibration approaches, assume that deviations are mainly due to measurement error of deviations and imply heteroskedastic, normally distributed error terms with zero mean, which would e.g. be warranted when the model is expected to be unbiased [Poeter and Hill, 1997; Hill et al., 1998].

Calibration by minimizing a statistically derived loss function is essentially equivalent to econometric least squares regression [Hansen and Heckman, 1996]. Analogous to regression analysis, calibration can be understood as a procedure to maximize the likelihood of the selected model. Likelihood based approaches to model selection and parameter estimation [Hobbs and Hilborn, 2006; Hartig et al., 2011] explicitly start from a formulation of the estimation problem as a stochastic model (even if the actual model is deterministic). They assume a probability distribution for ε, which may potentially also reflect systematic components like bias, autocorrelation or heteroskedasticity. Based on this probability function, a likelihood function can be derived, which is proportional to the probability of a model θ_i given the observed data (D).

$$\mathcal{L}(\theta_i|D) = p(D|\theta_i) \tag{3.2}$$

Using a likelihood function as goodness-of-fit measure also highlights another aspect of calibration: While the term goodness-of-fit often entails the notion that a model with a higher goodness-of-fit is the better model and should be preferred over others, using a likelihood function underlines that a model with a higher goodness-of-fit is not necessarily better, but merely more probable given the observed data, and it does not mean that the alternatives are impossible or even improbable. Using only the best-fitting model, however, renders all alternatives irrelevant and may unduly reduce or rather hide uncertainty – especially if the difference in likelihood between alternatives is small, the data quality is low and the epistemic uncertainty associated to the auxiliary assumptions on error terms is high [Beven and Binley, 1992].

As an alternative, Bayesian parameter inference allows the attribution of probabilities to different models or parameter combinations based on the likelihood function and Bayes' Theorem. Instead of selecting the one parameter vector that maximizes likelihood (or minimizes deviation), it assigns to each parameter vector a probability of being the correct simulator of the system. This then automatically translates into a probability distribution for the model outcomes [Reichert and Omlin, 1997; Hobbs and Hilborn, 2006; Hartig et al., 2011]. In Bayesian parameter inference, additionally to the likelihood function described above, prior knowledge about model parameters is summarized in a prior probability distribution for the potential parameter combinations. The likelihood function is then evaluated for each parameter vector (θ_i) by running the model and the posterior probability of a parameter vector is calculated using Bayes Law for discrete hypothesis [Hobbs and Hilborn, 2006]:

$$\mathrm{P}(\theta_i|D) = \frac{\mathcal{L}(\theta_i|D)\mathrm{P}(\theta_i)}{\sum\limits_j \mathcal{L}(\theta_j|D)\mathrm{P}(\theta_j)}$$

where $\mathrm{P}(\theta_i)$ is the prior probability for θ_i and $\mathcal{L}(\theta_i|D)$ is the likelihood of θ_i given the data.

Especially in hydrology, considerable debate has centered around the choice of the likelihood function, specifically whether formal or informal likelihood functions should be preferred (see Schoups and Vrugt 2010 for an overview). The attribute formal is used in this context to refer to likelihood functions that have been derived from explicitly formulated statistical models for residuals ε, which may range from the simple assumption of normality, over the inclusion of autocorrelation and heteroskedasticity [e.g. Schoups and Vrugt, 2010], to explicit distinction of different error sources with associated stochastic processes [e.g. Kuczera et al., 2006]. Proponents of the formal approach underline the possibility to statistically validate assumptions a posteriori given an explicit error model [e.g Stedinger et al., 2008]. Often, however, the modeler is not able to specify a formally derived probability distribution for ε. Beven et al. [2008] argue that the use of incorrect formal error distributions may substantially bias the estimation and lead to considerable underestimation of uncertainty, especially if residuals are autocorrelated or heteroskedastic. Formal likelihood measures will then overestimate the information content of data and overcalibrate and thus bias parameter estimation.

Spear and Hornberger [1980] pioneered an informal Bayesian-like procedure by defining a binary-valued likelihood function that just classifies model realizations as acceptable or inacceptable by comparing it to the behavior of the real system. All accepted models are treated as equally likely representations of reality and the distribution of results remained unweighted. Brenner and Werker [2007] argue for the use of a similar approach for agent-based models under the heading of 'abductive modelling'. Beven and Binley [1992] go one step further in weighting accepted model parameterizations using likelihoods that are often based on goodness-of-fit measures (see Beven and Freer 2001 for examples of informal likelihood measures). Criticism raised against informal likelihood measures point to their 'arbitrariness', lack of statistical validity of posterior distributions and an overestimation of uncertainty by informal likelihood measures compared to the correct error model in cases the later is known [Schoups and Vrugt, 2010]. On the other hand, the Generalized Likelihood Uncertainty Estimation (GLUE) framework that Beven and Binley [1992] propose is able to incorporate the whole continuum of likelihood measures from Spear and Hornberger 1980's binary function to a completely specified formal likelihood, while laying an emphasis on an explicit formulation of assumptions.

Identifiability and overfitting

Besides the underdetermination of the model due to epistemic uncertainty about the source of predictive inaccuracy, models may also be underdetermined by the amount and quality of data itself. To uniquely identify model parameters, the number of conditional equations derived from applying a model to a dataset has to be higher than the number of parameters. In addition to that, also sufficient variation in observations is required [Bellman and Åström, 1970; Cobelli and DiStefano, 1980]. For example, it is not possible to estimate a parameter that directly captures the development of a variable over time if all observations originate from the same point of time, even if thousands of observations are available. If we additionally consider the existence of error and epistemic uncertainty as discussed in the previous section, it becomes apparent that the number of parameters should be well below the number of observations in order to allow for sufficient degrees of freedom for the error and avoid overfitting the model. In an overfitted model, the parameters of the model are chosen to reproduce also the deviations present in the dataset, leading to an optimal fit in the calibration dataset, but deteriorating prediction in other situations [Forster, 2000]. Parameter estimation and model selection always have to deal with this trade-off between a perfect fit for the test sample and the risk of deteriorating predictive capacity for other samples. It is therefore necessary to understand the problem of calibration as a problem of maximizing the *expected* accuracy of prediction for any sample, rather than maximizing the accuracy of fit to the observed sample [Zucchini, 2000].

Parsimony, i.e. reducing the number of parameters and keeping models as simple as possible, is one way to guard against overfitting. However, this may be at odds with the requirement that parameters be invariant between different situations (discussed in section 3.1): Theoretical considerations about stability or variation of parameters outside the identification domain have to take precedence, and if theory requires the inclusion of further parameters, leaving them out to simplify parameter estimation would clearly invalidate the model from a theoretical point of view [Reichert and Omlin, 1997].

Identifiability can also be improved by adding further constraints to the calibration problem and thus increasing the number of defining equations. For example, in hydrological modeling, the high number of parameters is often caused by a detailed spatial resolution. Each node in a flow network may potentially differ in locally specific parameters. Identifiability can be improved by regularization, i.e. requiring that locally specific parameters be equal within certain zones, or differ only to a restricted extent from neighboring nodes, effectively smoothening the distribution of parameters over the landscape [Hunt et al., 2007].

If a formal likelihood function can be formulated and candidate models are nested, i.e. created by subsequently adding additional equations and factors to the model, or they can at least be ranked by complexity (numbers of free parameters), classical hypothesis testing could be used, with the null hypothesis that preferring a more complex over the simpler model does not increase accuracy, and only proceeding to the next more complex model if the null hypothesis is rejected. The significance level then corresponds to the weighting of goodness-of-fit vs. simplicity [Forster, 2000].

Classical hypothesis testing, however, lacks a deep theoretical underpinning when used for model selection and the use of the Akaike Information Criterion (AIC) is preferable [Forster, 2000]. The AIC is based on the Kullback-Leibler divergence, which measures the loss of information incurred by using the tested model instead of the "true model" that created the observations. Though we cannot know the true model, the AIC provides us with an approximately unbiased estimator[3] of relative K-L information for large data samples and good models that can be used to rank the models [Burnham and Anderson, 2004; Hobbs and Hilborn, 2006]:

$$AIC = -2\log\left(\mathcal{L}(\theta_i|D)\right) + 2K \tag{3.3}$$

where K is the number of parameters in the model. For small samples, i.e. $\frac{n}{K} < 40$, a modified version AIC_c should be used. As the modified version converges to the original one for large samples, AIC_c can always be preferred [Burnham and Anderson, 2004].

$$AIC_c = -2\log\left(\mathcal{L}(\theta_i|D)\right) + 2K + \frac{2K(K+1)}{n-K-1} \tag{3.4}$$

The information criterion is usually only calculated for different model structures (model selection in the narrow sense), where the parameters of each model structure have first been estimated by maximum likelihood estimation. They will not lead to different conclusions compared to pure likelihoods if the number of parameters is the same. For Bayesian parameter estimation, Burnham and Anderson [2004] suggest the use of an informative prior distribution that should depend on the number of observations and the number of parameters

[3] Actually, AIC is the unbiased estimator multiplied by -2 [Burnham and Anderson, 2004].

in a candidate model. In fact, they argue that the AIC can be understood as being derived from a Bayesian foundation similar to the alternative Bayesian Information Criterion (BIC), but with a prior distribution based on K-L divergence instead of the uniform prior used by the BIC . Both, AIC and BIC can also be used to judge whether the performance difference between two models is large enough to warrant the strict preference of one model over the other, or whether it seems preferable to base conclusions of a combination of several models [Burnham and Anderson, 2004], such as is done in Bayesian Model Averaging [Hoeting et al., 1999; Gibbons et al., 2008].

A traditional approach to guard against overfitting that does not depend on a formal likelihood function is cross-validation. Cross-validation is based on the idea that the expected predictive performance of a model can be assessed by observing the performance of a model predicting past data that has not been used for calibration. The observed data is partitioned into a calibration dataset and a validation dataset. In model selection, increasingly complex models can be fitted to the calibration dataset and observe the behavior of a goodness-of-fit measure in both datasets. While the goodness-of-fit for the calibration dataset should steadily increase, at a certain point of time the goodness-of-fit for the validation dataset will usually start to decrease or remain stable indicating the degree of complexity that should not be surpassed if one wants to avoid the risk of overfitting [Browne, 2000]. For candidate models that cannot be ranked by complexity or parameter estimation in the narrow sense, a similar criterion can be applied if the process of calibration involves an evolving reduction of the number of candidate models. Whittaker et al. [2010], for example, describe parameter calibration using genetic algorithms, where the cross-validation criteria is to stop optimization at the generation for which the goodness-of-fit for the validation dataset starts to decrease.

There are different approaches to the partitioning of the dataset into a calibration and validation dataset, including a jacknife procedure in which N-1 data points are used for calibration and the N for validation with every data point being omitted once. This leads to an expected value of the goodness-of-fit measure that is asymptotically equivalent to the AIC, with the difference that no explicit assumption on the likelihood has to be made [Browne, 2000]. The major problem with cross-validation methods is that results depend on the size of the dataset used and protection from overfitting vanishes the larger the dataset gets. Further, there is little guidance on how to partition the data, while this may strongly influence the results and make estimates highly variable [Shiffrin et al., 2008; Browne, 2000]. "[C]ross-validation says nothing about any best model for the population. What it does consider is to what extent a model has to be oversimplified to avoid the effect of random fluctuations due to a small sample." [Browne, 2000, p.114]

Bootstrap methods can at least document the variability in the goodness-of-fit measures, providing an idea on the stability of the result given different partitions of the dataset [Efron and Tibshirani, 1997]. Busemeyer and Wang [2000] argue for generalization tests, where the validation dataset is not obtained by dividing one sample, but by using an entirely different sample if possible from a different experimental design or by mimicking a directed extrapolation [Forster, 2000]. Related approaches are used in time-series, where the performance of a model in predicting always one step ahead is tested leading to a measure of accumulative prediction error [Wagenmakers et al., 2006].

Algorithms

The most basic and least formalized calibration algorithm is manual trial-and-error: Modelers start with their best guess of parameters, run the model, evaluate the goodness of fit, change the parameters, and run the model again to see whether the fit improves. They continue this procedure until they are satisfied with the fit [Hunt et al., 2007].

Calibration based on minimizing a squared loss function frequently employs nonlinear least-squares algorithms like the Levenberg-Marquardt-Algorithm, which, however, are only able to detect local minima [Poeter and Hill, 1997; Doherty, 2005]. Genetic algorithms provide an alternative for global optimization and other nonstandard loss functions [Duan et al., 1993; Yapo et al., 1998]. Tikhonov regularization can be combined with these algorithms to automatize regularization. The modeler then specifies only potential regularization constraints, while the actual extent of regularization is determined by the algorithm itself [Hunt et al., 2007]. Optimization algorithms for likelihood maximization include simplex searches, simulated annealing, and genetic algorithms, while the distribution of likelihoods over a parameter space, as necessary for Bayesian parameter estimation, can be achieved by rejection sampling, Markov Chain Monte Carlo, particle filters or Approximate Bayesian Computation [Hartig et al., 2011].

The choice of an algorithm thus has to depend on the calibration technique employed, but will also be determined by the feasible number of model runs and the degree to which model evaluation can be automatized.

3.2.2 Uncertainty and sensitivity analysis

A major concern of validation is to illustrate, how model uncertainty affects the conclusions that can be drawn from the simulation study. It may show that the conclusions are robust with respect to model uncertainty, or, if robustness cannot be established, it is important to clearly identify the conditions under which certain conclusions hold.

The literature on simulation modeling offers various definitions of uncertainty and sensitivity analysis. According to Helton et al. [2006], uncertainty analysis refers to quantifying the uncertainty in model outcomes, while sensitivity analysis is dedicated to determining the individual contribution of uncertain input variables to the uncertainty of outcomes.

Uncertainty analysis should document the full range of outcomes that are to be expected given the uncertainty in model assumptions. It should cover the full set of candidate models and parametrization remaining after calibration. Model outcomes are then presented either as an average of the candidate models include or indicating the full range of outcomes. If a Bayesian approach is used for calibration, the posterior probabilities for the models directly translate into probabilities for predictions [Helton et al., 2006], or at least allow the calculation of probability-weighted averages (Bayesian Model Averaging [Hoeting et al., 1999; Gibbons et al., 2008]).

Sensitivity analysis is employed at two different stages of the modeling process: (i) Before parameter estimation, a parameter screening can be used to determine the influence of parameters on goodness-of-fit and outcome measures in order to decrease the number of parameters included into the estimation and reduce the computational effort. Uninfluential parameters can be fixed at any theoretically satisfactory value. Parameters that have little influence on goodness-of-fit measures, but a considerable influence on outcome measures, may be excluded from parameter estimation, but should be included in later uncertainty analysis. (ii) After parameter estimation, when the model is run for predictive analysis, sensitivity analysis is used to identify parameters or exogenous variables which have the greatest influence on observed outcomes. On the one hand, this allows to reduce the tested parameter combinations for uncertainty analysis by including only those that really have an influence on model outcomes. On the other hand, it helps to prioritize future research into those processes that are relevant in determining conclusions [Saltelli et al., 2004].

Uncertainty and sensitivity analysis can be *local* or *global* in nature. In local sensitivity analysis, uncertain input factors are varied *ceteris paribus*, one-at-a-time (OAT), around a 'baseline' result and derivatives of the model function with respect to each input factor are calculated. While the results of local sensitivity analysis may be useful in some cases, they do not provide a reliable characterization of the uncertainty involved in the model except when the model is strictly linear (for first-order derivatives) or additive (for higher-order derivatives) [Saltelli and Annoni, 2010].

Model uncertainty may be very high, even after calibration and consequently the amount of potential parameter combinations to be included in uncertainty and sensitivity analysis may be very high, usually too high to test every candidate. Instead efficient experimental designs are required that allow a representation of the parameter space with a feasible number of model runs.

For uncertainty analysis, Latin-hypercube sampling (LHS) [McKay et al., 1979] can be used to cover the full parameter space with equal weight. Apart from global coverage, designs for sensitivity analysis need to ensure that outcome uncertainty can be unambiguously attributed to individual parameters. Factorial, fractional factorial and orthogonal designs [NIST/SEMATECH, 2012] are useful for this purpose, but may easily require too many model evaluations if the number of parameters is high.

An extensive overview and discussion of sampling-based sensitivity measures can be found in Helton et al. [2006]. Two of them, the elementary effects screening, a very efficient design to identify influential parameters, and the concept of variance decomposition by sensitivity indices are described in more detail in the next two subsections.

Elementary effects method

The elementary effects method for screening input factors[4] was developed by Morris [1991] and improved by Campolongo et al. [2007]. *Screening* refers to the fact that the method is able to order input factors by their importance in determining the output variance, without being able to quantify their contribution to uncertainty. It can therefore be used to select the most important input factors, which may then be assessed using more computationally intensive methods.

For the elementary effects method, the range of each input factor is represented by p integers that denote equally spaced levels. A k-dimensional vector of factor values is drawn by randomly choosing a level for each input factor. Then, step-by-step, the value of one input factor after the other is increased or decreased by one level (or a predefined multiple of a level) leading to a trajectory of $k+1$ different input factor vectors, which distinguish themselves only by one element from their predecessor. The whole procedure is repeated r times, each time starting from a different start vector leading to $r(k+1)$ sample points. This design allows calculating a distribution of r elementary effects for each of the k input factors by comparing the model result for those two vectors in the trajectory which differ only in this factor. While each difference calculated in this way constitutes only a *ceteris paribus* effect of the factor, the distribution of these local effects over the r trajectories constitutes a global assessment of the elementary effect under very different conditions [Saltelli et al., 2004].

From this distribution of elementary effects, three measures can be calculated: the mean (μ) and the standard deviation (σ) of the distribution and the mean of the distribution of the absolute effects (μ^*). μ^* can be used to rank the factors by importance, the greater μ^*, the

[4] The uncertain parameters and exogenous variables considered in a sensitivity analysis are in the following subsumed under the term input factors following the terminology of Saltelli et al. [2004].

greater the average influence of the factor on the outcome. A high standard deviation points to a high variation in the elementary effects, which may be due to a nonlinear relationship between the factor and the model output or strong interactions with other factors. A difference between the absolute of μ and μ^* indicates a nonmonotonous relationship between input factor and model results, i.e. at some points the effect has a negative sign, at others a positive one [Saltelli et al., 2004].

Variance decomposition

The elementary effects screening allows for a ranking of input factors by their influence onto the outcomes, but it is not suitable to quantify the contribution of the uncertainty in an individual input factor to the uncertainty in the outcomes [Saltelli et al., 2004]. Such a quantification would be provided by a complete variance decomposition, which means attributing a share of the total variance to each factor and to each possible interaction term between the factors :

$$V_y = \sum_{i}^{K} V_i + \sum_{i}^{K} \sum_{j=i+1}^{K} V_{ij} + \sum_{i}^{K} \sum_{j=i+1}^{K} \sum_{k=i+2}^{K} V_{ijk} + \cdots + V_{1,2,\ldots,K} \tag{3.5}$$

In this equation, V_i is the variance in the outcome y due to the first-order effect of factor x_i. V_{ij} is the variance in y due to the interaction between factors x_i and x_j. V_{ijk} is the variance in y due to the third-order interaction between x_i, x_j and x_k, and so on until $V_{1,2,\ldots,K}$, which indicates the variance in y that can be attributed to the Kth-order interaction between all factors [Helton et al., 2006].

Most of the times we are not actually interested in each single term in the decomposition. However, two measures can be derived from the equation, which are very useful for common purposes of a sensitivity analysis: the first-order sensitivity index (S_i) and the total effects sensitivity index (S_{Ti}) [Saltelli et al., 2004].

The first-order sensitivity index is calculated by dividing the first-order term for factor x_i by the total variance.

$$S_i = \frac{V_i}{V_y} \simeq \frac{V_i\left(E_{-i}(Y)\right)}{V_y} = 1 - \frac{E_i\left(V_{-i}(Y)\right)}{V_y} \tag{3.6}$$

S_i indicates the share of the total variance that is expected to disappear if input factor x_i is fixed at any value. This measure is very useful if the goal of the sensitivity analysis is to prioritize factors for further research with the objective of reducing the uncertainty in model results. It can, for example, be estimated by fixing x_i at n random values from its distribution and run a sample varying all other factors for each of the n fixed values. For each of the n samples, the variance $V_{-i}(Y)$ can then be calculated and forms the basis for deriving the expected value $E_i\left(V_{-i}(Y)\right)$ over all n samples. The later is an estimate for the sum of all terms in the variance decomposition except V_i, which can consequently be derived by subtracting the result from the total variance [Helton et al., 2006; Saltelli et al., 2004; Morris et al., 2008].

The total effects sensitivity index is the sum of all terms of the variance composition including x_i divided by the total variance:

$$S_{Ti} = \frac{V_i + \sum_{j \neq i} V_{ij} + \sum_{j \neq i, k \neq i,j} V_{ijk} + \cdots + V_{1,2,\ldots,K}}{V_y} \tag{3.7}$$

It represents the variance that is expected to remain if only factor x_i was uncertain and all other factors were fixed. It is useful if the task of the sensitivity analysis is to identify factors that can safely be ignored. Factors with low total effects indices have little influence on model outcomes and can be fixed at arbitrary values in order to reduce the dimensionality of the factor space, and consequently the necessary number of model evaluations [Helton et al., 2006; Saltelli et al., 2004].

Some convenient properties of the sensitivity indices can be derived from the variance decomposition: For linear models, S_i is equal to the square of the standardized regression coefficient (SRC) of factor x_i.[5] For additive models, which have no interactions, $\sum_i S_i = 1$ and $S_i = S_{Ti}$. The difference between first-order and total sensitivity indices for a factor can be interpreted as a measure for the influence of interactions involving this factor on the total model variance [Saltelli et al., 2004].

While the sensitivity indices provide an informative characterization of the contribution of individual factors to model uncertainty, their estimation can be very costly in terms of required model runs. Several authors have come up with different sampling plans for the efficient estimation of either or both sensitivity indices. Saltelli et al. [2004] devise a sampling strategy that requires $2N$ samples to estimate S_i, $(K+1)N$ samples to estimate S_{Ti}, and S_i, $(K+2)N$ samples to estimate both, where N is the sample size necessary to derive a converging estimate for a single index and might range between several hundreds and thousands of repetitions. Morris et al. [2008] develop a sampling strategy for the estimation of S_i based on replicated LHS, where the columns of the LHS are permuted according to an $OA(k^2, k+1, k, 2)$ orthogonal array. The orthogonal array consists of k^2 runs subdivided into k blocks. Morris et al. note that not all k blocks of this unbiased permuted column sample (UPCS) need to be used.

3.3 Evaluation of model behavior and predictive accuracy

Abandoning predictive accuracy as the one and only criterion of model validation and emphasizing its shortcomings does not mean the evaluation of model behavior should not play a role in model validation. What has to be kept in mind, however, is that the fundamental problem of underdetermination explained in the discussion on calibration also affects the evaluation of model behavior and gives rise to the dilemma that Beck et al. have formulated as a fundamental paradox of modeling:

> "The greater the degree of extrapolation from past conditions, so the greater must be the reliance on a model as the instrument of prediction; hence, the greater the desirability of being able to quantify the validity (or reliability) of the model, yet the greater is the degree of difficulty in doing just this." [Beck et al., 1997, p.238]

Under the modern paradigm of model validity, the role of the evaluation of predictive accuracy cannot be to determine whether the model matches reality, but merely whether the model is accurate and precise enough for a given purpose or whether it significantly differs from reality. A significant difference must not be understood in the statistical sense, in this case. Hypothesis testing to detect a statistically significant difference between model and reality is not very useful in this context since it is known in advance that the model

[5]The standardized regression coefficients are the coefficients β estimated by the regression model $\frac{y-\bar{y}}{\sigma_y} = \sum_i \beta_i \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ [Saltelli et al., 2004].

is necessarily wrong. Rather, a significant difference is one that would lead to substantively different conclusions for the problems the model intends to answer [Barlas, 1996; McCloskey and Ziliak, 1996].

If the accuracy of the model in predicting observed data is lower than the accuracy required for forecasts, the model may be of little use. On the other hand, in many studies the magnitude of interest is actually not the actual model prediction for a situation, but the difference in outcomes between different scenarios, e.g. the difference in welfare between a situation where policy is in place and the baseline without this policy. A simulation experiment allows controlling for all other model parameters, scenario assumptions and aleatory uncertainty (using common random numbers), so that the difference between two situations can be reduced to the policy intervention itself. The probability distribution of the difference may be more conclusive than comparing the two probability distributions of the outcome variable under the two scenarios [Law, 2007]. While a considerable variance of a welfare measure may be observed over repetitions within scenarios, the difference between scenarios may be positive for all repetitions, making one policy strictly preferable even though the model has only a limited capacity to predict the exact outcome of the welfare measure under this policy. While in these situations the accuracy is of minor importance, it is still a good idea to report it in order to facilitate a critical review of the model and allow conclusions as to where it might be improved.

There is a number of evaluations of model behavior that can be realized even if little or no observation data of the real system is available [Forrester and Senge, 1980; Rykiel, 1996]:

Extreme Conditions Test Often, real system behavior is known or can be logically induced for some extreme conditions. E.g. most plants will not germinate with zero soil moisture, farms go bankrupt if liquidity is zero. Model results should be consistent in those cases.

Face Validation Experts on the system can be asked whether they deem model behavior reasonable.

Turing Tests Experts on the system are given datasets of real world observations and datasets of model results and are asked whether they can identify, which one is the model result.

If observation data is available, a number of statistical measures can be calculated that help to document model accuracy and bias in order to compare it against a required accuracy. Measures of model error include mean squared error (MSE), root mean squared error ($RMSE$), and RSR, the ratio of $RMSE$ to the observation standard deviation. Measures of model bias like percentage bias ($PBIAS$) help to identify systematic over- or underprediction in the model and may serve as a basis for bias correction in case one can assume that the bias is invariant between observation and prediction domain [Moriasi et al., 2007].

A number of dimensionless statistical measures – e.g. the coefficient of efficiency, E [e.g. Nash and Sutcliffe, 1970], or the index of agreement d [Willmott, 1981] – have been developed to allow assessment of model accuracy in a more relative sense by comparing the accuracy of the model across different domains, with other models applied under different circumstances, or to a random guess [Legates and McCabe, 1999]. Also the AIC allows assessing how much better one model is compared to another model, although it does not lend itself to evaluate model performance against a 0-1 scale [Burnham and Anderson, 2004].

While the great advantage of these measures is their reduction of the error vector to one value, which facilitates comparisons against other models or reference scales, this aggregation may very well hide systematic patterns underneath and it is strongly advisable to conduct disaggregated analysis in order to detect e.g. autocorrelation, heteroskedasticities or correlation of errors with predictors. Like parameter estimation methodologies, model evaluation measures entail weightings of errors and their suitability depends on our knowledge or assumptions of the likelihood of certain errors under a given model [Moriasi et al., 2007].

For example, due to squaring, large errors influence MSE and $RMSE$ stronger than small errors, making them sensitive to outliers. Many authors prefer the mean absolute deviation (MAE) as a less sensitive and more natural measure of average error in case of uncertain error distributions [e.g. Willmott and Matsuura, 2005; Hyndman and Koehler, 2006]. Relative or logarithmic transformations of the errors may prove useful for certain data [e.g. as discussed by Krause et al., 2005]. In any case, errors should be corrected for heteroskedasticity and autocorrelation before calculation if necessary [Willmott et al., 1985]. This applies analogously to the coefficient of efficiency and the index of agreement, which are based on squared error terms in their original forms. Additionally, the coefficient of efficiency weighs the prediction error against the deviation from the observation mean which can be understood as a benchmark model. The observation mean may, however, not be the best benchmark predictor available: in the case of time-series, for example, a persistence model, or trend and seasonality-sensitive predictors are better alternatives [Legates and McCabe, 1999; Schaeffli and Gupta, 2007].

Agricultural economic models are often used to simulate the areas which farmers allocate to different production activities. These simulated areas are then often used for comparison to observed data. As the sum of all crop areas is constrained by the land resources of the producer, or the total agricultural area in the study region, deviations are not independent. Rather, the overestimation of the area of one activity necessarily leads to the underestimation of another. As a consequence, land use data have to be understood as categorical data: From this perspective, the available area is classified into land use classes through farmers' production decisions.

The classical measures for testing goodness-of-fit for categorical data is Pearson's χ^2-statistic [Zucchini, 2000]. Similar statistics include the likelihood ratio G^2, the Freeman-Tukey statistic, FT^2, and the Cressie-Read-power divergence statistic, CR^2 [see Voas and Williamson, 2001], all of which would be tested against the χ^2 distribution.

Apart from the general reservation against hypothesis tests for model validation, all of these measures require the assumption that the table of categories is not sparse, i.e. does not contain many cell values close or equal to zero. As an alternative, Voas and Williamson [2001] suggest using the standardized absolute error (SAE) as the most simple descriptive statistic of deviation.

$$SAE = \frac{\sum_i \left|Y_i^{obs} - Y_i^{sim}\right|}{T} \tag{3.8}$$

This statistic normalizes the total absolute error by the count of individual entities T that have been distributed over categories. Its lower bound is zero indicating perfect fit. The upper bound is $2 - 2 * \min\left(Y_i^{obs}\right)$, the largest error that can potentially be created.

Voas and Williamson note that an average SAE can be easily calculated for different simulations even if they have a different overall count. Since comparing two randomly created categorizations with the same overall count would result in an expected SAE of one, they suggest the use of $1 - SAE$ as a coefficient of efficiency for categorical data. A value less than zero would then indicate a fit worse than a random allocation.

Chapter 4

Validation and calibration of MP-based agent-based models in agriculture

Chapter 2 showed how multi-agent models differ in the way they model human behavior. In agricultural economics, agent-based models that are built to simulate the behavior of farming populations can draw on a tradition of farm modeling that has developed in two major strands: econometrics and mathematical programming models [Just, 1993]. Econometric models are estimated from observed data using well-defined statistical techniques, but are often restricted to simple functional forms and are not suitable for extrapolation in the presence of structural changes or innovative techniques that have not been observed yet.

In contrast, mathematical programming models are built on a theoretical description of the decision problem of the farmer with all its opportunities and restrictions and are therefore theoretically able to extrapolate into new technologies or policy conditions. On the other hand, it can be quite difficult to empirically parameterize and validate mathematical programming models [Buysse et al., 2007].

The present chapter is concerned with the application of the principles and approaches of validation, calibration, and uncertainty analysis discussed in chapter 3 to simulation studies using mathematical programming-based multi-agent models.

It will start by discussing the theoretical foundations for the generalizability of a model of agricultural production decisions, required to ensure the invariance of the model over different agents and structural changes (section 4.1). Section 4.2 analyzes in how far potential approaches to calibration are consistent and useful with the generic formulation of the model and knowledge about its properties. Section 4.3 shortly introduces the ODD protocol as a standard protocol for documenting ABM.

Finally, section 4.4 summarizes the conclusions and again underlines the importance of documentation and uncertainty analysis.[1]

4.1 The invariance of the model: generalizability of agricultural decisions

Section 3.1 identified the invariance of the model between different situations as a prerequisite for a valid prediction of the differences in endogenous variables between these situations. In models intended to understand the reaction of farmers to changing environmental conditions, the understanding of human behavior forms an essential part of the analysis. One of the crucial questions for the modeler is to judge, which patterns of human behavior can be generalized and expected to hold for the domain of prediction.

[1] Specific approaches to uncertainty analysis are not discussed in a separate section in this chapter, because general principles have been introduced in chapter 3 and choice of experimental designs or screening techniques depend very much on the size and run-time of the model, the size of the parameter space and the available resources for simulation.

4.1.1 Rationality and situational analysis

Agent-based models relying on mathematical programming techniques generally assume that people behave rationally, where rational behavior must not be understood in the narrow sense of profit maximizing often used in microeconomics. Rather, one can understand these models as employing a form of situational analysis.

The concept of situational analysis is a general concept for making sense of human behavior developed by Karl Popper [Hands, 2001; Caldwell, 1991], but was stated most clearly by Koertge [1975, 1979]:

"1.	Description of the Situation:	Agent A was in a situation of type C
2.	Analysis of the Situation:	In a situation of type C, the appropriate thing to do is X
3.	Rationality Principle:	Agents always act appropriately to their situations.
4.	Explanandum:	(Therefore) A did X."
		[Koertge, 1975, p. 440]

A more elaborate formulation, also due to Koertge [1975], underlines that situational analysis acknowledges that the agents'[2] perception of the situation and their appraisal of the correct action in this situation may differ from the analyst's view, or that the agents may be impeded to realize the action they intended to do:

"1.	Description of the Situation:	Agent A was in a situation of type C
2.	Dispositional Law	For all such problem-situations A would use appraisal-rule R
3.	Analysis of the Situation:	The result of appraising C using R is X
4.	Description of Agent's Competence:	A did not make a mistake in applying R to C
5.	Rational Appraisal Principle:	All agents appraise their situations in a rational manner.
6.	Explanandum-1:	(Therefore) A concluded X was the rational thing to do.
7.	Rationality Principle:	Agents always act on the outcome of their rational appraisals.
8.	Explanandum-2:	(Therefore) A did X."
		[Koertge, 1975, p. 445]

The rationality principle is central to situational analysis. Without it, the logical inference from situation to action breaks down. Whether and to what extent the rationality principle has empirical or metaphysical content has been subject to debate. Popper himself regards it as being false as a general statement, but a sufficiently good approximation to reality. His

[2] In this context, 'agent' does not refer to a computational agent, as in the rest of this thesis, but to the sociological term designating an acting individual.

advice is not to question the rationality principle, when an explanation based on situational analysis fares badly, but rather the assumptions on perceptions, personal goals and decision strategies [Caldwell, 1991]. Koertge [1975] has understood this as a pragmatic strategical advice, which suggests what is more likely to be true, and what would probably lead to the more interesting results. In other words, if we fail to explain the actions of an individual, we should first consider that we did not understand the situation the way the individual understood it, before we conclude he or she acted irrationally.

4.1.2 Rational economic behavior and recursive dynamic programming

While there are certainly situations, where human behavior can be better explained by other (e.g. neuro-psychological) theories, economic, and specifically production decisions generally seem suited to be analyzed in terms of situational analysis. Traditional microeconomic analysis can be subsumed under the concept of situational analysis, although it often uses a very narrow definition of the appropriate decision strategy equating it to profit maximization or cost minimization, and in the purest forms of neoclassical analysis does not consider deviations between individual's perceptions and objective observations of the situation [Hands, 2001].

In many cases, economic multi-agent models are explicitly built in order to overcome these restrictive assumptions of neoclassical economic theory. A more general conceptualization of rational economic behavior that encompasses several different approaches to model human behavior in agent-based models has been developed by Day [2008]. Summarized in short, Day emphasizes that

- a rational decision is always *bounded* by an agent's knowledge and intellectual capacity at the time of the decision;
- rational behavior is *intermittent*, i.e. times of rational and nonrational thinking alternate, and *recursive*, i.e. the agent does not make one plan till the end of his life, but takes and reevaluates decisions at specific points of time, and his behavior evolves over time.
- *strategic* and *tactical* decisions have to be distinguished: *Strategic* refers to decisions on 'lifestyle paradigms' that set a frame for further actions, are taken only from time to time and are not necessarily taken with resource constraints in mind, but rather define what kind of person one wants to be. *Tactical* decisions are decisions on concrete actions, which are restricted by economic considerations, but also by the lifestyle paradigm;
- complex decisions are taken employing *algorithms*, like prioritizing and satisficing, that do not necessarily lead to the objectively optimal solution to a problem.

Day concludes that *recursive dynamic programming* models are the best mathematical analog to this conceptualization of economic behavior. A recursive dynamic programming model is understood here in a general sense as a constrained optimization among pre-specified alternatives with respect to a defined goal that is taken and reevaluated at specific points of time (*recursive*), considers actions to be taken then and their consequences for the future (*dynamic*) and uses a defined algorithm to find the optimal action (*optimize*).

4.1.3 Generalizing recursive dynamic programming models

If we accept the general paradigm of a recursive dynamic programming problem as a model for rational economic behavior, a modeler that would like to completely explain or predict the behavior y of an individual i at a certain point of time t would need to have information on:

- the individual's goals and prioritization of goals at the time of the decision;
- the individual's perception of the range of alternatives available to him;
- the individual's perception of the contributions of these alternative activities towards his goals at the time of the decision;
- the individual's perception of the constraints on potential activities at the time of the decision;
- the individual's strategy to determine the appropriate alternative at the time of the decision.

A generic form for a model for such purposes could be formulated as:

$$y_{i,t} = f(x_t, x_{i,t}, \theta, \eta_i, \gamma_t, \gamma_{i,t}) \tag{4.1}$$

It includes (theoretically) observable, time-variant exogenous variables, some of which are specific to an individual ($x_{i,t}$), others general (x_t). Time-invariant model elements, parameters for short, again some of which are individual-specific (η_i) and others not (θ), as well as unobservable information, which may or may not be specific to the individual ($\gamma_{i,t}$, γ_t).

Prediction of behavior for other individuals and other points of time requires the assumption that the decision is significantly related to the observable characteristics of an individual and his environment at a specific point of time ($x_t, x_{i,t}$), and only insignificantly dependent on unobservable individual- and time-specific deviations. Consequently, the analyst will tend to build a model that relies on observable exogenous variables including observable individual characteristics, and time- and individual-independent parameters. There may be exceptions, e.g. one can imagine a situation where a pre-announced change in tariff rates will cause all individuals to assume a considerably lower price than observed in the previous year, deviating from the normal formation of price expectations. This could be incorporated using a time-specific parameter (γ_t). In other cases it may be possible to induce risk preference parameters from past behavior that can be assumed to be constant over time and then be included as individual-specific parameters (η_i). However, in general, one will have to understand the behavior of an agent as a combination of generalizable patterns of behavior that can be cast as a general model (f) and individual-specific decision patterns that are not accessible to the analyst (h). The linkage function (g) between these two "models" is not self-evident *a priori* (i.e. it is not clear whether it can, for example, be captured by an additive error term).

$$y_{i,t} = g\left(f(x_t, x_{i,t}, \theta), h(x_t, x_{i,t}, \eta_i, \gamma_t, \gamma_{i,t})\right). \tag{4.2}$$

4.1.4 Recursive dynamic programming models of agricultural production and land use

To answer research questions in the context of climate change adaptation as identified in chapter 1.4, researchers are typically interested in modeling farm income (π_i), production output ($\mathbf{q}_i$), land use ($\mathbf{a}_i$) and/or production input ($\mathbf{b}_i$) of a number N of producers or households ($i = 1, \ldots, N$). A typical agro-economic model predicts these outcome variables as a function of *inter alia* weather and other natural conditions (ν), product prices ($\mathbf{p}$), input prices ($\mathbf{w}$), assets and resources of the agent ($\mathbf{r}_i$), the household composition, or ownership structure and permanently employed labor ($\mathbf{l}_i$), the actions of other producers (ζ), and a set of production functions ($g_i(\cdot)$). The goals and the perceived contribution of the production function towards these goals are formulated as a utility function of the agent ($u_i(\cdot)$) [Hazell and Norton, 1986].

Under a recursive dynamic programming paradigm, the determination of the outcome variables has to be understood in at least two steps.[3] In the first step, at the beginning of the season, producers make production decisions intending to maximize their utility based on their expectations of yields, prices, available resources and other producers' activities (denoted by asterisks). In the second step, actual production and income are then determined by nature, markets and the outcome of other agents' behavior [Pope and Just, 2002].

$$[\mathbf{a}_i, \mathbf{b}_i] = \arg\max_{b,a} u_i\left(\pi_i\left(\mathbf{p}^*, \mathbf{w}^*, \mathbf{r}_i^*, \mathbf{l}_i^*, \mathbf{b}_i, \zeta^*, g_i^*\left(\mathbf{b}_i, \mathbf{a}_i, \mathbf{r}_i^*, \nu^*\right)\right), \mathbf{l}_i^*\right) \tag{4.3}$$

$$\mathbf{q}_i = g_i\left(\mathbf{a}_i, \mathbf{b}_i, \mathbf{r}_i, \nu\right) \tag{4.4}$$

$$\pi_i = \pi_i\left(\mathbf{p}, \mathbf{w}, \mathbf{r}_i, \mathbf{l}_i, \mathbf{q}_i, \mathbf{b}_i, \zeta\right) \tag{4.5}$$

Distinguishing the individual elements of the model by observability and specificness, we would assume that natural conditions ν_t like the weather are observable exogenous variables not specific to the agent.[4] Assuming we are modeling only a small region and abstracting for a moment from contract farming, futures, on-farm storage etc., the same would hold for prices ($\mathbf{p}_t$, $\mathbf{w}_t$).

Data on farm resources $\mathbf{r}_{it}$ and household composition as well as ownership structure and permanently employed labor $\mathbf{l}$ may be available from farm surveys or agricultural censuses and constitutes at least partly observable producer-specific information. On the medium to long run, certain resources of producers are, however, not fixed, but also subject to investment decisions, and thus need to be understood as endogenous.

$$\mathbf{r}_i = \arg\max_{r} u_i\left(\pi_i\left(\mathbf{p}^*, \mathbf{w}^*, \mathbf{r}_i, \mathbf{l}_i^*, \mathbf{b}_i, \zeta^*, g_i^*\left(\mathbf{b}_i, \mathbf{a}_i, \nu^*\right)\right) \mathbf{l}_i^*\right) \tag{4.6}$$

Land use $\mathbf{a}_{it}$ and input use $\mathbf{b}_{it}$ and production quantities are endogenous and producer-specific by definition, but are also observable allowing the estimation of crop yield functions (g) according to equation 4.4. Crop growth, as a biophysical process, is certainly a generalizable element of the model, as long as its representation allows producer-specific management and resources and time-specific weather as exogenous input. Ensuring the invariance of crop growth functions under climate change is an essential element for ensuring the invariance of the whole model under climate change.

The same is true for the profit function π, which is not merely a subtraction of costs from returns, but a production function including the technical coefficients describing the potential contribution of production inputs, farm assets, household labor and other resources to farm production, as well as the restrictions applicable to their use. The considerable content of agronomic and technical relationships in the production function makes it potentially generalizable. However, direct estimation of π according to equation 4.5 is usually not an option, because observations of individual accounting data is usually not available, and, more importantly, ensuring the invariance of the model as a whole requires a decomposition of the production functions into invariant parameters (e.g. the amount of land which

[3] In reality, producers may have the possibility to adapt their decisions more often, e.g. after learning more about the state of nature.

[4] They may spatially vary over the plots of different agents, but they do not depend on which agent owns or uses the plot.

can be ploughed with a certain type of tractor) and parts which may potentially change over time (e.g. the availability of the tractor class) and, specifically, with climate change (e.g. the number of days suitable for ploughing).

The part of the model that is potentially hardest to generalize is the utility function u of the individual farm decision-maker, the embodiment of his or her preferences. It can be assumed that somehow income, risk, status, work satisfaction etc. play a role, and it may be argued that the weighting of these factors can be related to $\mathbf{l}$, i.e. family farmers have different preferences to farm business investors, and young farmers different ones from farmers close to retirement without successor. Still, the utility function is certainly a likely source of deviations between a generalized model and individual decisions.

Similarly, the modeler will have to make assumptions, or better, develop a model, to derive the individual expectations $(.^*)$ of variables unknown at the time of the decisions additional to the formulation of the utility and production functions. Pope and Just [2002] show that it is generally impossible to estimate the supply and demand behavior of farmers consistently based on the statistical relationship between observed inputs and observed supply under the assumption of profit maximizing decisions if input choice affects output quantity and producers make errors in optimization, e.g. because output depends on weather and producers do not know weather in advance.[5] As a consequence, any estimation of a crop production decision model implicitly involves assumptions on expectation formation, and the explicit formulation of an expectation model increases model transparency and invariance.

4.2 Calibration: error distributions, likelihood functions and overfitting

The previous section already gave an impression of the complexity and the potentially large number of processes, variables, and parameters that have to be included in order to make a meaningful use of MP-based multi-agent models to simulate climate change adaptation in agriculture. It is not hard to imagine that it may be cumbersome and in many cases outright impossible to come up with prior empirical information to parameterize the model. Moreover, the suitable functional representations for important processes (e.g. expectation formation, risk management) are not clear are *a priori* and in some cases little theoretical knowledge has been established. The modeler can often not avoid making *ad hoc* decisions, or even develop completely new model approaches, leading to considerable model uncertainty.

Is there any chance to reduce at least some of the uncertainty by calibration? As the first of the following subsections (4.2.1) will show, approaches to calibrate MP models traditionally employed in agricultural economics are of little use, when it comes to calibrating MP-based multi-agent models for climate change adaptation. The remainder of the section will therefore discuss, how basic principles and general approaches of calibration presented in section 3.2.1 can be applied to these types of models.

4.2.1 Parameter estimation in traditional MP models in agricultural economics

Mathematical programming models were originally used as normative tools for the development of optimal farm production plans in agricultural economics, due to their capability of reflecting the complex interrelationships between multiple products and partly fixed re-

[5]The only exception is for a quadratic production function.

sources. Based on the assumption that farmers would generally act in a rational, optimizing way, they were then taken up also for descriptive analysis, e.g. the response of agricultural supply to price and policy changes, both for individual (representative) farms and whole agricultural regions, which are often modeled as a single regional farm despite the aggregation bias this entails [Buysse et al., 2007; Hazell and Norton, 1986].

When modelling an individual farm, the modeller has access to enough information to correctly include all relevant relationships into the model. If the model does not reproduce the observed decisions correctly, the modeler can gather further information, often discuss the differences with the farmer and include missing constraints or adapt coefficients to improve the fit. On a regional level, information is usually scarcer and technology and farm resources heterogeneous. Classical linear programming models have certain drawbacks in this situation: The number of activities in the solution is limited by the number of binding constraints. Given few empirically or theoretically justifiable constraints and many different production activities at regional level, linear regional models tend to show overspecialization on just a few crops. Further, solutions show 'jumpy' behavior when decision variables are varied: i.e. for small variations solutions do not change at all, while larger variations cause jumps to another locally stable solution. Consequently, it is rather hard to reproduce observed cropping activities at a regional level using a linear programming model and readily available information [Buysse et al., 2007].

Initially, a common remedy was the inclusion of artificial constraints chosen to make the data fit to the observations. However, these restrictions may be too rigid to ensure realistic model response under changing conditions and bear a large danger of overfitting. The other alternative may be the inclusion of nonlinear terms in the objective function, which may be rationalized by risk behavior, endogenous demand or by increasing marginal costs and decreasing marginal yields [Pope and Just, 1991; Just, 1993; Howitt, 1995].

Nonlinear cost or yield functions can theoretically be justified on a regional level by the extension of production to more marginal soils and less efficient farms as produced quantities grow. Consequently, they incorporate the unobserved heterogeneity that is not captured by the constraints developed from the information available to the modeler. Assuming the observed land allocation is the optimal allocation, Howitt [1995] showed that the nonlinear terms in the full model can be estimated from the dual values of calibration constraints that restrict the solution of a linear programming problem to the observed land use. This two step procedure ensures a perfect fit of the nonlinear model to the data and has found widespread use in agricultural economics under the name of positive mathematical programming (PMP) [Buysse et al., 2007].

In light of the principles of validation and uncertainty analysis discussed in chapter 3, PMP is associated to a number of problems:

- Nonlinear cost or yield functions can only be estimated for activities that have already been observed in a certain region. New production activities can only be parameterized on available information leading to potential inconsistencies if functions for established crops are based on available information and calibration, while those for new crops are only based on the former [Buysse et al., 2007].

- Extrapolation with PMP models rests on the crucial assumption that the conditions captured by the estimated nonlinear terms will not change in the new situation. If nonlinear terms are rationalized as marginal cost or yield functions, this assumption may be justified for analysis of supply response to world market price changes, however, it is hard to defend *a priori* for climate change. Rather, these changes are exactly what researchers are interested in.

A mathematical programming model estimated with the PMP techniques is much closer to an econometrically estimated parametric model than to a theory-based simulation model, albeit with a lack or severe shortcomings in its statistical foundation:

- PMP allows zero degrees of freedom for model error in calibration and only information from one dataset, e.g. one point of time, to be incorporated into the model unless separate dual values are estimated for each point of time [Buysse et al., 2007]: It is almost certainly overfitted to the dataset used to estimate it.

- Dual values that have been estimated in a setting without fixed resource constraints capture all production costs including fixed resources. If these dual values are later used in a full model with constraints, the fixed input factors are reflected twice and the resulting model does not correspond to its theoretical justification [Britz et al., 2003; Heckelei and Wolff, 2003].

To avoid the last two of the problems listed above, Heckelei and Wolff [2003] suggest using the Karush-Kuhn-Tucker optimality conditions of the nonlinear programming problem to estimate the dual and the other model parameters jointly from several observations in one step. They add additive error terms to the observed land use allocation (rationalized as measurement or optimization error) and use a generalized maximum entropy estimator (GME) maximizing the entropy in the probability distributions of the error terms to estimate the parameters. This setup is also able to accommodate further conditions like restricting the dynamic behavior of the model to be consistent with empirically estimated supply elasticities. The later does avoid overfitting on a single observation, but is not suitable for situations, such as climate change, where the estimated supply elasticities on which the calibration is based cannot be treated as invariant.

The idea has been taken further by Jansson and Heckelei [2011] for the estimation of aggregate regional PMP models from time-series of the CAPRI database. They use generalized maximum posterior density estimation with a binary likelihood function that is one if the model result exactly fits the data and zero otherwise. Given the extreme number of parameters, the model only becomes identifiable through the use of informative prior distributions for all parameters including the error parameters. The estimator then finds the combination of parameters and error terms that perfectly fits the data and is most likely according to the prior distributions. This combination of mathematical programming models and econometric estimation techniques has been termed econometric mathematical programming by Buysse et al. [2007] and parallels the development towards the use of statistical error models and likelihood functions to estimate parameters of deterministic models observed in other disciplines.

The problems associated with the original PMP technique make it unappealing for the use as a parameter estimation approach for MP-based multi-agent models, especially for applications analyzing regime shifts like climate change. Moreover, the original problem that motivated the use of nonlinear terms requiring calibration does not really exist in these models. In MP-based ABM, a separate mathematical programming problem is solved for each agent, potentially at many different points of time. Compared to an aggregate regional farm model 'jumpy' behavior and overspecialization constitute less of a problem, because a single farm may in fact react with abrupt shifts rather than smooth adaptation [Buysse et al., 2007] and typically grows only some of the crops grown in the region. Nonlinear terms in the objective function are therefore not needed to smoothen behavior and are only warranted if the theoretical derivation of the objective function does require it (e.g. for risk behavior). Smooth reaction at the aggregate level of the whole region is achieved by the heterogeneity of agents, which causes different agents to react differently in the same situation. The theoretically consistent way to achieve this demands a sufficiently detailed representation of agent heterogeneity and of the mechanisms that transmit heterogeneity to different outcomes [Berger, 2005].

4.2.2 Error distribution

Section 3.2.1 showed how the choice of an adequate loss or likelihood function for calibration needs to be informed by the expected distribution of prediction error. Equation 4.2 showed that a considerable part of prediction error in a generalized model of situational analysis will come from not capturing individually specific preferences and perceptions of reality.

It is not clear *a priori*, if one could express the linkage function g between the generalized (f) and the individually-specific (h) behavior pattern as a simple addition or multiplication and establish that h is truly inexplicable, i.e. unsystematic with respect to any observable variable, such that simple reformulations of the problem such as those below would be possible.

$$y_{i,t} = f(x_t, x_{i,t}, \theta) + \varepsilon_{i,t} \tag{4.7}$$

$$y_{i,t} = f(x_t, x_{i,t}, \theta) * \varepsilon_{i,t} \tag{4.8}$$

Is there any theoretical knowledge on the distribution of errors of microeconomic models? Stanley [1998] and Davis [2004] observe that economic theory has largely neglected theoretical derivation of the properties of expected error distributions and strongly encourage research into this topic. Still, a few efforts in this direction have already been made.

In econometric estimation of demand systems following random utility theory, consumer demand is understood as being derived from the combination of a common 'average' utility function and an individual-specific additive error term, allowing for individuals with the same observable attributes to have differing preferences. Note that the individual is still assumed to be rationally maximizing utility without error, only that the full utility function is not known to the researcher. Lewbel [2001] investigated whether such statistical demand functions can be expected to satisfy the rationality principles of revealed preference theory if it is assumed that individual demand functions are rational. They found that econometric demand functions show rationality and can be understood as 'average' demand functions if the unobserved preferences are not depending on income and prices and any observables that correlate with income have been controlled for in estimation. They also note that – unless unrealistically restrictive assumptions are made – the additive error terms of demand functions will be heteroskedastic confirming a result of Brown and Walker [1989]. McElroy [1987], Pope and Just [2002] and Kumbhakar and Tsionas [2011] investigate the consequences of optimization and measurement error in production functions for the error terms of cost/-cost share functions, respectively input demand functions derived as duals from production functions and find heteroskedastic error terms with nonzero means.

The results of these exemplary studies raise the suspicion that the error terms of the production decision model in equation 4.3 should not simply be treated as white noise. Like the cited authors, one could now try to analytically derive the form of the error distribution of the model from the error distributions of the potential error sources. These would be the error terms ($\varepsilon_p, \varepsilon_w, \varepsilon_\nu, \varepsilon_\zeta, \varepsilon_g$) associated to the expectation models for prices ($\hat{e}_p, \hat{e}_w$) , weather conditions ($\hat{e}_\nu$), the actions of other producers ($\hat{e}_\zeta$) and the crop growth function

($\hat{e}_g$). Further, it would be the error in the production function ε_π and in the utility function ε_u. Following Day [2008], the difference between the optimization algorithm employed by the modeler and the one employed by the farmer $\varepsilon_{\text{argmax}}$ needs to be considered. And, last but not least, the observed quantities might be subject to measurement error (ε_m). What additionally needs to be considered, when comparing trajectories of agent development, is that deviations in one year will transmit into subsequent years, leading to autocorrelated errors over the model run-time.

$$[\mathbf{a},\mathbf{b}] = \arg\max_{b,a} u\left(\hat{\pi}\left(\hat{e}_p+\varepsilon_p, \hat{e}_w+\varepsilon_w, \mathbf{r_i}+, \mathbf{l_i}+, \hat{e}_\zeta+\varepsilon_\zeta, \mathbf{b}, \hat{e}_g\left(\mathbf{b},\mathbf{a},\hat{e}_\nu+\varepsilon_\nu\right)+\varepsilon_g\right), \varepsilon_\pi\right) + \varepsilon_u + \varepsilon_{\text{argmax}} + \varepsilon_m \tag{4.9}$$

Even if defensible assumptions on the error distributions of partial models could be established, an analytical derivation of the distribution of the aggregate error vector over producers is usually not possible because the optimization function is analytically intractable. Especially if the decision model is formulated as a (mixed integer) linear programming problem, the optimal solution can only be derived numerically.

Some qualitative expectations can sometimes be derived based on theoretical considerations: E.g., for a model that does not model risk management in decision making, we may expect a bias towards more risky crops in the simulation compared to observations. In an irrigation setting, models that do not model deficit irrigation can be expected to yield smaller areas of irrigated crops than expected.

Four potential options to deal with the problem of uncertain error distributions for parameter estimation can be identified:

Numerical simulation Using any efficient Monte-Carlo integration technique, one could repeatedly sample from the submodel error distributions, solve the model and observe the resulting distribution of ε_f. The simulated distribution would provide a nonparametric estimate of the distribution of the aggregate error associated with a production decision model of the type used. However, this would involve a large number of model evaluations that could easily surpass the total number of feasible model evaluations: Apart from having to cover the joint error distribution of all submodels, the simulation would also have to cover the full space of possible combinations of exogenous variables and model parameters, as long as one cannot rule out that the aggregate error term is correlated with any of the later.

Direct estimation with disaggregated errors One could follow the approach of Jansson and Heckelei [2011] and abandon the concept of an aggregate error term and estimate the model in the disaggregate formulation of equation 4.9, including individual submodel error parameters for every producer. As mentioned above, Jansson and Heckelei [2011] used a binary likelihood function that was one only for an exact fit and zero otherwise and identified one best fitting parameter distribution by maximizing the posterior density. The problem became identifiable only by using informative prior distributions for the error parameters. In other words, the estimator finds the combination of parameters and error terms that perfectly fits the data and is most likely according to the prior distributions. Individual error terms are estimated for each producer, but would be discarded for further modelling, because there is little theoretical reason to expect them to be constant over time. Apart from the fact that this approach also involves a potentially prohibitive number of model evaluations, confidence in the prior information is not necessarily high (otherwise the motivation for calibration was less strong) and identification of a single best fitting parameter set seems hardly warranted given the high epistemic uncertainty in both the model and the calibration procedure.

Aggregation Instead of looking for a likelihood function for the individual decision models of farmers, one could instead calculate the sum of outcomes of all individuals and make use of the central limit theorem assuming that the errors in the individual optimization problems are identically and independently distributed, and consequently their sum will be normally distributed [Hartig et al., 2011]. Independence of decisions is ensured even if interactions between agents are modeled as long as the individuals take their decision based on expectations formed from past observations of other individuals behavior and the decision outcome can be observed before the next interaction takes place [as e.g. described in Berger, 2001]. Still, any expected bias in the individual decision model would transmit to the mean of the normal distribution. If the outcome variables of interest are land uses, the errors in different land use categories are by definition not independent, because the overall land use area is restricted by the overall agricultural area of the agents. Further, recurring to the comparison of aggregate outcomes strongly reduces the number of identifying conditions in the calibration problem compared to the use of disaggregate datasets. Still, if disaggregate data is not available, the modeler has no other choice.

A Bayesian-like approach with robust, informal likelihoods One could follow the approach of Spear and Hornberger [1980] and formulate 0-1 likelihoods based on acceptability criteria that can be defined based on largely qualitative expectations of model behavior, or robust measures of goodness-of-fit. These measures can then be employed in an informal Bayesian framework: They are used to identify potential parameter settings, which are then treated equally probable for scenario analysis. Similar approaches have been suggested by Brenner and Werker [2007] and Deichsel and Pyka [2009] for the construction of agent-based models in social sciences and economics in general.

4.2.3 Avoiding overfitting

The second important element of a valid calibration strategy besides a suitable error distribution is to avoid overfitting the model to noise in the calibration dataset. A first basic decision for parsimony that helps to attain this goal is to restrict the ambitions to finding a model that only represents the generalizable part of the decision of farmers, as discussed above. Trying to estimate individual parameters or unobserved characteristics for each individual decision-maker amounts to essentially fitting an individual model for each farm, which will hardly be identifiable with the data usually available. In special cases, where individually specific parameters, e.g. risk aversion coefficients, are to be estimated, one can at least regularize them, e.g by restricting them to satisfy an expected distribution in the population.

While the use of the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) will not be possible as long as an explicit likelihood function for the model cannot be formulated, the nature of multi-agent models seems quite amenable for cross-validation: If a disaggregate farm dataset is available the dataset can be used for bootstrapping or other splitting procedures to create calibration and validation datasets. Otherwise, observations of aggregate data at different points of time can be used. Cross-validation does, however, require a nested structure of models with increasing complexity, or at least a stepwise advancing calibration procedure in order to be able to stop calibration when the goodness-of-fit in the validation dataset starts to decrease. One option for stepwise, manual calibration of nonnested models could be the use of a rule that allows reducing the range of parameters only if this improves the fit in all test datasets.

4.2.4 Modularity

As can be seen, assuring suitable conditions for the calibration of the decision model is not straightforward. However, recalling equations 4.3-4.5, the decision model consists of several submodels (e.g. the crop growth function and other production functions), for which suitable conditions may be more easily established. This modularity should be exploited to reduce uncertainty in the submodules and consequently the prior uncertainty of the decision model. Similarly, the production decision problem of individual agents is only one part in a multi-agent model, which often includes agent-agent interactions and the evolving of natural resources over time. Again, if suitable conditions for the calibration of individual processes can be found, these should be exploited [Troost et al., 2010].

4.2.5 Interactive validation

Since the concept of agent-based models rests on the representation of individual decision-making at the farm level, they lend themselves to interactive modeling sessions with experts, stakeholders and especially farmers themselves. Core functions and restrictions underlying the model, and predictions of behavior for typical, exemplary situations can be discussed in face validation workshops with farmers. Such sessions provide qualitative feedback on the realism of assumptions and model behavior and potentially point to important processes and restrictions still absent from the model [Berger et al., 2010]. Turing tests or interactive simulation experiments may be one element of validation workshops.

4.3 Model documentation

Agent-based models can get very complex and often not be summarized with a few equations. A standardized description sequence can help the reader understand the model and support the modeler in formulating a complete documentation, or even provide a starting point for formulating a new model [Grimm et al., 2010].

Grimm et al. [2006] have developed a standard protocol (ODD) for individual and agent-based models. The acronym ODD stands for Overview, Design concepts and Details and refers to the three main sections of the protocol. The overview section is intended to allow the reader a quick understanding of the purpose, entities and major processes of the model. The second section documents, how key concepts of the field of complex adaptive systems (e.g. heterogeneity, emergence, interaction) have been considered in the design of the model. The details of the implementation (e.g. equations, algorithms, data) are provided in the eponymous third section.

While the original ODD was developed for individual-based models, the ecological counterpart of agent-based models, Müller et al. [2013] developed an adaptation for the social sciences, which facilitates the representation of human decision making.

4.4 Conclusions

In light of the discussion of the basic principles of model validation in chapter 3 and their application to mathematical programming-based multi-agent models in this chapter, and especially after highlighting the difficulties in establishing convincing error distributions and calibration techniques, a major conclusion for constructing valid agent-based models can be formulated using the words of Browne:

> "No mechanical data analytic procedure for evaluating model fit should [.] either replace or override human judgment." [Browne, 2000, p.110]

Human judgment requires critical review by others and a comprehensive model documentation is a basic requirement for critical appraisal. The ODD+D protocol provides an increasingly accepted structure for the description of the model itself. However, the documentation should also cover critical topics including the invariance of the model, the procedures used for calibration, the predictive accuracy of the model, and ultimately also a quantification of uncertainty and its implications for model conclusions.

Uncertainty can potentially be reduced by calibration, although in most cases theoretically well-founded measures of goodness-of-fit will not be available, model bias is likely and the complexity of models causes a real danger of overfitting. A reduction of candidate process representations and parameter combinations to a single, best-fitting model is therefore not warranted. On the long run, numerical experiments simulating the potential error distributions of models would help further the knowledge on expected error distributions. On the short run, however, a Bayesian approach that identifies a number or distribution of candidate models using qualitative likelihood functions based on stylized facts or robust goodness-of-fit measures and is possibly combined with cross-validation should be the method of choice.

Nevertheless, the current difficulties presented to empirical validation and calibration by lack of data, lack of knowledge on the error distribution and the danger of overfitting should not discourage the comparison of model predictions to observations. Observed biases in the model can point to unrecognized errors in the implementation or the omission of important processes and the modeler is well advised to thoroughly analyze the model. This includes the use of Turing Tests and other forms of interactive validation.

Given the difficulties, no automatized procedure can be recommended to modelers. Rather, they will have to repeatedly run the model and evaluate the results themselves to iteratively improve the model. After correcting an error or including a new process, calibration will have to be repeated starting with the full set of candidate models in order to ensure a truly global assessment. The use of efficient experimental designs and screening procedures will greatly facilitate these procedures.

Due to such a cautious use of calibration, the remaining uncertainty will be high, represented by many candidate models or a large parameter space. For scenario analysis, screening techniques and experimental designs can be employed to efficiently represent this uncertainty in a feasible number of model runs. Simulation results are then presented as ranges or distributions over the uncertainty space. Conclusions can be assessed with respect to their robustness to different parameterizations. If they are not robust, sensitivity analysis can be employed to determine the parameters with the highest influence on the conclusions in order to concentrate research efforts on the processes associated to these parameters.

Chapter 5

Enhancing the MPMAS modelling framework

MPMAS is an agent-based modeling software in the agricultural economics tradition of recursive farm modeling and adaptive microsystems [Schreinemachers and Berger, 2011; Berger and Troost, 2012]. The main class of agents in MPMAS are farm households, each of which runs through a typical sequence of actions in each cropping season: Based on past experience and available information, the agents form expectations about future conditions (e.g. prices, yields). Based on these and their knowledge about their current situation, the agents first decide on investments into assets (e.g. machinery, stables, etc.), and then on the production plan for the coming season. The actual physical and economic outcomes of production are determined and the agents react to the observed outcome deciding on the usage of produce and income, whether to sell assets to retain solvency, and whether to continue farming or leave the agricultural sector.

The core element of the model and the major link to the agricultural economics tradition is the use of mathematical programming to represent and solve the decision problems of farm agents. This basic setup can be enhanced by a number of modules to represent interactions between agents (e.g. markets, innovation diffusion, information) and between agents and their biophysical environment (e.g. irrigation and hydrology, crop growth and landscape models). These interactions will usually influence both, the outcome of the agents' decisions as well as the expectations they have while planning.

5.1 Challenges of modeling climate change adaptation with MPMAS

In general, MPMAS as a modeling framework seems well-suited for studying climate change adaptation (cf. section 2) It implements the recursive-dynamic programming paradigm and allows flexibility with respect to the choice of the utility function, knowledge and expectations of agents and the level of detail and scope of the conditions of agricultural production. It includes investment decisions, demographical development and land markets and provides interfaces for the integration with biophysical models [Schreinemachers and Berger, 2011].

While the representations of risk management and learning that have so far been used in MPMAS can certainly be improved, a more fundamental problem pertains to the validation, calibration and uncertainty analysis of MPMAS models. Modelers have so far used regressions of simulated on observed values as simple goodness-of-fit measure at various levels of aggregation and comparison with stylized facts for empirical validation [see e.g. Schreinemachers et al., 2009, 2010]. Uncertainty analysis has mostly been confined to testing different initial populations. Progress towards the more comprehensive validation frameworks advocated in the previous chapter has so far been hampered by technical obstacles presented by the established interfaces, tools and procedures for working with MPMAS:

Pre- and postprocessing for massive simulations The core of the MPMAS model is the `mpmas` executable. The executable reads input data from plain text files and produces plain text files as output. The content of these files is purely numerical and follows a fixed format. All numbers receive their meaning only by the position in the file. While this format improves the efficiency of simulation runs, it makes it impossible for the modeler to directly work with the files in MPMAS format if the model application has any meaningful size.

In the established preprocessing procedure, a set of Excel spreadsheets containing a commented version of the input file format provides the main interface for model preparation and a Visual Basic for Applications (VBA) macro (`mpmas.xla`) is used to transform the Excel content into plain text files in MPMAS format. This setup, further referred to as *MpmasExcel*, has certain drawbacks that apply especially for large and complex models, massive numbers of simulation runs, or frequent revisions of the model setup:

- The Excel spreadsheet content follows the format of the input files, which is optimized for the internal memory structure of MPMAS, but not for an efficient data processing by the user. Information referring to model entities (e.g. activities, constraints) is spread over different files and most often linked via references based on the position of the element in a list or matrix. Any addition or deletion of a model element will therefore trigger updating of links in many different input files, of which users have to take care themselves. Certain elements (e.g. selling activities) also have a required position that has to be respected while updating.
- The VBA conversion macro works quite slowly and may run into memory problems, when the size of the decision model grows very large. It is very cumbersome and error prone to work with Excel spreadsheets with several thousand columns and rows.
- The MPMAS model offers a lot of different modules and features, many of which users will not employ for their specific application. Nevertheless, the input format expects default input information for these modules, confusing especially new users.
- The *MpmasExcel* setup is tied to the MS Windows operating system, while Linux is the preferred operating system for MPMAS, especially for massive simulation runs.
- Post-processing of output files is usually performed using Stata scripts, which users adapt for their own applications and need to update every time the model is updated.

Sampling of agent-populations Agent populations for MPMAS simulations are usually randomly generated based on cumulative distribution functions, which are traditionally combined with clustering and constraints in order to achieve consistent joint distributions of different agent characteristics. The `mpmas` executable itself offers a lottery feature which is able to sample characteristics from individual (marginal) distribution functions for clusters. Constraints on asset distribution (e.g. ensuring ownership of a perennial plantation is only assumed for agents that have the water rights to irrigate it), on the other hand, are hard-coded and thus require intervention by a programmer for any change. Joint distributions between variables within clusters cannot be represented.

Solver performance The `mpmas` executable uses the IBM OSL solving library to solve the mixed integer programming problems representing agents' production decisions. An efficient solving of large decision problems with many integer variables requires an optimal use of the pre-solving tools offered by the library. The `mpmas` executable makes only rudimentary use of these features, leading to high solving times and even crashes.

5.2 *MpmasMySQL*: A new set of tools for pre- and postprocessing for MPMAS simulations

In the course of the work for the present thesis, a new software toolbox for pre- and postprocessing of MPMAS simulations, including the sampling of agent populations, has been created that provides a full alternative to the *MpmasExcel* setup. As it is centered around MySQL databases for data storage, the model setup was given the name *MpmasMySQL* (though the use of MySQL is facultative: tables in plain text files can also be used, but provide less flexibility).

The toolbox consists of three tools:

`mpmasql` is used to prepare input files for MPMAS or `mqlmatrix` as well as Stata/R scripts to import result files;

`mqlmatrix` is a console-based interface to view and solve MIP problems intended for testing and debugging decision problems of individual agents;

`mpmasdist` can be used to create agent populations based on statistical distribution functions and theoretical assumptions, including the random spatial distribution of agricultural plots.

Figure 5.1 provides an overview of the modeling process using the *MpmasMySQL* toolbox. Any empirical information gathered in the field or from secondary sources is stored in a relational database (e.g. MySQL) in a structured and normalized form. Spatial data should be processed and stored using GIS software and linked to the database. Often detailed information about all the individuals to be represented as agents in the model is not available. In those cases, `mpmasdist` is used to create synthetic, but representative agent populations using statistical information and heuristic rules.

The design of the MPMAS model is laid out in four text files defining the model structure and containing references to information in the database, but not the data itself. For input preprocessing, `mpmasql` combines the model design in the configuration files with the information stored in the database to create the MPMAS input files. At the same time, it generates import scripts for Stata and R, which can be employed by the user to automatically import the simulation results into the respective statistical package.

All of the tools are written in Perl 5 for Linux and together with the use of the statistical software R allow a full simulation cycle based completely on open-source software, allowing a wider use of the model and improving the reproducibility and transparency of simulation results.[1]

5.2.1 `mpmasql` - a new user interface for MPMAS input files

The concept for the new routines for handling model input was developed having the iterative nature of simulation analysis in mind, which – in our view – consists of a continuous process of learning. It usually proceeds by the stepwise enhancement of a model from a simple, basic version to more complex implementations, and likewise from scarce, often assumed information to representative, complete datasets. Such a process involves many revisions and enhancements of the model setup and a frequent addition of data or submodules.

[1] Currently still with the exception of the IBM OSL solving library, which is proprietary, but will be replaced with the SYMPHONY package in the future.

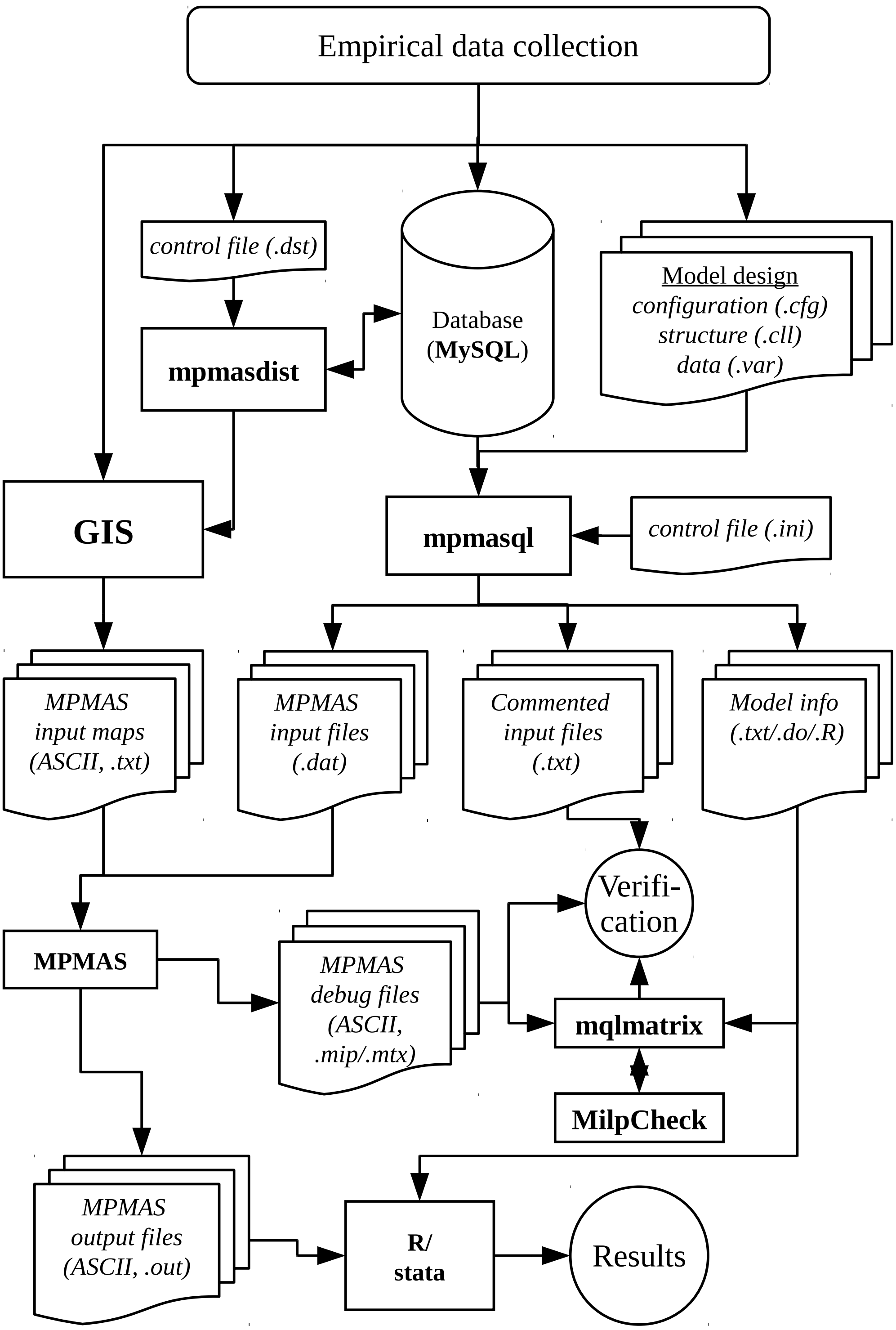

Figure 5.1: *Overview of the MPMAS modeling process with the MpmasMySQL setup.*

Concepts

Usually, data gathered for simulations is not used directly in the model, but undergoes several transformation steps (e.g. selection, discretization, rounding, aggregation, statistical analysis, formatting) and often assumptions have to be made to fill data gaps. Any new data or revised assumptions will require to go through the full process of transformation again. It is therefore preferable to store the data in the original format and additionally store the transformation instructions, so new model input can easily be updated, when either of the two is changed. This separation of data and transformation instructions is one of the core concepts underlying the *MpmasMySQL* preprocessing setup.

A second core concept is the use of classes borrowed from object-oriented programming: Information related to a real-world object, e.g. a 55 kW tractor, is often associated to several different elements of an MPMAS model, e.g. an investment activity and a capacity constraint in the decision model and an asset in the agent's account. The structure of these model elements is usually similar for the members of a class, e.g. the class of tractors, or the class of machinery. The differences between members arise mainly due to different values for the same parameters. In `mpmasql`, the element structure is defined once for each class and filled with references pointing to the specific data for each class member. Third, `mpmasql` works with persistent, alphanumeric identifiers that do not depend on the order of elements in the MPMAS input files. `mpmasql` takes care of positioning model elements in the MPMAS input files and updating numeric, position-related links. Fourth, `mpmasql` strives to be user-friendly in the sense that it requires input only for submodules actually used by the user and provides hidden default values for all other submodules. This also means that updates to the `mpmas` executable require less maintenance work for model users: Any changes in the MPMAS input format that do not directly concern input used by the modeler require no changes to `mpmasql` configuration files. The necessary changes are implemented in `mpmasql` by its maintainer and the users will only have to update their installation of the `mpmasql`-toolbox to use the new executable.

Technical realization

`mpmasql` consists of a collection of Perl 5 modules[2] for Linux. This section presents only a short overview of its working principles. A user manual including a small tutorial is delivered with the tool.

Configuration files Using the preprocessing algorithm requires the generation of four configuration files. The control file (.ini) provides the names of the other configuration files, the information necessary to connect to the database, and the paths were output files are to be written. It further lists the names of the scenarios to be created as well as the changes to the basic model setup to be applied for each scenario. The configuration file (.cfg) contains basic information on the model setup, like the number of simulation years, the number of spatial subdivisions of the model area, the path to map files, and the submodules used.

The structure file (.cll) defines the core of the user-defined part of the model structure, i.e. all the elements of the production decision matrix and related links to farm assets, market prices, crop growth models, *et cetera*. It uses a class-based structure: each class is made up of model elements (e.g. activities, constraints, assets) with a list of attribute fields (e.g. the type of equality sign of the constraint or the coefficient of an activity in a certain constraint). The definition of classes, elements and attributes is given in a special-purpose functional

[2]More specifically, it requires version 5.10 or higher.

macro language. This macro language provides placeholders for instance and element identifiers, references to values stored in tables, basic arithmetic operations and control structures, which can be used to distinguish attribute values and even the number and names of attribute fields or elements between class members.

The fourth file, the data file (.var) provides the data references necessary to fill the model structure. It contains three types of definitions: (i) parameters and data in table format, which are more or less communicated to MPMAS as is, and have to follow a pre-specified format; (ii) variables and tables declared and defined by the user, which can be referenced in the structure file, and (iii) the lists of members (instances) for each class defined in the structure file. Data may be provided directly in the file, calculated from previously defined variables, or formulated as a query to the database, which is then used to retrieve data during the conversion process. All of the data definitions may use functions defined by the `mpmasql` macro language. As a consequence, users have full flexibility in designing their database, even for input expected in specific formats. They are free to use any database structure as long as they are able to create the required table format with one SQL query.

Conversion process Figure 5.2 shows the steps performed by `mpmasql` during the conversion process. At the beginning, the model control file and subsequently the other configuration files are read and the connection to the database is established. The class, variable and table definitions are loaded into memory and the macro language is translated into Perl code with place holders. Then the program starts to loop over the given scenario list and the following steps are performed for each scenario: First, all data definitions for parameters, variables, tables and instance lists are updated with scenario specific instructions if necessary. The user-defined tables are then created in memory by retrieving data from the database using the updated queries. The program then loops over all instances of every class and prepares all model elements and their respective attributes by evaluating the prepared Perl code after replacing the place holders with the respective instance, element and attribute names. The Perl code will usually be used to calculate attribute names and values from the user-defined variables and tables.

After the core model elements have been created in memory in this way, `mpmasql` loops twice over all type of MPMAS input files it has to create. During the first loop a file object is created and all preparations, which also concern other file types are realized. For example, for `MILP.dat` the activities and constraints of the decision matrix are sorted and numbered, and for `Network.dat` the asset objects are sorted and numbered. In the second loop, the structure of the file is created, filled and written to the output folder. If requested by the user, a commented copy of the file is also written. After all files for all scenarios have been created, `mpmasql` creates a batch file containing the program invocation for each scenario and, if requested, an import script for R or Stata.

5.2.2 `mqlmatrix` - a tool to test decision problems

`mqlmatrix` provides an interface to browse, temporarily change and solve decisions matrices using the standalone solver of MPMAS (MilpCheck). It allows modelers to check the decision matrix for errors and test corrections with individual decision problems, without having to run the full `mpmasql` and `mpmas` modeling cycles. Although the modeler could open the matrix (or its commented version) in a spreadsheet program, MPMAS matrices can easily grow large and hard to handle as a whole in spreadsheets. `mqlmatrix` can read a number of different matrix formats that can either be produced as debugging output of `mpmas` or directly created using `mpmasql`. As all of these files only contain numbers and no descriptions,

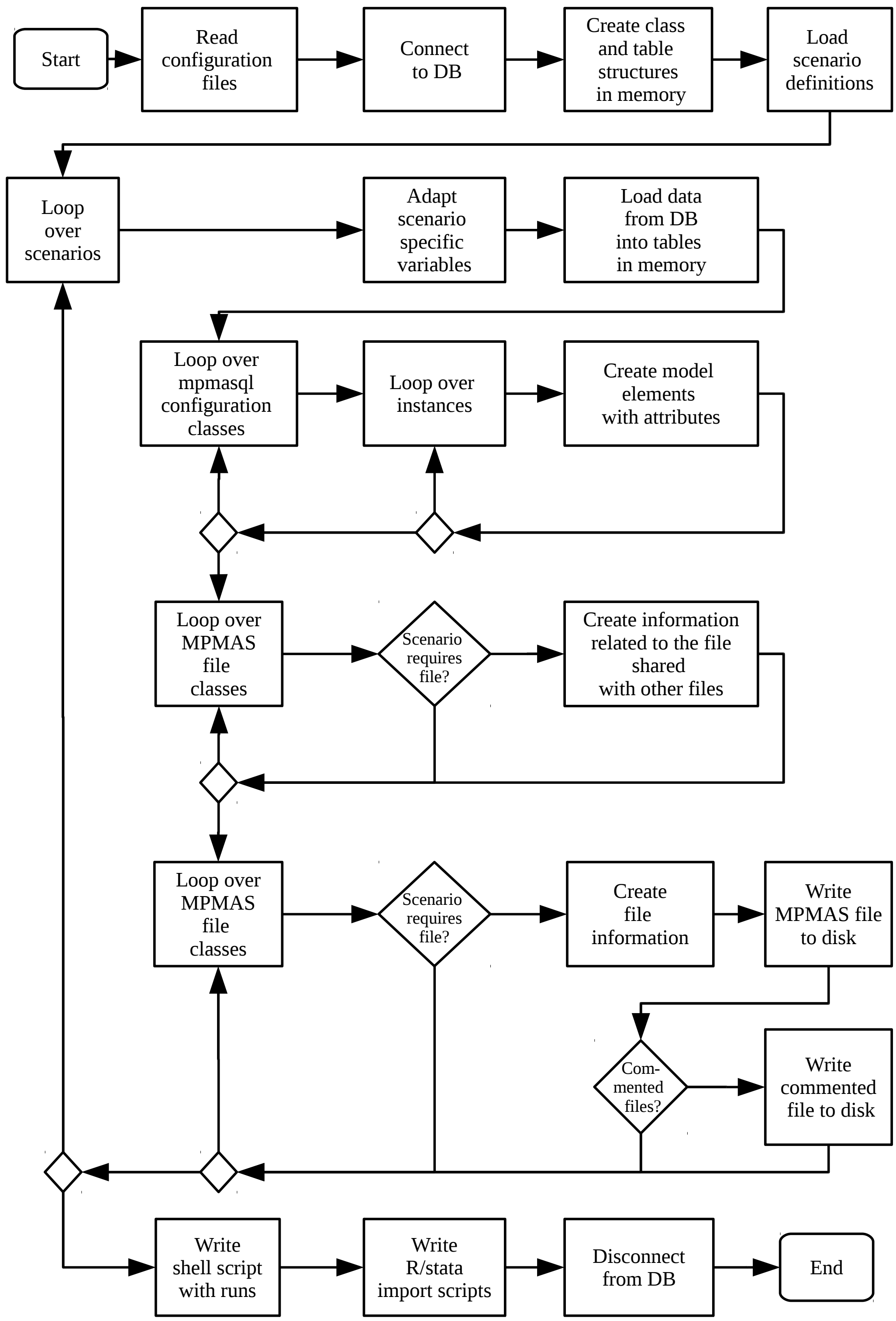

Figure 5.2: *Flowchart of the MPMAS file creation process performed by* `mpmasql`*.*

`mqlmatrix` relies on two special files automatically produced by `mpmasql` for user information on activities and constraints. These files can also be easily created by hand, allowing the use of `mqlmatrix` also in the *MpmasExcel* setup.

5.2.3 `mpmasdist` - a tool to initialize agent populations

`mpmasdist` is a flexible tool to initialize agent populations. It allows the user to sequentially build up agent populations by combining sampling procedures and rules-based allocation. These populations can then be read directly by `mpmas`, instead of using the in-built lottery feature of the executable. This allows the separation of agent creation from the main `mpmas` executable ensuring consistency of agent populations over different scenarios and granting the user full flexibility in the application of suited rules and sampling procedures without requiring the interference of a programmer to `mpmas` code.

The initialization of agent populations can be subdivided into three phases:

1. Create an agent population assigning values of observed characteristics based on statistical information, i.e. recreate a sample from aggregated information.
2. The distribution of agents' farmsteads and plots in space.
3. Transform observed characteristics into model variables, including the inference of unobserved variables.

Recreate statistical and theoretical properties of populations

Usually, empirical agent-based models are employed with synthetic, but representative populations, i.e. the agent population does not correspond to an exact observation of each individual in the real-world population, but it is artificially created ensuring that its statistical properties correspond to the statistical properties of the real-world population [Berger and Schreinemachers, 2006]. This is usually done due to a lack of data on the full population, e.g. when the statistical properties of a population have to be inferred from a small sample interviewed in a farm survey or when privacy restrictions prevent the one-to-one use of statistical data. Even if full-resolution data are available, modelers may still prefer synthetic populations in order to allow for a generalization of results over the specific population observed.

Recreating the statistical distribution of an agent attribute in a synthetic agent-population is straightforward using a standard procedure in Monte-Carlo Simulations called *inverse transform method* [Law, 2007]: The modeler estimates empirical cumulative distribution functions from the real-world population and uses their inverse (the quantile function) to distribute values according to random draws from a uniform distribution [Berger and Schreinemachers, 2006]. It is more challenging, however, to consistently recreate the covariance structure of different attributes in a population. A neglect of covariance between attributes, however, will bias the resulting distribution and consequently the model outcomes [Saltelli et al., 2004]. Estimating and sampling from parametric joint distribution functions is often complex or not possible [Law, 2007] and modelers will tend to use non- or semi-parametric approaches like hierarchical trees, clustering or copulas.

In hierarchical trees, variables are ordered. While the probability distribution of the first variable is explicitly given, the probabilities of subsequent variables are expressed as functions of previous variables [Law, 2007]. When using clustering as suggested e.g. by Berger and Schreinemachers [2006], agents are classified according to the variables that show highest correlation with all other resources, potentially using common clustering algorithms.

(Factor analysis can be used to reduce the dimensionality of the problem.) Distribution functions for each attribute are then estimated within each cluster, i.e. attributes are treated independently within each cluster. As a third option, copulas [Schweizer and Sklar, 2011] can be used to express the relationship between the uniform abscissa of the quantile functions of several variables. Both, the copula itself and the marginals, can be parametric or non-parametric in all thinkable combinations. To recreate the distribution, a sample is then first drawn from the copula and then translated into values of the variables by reading from the marginals. The choice of a specific approach will very much depend on the characteristics of the population and the available data.

Apart from stochastic covariance between variables, there may also be deterministic theoretical constraints that have to be respected during the assignment of agent characteristics in order to avoid inconsistent agent populations. These theoretical constraints can be used in cases where joint distributions of variables cannot be observed (e.g. if distribution functions for water availability and apple plantations come from different sources, apple plantations can be constrained to be allocated only to agents that have enough water to sustain them), but they may also be important if a joint distribution function is known: For example, shares of land classes need to add up to one or legal restrictions constrain the number of animals based on the size of the land holding. The restrictions provided by the estimated joint distribution may not be strict enough to ensure these relations. For example, it may assign a probability of 80% for the share of arable land to lie between 40-60%, and of 20% for it to lie between 0 and 40%, if the share of grassland lies between 30 and 50% of the total agricultural area. If the grassland share of an agent was randomly determined to be 45%, arable shares of more than 55% could occur according to the estimation, although they are clearly impossible from a theoretical point of view.

In the traditional approach used in MPMAS, theoretical constraints are implemented using rejection sampling. The `mpmas` lottery algorithm loops over each agent and each variable, and independently draws a value from the distribution function of a variable in the agent's cluster. The value drawn from the distribution function is tested for compliance with the theoretical constraints. If it complies, it is assigned to the agent, if not, a new value is drawn. This procedure is repeated until a suitable value has been found (or a pre-specified maximum number of iterations has been surpassed). As a consequence of this algorithm, the distribution in the agent population is biased towards 'less demanding' characteristics, because in effect the values are drawn from truncated distributions. A mixture of rescaling of input distributions and rejection of too inconsistent samples largely based on trial and error was used in the past in order to ensure consistency of populations [Berger and Schreinemachers, 2006; Troost, 2009]. Still, filtering of populations using statistical tests does not overcome this problem: It will only exclude those samples, whose likelihood of being a good representation of the true population falls below a certain threshold, but not correct for the bias in the accepted populations.

One way to overcome this problem is to first draw N values from the estimated distribution ensuring that the whole of the distribution is properly represented and then randomly distribute these values among the agents respecting the specified theoretical constraints. In case a single, one-sided theoretical constraint has to be respected (e.g. the value assigned to the agent has to be smaller than a certain characteristic of the agent), the random allocation can follow a simple algorithm:

1. Order the drawn values from 'most demanding' to 'least demanding', i.e. in ascending order in case of a greater-than and in descending order in case of less-than constraint.

2. Starting from the 'most demanding', one can then randomly assign each value to one

of the agents for whom the constraints are fulfilled and who has not been assigned a value yet.

In case of a complex set of constraints that define both upper and lower bounds, it will be necessary to use a matching algorithm. For example, the well-known Hungarian Method or Kuhn-Munkres algorithm [Munkres, 1957] could be used in combination with a random cost component: Each potential combination of an agent with a value is associated with a cost. This cost consists of two parts: a deterministic and a random component. The deterministic component should be zero, when the constraint is fulfilled and positive if the constraint is not fulfilled. The cost can also be used to reflect the severeness of a bad match, in order to prefer slight violations of constraints over more severe ones in case a complete fulfillment of constraints is infeasible. The second cost component should be a random value that is comparatively small compared to the deterministic part, such that it does not overrule constraint penalties. It is added to the deterministic component and ensures a unique and random solution of the matching problem, which usually contains many feasible value matches for many agents.

In the newly developed `mpmasdist`, agent populations can be sequentially created and endowed with attributes. In each step, different allocation mechanisms including fixed rules and sampling from probability distributions can be combined with an arbitrary number of constraints and a selected algorithm, including the traditional `mpmas` algorithm, the simple order-based algorithm and the Hungarian matching algorithm.

Spatial distribution

In most applications, the modeler will want to determine the spatial location of farms and their land holdings in a map of the model area. Again this will often have to be done by some kind of random allocation, while considering that only part of the area is actually a likely place for a farmstead or plot.

`mpmasdist` implements a two stage approach: In the first step, farmsteads are randomly allocated over the area declared suitable by the modeler. Each eligible, unoccupied plot is equally likely to become the farmstead of an agent. In the second step, plots are randomly allocated again only considering area declared suitable by the modeler. To ensure realistic patterns of plot allocation, the algorithm repeatedly loops over all agents, each time only attempting to allocate a part of an agent's land as close a possible to the agent's farmstead or existing plots creating patterns with a realistic balance between a completely scattered, random allocation of plots over the area and a total agglomeration of plots around the farmstead.

Transform observed characteristics into model variables

The third step is rather specific to the given dataset and the conceptualization of the model for a specific application, and is not discussed in detail here. The modeler will usually employ the 'fixed' allocation routines of `mpmasdist` for this task.

5.3 Solver optimization

The `mpmas` executable was adapted to allow a flexible use of the pre-solving tools offered by the OSL library. Users can now increase the efficiency of solving by adapting the sequence of pre-solving tools that produce the best results for their mixed integer programs, in addition to the priority and pseudo-costs settings already provided by earlier versions. The new feature also allows logging the computational time required by different pre-solving tools and the main solver to help the modeler identify a suitable solving strategy.

Based on personal experience, I would suggest a new default configuration that uses scaling, crashing and the LP pre-solver as a standard. If any of these fails, it is simply skipped in a second attempt. If the LP pre-solver is used, post-solving is necessary, but may cause problems. `mpmasql` will try both post-solving modes. If both fail the problem is re-initiated and solved without LP pre-solver, but with the MIP pre-solver.

Further, the OSL library very infrequently causes segmentation faults that are not systematically related to the input data. When solving many, complicated problems, this will generate uncontrollable failures of individual scenarios during simulation. As the library is closed-source legacy code, there is no way to correct the error. Instead, a signal handler was created that circumvents the problem by catching the segmentation fault and re-initiating the solver allowing simulation to continue.

5.4 Using MPMAS on the `bwgrid` cluster

The bwGRiD [2013] initiative offers grid computing facilities to members of 10 universities in Baden-Württemberg. `mpmas` was successfully run on these computers allowing simulation of several hundred scenarios in parallel. A collection of batch scripts helps efficiently using this resources by automatically queuing simulations on a requested number of nodes and processors of the bwgrid and importing output files for runs listed in the batch scripts and R import scripts created by `mpmasql`.

Part II

Agent-based modeling of climate change in practice: the MPMAS Central Swabian Jura model

Chapter 6
Research context and objective

The methodology discussed and the technical solutions developed in the first part of the present thesis are applied in a case study in the Central Swabian Jura. The research is part of a larger effort of scientists at the University of Hohenheim and the Helmholtz Center Munich to improve the understanding of potential effects of climate changes on agricultural landscapes and farming. To this end, the project DFG FOR 1695 'Regional Climate Change'[1] integrates the work of meteorologists, geophysicists, soil and plant scientists and agricultural economists, who conduct field measurement campaigns, laboratory experiments and farm surveys and develop or enhance simulation models for relevant biophysical and economic processes. In a sequence of several steps, the disciplinary models will be combined into an integrated land model system. The research project is conducted in two different study regions, the Kraichgau and the Central Swabian Jura, to allow a comparison of results between two rather different landscapes and agricultural sectors. Agro-economic research in the project focuses on the potential and probable adaptations of farmers to changing climatic conditions. Adaptations to observed weather applied in crop management during the season are evaluated using the FarmActor model [Aurbacher et al., 2013]. Experimental economic approaches and surveys are employed to examine farmers' strategies to deal with production uncertainty and to understand processes of learning and anticipation of future conditions.

The present work contributes to the project by establishing an agent-based model of agricultural production in the Central Swabian Jura using the MPMAS modeling framework. This model will finally constitute the socioeconomic component of the land model system and to this end will be coupled with an integrated atmosphere-land surface-crop model. It will form the basic framework to incorporate the modeling approaches for crop management adaptation, learning and risk management that are being developed within the project context. The objective of the present work is to design a recursive-dynamic agent-based model of year-to-year agricultural production and investment decisions on the Central Swabian Jura and show that the model is suitable for the analysis of climate change adaptation and environmental policies.

After a short introduction to the study area and the possible impacts of climate change expected to affect it in the remainder of this chapter, chapter 7 describes the design of the model. To model the effect of climate change on crop yields in the area, MPMAS has been integrated with the modeling package Expert-N, which models soil and plant processes at plot level (see section B.4). Apart from yield effects, the MPMAS model was designed to reflect several other pathways in which climate change may alter conditions for farming in the area, including changes in available days for fieldwork, changes in crop rotation options, price changes and policy adaptations. The model is parameterized based on secondary datasets as well as a farm survey and expert interviews. Following the discussion in the first part of this thesis, parameter uncertainty is reduced in a conservative calibration approach comparing model data to three observation years and the model is subjected to a Turing test

[1]including its predecessor, the DFG Integrated Project (PAK) 346 'Structure and Functions of Agricultural Landscapes'

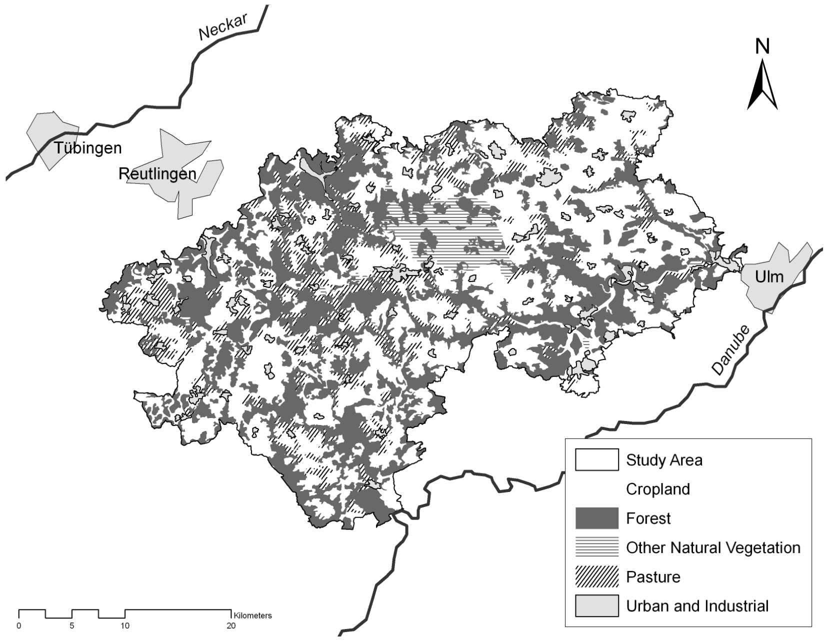

Figure 6.1: *Study area (land use adapted from CLC2006)*

(chapter 8). Screening and experimental designs help addressing and communicating the remaining parameter uncertainty.

The usefulness of the model is demonstrated by analyzing the potential impact of climate change on short-term production decisions under different price scenarios, specifically focusing on the relative importance of the three pathways of climate change impacts (sections 9.3 - 9.5). Illustrating that the analysis need not be restricted to a few price scenarios, section 9.6 examines the impact of the climate change scenario on regional agricultural supply functions. Section 9.7 analyzes the interactions between the biogas support provided by the German Renewable Energy Act (EEG) and the agri-environmental policy scheme MEKA providing an example for potential conflicts between a policy measure motivated by climate change mitigation (the EEG) and more general environmental policy goals, such as the ones incorporated in the MEKA scheme. Chapter 10 finally examines the stability of results when extending the analysis to recursive-dynamic simulations and chapter 11 discusses the findings.

While some of the design choices during model development have to be understood in the context of the wider project, the work presented here can and should also be seen as a case study of analyzing the effects of climate change on agriculture in its own right using an agent-based model of agricultural decision-making and relying on the methodologies discussed in the first part of this thesis.

6.1 The study area: Central Swabian Jura

The Central Swabian Jura is part of the South German Scarplands and located between Stuttgart and Ulm, roughly between about N 48°12′, E 9°7′ and about N 48°34′, E 9°57′. It is a hilly plateau sloping from a high escarpment in the Northwest towards the Danube river, with heights ranging between 650-850 m a.s.l. The area is characterized by shallow soils over Jurassic limestones and a harsher climate (mean annual temperatures around 7 °C, mean annual precipitation 800-1000 mm) compared to surrounding landscapes.

The whole area comprises about 1,300 km^2. According to the 2006 CORINE datasets [CLC2006, 2009], about 54% of the area showed some agricultural land use, about 37% were covered by forest, 5% were urban or industrial areas and 4% were natural grass-, heath- or marshland. The agricultural land can further be subdivided into arable land (22% of total area), pastures (11%), mixed cultivation patterns (18%), and mixed agricultural and natural area (4%). Soils of the area are mainly Rendzic Leptosols, Cambisols, Luvisols, Regosols and Anthrosols (see tab. C.2 in the appendix). Administratively, the region is subdivided into 27 municipalities, eleven of which belong to the Alb-Donau district, while the others are part of the district of Reutlingen. The study area covers about 78% of the area of the district of Reutlingen and about 33% of the Alb-Donau district.

6.2 Agriculture in the study area

Statistical information about the agricultural sector of the area was derived from the FDZ 2010 panel dataset, which integrates the Agricultural Census of 1999 and the public Farm Structure Surveys of 2003 and 2007. For the following description, only those farms in the panel that have their farmstead within the municipalities that belong to the study area were considered. This does not necessarily mean that the aggregated land use of these farms is equivalent to the land use observed in the study area, as these farms may cultivate plots outside the study area and farms from outside may cultivate plots within the study area.

Table 6.1: *Farms in the study area*

	FADN Class			
Year	Not represented	Part-time	Full-time	Total
1999	993 (40%)	534 (22%)	933 (38%)	2460
2003	888 (43%)	559 (27%)	606 (30%)	2053
2007	812 (46%)	432 (24%)	533 (30%)	1777

Table 6.1 shows the number of farms registered in the study area in each observation year. There has been a constant decline in farm holdings by nearly 30% in the eight years mirroring the general trend of declining farm numbers observable in Germany [destatis, 2012*d*] and most of Western Europe [Breustedt and Glauben, 2007]. In the table, farms are classified according to the Farm Accounting Data Network (FADN) scheme, which distinguishes full-time farms, part-time farms and those that are not included into the FADN. Unfortunately, the classification methodology changed between 1999 and 2003, introducing a structural break into the data.[2] The classification rules valid in each year are shown in tab. D.1.

Table 6.2 shows the agricultural land use in the study area. Winter wheat and summer barley are the crops with highest area share, followed by winter barley and winter rapeseed. Overall, the crop distribution appears quite stable during the eight years. As a major change, a near-doubling of silage maize areas and a strong decline in the area of fallow and other crops can be observed. Minor changes are a reduction of summer barley area in 2007 and an increase of winter wheat area in the order of 6-7%.

With respect to animal production, table 6.3 shows a decline of cattle husbandry, both in livestock numbers and cattle farms. As the number of farms is decreasing faster than the total

[2] I decided to use the classification valid in each year instead of a unifying one, because the data was also used for model parameterization (chapter 8.2) and in this way model results could be compared to FADN data for the given year.

Table 6.2: *Agricultural land use in the study area*

	Area			Share [%]		
Land use	1999	2003	2007	1999	2003	2007
Summer barley	7,478	7,501	6,940	7.2	7.3	6.9
Winter wheat	6,605	6,606	7,198	6.4	6.5	7.2
Winter barley	3,784	4,211	3,799	3.7	4.1	3.8
Winter rape	3,286	3,282	3,109	3.2	3.2	3.1
Oats	2,077	2,186	1,671	2.0	2.1	1.7
Triticale	954	1,185	1,268	0.9	1.2	1.3
Silage maize	1,724	1,881	3,553	1.7	1.8	3.5
Clover & alfalfa	2,705	1,768	2,865	2.6	1.7	2.8
Grain legumes	523	466	262	0.5	0.5	0.3
Fallow	1,877	2,079	832	1.8	2.0	0.8
Other arable*	2,434	1,887	1,573	2.4	1.8	1.6
Forest	40,961	40,679	38,111	39.6	39.8	37.9
Grassland	23,834	23,437	22,847	23.0	22.9	22.7
Extensive†	2,883	3,187	3,419	2.8	3.1	3.4
Permanent crops	65	51	44	0.1	0.0	0.0
Other‡	2,249	1,798	3,169	2.2	1.8	3.1
Total	103,440	102,205	100,661	100.0	100.0	100.0

* Potatoes, CCM, grain maize, sugar beets, other cereals, other field forage, industrial crops, vegetables

† Extensive grassland (Hutungen, Almen, Streuwiesen) and agricultural areas permanently taken out of production

‡ Buildings, yards, house gardens, roads, water, unused and barren land etc.

Table 6.3: *Livestock husbandry in the study area*

	Heads			No. farms		
	1999	2003	2007	1999	2003	2007
Dairy cows	15,582	13,741	12,591	786	539	408
Mother cows	2,148	1,853	1,746	257	197	157
Other cows	2,811	2,543	2,018	609	453	360
Cattle < 1 year	15,111	13,564	13,214	1,047	782	620
Cattle 1-2 yrs	9,921	9,461	9,515	1,039	753	613
Male cattle > 2 yrs	236	196	207	122	97	109
Sows	12,578	11,492	11,280	306	204	154
Fattening pigs > 50 kg	27,056	31,461	29,822	882	623	479
Other pigs	58,662	62,072	56,040	825	607	450
Chicken	-	-	-	22	25	11
Laying hens	-	-	-	41	25	14
Other poultry	5,963	6,255	1,474	46	41	34
Horses	2,664	2,798	2,865	333	323	301
Sheep	23,056	29,394	21,945	210	192	162

number of cattle reared, average cattle numbers per farm have risen, e.g. from about 20 to 31 dairy cows per dairy farmer. A similar development can be observed for pig rearing, where e.g. the average number of fattening pigs per farm has risen from 30 to 62 heads per farm. In this case, the total number of pigs has even slightly risen, while the number of breeders declined by 54%.

6.3 Potential effects of climate change in the study area

AOGCM can only give a very coarse picture of the potential climatic development at a regional resolution, such as required for our study area. The AR4 divides the world into 30 subcontinental regions of a scale that still allows meaningful interpretations of AOGCM results. The study area falls just onto the Southern edge of the Northern Europe (NEU) region. Table 6.4 shows temperature and precipitation changes between 1980-1999 and 2080-99 as simulated by 21 AOGCM under the A1B scenario: Temperature increases are predicted for all seasons, but strongest in winter. Likewise precipitation is expected to increase in winter, while the expectations on summer precipitation are more ambiguous for the full region due to geographical variation: Precipitation is generally expected to increase to the north and to decrease to the south of N 55°[Christensen et al., 2007].

In Central Europe, these changes in means are expected to be accompanied by an increased risk of droughts and an increase in frequency and magnitude of high precipitation events in winter. The effect on summer extreme precipitation is unclear, due to a decreased number of precipitation days in combination with an increased rainfall intensity. Interannual and day-to-day variability in temperatures are likely to increase in summer, but decrease in winter. The number of frost days, duration of snow cover and snow depth are likely to be reduced all over Europe [Christensen et al., 2007].

Table 6.4: *Projection of temperature and precipitation change 2080-2099 compared to 1980-1999 for Northern Europe (N 48°, W 10°to N 75°, E 40°,) from 21 AOGCMs as reported in AR4 [Christensen et al., 2007, tab. 11.1]*

	Temperature response (°C)					Precipitation response (%)				
Season	Min	25%	50%	75%	Max	Min	25%	50%	75%	Max
DJF	2.6	3.6	4.3	5.5	8.2	9	13	15	22	25
MAM	2.1	2.4	3.1	4.3	5.3	0	8	12	15	21
JJA	1.4	1.9	2.7	3.3	5.0	-21	-5	2	7	16
SON	1.9	2.6	2.9	4.2	5.4	-5	4	8	11	13
Annual	2.3	2.7	3.2	4.5	5.3	0	6	9	11	16

The WETTREG [2010] project has conducted statistical downscaling experiments using the WETTREG weather generator [Kreienkamp et al., 2013] for the whole of Germany based on global projections by the climate model ECHAM5/OM for the SRES scenario A1B. Table 6.5 compares climatic averages of observed and projected data (two selected realizations) for the Stötten weather station lying just North of the study area at similar altitudes (N 48° 40′, E 9° 51′, 734 m a.s.l.). An increase in average temperature, vegetation days (> 5°C) and summer days (> 25°C), and a decline in summer precipitation can be observed.

Iglesias and Rosenzweig [2010] expect generally positive yield effects for wheat and negative effects for maize for the whole of Germany. Positive effects on wheat are however nearly exclusively due to the CO_2 fertilization effect, while the pure climate effect is negative with only a few exceptions (tab. 6.6). Global or national assessments of yield effects are, however,

Table 6.5: *Differences in climatic averages between observed weather and WETTREG projections for Stötten weather station. (Own calculations based on data from DWD 2011 and WETTREG 2010)*

	Observed		WETTREG Projections 2501300	2501344
	1951-1980	1981-2010	2000-2030	
Average temperature [°C]	6.7	7.5	8.3	8.3
Number of days with avg. temperature > 5°C	213	212	224	226
Number of frost free days	263	264	267	268
Number of days with a max. temperature > 25°C	7	14	19	20
Precipitation Spring [mm]	250	269	243	232
Precipitation Summer [mm]	346	320	286	296
Precipitation Autumn [mm]	228	249	269	247
Precipitation Winter [mm]	217	224	235	215

only of limited help for regional impact assessments, especially in an area lying at the margin of production suitability for silage maize. An increase of temperature could in this case improve the profitability and reduce the riskiness of maize production. Assessing the impact on crop yields is therefore a core research interest of the joint project and the simulations will draw on crop yield simulations conducted with the Expert-N modeling package within the project. Besides yield effects, an increase in vegetation days might trigger shifts in growing periods and the timing of sowing and harvesting. This could affect the time slots available for field work, both positively and negatively. In an expert interview, farmers pointed to the fact that currently growing rapeseed after wheat is not possible due to overlapping wheat harvest and rapeseed sowing days, but a small shift in wheat harvest days might make this feasible in the future.

Table 6.6: *Expected changes in crop yields (%) for Germany compared to 1970-2000 estimated by Iglesias and Rosenzweig [2010] using DSSAT simulations and HadCM3 projections*

		Yield change (%)			CO_2 fert. effect (%)			Climate only effect (%)		
Crop	Scenario	2020s	2050s	2080s	2020s	2050s	2080s	2020s	2050s	2080s
Wheat	A1FI	4.19	9.28	7.48	4	11	18	0.18	-1.55	-8.92
	A2a	5.67	9.2	13.14	4	10	18	1.61	-0.73	-4.12
	A2b	3.49	8.92	13.15	4	10	18	-0.49	-0.98	-4.11
	A2c	3.34	9.07	13.51	4	10	18	-0.63	-0.85	-3.81
	B1a	1.61	5.28	6.86	3	6	8	-1.35	-0.68	-1.06
	B2a	3.66	5.13	7.24	3	6	11	0.64	-0.82	-3.39
	B2b	3.16	5.5	8.85	3	6	11	0.16	-0.47	-1.94
Maize	A1FI	-1.68	-1.37	1.19	1	4	8	-2.65	-5.16	-6.31
	A2a	-1.16	-1.32	0.45	1	3	7	-2.14	-4.19	-6.12
	A2b	-0.72	-1.66	0.77	1	3	7	-1.70	-4.52	-5.82
	A2c	-1.48	-1.81	0.44	1	3	7	-2.46	-4.67	-6.13
	B1a	-1.71	-2.83	-3.21	0	1	2	-1.71	-3.79	-5.11
	B2a	-2.60	-3.19	-1.42	0	1	4	-2.60	-4.15	-5.21
	B2b	-2.55	-3.13	-1.35	0	1	4	-2.55	-4.09	-5.14

Note: Letters a-c in Scenario names denote different ensemble members of the HadCM3 simulations.

Chapter 7

Model design & parameterization

The simulation model for this study has been implemented using the multi-agent software package MPMAS. A description of model equations and software architecture of MPMAS following the ODD protocol can be found in Schreinemachers and Berger [2011]. Here, they are only briefly summarized and the focus lies on the specifics of the MPMAS implementation for the Central Swabian Jura, which are grouped under the section headings used for the ODD+D protocol [Müller et al., 2013], the adaptation of the ODD protocol for models of human decision making. Details about the implementation of the submodels are given in the appendix B.

7.1 Overview

7.1.1 Purpose

The model has been designed to analyze the adaptation of agricultural production decisions to potential effects of climate change. It should be capable of simulating the vulnerability of different types of farms and highlight the effects of climate change on the effectiveness of existing policies, specifically agri-environmental measures and biogas support. As the model is to be tested against observation data from 1999 to 2007, it needs to include relevant policy regulations valid during this time span. The model should provide insight into the importance of different climate related impacts, specifically the influence of yields, changes in available field working time due to meteorological conditions, changes in rotation options and market prices. The model is not, however, expected to provide an accurate forecast of future development, i.e. answer a 'how will it be?' type of question, but rather improve the understanding of the influence of relevant processes, e.g. agent heterogeneity, expectation formation, and land market transactions and help to explore potential feedbacks on land surface processes.

7.1.2 Entities, state variables and scales

Every full-time farm of the study area is represented by an individual model agent. The state of the agent is characterized by individual household composition, asset ownership, soil endowment and current expectations. The state of the household includes gender, age, the status of household members, and the expected remaining lifetime of the farm (until the retirement of the household head, respectively their potential successor). The state of assets includes the age and time value of tangible assets and intangible assets (quotas and entitlements) as well as equity, cash and liabilities. Expectations are related to expected future values of prices, average yields and household composition. Agricultural land is characterized by different soil types and represented at a resolution of one hectare.

7.1.3 Process overview and scheduling

Figure 7.1 summarizes the sequence of actions that is repeated for every agent in every simulation period. Agents start each season by forming expectations about future conditions (e.g. prices, yields). In the first simulation period expectations are specified by the modeler, in later periods they maybe be updated if learning processes are simulated. In the next step, the agents decide on investments into assets (e.g. machinery, stables, etc.) by solving the decision problem for an expected average year of the future, and the chosen investments are then implemented, i.e. the state of the agents' assets is updated accordingly.

Agents then make a production plan for the current season and the actual physical and economic outcomes of production are determined. Two model versions have to be distinguished: (i) In the simple case, the outcome corresponds to the production plan. (ii) For recursive-dynamic simulations, yields (and potentially also prices) might be different every year. In these cases, the agents may solve a third harvest decision problem in order to adapt production and sales decisions to the obtained result. (This case was not used in the present thesis.)

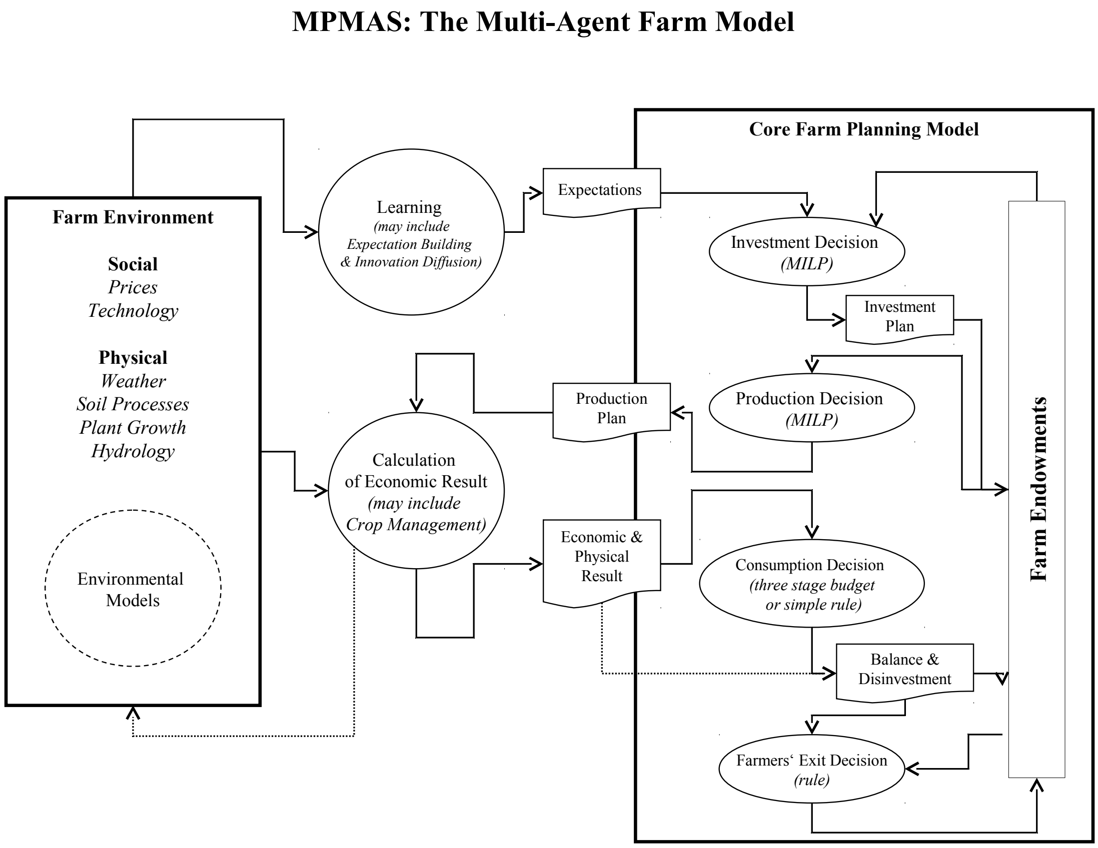

Figure 7.1: *The seasonal cycle of agent decisions and submodel invocation in MPMAS.*

The model calculates income, cash flow, debt service and rental payments, and increases the age of assets and household members. Assets that have reached the end of use life are removed from the list of assets and the agent decides on withdrawals for consumption. In case of cash shortage, the agent attempts to sell land to retain solvency (if land markets are activated) or shuts down the farm if bankruptcy cannot be avoided. Finally, the model determines whether household members die, retire or give birth. If the household head is scheduled to retire, retirement and continuation of the farm depend on the willingness of a potential successor to continue the business.

If the land market is simulated, agents decide whether they bid to rent in land between updating their expectations, but before deciding on investments. Bids for available plots are compared and the plot is assigned to the highest bidder. The characteristics of land do not change throughout the simulation, except for ownership and renter information in case land markets are used. Rental contracts have the same standard length depending on scenario assumptions, and reenter the rental market once the contract expires.

7.2 Design concepts

7.2.1 Theoretical and empirical background

The model rests on the traditional agricultural economics approach of representing farm decisions as mathematical programming problems to determine an optimal set of activities given technological and resource constraints specific to the farm [Hazell and Norton, 1986]. In a wider sense, it belongs to the class of recursive-dynamic programming models representing economic decisions as described by Day [2008]. Besides the economic considerations of maximizing expected farm income while ensuring liquidity and long-run survival of the farm, agents have a preference for employing their own children (see appendix B.2.3) and are assumed to comply with good farming practice and agri-environmental regulations.

Technical coefficients are based on standard references for farmers provided by extension services [e.g. KTBL, 2010; LfL, 2010, 2011], expert interviews and a farm survey. Farm census data and official demographical statistics are used to initialize the agent population (see section 8.2). Price data is derived from various statistical databases [LEL, 2010, 2011*a*,*b*; KTBL, 2010; destatis, 2012*e*] (see section 8.1). Crop yield information is taken from statistics and simulations with a crop growth model that was calibrated on multi-year field observations in the area [Calberto Sanchez, 2015; Aurbacher et al., 2013](see section 8.1).

7.2.2 Individual decision making

Farmers usually dedicate themselves to a number of mutually interdependent production activities. Figure 7.2 shows a rather aggregate representation of the conceptual model of the farmer decision problem. The basic assumption of the decision model is that farmers maximize expected total farm income by choosing an optimal combination of production activities (shaded in the figure). Crop production, grassland use, animal production and biogas production constitute the major alternatives, but decisions also have to be taken on the selling and buying of products and inputs, field work, investments and application for agricultural and agri-environmental support schemes. The choice of activities is constrained by a number of restrictions and balances including the manure balance, the time budget, crop rotation, the financial balance, the balance of products and inputs (yield, feedstock) and restrictions imposed by policy regulations or subsidy conditions. This coarse conceptualization is mathematically represented as a mixed integer programming problem (MIP) in the model (for details see appendix B.1), which serves as the basis for the three decision problems: investment, pre-season and post-season production decision. While the general structure of the problem remains the same in all three decision stages, investment activities are only included for the investment stage, and land use decisions cannot be reversed at

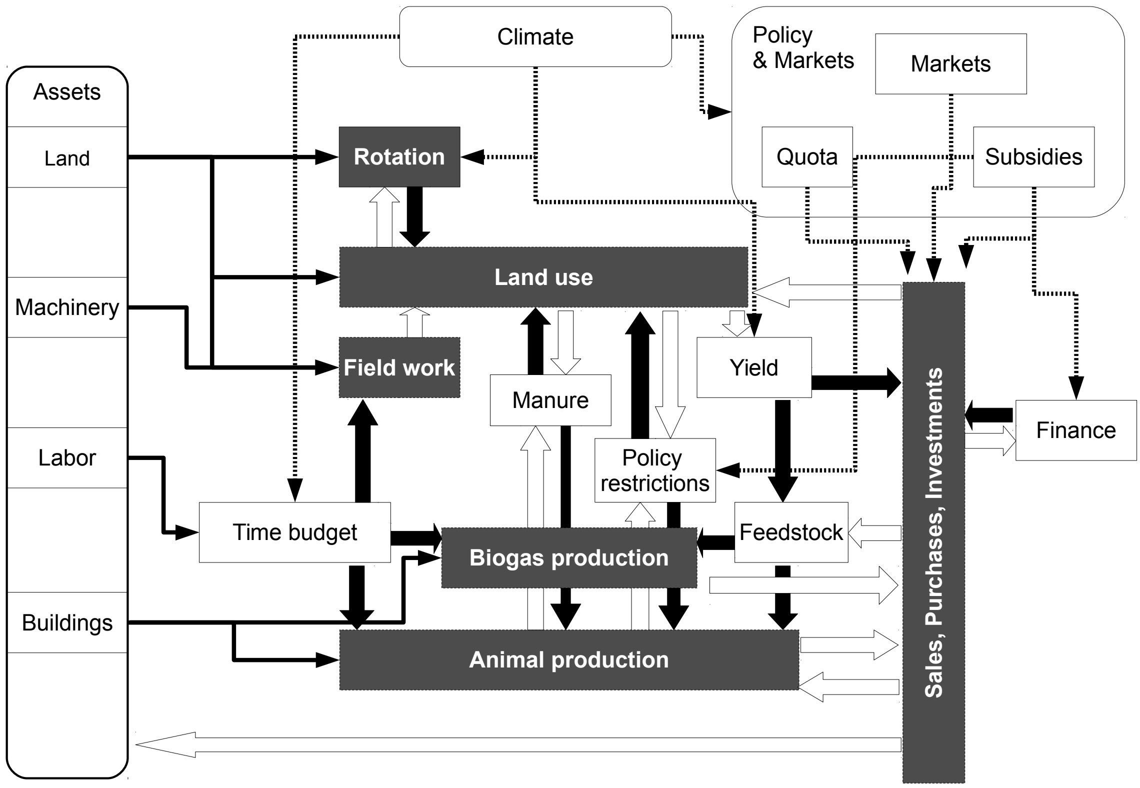

Figure 7.2: *Interdependencies of farm activities (shaded) in the decision problem of farm agents and the impact of climate change.*

post-season stage.

7.2.3 Learning, individual sensing and individual prediction

In the scenarios used in this thesis, long-term averages of yields, prices and environmental conditions are used and the model abstracts from the process of learning. It is assumed that the adaptation of knowledge has taken place and agent expectations coincide with outcomes.

7.2.4 Interaction and collectives

In some of the scenarios, agents interact on land markets. Collective actions of agents are not considered.

7.2.5 Heterogeneity

The structure of the agents' objective function and constraints is identical for all agents: it is a comprehensive representation of production technology and local conditions for agricultural production. Heterogeneity is introduced into the decision module by different household compositions and resource availabilities of individual agents, e.g. the amount and type of available farm labor and land as well as the machinery and buildings owned at the start of the simulation.

These differences in starting conditions affect the profitability of production options, and in this way produce heterogeneous agent behavior. Soil types determine attainable crop yields and the tractor-power required for field work. Existing machinery and buildings are associated with sunk costs. Profitability of crops can differ given the possibility of selling or use for feeding. Household composition determines the amount of household labor available and affects the planning horizon for investment decisions. The household head's age or his/her potential successor's age determines the expected remaining operating time of the farm, i.e. the maximum lifetime considered in agent investment calculations. Farm succession is an important topic in family business and requires some additional rules for implementation in dynamic simulation models. Here, it is assumed that agents are glad to employ their potential heirs on the farm and are even willing to forgo own-income if a major investment or expansion of the farm is necessary to employ their successors. In the MIP decision problem, agent household heads have to remunerate their adult children's work on the farm, but they do not consider this a cost as long as their own minimum income expectation is met.

Further, the model structure exhibits economies of size at farm level. In Southwest Germany with rather small farm sizes and indivisible tractors and other field work implements, the capacity-to-cost ratio usually declines with increasing capacity, which was considered accordingly. Building costs and most animal-related works are implemented using fixed and size-dependent costs, leading to decreasing average cost functions. Certain policy schemes, however, include special regulations for smaller farms leading to dis-economies of scale: under EEG regulations, for example, guaranteed biogas electricity prices decrease with volume. Again, these farm-level effects were considered accordingly.

7.2.6 Stochasticity

The model is deterministic.

7.2.7 Observation and emergence

Observed total land use and crop production in the study area emerge as the sum of individual agent decisions. In principle, the full state of all agents is accessible and the individual courses of actions can be traced with log and debugging tools. Analysis mostly focuses on agents' production activities, income, asset ownership and household composition.

7.3 Details

7.3.1 Implementation

The model has been implemented using the MpmasMySQL setup developed in chapter 5. Files are available in the supplementary material to this thesis.

7.3.2 Initialization

The initial state of the model is defined by the initial agent population including their assets, household members, owned and rented land and expectations. Initialization differs between scenarios and repetitions for uncertainty analysis and is consequently described in more detail in the description of simulations.

7.3.3 Input data

The exogenous inputs that determine the course of the simulations over time are prices and crop yields. A second set of exogenous conditions that (may) change depending on the simulation year is defined by policies, e.g. European Union (EU) regulations and support, MEKA agri-environmental measures, and biogas support. Details differ between scenarios and are provided with the description of individual simulations.

7.3.4 Submodels

The three modules for farm decisions, farm household demographics and land rental markets as well as the integration with the crop growth model Expert-N is described in detail in chapter B in the Appendix.

Chapter 8

Validation and calibration of the short-term production decision

One of the basic assumptions of the model[1] is that the short-term production decision ($\mathbf{a}$) of farmers (i) for a given year (t) can be predicted with reasonable accuracy if one knows their asset endowments at the beginning of the season (B_t) and their knowledge or expectations of major production parameters. These can be divided into parameters (θ) that are expected to remain constant over time/between scenarios and those that constitute exogenous variables that may potentially change over time/between scenarios. In the case of the present study, the exogenous variables considered are crop yields ($\mathbf{c}$), prices ($\mathbf{p}$), rotational constraints ($\mathbf{r}$), field working days ($\mathbf{c}$) and policy regulations ($\mathbf{z}$).

$$\hat{\mathbf{a}}_{\mathbf{t}} = f\left(\mathbf{B}_t, \mathbf{p}^*_t, \mathbf{y}^*_t, \mathbf{r}^*_t, \mathbf{c}^*_t, \mathbf{z}^*_t, \theta\right) \tag{8.1}$$

An empirical test of the model requires simultaneous observations of production decisions, exogenous variables and asset endowments at the beginning of the period. It would be advantageous to have several such consistent data points in order to have some control against overcalibration to one specific situation. FDZ [2010] includes observations of land use and animal stocks at the farm level for the years 1999, 2003 and 2007. More specifically, the statistical surveys have been conducted always in May of the respective years and reflect production in the cropping seasons 1998/1999, 2002/2003 and 2006/2007. As total land endowment and its partition into grassland, arable and forest land, as well as stable places induced from animal stocks can be considered asset observations exogenous at the beginning of the season, and the particular crop choice and intensity of grassland use as well as actual stocking rates can be considered major endogenous outcomes of the production decision, the panel dataset provides a good, but incomplete basis for a calibration and validation dataset.

The major challenge in using the panel are the privacy restrictions tied to the use of the dataset: Although information is recorded at farm level, individual observations are not directly accessible to the modeler. This problem is circumvented by estimating statistical distributions that are as aggregated as necessary to comply with privacy restrictions, but disaggregated enough to allow for a randomized creation of a model population that preserves the relevant statistical properties of the observed population. The gaps in the information on the initial farm assets – the panel does neither contain information on soil types, machinery ownership, household composition nor cash reserves – are closed by combining it with additional information from land use and soil maps, general demographic data, expert information and survey results relating machinery ownership to farm size and structure. Data on the exogenous variables have been gathered from other sources.

Figure 8.1 gives an overview of the process of testing the short-term production decision in the model, which will be described in more detail in the present chapter. It starts by sub-

[1] For a journal publication on the combined framework for calibration and sensitivity analysis developed here, please refer to Troost and Berger [2014].

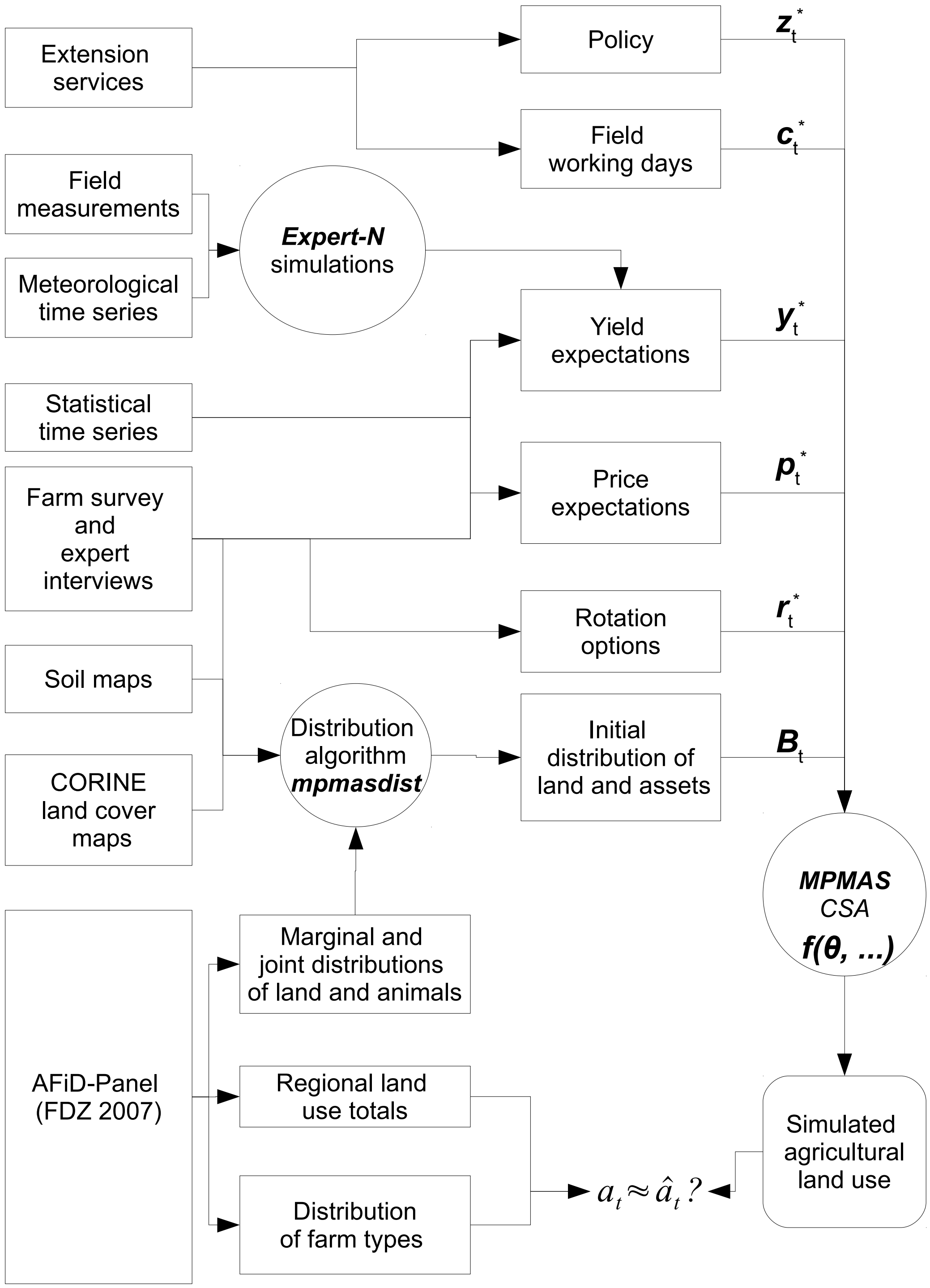

Figure 8.1: *Generating a dataset for the empirical validation of short-term production decisions in the model ($t \in \{1999, 2003, 2007\}$).*

stantiating the choice of data for the exogenous variables at each point of time (section 8.1) and then explains the process of generating starting values for the initial farm population (section 8.2). It presents the results of a Turing test of the model, which was conducted with farmers and agricultural experts from the study area (section 8.3), before describing the actual calibration (section 8.4) and validation experiments (section 8.5).

8.1 Exogenous variables

During the time covered in validation simulations, the regulations associated to the three major policy regimes considered in the model, the farm support under the EU Common Agricultural Policy (CAP), the support for agri-environmental measures under the Compensation Scheme for Market Easing and Landscape Protection (*Marktentlastungs- und Kulturlandschaftsausgleich*, MEKA) and the biogas production support under the German Renewable Energy Act (*Erneuerbare Energien Gesetz*, EEG), changed multiple times: The CAP regulations applicable to the first season simulated (1998/99) still date from the 1992 MacSharry reforms. From 1999/2000 on, the changes under the Agenda 2000 applied. The CAP Mid-term Review (MTR) of 2003 enacted regulations applying from seasons 2004/05 on, which were only slightly adapted under the CAP Health Check in 2008. There have been three phases of the MEKA program: MEKA I from 1994 to 1999, MEKA II from 2000 till 2006, and MEKA III from 2007-2013, while the EEG was first established in 2000 and has been subject to revisions in 2004, 2009 and 2012.

The different policy regulations valid in each of the selected years are described in the description of the farm decision model (see sections B.1.5, B.1.9 and B.1.10 in the appendix), and the respective setting was chosen for each observation year. The choice of crop yields, prices, available field work days and crop rotation options for each observation year is explained in the remainder of this section.

8.1.1 Expected crop yields

Fellow researchers in the projects use the Expert-N model package to simulate crop yields under current and future climate conditions, in order to assess the influence of climate change on crop yields. The Expert-N configuration uses the CERES model for winter wheat, barley, and silage maize, and the GECROS model for winter rapeseed. The study area specific parameterization was calibrated and validated against leaf area index (LAI) and phenological observations at the three field measurements sites in the years 2009-2011 by Calberto Sanchez [2015] and Aurbacher et al. [2013]. As a consequence, the simulated yields reflect current technology, which may cause a bias when used in the calibration and validation process for the 1999, 2003 and 2007 observations. Alternative yield sets derived from public yield statistics were therefore also included into the calibration process in order to avoid overfitting of model parameters to a potentially biased, simulated yield set.

Simulated yields for current climate

The Expert-N model parameterization was calibrated and validated against LAI and phenological observations at the three field measurements sites in the years 2009-2011. The calibrated model was then used to predict yields for each combination of soil type and management included in the model for each season between 1951 and 2010 using the corresponding record from the meteorological time series of Stötten weather station [Calberto Sanchez, 2015]. As the information on weather data was limited to a single weather station and the crop model does neither fully reflect exposition and slope effects nor incorporate plant pest models, only the spatial distribution of soil types introduces spatial heterogeneity of yields in the area. Using LUBW [2007], ten soil mapping units were identified by the soil scientists of the research project and linked to reference soil profiles of the Institut für Bodenkunde und Standortlehre der Universität Hohenheim to obtain the relevant soil characteristics for modeling [Calberto Sanchez, 2015]. Tables C.2 and C.3 in the appendix give an overview of the importance of each soil class in the study area and the structure of reference profiles linked to each soil class.

To keep things simple at this stage, management practices were limited to the typical management applied to a crop in the area: While a distinction is made between fertilization schemes with cattle manure, pig manure or only mineral fertilizer, and between low tillage and ploughing, these management practices were designed such that the yield differences between them should be minimal. Table 8.1 shows the average yields over the full 60 year time series for each crop and soil type. It also reports the plain average and coefficient of variation over the nine arable soil types and the average weighted by the proportion of soils in the arable[2] land of the study area.

Table 8.1: *Yield results of Expert-N simulation, averaged by soil type over 60 years of observed weather and six management practices (Source: Own calculations based on simulation results of Calberto Sanchez 2015).*

Soil	Crop				
	Silage maize	Summer barley	Winter barley	Winter rape	Winter wheat
0	424.8	46.5	57.8	34.3	80.8
1	380.1	46.6	59.1	33.7	80.6
2	458.2	49.3	65.2	31.6	81.3
3	415.1	47.6	61.5	34.0	80.7
4	475.7	51.7	72.5	35.3	81.3
5	437.4	48.5	61.3	34.3	80.8
6	360.7	45.5	55.7	32.7	77.4
7	461.7	51.0	68.5	34.4	81.0
8	476.4	51.7	70.0	30.3	81.2
Avg	432.2	48.7	63.5	33.4	80.6
CV	9.0%	4.6%	8.7%	4.5%	1.4%
wAvg	422.9	47.3	59.9	33.5	80.5

Note: For soil codes refer to tab. C.2 in the appendix.

Yield scenario xn3 uses these long-term average yields as expected yields for each crop production activity in each of the years, ignoring any technology-induced yield difference between the years. For wheat, a yield reduction of 20% is considered in the model for wheat grown the second year on the same plot compared to wheat grown after other crops. As the observed or simulated yield is assumed to represent the area-weighted average wheat yield in the area, the first year wheat yield is increased using the scaling factor *wheat_normal*, which is subject to calibration.

[2]Here the category includes the arable and mixed cultivation pattern classes of table C.2.

Alternative statistical yield sets for calibration-validation

For the calibration and validation of the short-term production decisions it was important to infer the yield farmers' expected during their production decision at the beginning of the year, which does not necessarily correspond to the real yield obtained by farmers at harvest.

In a farm survey[3], conducted between August and October 2010, farmers were asked to describe their expectation for wheat, barley and rapeseed yields as a triangle distribution. Table 8.2 shows mean, standard deviation, minimum and maximum of the modus of the triangle distribution over responding farmers practicing conventional farming on the Central Swabian Jura. These figures are not statistically representative of all farmers in the study area due to the rather low number of respondents and the nature of the survey sample, but these numbers are valuable as a first impression to derive calibration input.

Table 8.2: *Medium yield expectations [dt/ha] in the farm survey (Sep/Oct 2010).*

Crop	N	Avg	Sd	Min	Max
Bread wheat	14	73.9	4.77	67.5	80
Fodder wheat	14	78.6	8.36	70.0	95
Organic wheat	4	41.3	8.10	35	53
Malting barley	5	61.6	11.63	50	80
Winter rapeseed	13	39.7	4.52	30	46

As a second source of information, the online database of the statistical office of the State of Baden-Württemberg [Statististisches Landesamt Baden-Württemberg, 2012] provided yield averages for the two study area districts ranging back until 1983. As depicted in figs. 8.2 and 8.3, there was a long-term trend of increasing yields for most crops, maybe with the exception of summer barley in the Reutlingen district and silage maize in the Alb-Donau district, combined with considerable interannual variability. Silage maize yields in Reutlingen experienced an abrupt upward shift around 1998 from stable levels below 300 dt/ha to stable levels above 400 dt/ha. Farmers' yield expectations can be assumed to average out interannual variability, but do reflect long-term yield development. As an approximation, the average of the yield of the a years preceding the respective year of harvest was calculated. Table 8.3 shows the results for $a = 3$ and $a = 6$, reflecting two types of averages that are rather more and rather less sensitive to short-term fluctuations in observed yields.

Yields in the Reutlingen district were consistently lower than in the Alb-Donau district. Despite the fact that one would generally expect the Reutlingen district to be more representative of the study area (cf. section 6.1), the farm survey results for wheat, rapeseed and barley seem to be more consistent with the pre-2011 averages in Alb-Donau rather than Reutlingen. While a slight majority of survey respondents (8 out of 14, resp. 13) was located in the Alb-Donau district, there was no significant difference in yield expectations for wheat and rapeseed between the two districts in the survey, when using a t-test for mean comparison. Even the lowest expectation for wheat mentioned in the survey was higher than the long and short-term averages recorded in the statistics for Reutlingen. For malting barley, only one respondent was from Reutlingen, but his answer also lay well within the range of answers provided by the four farmers from Alb-Donau.

Based on this information, a set of three uniform yield scenarios (tab. 8.4) was generated: one modeled on the Reutlingen yields (*urt*), one based on the Alb-Donau time series (*uad*), and one mixing summer crop yields from Reutlingen with winter crop yields from Alb-

[3]The farm survey was conducted by Marius Eisele in the course of his Master thesis.

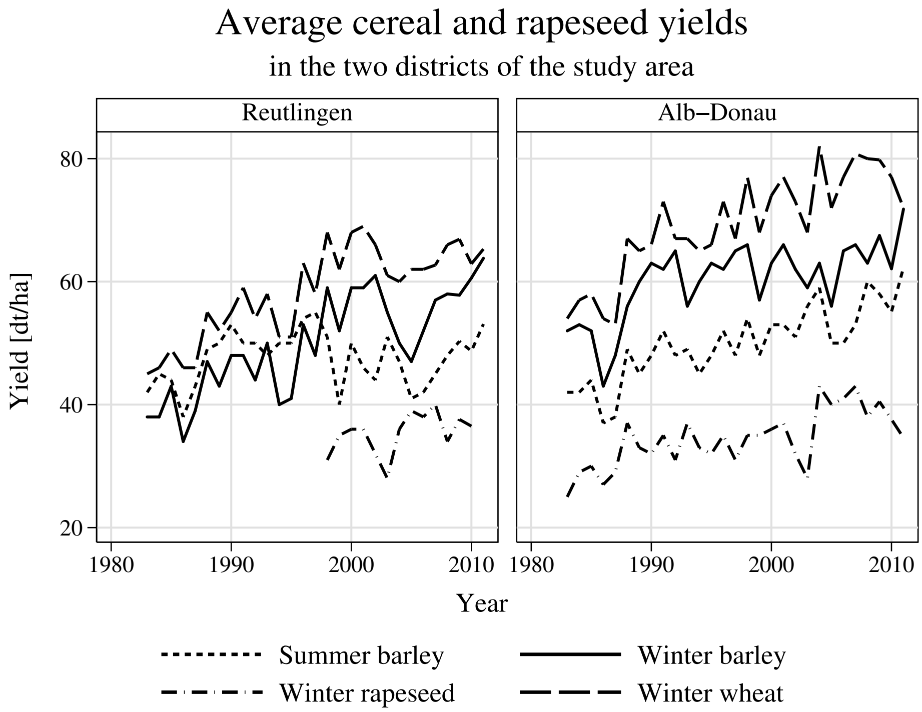

Figure 8.2: *Average wheat, barley and rapeseed yields in the two study area districts, 1983-2011 [Statististisches Landesamt Baden-Württemberg, 2012].*

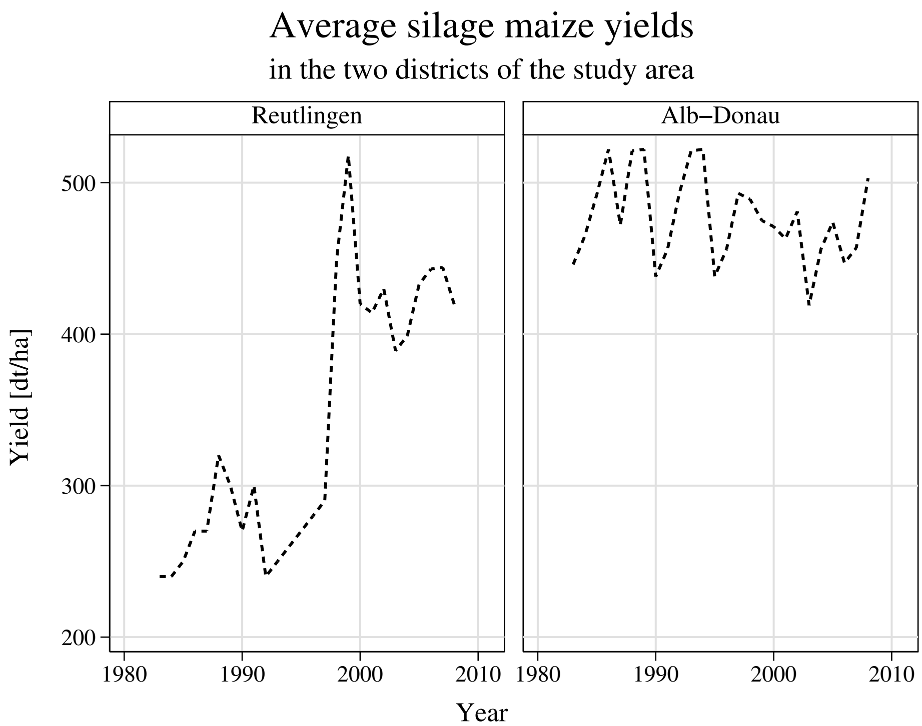

Figure 8.3: *Average silage maize yields in the two study area districts 1983-2011 [Statististisches Landesamt Baden-Württemberg, 2012].*

Table 8.3: *Yield average [dt/ha] of the a years preceding the years of observation.*

		Reutlingen				Alb-Donau			
	a	1999	2003	2007	2011	1999	2003	2007	2011
Silage maize	3	339.7	421.3	425.0	419.0	479.3	471.7	459.0	503.0
	6	299.8	420.2	418.0	434.8	486.5	478.7	456.7	470.3
Summer barley	3	53.3	46.7	43.3	49.0	51.3	52.3	53.0	57.7
	6	51.3	47.7	45.2	45.8	49.3	51.2	53.2	54.4
Winter barley	3	53.3	59.7	49.7	58.8	64.3	63.7	61.3	64.2
	6	48.5	56.3	54.0	55.4	62.0	63.2	61.8	63.3
Winter rapeseed	3	31.0	34.7	37.7	36.0	33.7	35.0	41.3	38.7
	6	31.0	34.0	34.8	37.5	33.8	34.3	36.8	40.0
Winter wheat	3	63.0	67.7	61.3	65.3	72.3	74.7	77.0	78.9
	6	58.2	65.2	63.3	63.8	69.2	72.7	74.8	77.8

Source: Own calculation based on Statististisches Landesamt Baden-Württemberg [2012]

Donau (*umx*). They are qualified as uniform, because they assume the same yield irrespective of soil type or type of fertilization (manure, mineral).

Table 8.4: *Alternative yield scenarios*

		Year			
Scenario	Crop	1999	2003	2007	2011
urt	Silage maize	340	420	420	420
	Summer barley	50	45	45	45
	Winter barley	55	55	55	55
	Winter rapeseed	31	34	34	37
	Winter wheat	63	63	63	63
uad	Silage maize	470	470	470	470
	Summer barley	50	51	53	54
	Winter barley	63	63	63	63
	Winter rapeseed	34	34	37	38
	Winter wheat	73	75	77	79
umx	Silage maize	340	420	420	420
	Summer barley	50	45	45	45
	Winter barley	63	63	63	63
	Winter rapeseed	34	34	37	38
	Winter wheat	73	75	77	79

8.1.2 Prices

The price information required for the model comprises producer prices for crops and animal products, purchase prices for consumable inputs, buying prices and maintenance cost of investment goods and wages for hired labor. Producer prices for major crop and animal products were taken from the regional statistical time series in LEL [2010, 2011*a*,*b*]. For other products and inputs, a time series was constructed by combining prices reported for the year 2009 from KTBL [2010] with the corresponding price indices from destatis [2012*e*]. Figure 8.4 shows the development of the producer prices of the crops most relevant for the study area between 1995 and 2011.

The seasonal nature of agricultural production requires a distinction between products

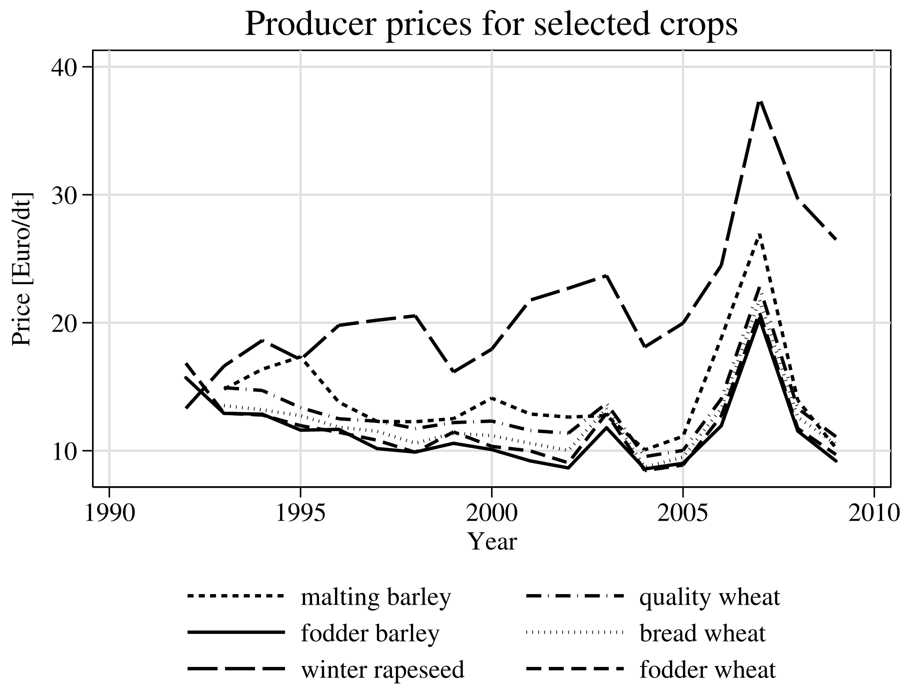

Figure 8.4: *Development of producer prices for the major crops of the area, 1993-2009. Illustration based on data from LEL 2010.*

whose prices are assumed to be known at the time of decision making (including investment goods, fuel and most other input prices) and those whose prices are not known at the time of the production decision, including prices for crops and – to a lesser extent – also for animal products. Similar to crop yield expectations, assumptions on the formation of expectations are required to infer the product prices used during production planning. During the farm survey, farmers were also asked to describe their long-term expectation for producer prices as triangle distributions. Table 8.5 shows for each product the mean and the range of answers over all respondents, which were asked for the price they expected to see most frequently in the following years. They were also asked for the lowest and the highest price they would expect to observe in the following years.

In a second question, farmers were asked for the wheat price they would specifically expect for 2011, with answers shown in table 8.6. Except for bread wheat of quality A, the answers differ very little from the long-term expectations discussed above.

As in the case of crop yields, statistical price averages were calculated over the three, respectively six years preceding each point of observation used in the calibration experiments (tab. 8.7). Comparing the results for 2010 – the most recent year for which data was available – with the expectations recorded in the farm survey, it seems that the 3-year average looks much more consistent with farmers' responses than the 6-year average. Based on this observation, the calculated 3-year averages were used as proxies for expected prices in our simulation.

Table 8.5: *Survey farmers' long-term price expectations (Lowest, most frequent and highest points of triangle distribution).*

			Triangle distribution of expected price [€]					
			Most frequent		Lowest		Highest	
Product	Unit	N	Mean	Range	Mean	Range	Mean	Range
Fodder wheat (C)	dt	9	13.98	[11; 16.5]	10.06	[7; 14]	20.56	[17; 25]
Bread wheat (B)	dt	4	16.75	[15; 20]	12.25	[9; 18]	23.75	[20; 25]
Bread wheat (A)	dt	6	15.67	[13; 18]	10.67	[9; 16]	24.33	[22; 26]
Bread wheat (E)	dt	4	19.50	[17; 25]	11.00	[8; 16]	43.00	[22; 100]
Malting barley	dt	5	18.00	[15; 23]	11.20	[6; 20]	37.20	[24; 80]
Winter rapeseed	dt	13	30.77	[25; 35]	23.58	[19; 29]	47.38	[35; 120]
Milk	100 l	14	31.14	[28; 35]	22.69	[18; 27]	38.77	[35; 45]
Beef	kg	5	3.06	[2.85; 3.3]	2.66	[2.5; 2.8]	4.02	[3.1; 6]
Pork	kg	10	1.38	[1.3; 1.5]	1.16	[1.1; 1.4]	1.71	[1.5; 1.9]
Piglet 25kg		2	52.75	[45.5; 60]	32.00	[30; 34]	70.00	[60; 80]

Table 8.6: *Survey farmers' wheat price expectations for 2011.*

Category	N	Mean	Range
Fodder wheat (C)	11	13.77	[10.5; 17]
Bread wheat (B)	4	16.50	[15; 18]
Bread wheat (A)	6	17.08	[15; 20]
Bread wheat (E)	4	18.13	[15.5; 20]
Organic bread wheat (E)	4	39.50	[36; 44]

Table 8.7: *Price average of the x years preceding the years of observation (Source: Own calculations based on data from LEL 2010, 2011*a,b*)*

		Price [€]			
Product	x	1999	2003	2007	2010
Malting barley	3	12.79	13.20	13.34	17.03
	6	14.48	12.78	13.04	15.19
Fodder barley	3	10.58	9.32	9.85	13.69
	6	11.52	9.77	9.87	11.77
Winter rapeseed	3	20.18	20.80	20.85	31.25
	6	18.82	19.88	21.78	26.05
Bread wheat	3	11.32	10.59	10.51	14.97
	6	12.24	10.88	10.91	12.74
Quality wheat	3	12.18	11.76	11.21	15.73
	6	13.26	11.92	11.71	13.47
Fodder wheat	3	10.71	9.80	10.02	14.08
	6	11.64	10.26	10.34	12.05
Piglets	3	45.75	46.97	42.67	42.45
	6	46.49	43.23	42.67	42.56
Pork	3	1.40	1.43	1.43	1.45
	6	1.44	1.35	1.40	1.44
Beef (young bulls)	3	2.70	2.43	2.92	3.12
	6	2.70	2.58	2.69	3.02
Milk	3	0.295	0.311	0.286	0.309
	6	0.294	0.304	0.294	0.297

8.1.3 Field work days

KTBL [2010] provided estimates of available field working days in each half month of the growing season. Estimates are specific to the weather sensitivity level of field work, 12 agro-climatic subregions and the probability of occurrence (60%, 70%, 80%, 90%), respectively grain water content (14%, 16%, 18%) for cereal harvest activities. The study area falls into three of these subregions. Areas above 700 m fall into region 4, areas below 700 m into region 5 (The Hochalb, also above 700 m, falls into region 2). As the current model design does not allow to distinguish different climatic regions in the model area[4], only the values of one of the regions could be used in a simulation run, and both, 4 and 5, were tested in the calibration experiments.

Further, two levels of probability of occurrence, 60% and 80% (respectively 16% and 14% grain water content at 80% probability for cereal harvest) were tested as well as the parameter relating the potential to hire work of a certain type in a work season to the suitable field working days in the corresponding time span. The resulting number of suitable days for each field work season is shown in table C.1.

8.1.4 Rotation options

The compatibility of crops as direct neighbors in the crop rotation was obtained through expert interviews and recorded in the compatibility matrix shown in figure 8.8. It can be assumed that these relationships have not changed in the time period considered for calibration and validation and the same matrix was used for all three points of time.

[4]This is theoretically possible by distinguishing soils not only by soil type, but also by climatic region. I chose not to do so for simplification.

Table 8.8: *Compatibility of crops in rotation*

	Following crop								
Preceding crop	FA	FG	SM^i	SM	SB^i	SB	WW	WB	WR
FA	1	1	1	1	1	1	1	1	1
FG	1/2	2/3	0	1/2	0	1/2	1/2	1/2	0
SM	1	1	0	X	1	1	1	0	0
SB	1	1	1	0	1/2	0	0	1	1
WB	1	1	1	0	1	0	0	0	1
WR	1	1	1	0	1	0	1	1	0
WW	1	1	1	0	1	0	1/2	0	0*

Crops:
FA: fallow; FG: field grass; SM: silage maize; SB: summer barley; WW: winter wheat; WB: winter barley; WR: winter rapeseed; SM^i: silage maize with intermediate; SB^i: summer barley with intermediate.
Coefficients:
0: incompatible; 1: compatible; X: uncertain, subject to calibration (*maize_on_maize*);
1/2: maximum half of the area can be considered, e.g. wheat can directly follow wheat only once, then another crop has to be grown before wheat can be grown again
2/3: Field grass is a semi-permanent culture that is usually kept 2-3 years on the same field. So at maximum half the area can be considered preceding crop for other crops and at maximum 2/3 can be considered preceding crop for next year's fields grass.

0*: 0 in calibration and baseline, 1 if climate change shifts crop management dates

8.2 Initializing the agent population

The agent population consists of the full-time farms listed in the FDZ [2010] for each observation year. Farms were classified into full-time farms and others using the classification rules used for the German Farm Accounting Data Network (FADN).[5] Simulations have been restricted to the group of full-time farms due to constraints on data and resource availability, and also because the basic assumption of income maximization in the farm decision model is less convincing for part-time and hobby farmers. Nevertheless, the non-full-time farms were also included in the distribution algorithm in order to achieve more realistic results in the spatial distribution of plots.

Since privacy restrictions prevented the direct use of the farm information contained in the panel, it was used to derive marginal and joint distributions of farm areas, arable, grassland and forest shares and animal numbers. This statistical information was extended by aggregate statistical information and theoretical rules and formed the basis for a stepwise distribution algorithm for agent asset endowments. Next, the farm area allocated to the agent was spatially distributed over the map defining the soil type distribution for each agent. Finally, the demographic composition of farm households was generated based on a random sampling from general demographic information for Germany.

8.2.1 Estimation of distributions

Estimation of marginal distribution

Marginal distributions for each variable v representing a household characteristic were estimated as empirical inverse cumulative distribution functions $icdf_v(p)$ for each population at a resolution of 0.01 for p. Percentiles for 0% ($p = 0$) and 100% ($p = 1$) had to be excluded due to anonymization requirements. (The statistical office is not allowed to report the minimum and maximum of a variable.)

To arrive at complete continuous distribution functions, linear interpolation was used to infer values between the centiles and values for $p = 0$ and $p = 1$ were imputed according to the following rules:

1. $icdf_v(0) = 0 \,\forall\, v$
2. For certain variables representing shares of aggregated land use groups (e.g. share of arable land), one could safely set $icdf_v(1) = 1$.
3. If information on the population total was available for a variable (e.g. for the total agricultural area or the the total number of dairy cows), it was assumed that due to the construction of the percentiles, summing over the values assigned to N agents (equally distributed over p) should equal the population total observed in the area. Given estimation and interpolation, the values for roughly the first $0.99N$ agents were hence known. The values for the remaining $0.01N$ had to add up to the difference between observed total and total of $0.99N$ agents. $icdf_v(1)$ could then be calculated assuming linear progression among the last $0.01N$ (see appendix section D.3 for calculations).
4. For variables where neither theoretical values nor a population total were available, $icdf_v(1)$ was extrapolated using the slope of the linear interpolation between $icdf_v(.98)$ and $icdf_v(.99)$

[5] There has been a change of classification rules between 1999 and 2003 (see table D.1). In order to allow comparability with FADN data, for each observation year the rules that were valid in that year were used.

Estimation of the joint distribution

The joint distribution was estimated as an empirical copula that was expressed as a frequency distribution, in contrast to the canonical form of a copula [Schweizer and Sklar, 2011], which uses a cumulative distribution. Privacy restrictions allowed estimation of this copula $f(c)$ only at the resolution of quintiles. For each farm i, a vector $\mathbf{c_i}$ was created with a column $c_{i,v}$ for each variable v containing the quintile of this variable to which the farm belongs. E.g. $\mathbf{c} = (1, 3, 1, 5)$ denotes that the associated household falls into the first quintile of the first variable, into the third quintile of the second variable, into the first quintile of the third variable and into the fifth quintile of the fourth variable. A value of zero was used to aggregate all lower quintiles whose upper value was zero in order to reduce the number of distinct vectors. A table containing the frequency $f(\mathbf{c})$ of the different vectors observed in the population then served as the copula.

Privacy restrictions further meant that a full frequency distribution of quintile associations could only be estimated for the vector $\mathbf{d} = (c_1, c_2)$ containing only the first two dimensions (total agricultural area of the farm and share of arable land), while the frequency distribution of the complete vector $\mathbf{c}$ could only be estimated from a 85% sample of the entire population.

Comparing the quintiles for the marginals of the 85% sample with the marginals estimated from the full population, the former were found to be an acceptable representation of the later.[6]

As a result of the sampling, the estimated number of households ($\hat{h}_c$) for a given quintile combination (c) was smaller or equal to the number of households (h_c) that were actually associated with c. This also led to quintile combinations not being reported at all, because their $\hat{h}_c$ was zero, although the true h_c was greater. (According to the FDZ statistician that roughly affected 10% of all c with a $h_c > 0$.) This had to be taken into account during the creation of the agent population by allowing all quintile combinations (including the ones reported as zero) to contain a higher number of agents than reported.

In a first sampling attempt, the forest area was largely underestimated and arable and grassland areas were consequently overestimated. The uppermost quintiles of total area and forest share span relatively wide ranges (for example, 80 - 2393 ha, respectively 10 to 100% for the full-time farms in 1999). Given this result and the expert information that there was no farm with more than 500 ha of agricultural area in the region, it was assumed that the largest enterprises are rather forestry than agricultural enterprises and the quintile distributions was changed such that the uppermost 2% of both the total area and the forest share distributions are now associated with each other. For the out-of-sample farms, the non-forest area was restricted to a maximum of 500 ha. This led to a satisfactory forest area (increase of about 30,000 ha compared to no constraint).

8.2.2 The distribution algorithm for farm endowments

The estimated distributions were combined with theoretical constraints in order to ensure compatibility with the model. These were also necessary to avoid unrealistic combinations, given the fact that estimated quintile associations deliver only a relatively coarse representation of the joint distribution function and 15% of the agent population was not subject to the full joint distribution at all. [7]

[6] This approach worked due to the high sampling fraction. For smaller fractions, one should probably re-estimate the marginals for the sample and later project the quintile association onto the original marginal.

[7] Since the copula could only be estimated for 85% of the full population, the joint distribution for 15% of agents was considered unknown.

This required a stepwise sampling procedure using different techniques at different steps of the process, which were implemented using `mpmasdist` and are described in the following:

1. The algorithm created an agent population of size N and randomly distributed the different observed realizations of the vector d according to its frequency distribution among the agents.

2. It randomly distributed the different observed realizations of the vector c according to its observed frequency distribution among $0.85N$ agents, making sure the first two dimensions of the selected c fit the previously allocated vector d.

3. $0.15N$ agents remained without c imposing no statistical restriction on the joint distribution of characteristics for these agents (except for the farm size and share of arable land reflected in d).

4. Next, looping over farm size quintiles, random farm sizes were distributed among the $0.2N$ agents associated to each respective quintile according to the corresponding partial marginal distribution. A theoretical constraint ensured that the allocated arable land resulting from multiplying the allocated farm size with the minimum share of arable land of the agents defined by c did not surpass 500 ha. The simple, order-based distribution algorithm described in Ch. 5.2.3 was used here and in the following steps unless otherwise noticed in order to ensure covering the full range of the distribution function.

5. Similarly, the arable, grassland and forest shares were allocated within each quintile, making sure that the sum of these was close to one and the resulting non-forest area not greater than 500 ha.

6. At this point, the resulting grassland and arable land ownership was used to spatially allocate plots in the study area to each agent as described in the next section and thus defined the soil composition of the land owned by the agents.

7. Again looping over quintiles, the observed animal numbers were randomly distributed to the agents. The basic restriction was the total animal-to-land ratio, which had to be lower than *gvpha* livestock unit (LU) per ha, where *gvpha* was assumed to lie between 2 and 3 and subject to calibration. Further, the number of calves and heifers was expected to be characteristically related to dairy cows, and the number of farrows to be dependent on the number of sows, respectively fattening pigs. Specifically, the algorithm used the following steps (separately for the agents with and without associated c vector):

   (a) Dairy cows were randomly allocated, ensuring the animal-to-land ratio was respected taking into account the expected minimum number of young animals entailed by the number of dairy cows (0.35 calves and 0.35 heifers per dairy stable place) and the minimum numbers of other animals defined by the quintiles associated with the agent.

   (b) Medium-aged cattle and calves were allocated using the Hungarian Method with random component (see chapter 5.2.3), where the deterministic cost component was set to infinity if the animal-to-land ratio was violated, to zero if the ratio of young animals to dairy cows was greater or equal 0.35, and to $\ln\left(\frac{1}{0.35dairy-young+1}\right)$ otherwise.

(c) Mother cows, horses, fattening pigs, sows and sheep were allocated subsequently ensuring the animal-to-land ratio was respected taking into account the already determined numbers of other animals, respectively the minima defined by the quintiles associated with the agent.

(d) Other pigs (i.e. mostly farrows) were then distributed using several loops: First, it was attempted to distribute values only to agents that had both sows and fattening pigs. Then, it was attempted to distribute the remaining values to agents which had either sows or fattening pigs. Third, values are allocated to those agents, who neither had fattening pigs nor sows, but were supposed to have farrows. In the first two attempts, values were accepted if they lay in a range of $\pm$ 15% of a third of the number of fattening pigs plus 6.21 times the number of breeding sows, reflecting the typical relation of stable places and turnover times of the production activities.

8. The statistical information on animal numbers that had been randomly allocated to the agents was transformed into model assets:

   (a) The livestock numbers were transformed into corresponding types and quantities of stable capacities.

   (b) For dairy cows, stable places were assumed to be in stanchion stables up to a number of 40 cows, above this cubicle loose-housing stables were allocated. Up to 10 dairy cows, a bucket milking machine was assigned, up to 40 cows a milking pipeline, and above 40 cows usually a herringbone milking parlor. Alternatively, between 60 and 160 cows an automatic milking system (AMS) was allocated with 10% probability [Harms and Wendl, 2009] and above 160 dairy cows a rotary milking parlor was allocated with 50% probability.

   (c) Agents received milk and manure storage facilities, feeding equipment as well as milk quotas corresponding to the amounts required according to the model assumptions.

9. The number of biogas plants to be allocated in each of the years was inferred based on the results of the farm survey, which asked for the capacity and year of establishment of biogas plants currently installed, and scattered information found in Fachagentur nachwachsende Rohstoffe e.V. [n.d.], Dederer and Messner [2011] and Hartmann [2008]. At maximum 17 biogas plants with capacities ranging from 75 to 400 kW in 2007, nine biogas plants with capacities ranging between 40 and 420 in 2003 and four biogas plants ranging between 40 and 420 kW in 1999 were allocated. These biogas plants were randomly distributed among those agents with the theoretical ability to produce feedstock for an electricity production using at least 80% of the plant capacity, taking into account the arable land, grassland and animals owned by the agent.

10. Tractors and other machinery were distributed according to rules developed based on the machinery endowments observed in the farm survey, expert information and model assumptions. The rules related the amount of arable land, grassland, expected manure to be spread and animals owned to certain combinations of tractors and implements as shown in table D.2 in the appendix.

11. Finally, relevant EU CAP entitlements had to be distributed. For the years 1999 and 2003, mother cow quotas were allocated by simply assuming agents own quotas corresponding to the mother cows they own. Milk quotas had been handled already as

explained above. For 2007, single farm payment entitlements had to be allocated. This was done by allocating grassland, arable and set-aside entitlements according to the land endowments of each agents, and determining their values according to the regulations, assuming agents obtained all premiums they could have potentially received in 2003 given their current (i.e. 2007) asset and land ownership.

8.2.3 Spatial distribution of farms

CORINE land cover maps – more specifically CLC2000 [2004] for 1999 and 2003, and CLC2006 [2009] land cover maps for 2007 – provided information on the basic spatial extent of urban, arable, grassland, forest and other natural areas. For our purpose, the original 47 land use categories of the CORINE datasets were aggregated into 13 categories shown in table 8.9.

Table 8.9: *Land use categories used for the spatial allocation of agents*

Code	Description	CLC Codes
0	Urban	111-112
1	Industrial & traffic	121-142
2	Arable	211-213
3	Permanent crops	221-223
4	Pasture	231
5	Mixed cultivation patterns	241, 242, 244
6	Agriculture & natural vegetation mixed	243
7	Forest	311-313, 323-324, 990
8	Heathland	322
9	Natural grasslands	321
10	Wetlands	411-423
11	Water	511-523, 995
12	Rocks	331-335

The spatial distribution of plots proceeded by first randomly distributing farmsteads over the plots classified as urban or arable (0 or 2). Then the forest, arable and grassland area previously determined for each agent was randomly distributed using the `mpmasdist` spatial allocation mechanism (see section 5.2.3), which divided the area owned by an agent into random-sized plots and sequentially placed these plots as close as possible to the farmstead or any other previously allocated plots of the agent. Forest area could be placed on plots of category 7 only. While categories 5 and 6 were considered suitable for both arable and grassland, plots were allocated to categories 2, respectively 4 first until all of these were used. Only after that, plots of category 5 and 6 were included into the distribution process.

The distribution mechanism in its current implementation took several days of run time to complete, such that only a limited number of different spatial distributions were generated.

The resulting agent property maps could then be overlaid with the soil maps described in section 8.1.1 to determine the composition of soil types on each agent's land. Additionally, the ownership of the land used by the agent had to be determined. Following estimations based on the statistics provided in Betzholz [2011] the share of rented land was set to 0.4 for agents with less than 20 ha of farm area. For agents using between 20 and less than 45 ha the rental share was calculated as $0.2886 + 0.0062 area$, and for agents using more land the formula $e^{-1.5096+0.2487\ln(area)}$ was used. A corresponding number of cells was then identified as rented, starting with the one furthest away from the farmstead.

8.2.4 Household composition

The generation of realistic household compositions started by randomly determining the age of the household head (age_{hh}) and whether he is married or not based on the statistical distributions reported in destatis [2011] and destatis [2012*a*]. The age of the household head's wife was drawn from the normal distribution $N(age_{hh}, 2)$. The number of children was estimated by randomly determining whether the household head's wife gave birth for each age between 15 and her current age. The probability of giving birth at each age was taken from destatis [2012*b*], but was proportionally increased by a factor *birth_factor_past* as the statistical data used cover only a relatively recent period and birth rates in the past have probably been higher. The career path and gender of children were determined using the the same coefficients that were used in the model (see section B.2.1). The procedure of determining marriage status and potential descendants (i.e. grandchildren of the household head) is repeated for each child.

The presence of the household head's retired parents was determined by first individually drawing their potential age from the normal distribution $N(age_{hh} + 28, 2)$ and then using the mortality information from destatis [2012*c*] to determine whether they actually reached this age or died in the past.

8.3 Turing test

Before conducting the actual validation and calibration experiments, the model was exposed to a Turing test in order to test the plausibility of the model predictions, and to obtain an impression of the accuracy of expert predictions of farmers' land use decisions. Four different farms from the PAK346 farm survey were selected, one mixed bull-fattening and dairy farm, one dairy farm, one pig producing farm, and one biogas and dairy farm, which were considered representative for the farms in the survey.[8]

For each sample farm, the model was run with various parameter combinations to obtain different cropping plans for the seasons 2009/2010, the year the survey was conducted. Five simulated plans spanning the whole spectrum of predictions were selected and mixed with the actual land use as responded in the survey.

The resulting mix of six potential cropping plans was presented to 22 farmers and extensionists: Eight from the study area Central Swabian Alps, ten from the project's other study area Kraichgau and two from other parts of Baden-Württemberg. The agricultural experts were asked to identify the cropping plan that they believed was the one actually used by the interviewee in the season 2009/2010 and to strike out those plans that they considered too unrealistic to be used by a farmer in reality.

Each set of plans was accompanied by basic information about the farm: the area of arable and grassland used by the farm, the number of dairy cows, fattening bulls, fattening pigs and sows owned and whether a biogas plant existed on the farm. (This information corresponds to the "hard" farm-specific data that was also fed into the model. All other model input data is either not farm specific or derived from general rules, both of which experts should replace with their own experience.)

As figure 8.5 shows, only very few of the participants considered the actual production plan the most realistic one for any of the four farms. Moreover, for all farms at least seven participants, i.e. nearly a third, ruled out the actual cropping plan as unrealistic (not counting those participants who misunderstood the task and crossed out every alternative except the

[8] This selection was done by Viktoriya Latysheva in the course of her Master thesis.

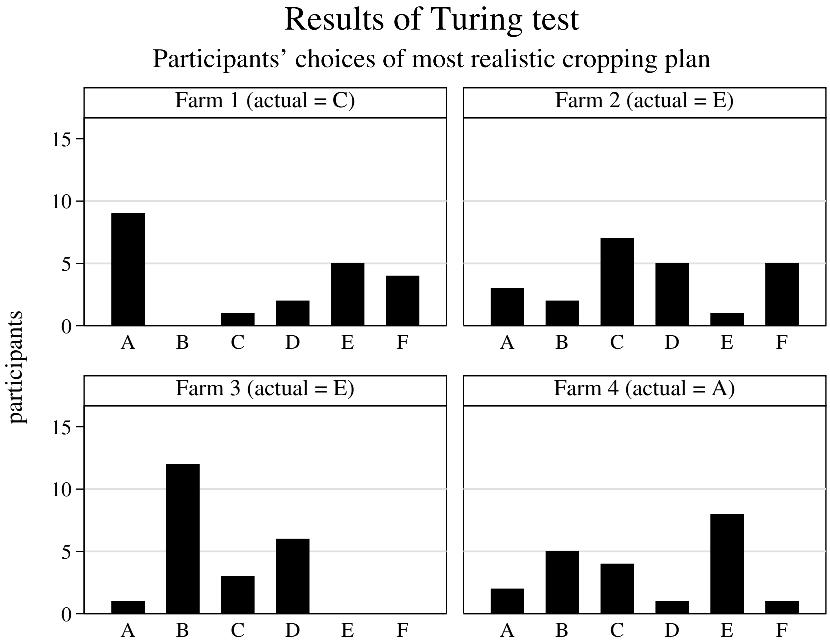

Figure 8.5: *Results of the Turing test with 22 participants guessing the observed production plan of four farms among six suggested ones (five simulated).*

one considered the actual one). Differences between participants from different study areas could not be discerned. In the discussion after the test, participants could mention peculiar characteristics of each actual plan that struck them as odd and kept them from choosing it. Interestingly, except for farm 2, the alternative considered realistic by most participants are the closest or second closest to the actual one according to goodness-of-fit. Participants also mentioned, that not enough basic information about the farm was given to arrive at more accurate results. While I do not want to overinterprete the results, the test shows at least that the model is capable of producing results that hold up to expert scrutiny, and that, given only the same farm specific information available to our modeling framework, experts might come to similar predictions as our model.

8.4 Calibration experiments

Following the conclusions drawn in chapter 4, the process of calibration followed a sequential, Bayesian-like approach which was not intended to identify a single, best parameter combination, but only to reduce the model uncertainty as far as considered possible without running the danger of overfitting. The whole process – which was as much about calibration as about technical verification of the model – can be subdivided into two phases:

The first phase can be characterized as an informal search for errors and significant omissions and comprised numerous iterative steps. Each step would usually include an elementary effects screening and then a full factorial including the most relevant parameters (i.e. those with the strongest effects on goodness-of-fit). The distribution of goodness-of-fit over the factorial as well as the predicted land use, farm type patterns and livestock numbers was then examined. Whenever the distribution did not cover the true value, i.e. there was no combination of parameters that was at all able to reproduce the observations, this prompted the identification of an error in the model implementation, or a reconsideration of a theoretical or empirical aspect of the model, usually leading to the introduction of another parameter and its inclusion into the testing procedure.

The second and final phase was then the actual calibration, in itself not different from any of the steps of the first phase, but including again all parameters tested or introduced during the previous experiments, globally varying them, and formally applying the calibration criteria. Only this final phase is reported here as it implicitly summarizes the previous process.

8.4.1 Parameter variation

Parameter settings tested during the calibration and validation experiments are listed in table 8.10 and explained in the following:

Table 8.10: *Parameter settings tested during the calibration and validation experiments.*

Initial agent population	*gvpha*	2, 2.5, 3
	popseed	4 different seeds
	birth_factor_past	[1; 1.05]
	potsuc_prob_male	[0.5; 1]
Yields	yield set	urt, uad, umx, xn3
	maize_yc	[0.75; 1]
	wheat_normal	[1; 1.1]
	wps	yes/no
	wps_coef	[1;1.3]
Crop rotation	*maize_on_maize*	0, 1/2, 1
	maizerotlimit	[0.4;0.6]
Field working days	*clregion*	4, 5
	fielddayprob	60%, 80%
Contracted field work	*workforhirecoef*	[0; 1]
	proptohire	[0.5; 2]
Cattle feeding	*pasturelabor*	[1; 3]
	pastureloss	[0.1; 0.4]
	freshgrasslabor	[1; 3]
	freshgrassloss	[0; 0.2]
Markets	*trade_yf_cattle*	no/yes
	trade_smaize	no/yes
	nawarosale	no/yes
	biertreber	yes/no
	kwkyno	yes/no
Manure	*high_manure_maize*	yes/no
	manure	[1; 1.5]
Farm household	*ihorizon_type*	4 different versions

A total of 48 different initial agent populations were generated for each year using four different seed values for the random generator, three different settings for the *gvpha* limit used during the random allocation of livestock, two different values for *birth_factor_past* and two different values for *potsuc_prob_male*, the probability of a male descendant to pursue a career in farming or not (see section 8.2).

Parameters related to crop yields include the four yield sets, the scaling factor for first year wheat (*wheat_normal*), a scaling factor for silage maize yields (*maize_yc*) to reflect the uncertainty of maize production, a parameter that controls the inclusion or exclusion of whole-plant silage production (*wps*) and a scaling factor for whole-plant silage yields (*wps_coef*). The production of whole-plant silage had to be considered an innovation and there was no data on the diffusion of this technology in the observation years and little information on crop yields. Two parameters affect the potential maize area of an agent: *maize_on_maize* controls the number of years maize can be grown after itself, and *maizerotlimit* constitutes the upper limit for the total share of maize in the crop rotation. Parameters related to field working days include the KTBL climate region (*clregion*) and the certainty level of the assumed field working days (*fielddayprob*). The *workforhirecoef* scales the price for contracted field work between the maximum and the minimum of the range given in KTBL [2010], while the *proptohire* coefficient scales the availability of hired field work per hour with suitable weather (see B.1.3). The *pasturelabor* and *freshgrasslabor* coefficients scale the amount of labor necessary for pasturing, and the *pastureloss* and *freshgrassloss* coefficients indicate the share of pasturing, respectively harvest losses.

Five parameters result from omitting the implementation of local resource markets: The *trade_yf_cattle* parameter controls whether farmers can buy young female cattle (3-months old calves and heifers) or whether they have to raise them themselves. While in reality, of course, there is a market for female cattle, the model is not able to reflect the quantity effects on the market. This leads to too many farms expecting to buy female cattle from the market without anyone producing them. Similarly, the *trade_smaize* parameter controls, whether farmers can sell silage maize on the market or only use it on their own farm. The *nawarosale* parameter controls whether there is any demand for crops produced on set-aside land under the NaWaRo regulations (see appendix B.1.9) or whether these can only be used in own biogas plans. The *biertreber* parameter controls whether brewery by-products are generally available as fodder. The *kwkyno* parameter controls whether there is unlimited or no demand for the surplus heat of biogas plants (see appendix B.1.5).

Two parameters are related to the maximum amount of manure that can be applied to a crop. The production activities that require manure assume a standard amount of manure use, which effectively creates an upper limit of manure application to each crop. The *manure* parameter scales this upper limit on manure use of all production activities in order to test whether the assumed standard amounts may be too low. The *high_manure_maize* is specific to silage maize production. It controls the inclusion of specific silage maize production activities that assume a manure amount of 30 m^3 instead of the standard 20 m^3 (with the complementing mineral fertilization reduced).

Last, the *ihorizon_type* represents four different implementations of the influence of farm household composition on the production decisions of the farm: In the simplest version, the investment horizon is independent of the farmers age and ζ_{H2ut} (cf. section B.2.3) is equal to zero, i.e. the farm manager derives no utility from employing potential successors. In the second version, the investment horizon remains independent of the agent household head's age, but the agent farm manager derives utility from employing potential successors, i.e. ζ_{H2ut} is equal to one. In the third version, the investment horizon depends only on the age of the current household head, while in the fourth version it depends on the age of the potential successor with highest priority. ζ_{H2ut} is equal to one in both cases.

8.4.2 Goodness-of-fit measures

To check the predictive accuracy of our model, the simulated land use decisions was compared to the panel observations using three indicators: the total land use in the study area, the total livestock numbers in the study areas, and the classification of farm according to the EU farm typology [Sauer and Hardeweg, 2006], specifically at the level of principal type of farm (PTOF). While the first two indicators reflect the aggregate response of the agricultural sector, the distribution of farm agents over PTOF classes reflects the combination of different production activities within one farm, and indicates whether the heterogeneity of farm setups has been well reproduced.

The farm classification by PTOF is shown in table 8.11. Total land use of full-time farms in the study area obtained from FDZ [2010] for the three observation years is shown in table 8.12. Due to the privacy constraints, several original crop categories had to be aggregated. Not all crop categories have been included into the model. For goodness-of-fit comparison, they were associated with the closest representative in the model in terms of cultivation pattern and use. Both, the distribution of total area over land uses and the distribution of farms over farm types are restricted by an overall total and thus constitute categorical data: For these, the model efficiency based on the standardized absolute error ($ESAE$, see chapter 3.3) was used as goodness-of-fit measure. For the livestock numbers (Tab. 8.13) a standard model efficiency was used.

Table 8.11: *Overview of farm classification for goodness-of-fit calculation.*

GTOF	PTOF	Code	1999	2003	2007
Field crops	Specialist cereals/oilseeds	13	40	25	18
	Specialist mixed field crops	14	0	0	5
Horticulture	Horticulture	20	14	11	9
Permanent crops	Permanent crops	30	7	5	5
Ruminant	Dairy	41	359	243	212
	Cattle fattening/raising	42	18	15	20
	Cattle mixed	43	71	49	38
	Grazing livestock mixed	44	28	27	18
Granivore	Granivore	50	72	58	44
Mixed field crop	Mixed field crop	60	17	9	8
Mixed livestock	Mixed ruminant	71	74	40	24
	Mixed granivore	72	42	22	20
Mixed	Mixed crop/ruminant	81	73	60	65
	Mixed crop/livestock	82	118	42	47

Table 8.12: *Overview of aggregate crop areas used for goodness-of-fit calculation.*

		Area		
FDZ 2007 categories	Model category	1999	2003	2007
Summer barley		4,822	4,266	4,188
Oats, mixed summer cereals		1,290	1,163	866
	Summer barley	6,112	5,429	5,054
Winter barley, mixed winter cereals	Winter barley	3,153	2,962	2,784
Winter wheat		4,965	4,450	4,997
Triticale		795	908	934
	Winter wheat	5,760	5,359	5,931
Winter rapeseed	Winter rapeseed	2,673	2,220	2,184
Silage maize	Silage maize	1,621	1,720	3,030
Field grass		32	22	183
Clover, alfalfa		2,086	1,197	1,483
Other field forage		546	271	100
Fodder peas, fodder beans		373	279	149
	Field grass	3,038	1,770	1,915
Fallow	Fallow	1,047	1,038	586
Total Arable		**23,403**	**20,497**	**21,483**
Pasture	Pasture	507	486	298
Meadow		15,906	14,138	13,868
Mown pasture		1,140	1,341	1,297
	Meadow	17,045	15,479	15,165
Total Grassland		**17,552**	**15,964**	**15,463**
Total		**40,956**	**36,462**	**36,946**

Table 8.13: *Overview of livestock numbers used for goodness-of-fit calculation.*

	Area		
	1999	2003	2007
Dairy cows	14,531	12,307	11,711
Cattle > 1yr, < 2 yrs	8,073	7,377	7,824
Cattle < 1yr	12,766	10,632	10,906
Mother cows	1,027	796	652
Fattening pigs	22,545	24,663	24,172
Sows	12,021	10,153	10,423
Piglets (> 8kg)	16,754	20,369	20,667

8.4.3 Reducing the parameter space

The parameters described above (section 8.4.1) all reflect some uncertainty about the best representation of reality in the model. To reduce parameter uncertainty, a stepwise procedure was employed to exclude parameter settings that consistently performed inferior than their alternatives. To guard against overcalibration, only those settings were excluded whose inferiority was consistently observed in all three observation years. As the observation years differ by several structural breaks (especially with respect to the policy setting, but also with respect to price levels), there is a reasonable chance that parameters fulfilling these conditions can be considered invariant and present a good choice also for scenario analysis.

There were two exceptions to this rule: First, parameters affecting the initial agent population could differ from year to year. Here, stable parameters were not so important, but rather the focus was selecting an agent population that best represented the real-world farming population in a given year. Second, in the case of the yield data, it was clear *a priori*, that the 'xn3' set of yields would be used during the scenario analysis: Expert-N was required in order to simulate future yields and therefore yields calculated by Expert-N had to be used also for the baseline to ensure consistency. The other yield sets were included into the calibration experiments to guard against overcalibrating the other parameters to this specific simulated set of yields, which was subject to considerable uncertainty itself. Analogous to the condition that a parameter setting was only excluded if it consistently performed inferior in all years, it was also excluded only if it consistently performed inferior for all yield sets. Yield sets were therefore never fixed in any of the screening steps.

Even if only two factor levels for each of our parameters were considered, $3 \cdot 2^{26} > 200\,\mathrm{Mio.}$ runs would have been required to run a full factorial design for our setting. With a model run time of at least 30 minutes for one simulation period, this is clearly infeasible. Rather, two rounds of elementary effects screening were used followed by a full factorial with the remaining parameters.

The first elementary effects screening used ten repetitions for each elementary effects and thus required $10 \cdot (26+1) \cdot 3 = 810$ model runs. The design was created in *R* using the morris() function of the 'sensitivity' package [Pujol et al., 2012], which includes the space filling improvements of Campolongo et al. [2007] and allows for choosing a different number of levels for each parameter. The later came very handy in this case, as the majority of parameters was discrete, many with only two defined levels. Based on the sample, Morris sensitivity measures were calculated to assess the importance of each parameter in determining the three goodness-of-fit measures (see 8.4.2)

Parameters were then grouped into three groups: Those parameters that showed little or no effect (low μ^* and low σ^*, see p. 45) on neither goodness-of-fit nor income or land use could be fixed at their theoretically most convincing values for the next steps as there was little hope to gain much insight on them in the calibration procedure.

The second group were parameters for which a clearly superior setting could already be identified in the screening. This was indicated by a very low difference between the absolute value of μ and μ^* in the most simple case. The sign of μ then indicated whether the parameter was to be fixed at the lower or upper end of the range. This applied mostly in the case of binary parameters. In other cases, a closer analysis of the sample points revealed that moving away from a certain value consistently deteriorated goodness-of-fit. These parameters could then be fixed at the identified value for the subsequent steps of the calibration.

The analysis was continued with the third group of parameters showing important, but ambiguous effects on goodness-of-fit. A second elementary effects screening ($195 \cdot 3 = 585$ runs) was used to reassess their importance after fixing the parameters of the first two groups.

The same procedure as above was repeated and then, as a third step, a full factorial with the six most important unfixed parameters ($324 \cdot 3 = 972$ runs) was simulated.

Table 8.14 shows the results of the parameter fixing. Parameters that could be fixed based on the empirical results are *gvpha*, *popseed*, *pasturelabor*, *pastureloss*, *fielddayprob*, *trade_yf_cattle*, *trade_smaize* and *nawarosale*. The potential range for *maize_yc* could at least be reduced. Parameters that were fixed temporarily due to insignificant effects on goodness-of-fit remained part of the reduced parameter space, since insignificant effects on goodness-of-fit do not rule out important influences on the effect of climate change or policy analysis.

Table 8.14: *Parameter fixing during the calibration and validation experiments.*

Parameter group	Parameter	EE1	EE2	FF
Initial agent population	*gvpha, popseed*	-	-	3 comb. per year
	birth_factor_past	(1)	(1)	(1)
	potsuc_prob_male	(1)	(1)	(1)
Yields	yield set	-	-	(xn3)
	maize_yc	-	(0.75)	0.8-0.9
	wheat_normal	-	(1.05)	-
	wps	-	0	0
	wps_coef	(1.15)	(1.15)	(1.15)
Crop rotation	*maize_on_maize*	(3/4)	(3/4)	(3/4)
	maizerotlimit	-	(0.5)	(0.5)
Field working days	*clregion*	(4)	(4)	(4)
	fielddayprob	-	80%	80%
Contracted field work	*workforhirecoef*	-	-	-
	proptohire	(1)	(1)	(1)
Cattle feeding	*pasturelabor*	3	3	3
	pastureloss	0.4	0.4	0.4
	freshgrasslabor	(3)	(3)	(3)
	freshgrassloss	(0.1)	(0.1)	(0.1)
Markets	*trade_yf_cattle*	0	0	0
	trade_smaize	0	0	0
	nawarosale	0	0	0
	biertreber	-	(1)	-
	kwkyno	(0)	(0)	(0)
Manure	*high_manure_maize*	-	(1)	-
	manure	-	-	-
Farm household	*ihorizon_type*	(2)	(2)	(2)

Values show fixings applied based on the results of each experiment.
Values in parenthesis denote temporary fixing at theoretical values due to less significant effects on goodness-of-fit
EE: elementary effects screening
FF: full factorial

8.5 Empirical validation

The reduced parameter space constituted the basis for further scenario and policy analysis. For most scenario analysis, it would still be too large to be fully explored through simulation and a specific screening was therefore used for each further application of the model in order to identify those parameters that had the greatest influence on the outcome variables relevant for each analysis.

Before conducting any specific scenario analysis, though, this section compares simulation outcomes to land use, farm type and farm accounting data to allow a detailed impression of the empirical performance of the model and the reduced parameter space.

The Morris designs of the calibration could not be used as a basis for this comparison, because applying the parameter fixing would have reduced the design to too few replications for this purpose. Also the full factorial design of step three was not suitable as it misses parameters that might have an important influence on income.

However, at least the Morris runs of calibration step 1 can be reused to calculate elementary effects on average income, income quartiles and crop areas. Based on this screening, the six unfixed parameters with highest influence on income and land use were selected and included into a full factorial design with 96 runs, while fixing the other parameters at their calibrated or theoretically most convincing values.

8.5.1 Goodness-of-fit of land use and farm type predictions

Figure 8.6 shows the distribution of the goodness-of-fit indicators over these 96 runs. The $ESAE$ for the farm classification by EU typology ranged between 0.65 and 0.75, with values improving when going back in time. The model efficiency (ME) for livestock numbers ranged between 0.9 and 0.95 for 2003 and 2007, and between 0.8 and 0.95 for 1999.

The $ESAE$ for land use lay between 0.8 and 0.87 for the year 2007. It decreased when going back further into the past. This can be explained by the use of the simulated yield set ('xn3'), which does not account for yield trends over time. The statistical yield sets ('uad', 'urt'), which do account for different yield expectations in the past, perform better, as can be appreciated in the results of the third step of the calibration procedure (see figure E.1 in the appendix).

A comparison of the predicted and observed areas for individual crops in 2007 as shown in table 8.15 reveals a tendency to consistently overestimate wheat, rapeseed and fallow areas, and an underestimation of silage maize and summer barley areas. The bias in the silage maize area is consistent with the omission of silage maize trade as a result of the calibration. (Note: Histograms showing the full distribution of land use area predictions for all three years can be found in figures E.4 and E.5).

The discrepancy in the farm type distribution as shown in table 8.16 can be explained mainly as a consequence of the omission of horticulture, fruticulture as well as sheep and other forms of extensive grassland use from the model. Agents falling into these categories are predominantly absorbed by the class of field crop specialists.

8.5.2 Comparison with FADN data

To validate the economic outcomes of the simulations, the predicted accounting data for 2007 were compared to publicly available average accounting data from the Farm Accounting Data Network, which are representative for the state of Baden-Württemberg for the cropping

Table 8.15: *Comparison of predicted and observed land use in 2007 (Prediction shown as average and standard deviation over 96 simulation runs).*

	Predicted			Observed
Winter wheat	6,577	±	207	5,931
Winter rapeseed	3,001	±	92	2,184
Summer barley	3,898	±	185	5,054
Winter barley	2,567	±	141	2,784
Fallow	1,544	±	11	586
Silage maize	2,351	±	200	3,030
Field grass	2,005	±	360	1,915
Meadow/Mown pasture	14,558	±	152	15,165
Pasture	598	±	81	298

Table 8.16: *Comparison of predicted and observed farm classification in 2007 (Prediction shown as average and standard deviation over 96 simulation runs).*

PTOF				
Code	Description	Predicted		Observed
13	Specialist cereals/oilseeds	78	± 5	18
14	Specialist mixed field crops	4	± 2	5
20	Horticulture	0	± 0	9
30	Permanent crops	0	± 0	5
41	Dairy	190	± 6	212
42	Cattle fattening/raising	17	± 3	20
43	Cattle mixed	22	± 4	38
44	Grazing livestock mixed	3	± 1	18
50	Granivore	46	± 6	44
60	Mixed cropping	7	± 2	8
71	Mixed grazing livestock	31	± 3	24
72	Mixed granivore	36	± 2	20
81	Mixed crops grazing livestock	56	± 8	65
82	Mixed crops livestock	43	± 4	47

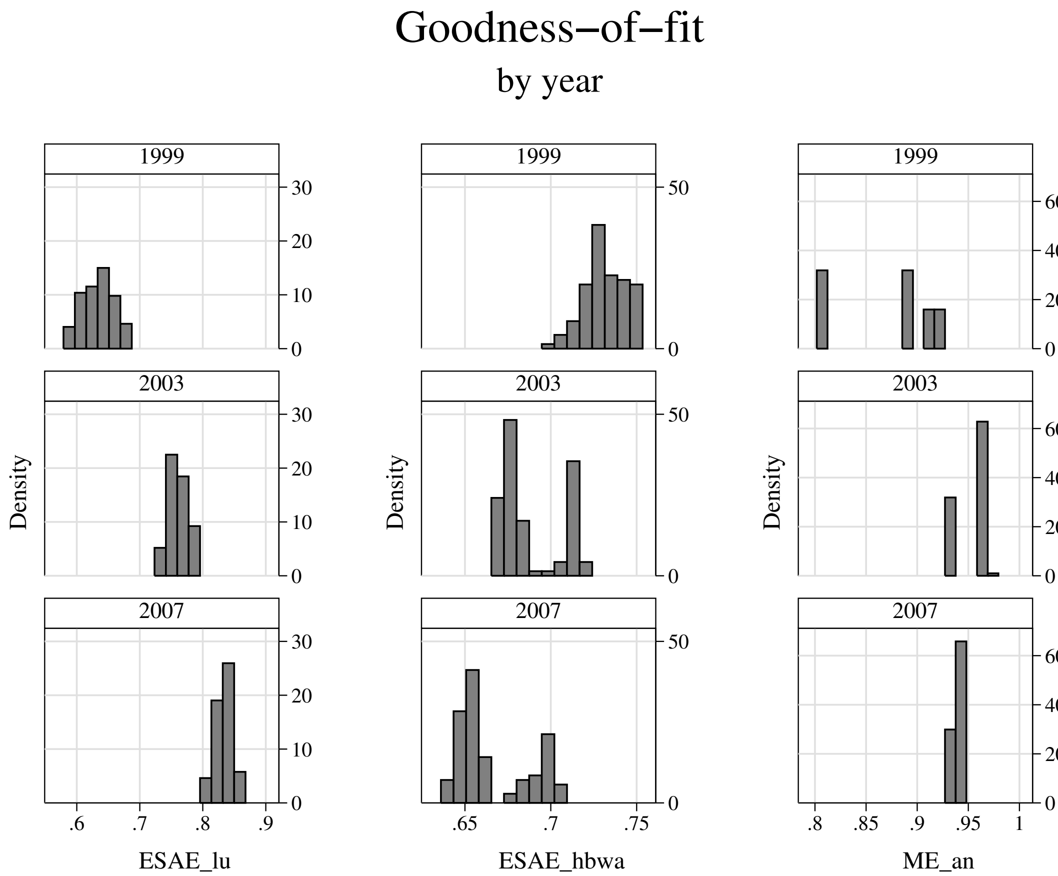

Figure 8.6: *Goodness-of-fit for total land use, farm type distribution and animal numbers for the three observation years.*

season 2006/2007 [BMELV, 2013] and grouped by general type of farm (GTOF), the highest level of the EU farm typology. This comparison could only be informal and coarse as the reference data are representative for a much wider area stretching many different agricultural zones.

Table 8.17 shows a comparison of basic characteristics between the simulated farm population and the statistical sample which constitutes the basis for the FADN data. The most striking difference can be observed for the class of field crop specialists, where the simulated farms show a much lower labor use than the state average. As said before, the model class of field crop specialists absorbed farms that could not be modeled correctly because of the omission of relevant activities from the model. Especially the smaller of these agents employed very low amounts of labors and did not show characteristics of a full time farm *ex post*. The table therefore also includes a separate column including only those agents showing characteristics of a full-time farm *ex post* (ESU ≥ 16 and workforce ≥ 0.75 person-year).[9]

[9] 0.75 py was used as a relaxed criterion for full time employment, because the model does no account for administrative overhead

Table 8.17: *Basic farm characteristics averaged by farm type (GTOF). Mean and standard deviation over 96 runs. Observation year 2007.*

		All	Full-time	BW average[10]
Field crop specialists				
No. farms		82.8 ± 6.1	29.7 ± 5.7	
ESU	€	34.7 ± 4.2	56.7 ± 9	62.5
UAA	ha	73.9 ± 7.9	116.6 ±16.2	72.7
Grassland	ha	19.5 ± 2.2	27.4 ± 3.8	7.0
Arable	ha	55.0 ± 6.6	90.5 ±14.5	63.7
Workforce	py	0.9 ± 0.1	1.8 ± 0.2	2.1
Ruminant specialists				
No. farms		231.4 ± 4.0	217.7 ± 4.3	
ESU	€	68.2 ± 1.5	72.0 ± 1.1	55.3
UAA	ha	64.2 ± 1.2	67.2 ± 0.6	52.8
Grassland	ha	35.6 ± 0.9	36.8 ± 0.5	31.9
Arable	ha	28.9 ± 0.6	30.6 ± 0.4	20.9
Workforce	py	2.4 ± 0.0	2.5 ± 0.0	1.5
Granivore specialists				
No. farms		46.2 ± 5.8	33.8 ± 5.2	
ESU	€	101.4 ± 5.9	121.4 ± 7.9	92.3
UAA	ha	58.0 ± 4.4	68.0 ± 5.6	48.2
Grassland	ha	13.7 ± 2.5	14.0 ± 3.0	3.8
Arable	ha	44.4 ± 3.2	53.9 ± 4.0	44.3
Workforce	py	1.5 ± 0.3	1.9 ± 0.2	1.9
Mixed field crop				
No. farms		6.9 ± 2.0	5.8 ± 2.0	
ESU	€	58.0 ±12.5	62.4 ±13.2	65.0
UAA	ha	90.7 ±15.1	95.9 ±16.9	48.3
Grassland	ha	24.2 ± 7.8	24.7 ± 8.7	5.6
Arable	ha	66.5 ±13.6	71.2 ±15.8	38.3
Workforce	py	2.0 ± 0.2	2.4 ± 0.2	2.1
Mixed animal				
No. farms		66.9 ± 2.7	65.3 ± 2.4	
ESU	€	82.3 ± 5.5	83.4 ± 5.1	59.5
UAA	ha	69.2 ± 7.1	69.7 ± 6.2	50.6
Grassland	ha	28.1 ± 3.5	27.9 ± 3.2	17.5
Arable	ha	41.4 ± 3.8	42.0 ± 3.7	32.8
Workforce	py	2.4 ± 0.1	2.5 ± 0.1	1.6
Mixed				
No. farms		91.8 ±13.1	76.2 ±10.7	
ESU	€	61.4 ± 5.3	67.2 ± 5.6	65.0
UAA	ha	88.4 ± 7.1	97.7 ± 7.4	64.8
Grassland	ha	28.9 ± 2.4	32.5 ± 2.6	9.9
Arable	ha	59.6 ± 5.1	65.3 ± 5.6	54.0
Workforce	py	2.0 ± 0.1	2.2 ± 0.1	1.7

[10] The reference is taken from BMELV [2013]. Data are representative for the state of Baden-Württemberg in 2006/2007

Predicted average labor use of ruminant specialists and mixed animal farms was higher than the state average, but utilized agricultural areas (UAA) were generally higher as well, except for field crop specialists. Except for ruminant specialists, grassland shares of model agents were higher than the state average. This reflects the mountainous character of the study area, though the imprecision of the estimated joint distribution functions may also have had an influence.

Table 8.18 lists four income indicators for each farm type: the average profit per hectare, the average value added per farm, the average profit plus labor costs per person employed and the average withdrawals for personal consumption and personal taxes per hectare.

In general, the predicted farm income indicators for ruminant and mixed animal farms were higher than state averages, while the predicted income indicators for field crop and granivore specialists were lower than observed state averages. Due to the differences in farm structure, it remains uncertain whether this merely reflects the specific characteristics of the study area or hints at misspecifications in the model. At least, the table shows that the income predictions do not seem completely out of touch with reality.

Table 8.18: *Mean and standard deviation of mean income indicators per farm type (GTOF) over 96 runs. Observation year 2007.*

		All	Full-time	BW average[11]
Field crop specialist				
Profit	€/ha	267 ± 32	308 ± 30	538
Value added	€	36,847 ± 6,575	65,596 ±13,926	64,907
Profit + labor cost	€/py	70,613 ±64,150	26,131 ± 4,885	24,097
Withdrawals	€/ha	392 ± 16	421 ± 30	366
Ruminant specialist				
Profit	€/ha	1,029 ± 42	1,114 ± 21	629
Value added	€	92,468 ± 2,200	98,097 ± 1,731	42,810
Profit + labor cost	€/py	31,718 ± 993	33,521 ± 640	22,494
Withdrawals	€/ha	800 ± 14	841 ± 15	592
Granivore specialist				
Profit	€/ha	448 ± 55	536 ± 68	947
Value added	€	57,143 ± 3,586	68,714 ± 5,890	64,158
Profit + labor cost	€/py	20,647 ± 5,678	23,593 ± 4,678	26,741
Withdrawals	€/ha	648 ± 43	653 ± 48	896
Mixed field crop				
Profit	€/ha	483 ± 69	510 ± 61	783
Value added	€	68,477 ±17,099	74,889 ±17,512	59,052
Profit + labor cost	€/py	29,375 ± 8,434	27,011 ± 8,908	23,114
Withdrawals	€/ha	488 ± 63	456 ± 76	673
Mixed animal				
Profit	€/ha	923 ± 50	931 ± 56	615
Value added	€	92,545 ± 6,412	93,949 ± 6,291	42,496
Profit + labor cost	€/py	30,304 ± 1,731	30,160 ± 2,108	20,247
Withdrawals	€/ha	767 ± 27	773 ± 32	644

[11] The reference is taken from BMELV [2013]. Data are representative for the state of Baden-Württemberg in 2006/2007.

Table 8.18: *Income indicators per farm type (GTOF) over 96 runs, 2007 (cont.)*

		All	Full-time	BW average
Mixed				
Profit	€/ha	588 ± 17	617 ± 19	588
Value added	€	79,310 ± 6,678	89,260 ± 7,212	54,806
Profit + labor cost	€/py	37,094 ± 5,047	33,412 ± 1,893	23,303
Withdrawals	€/ha	519 ± 12	528 ± 16	512

Table 8.19 presents a more detailed breakdown of the income statement into individual items. Similar to profits, predicted operational revenues per hectare were higher than observed state averages for ruminant and mixed animal farmers, and lower for field crop and granivore specialists as well as mixed farms, but so were materials costs. Depreciation was higher for animal farmers, and matches quite well for crop farmers. Simulated average land rental costs per hectare were lower than observed ones, which may be explained by the higher grassland share in the study area.

For field crop dominated farm types and granivore specialists, simulated per-ha labor costs were much lower than observed ones, while they were higher for ruminant dominated farm types. The former may be explained by an extensive use of hiring of field work services in the simulation, which would not be accounted for as labor cost.[12] Another explanation may be the dominance of less labor and input intensive crops in the study area compared to e.g. sugar beets or potatoes in other areas. A general underestimation of labor requirements for field work may however also not be ruled out at this stage of model development.

The comparison of the financial result, i.e. the balance between paid and received interests, shows a much more negative balance in the simulation than in the reference data. In a certain respect, this is consistent with the higher per-ha depreciation for animal farms and points to a more capital intensive production, but may also be due to higher cost of financing or lower shares of own financing than in reality.

Table 8.19: *Detailed income statement. Mean and standard deviation of farm type (GTOF). Mean over 96 runs. Observation year 2007.*

		All	Full-time	BW average[13]
Field crop specialist				
Operational revenue	€/ha	823 ± 42	930 ± 70	1,347
EU direct payments	€/ha	250 ± 5	268 ± 6	314
Materials cost	€/ha	382 ± 23	435 ± 33	833
Depreciation	€/ha	265 ± 13	262 ± 5	228
Labor cost	€/ha	2 ± 1	5 ± 3	177
Land rental cost	€/ha	106 ± 2	128 ± 4	138
Balance Interest	€/ha	-50 ± 2	-57 ± 5	-40
Ruminant specialist				
Operational revenue	€/ha	2,341 ± 37	2,370 ± 44	1,554
EU direct payments	€/ha	374 ± 6	377 ± 5	327
Materials cost	€/ha	700 ± 15	677 ± 13	914

[12]Note that the model does not include hiring in of machinery, but only hiring in of full field work service, which appears in the material costs account.

[13]The reference is taken from BMELV [2013]. Data are representative for the state of Baden-Württemberg in 2006/2007.

Table 8.19: *Detailed income statement 2007 (cont.).*

		All	Full-time	BW average
Depreciation	€/ha	592 ± 17	569 ± 5	374
Labor cost	€/ha	65 ± 6	68 ± 7	36
Land rental cost	€/ha	86 ± 1	88 ± 1	99
Balance Interest	€/ha	-221 ± 6	-208 ± 3	-50
Granivore specialist				
Operational revenue	€/ha	4,098 ±234	4,119 ±243	5,660
EU direct payments	€/ha	263 ± 13	282 ± 18	284
Materials cost	€/ha	2,808 ±203	2,831 ±195	4,146
Depreciation	€/ha	625 ± 33	565 ± 41	558
Labor cost	€/ha	7 ± 7	10 ± 10	101
Land rental cost	€/ha	104 ± 3	113 ± 3	208
Balance Interest	€/ha	-364 ± 26	-339 ± 33	-101
Mixed field crop				
Operational revenue	€/ha	1,420 ±118	1,443 ±131	2,197
EU direct payments	€/ha	349 ± 48	371 ± 48	286
Materials cost	€/ha	743 ±109	756 ±127	1,158
Depreciation	€/ha	315 ± 16	313 ± 12	315
Labor cost	€/ha	12 ± 8	15 ± 11	233
Land rental cost	€/ha	120 ± 11	123 ± 13	196
Balance Interest	€/ha	-92 ± 8	-91 ± 11	-37
Mixed animal				
Operational revenue	€/ha	3,113 ±194	3,117 ±207	2,416
EU direct payments	€/ha	345 ± 13	348 ± 14	315
Materials cost	€/ha	1,591 ±130	1,587 ±138	1,751
Depreciation	€/ha	525 ± 27	525 ± 23	393
Labor cost	€/ha	59 ± 9	60 ± 9	38
Land rental cost	€/ha	97 ± 3	97 ± 3	152
Balance Interest	€/ha	-241 ± 16	-243 ± 17	-54
Mixed				
Operational revenue	€/ha	1,723 ± 64	1,659 ±101	2,154
EU direct payments	€/ha	369 ± 21	390 ± 25	317
Materials cost	€/ha	901 ± 60	835 ± 86	1,574
Depreciation	€/ha	339 ± 6	330 ± 8	297
Labor cost	€/ha	22 ± 2	26 ± 2	52
Land rental cost	€/ha	111 ± 2	115 ± 2	155
Balance Interest	€/ha	-122 ± 4	-115 ± 4	-48

8.5.3 MEKA participation

Finally, predicted participation in the agri-environmental program MEKA was compared to observed participation as shown in table 8.20. For reference, model predictions can only be compared to MEKA application data at district level obtained from the ministry of agriculture of Baden-Württemberg (MLR). The underlying population is not compatible to the model population as the model area is smaller than the districts (cf. Ch. 6) and the agent population consists only of full-time farms. Nevertheless, the data at least provide an upper bound for the MEKA participation in the study area.

Given this upper bound, predicted participation in the diversification measures (A2, A7) was definitely overestimated, and so was participation in the biodiversity-focused extensive grassland measure (B4). The other extensive grassland measures (B1, B2) were overestimated in 2007, while they were within bounds in 2003. The relative development in participation in B1 and B2 from 2003 to 2007 was well captured in any case.

The overestimation in MEKA participation can be attributed to the cost of bureaucracy and the loss of flexibility due to the obligation to maintain the measure for 5 years, which were both not well represented in the model and may especially apply to measure A2/A7. In the case of B4, the model, additionally, does not consider any natural conditions that might be necessary to fulfill the biodiversity criteria.

Table 8.20: *Predicted MEKA participation in the validation runs. (Mean and standard deviation of registered area and participating farms. Note: The combined total observed participation in the districts Reutlingen (RT) and Alb-Donau (UL) is given as a theoretical upper bound only, as the underlying populations are different.)*

		Predicted		Observed RT/UL	
		Area [ha]	Farms	Area [ha]	Farms
2007	MEKA III				
	A2	14,662 ±661	331 ±21	9,668	312
	B1	5,916 ±288	258 ± 7	5,498	372
	B2	9,050 ±204	220 ± 8	8,482	241
	B4	5,351 ±121	447 ± 9	3,049	249
2003	MEKA II				
	A7	10,117 ±568	262 ±21	8,678	405
	B1	16,177 ±439	544 ± 5	19,541	1,070
	B2	6,495 ±587	164 ± 4	8,952	334
	B4	5,390 ±150	279 ± 8	4,647	499
1999	MEKA I				
	3.1h	6,270 ±139	342 ±12	-	-
	3.1l	7,627 ±614	313 ±10	-	-
	3.1m	3,815 ±120	145 ± 4	-	-
	3.2.1	4,073 ±284	266 ± 8	-	-
	3.2.2	837 ±172	121 ± 8	-	-

Chapter 9

Climate change effects on short-term production decisions

For a first assessment of the potential effects of climate change on the agricultural sector of the study area, the effects of a sudden, anticipated shift in climate on the production decision of farms in the area were simulated.[1] To this end, two climate adaptation scenarios S0 and S3 were designed. While S0 assumed current climate conditions as in the calibration and validation experiments, S3 assumed a climate shift that consisted of three separate effects:

- A *shift in rotation options* allowing for the production of winter rapeseed after winter wheat, which is rarely possible in the study area under current climate conditions due to overlapping harvesting and sowing dates (see chapter 6.3).
- A *shift in suitable field work days*: The estimated distribution for climate zone 7 was used as an approximation for a potential climate change effect (tab. C.1). Zone 7 includes the regions at lower elevations adjacent to the study area.
- *Shifts in crop yield averages*: Calberto Sanchez [2015] simulated yields for the years 2000-2030 using the WETTREG [2010] projection for the Stötten weather station. Average crop yields over both runs described in section 6.3 and all thirty simulated years were used as a future climate scenario, while the baseline was based on the yields for the years 1981-2010 simulated with the observed time series. Table 9.1 shows the relative yield changes resulting from the simulations compared to the baseline for each crop and soil type.

Table 9.1: *Relative change of yields in climate change scenario.*

Soil	Silage maize	Summer barley	Winter barley	Winter rape	Winter wheat
0	-2.6	-6.9	-1.2	-3.5	8.7
1	4.8	-8.5	-1.7	5.3	7.5
2	-3.0	-0.8	-7.0	12.2	20.1
3	0.5	-5.4	-1.7	-1.6	11.7
4	-4.3	11.9	-14.6	6.6	24.6
5	-0.4	-3.2	-2.4	1.0	14.6
6	1.1	-19.7	-1.5	8.5	2.4
7	-2.1	7.2	-10.8	9.3	18.8
8	-4.5	12.1	-12.1	24.1	22.7

This climate adaptation scenario S3 should not be interpreted as an accurate local prediction of global warming in the study area. Apart from the inherent uncertainty in climate change projections, section 1.1 discussed the limitations of statistical downscaling. The

[1] For a condensed account of the most important results, please refer to Troost and Berger [2014].

changes in availability of field work days were completely based on assumptions due to the current lack of projections for the study area. The feasibility of future wheat-rapeseed rotations was mentioned as one potential consequence of shifted sowing dates by experienced farmers from the study area. With these caveats in mind, I believe that the magnitudes of change assumed are sufficiently realistic and provide useful scenarios to test the robustness of estimated climate change effects against different potential model parameterizations.

The impacts of these assumed climatic changes were analyzed in combination with three different price scenarios. In the baseline price scenario ('B'), the average of the prices observed between 2000 and 2009 (converted to 2009 real terms) was used for goods with variable prices (crops, fuel, animal products, fertilizer) and 2009 prices for goods with a rather trend-dominated development (most investment goods). For Scenario 'O', real price changes were calculated based on the projections provided in the OECD-FAO Agricultural Outlook 2012-2021 [OECD/FAO, 2012], which were then multiplied with the baseline average. As a third scenario ('X'), an extreme price scenario based on the 2007 agricultural price surge was created using crop prices from 2007/2008 and fertilizer prices from 2008/2009. (The crop price increase in 2007 prompted a price increase in agricultural inputs that followed with a one year gap and would also be expected in case prices stabilized at the higher levels). Price coefficients applied in scenarios 'O' and 'X' are shown in table 9.2.

The resulting six combinations of climate and price scenarios were simulated using the 2007 farming population generated for the validation experiments. In all scenarios, continuation of the EU, EEG and MEKA policies as valid in the season 2012/2013 was assumed.

Table 9.2: *Relative price change for selected items in the future price scenarios compared to baseline price scenario (B).*

	Scenario	
	O	X
Winter wheat	+ 8.2%	+72.6%
Summer barley	+ 8.2%	+80.6%
Winter barley	+ 8.2%	+77.3%
Winter rapeseed	+28.4%	+49.5%
Grain maize	+45.3%	+56.9%
Pig meat	±0%	-2.3%
Bull meat	±0%	+5.4%
Milk	±0%	+10.1%
Soy meal	+26.2%	+25.6%
Rapeseed meal	+26.2%	+25.6%
Mixed feeds	+27.3%	+27-29%
Fuel & Energy	+35%	+7-16%
Fertilizer	+30%	+7-11%

The following section first presents the parameter screening and experimental design chosen, before the baseline results are presented in the subsequent section. The analysis of price and climate scenario results is then approached in three steps: Section 9.3 presents changes in crop areas, animal numbers and farm type distribution from an aggregate perspective at regional level. Section 9.4 assesses the potential effects on investments in biogas plants and participation in the *Marktentlastungs- und Kulturlandschaftsausgleich* (MEKA) scheme of agro-environmental payments relevant for environmental policy analysis. Section 9.4 is dedicated to a disaggregate perspective analyzing how future changes might influence different types of farms, especially potential effects on farm income.

9.1 Parameter screening and experimental design

After the calibration experiments described in section 8.4, the potential parameter space still comprised 19 parameters. To reduce the computational burden for scenario simulations, an elementary effects screening was used to reduce the parameters to the ones most important in determining the effect of the future scenarios represented by the scenario with OECD price projections and all climate change effects. Differences were assessed with respect to the aggregate absolute difference in land use, livestock numbers, installed biogas capacity, MEKA participation and income indicators.

Those parameters that were among the first five most important ones in determining any of the differences were kept for further analysis. (These were *starting_population*, *clregion*, *proptohire*, *workforhirecoef*, *freshgrasslabor*, *wheat_normal*, *biertreber*, *birth_factor_past*, *potsuc_prob_male*, *overmanure*, *kwkyno*).[2]

These eleven parameters were used to construct a replicated Latin-hypercube sample following the unbiased permuted column sample (UPCS) scheme suggested by Morris et al. [2008] (cf. section 3.2.2). The UPCS was constructed from an OA(121, 12, 11, 1) orthogonal array using six subarrays. In total, this results in 66 runs for each scenario.

9.2 Baseline

The total crop areas, livestock and farm type numbers simulated for the baseline scenario (BS0) are shown in tables 9.3 and 9.4. A comparison with the model predictions for the 2007 validation year presented in section 8.5.1 revealed notable shifts in crop areas, especially for fallow, which drops to a simulated total area of 56 ha on average. The differences between simulation runs for 2007 and 2012 were primarily due to the abolition of the set-aside requirement for EU direct payments in 2008 captured in the model and to a lesser extent to the slightly different price expectations.

Table 9.3 shows comparatively high standard deviations for the winter wheat and winter barley areas over the 66 repetitions. As the experimental design was structured as a UPCS, first-order sensitivity indices could be calculated following the suggestions of Morris et al. [2008] to get an idea of the determinants of this rather large variation. Results suggested that 45% of the variation in wheat area and 40% of the variation in winter barley area could be explained by the variation in the *proptohire* coefficient. Figure 9.1 relates the predicted areas to the parameter setting of *proptohire* and *clregion* and shows that an increased availability of foreign field work services as represented by *proptohire* shifted the balance from winter barley to winter wheat. While the *clregion* parameter showed little influence at levels of *proptohire* > 1, at lower levels the choice of *clregion* = 5 increased wheat and decreased winter barley areas compared to *clregion* = 4. The variance in livestock numbers was mainly due to the chosen starting population.

9.3 Effects on aggregate land use and the farm type distribution

The analysis in the present section starts by first presenting the potential differences in agricultural production arising only from the different price scenarios under current climate. Second, the changes occurring when all three aspects of climate change were applied in combination are described for all three price scenarios. Third, the combined effect of climate and

[2]Parameters were explained in section 8.4.1

Table 9.3: *Land use and livestock in the baseline scenario BS0 (Mean and standard deviation over 66 simulation runs).*

	Total area [ha]	
Crop	Mean	Sd
Winter wheat	7,310	± 830
Summer barley	4,109	± 180
Winter barley	2,473	± 628
Winter rapeseed	3,788	± 156
Silage maize	2,864	± 293
Field grass	1,300	± 339
Fallow	56	± 130
Meadows/Mown pasture	14,502	± 157
Pasture	644	± 124
	Total [heads]	
Animal	Mean	Sd
Dairy cows	10,564	± 198
Heifers	9,007	± 172
Fattening bulls	4,985	± 238
Calves	2,542	± 44
Mother cows	14	± 5
Sows	11,138	±1,625
Piglets (8 - 28 kg)	17,459	±1,664
Fattening pigs	31,527	±1,656

Table 9.4: *Farm types in the baseline scenario BS0 (Mean and standard deviation over 66 simulation runs).*

PTOF		No. farms	
Code	Description	Mean	Sd
13	Specialist cereals/oilseeds	78.6	±5.7
14	Specialist mixed field crops	5.9	±2.0
41	Dairy	183.5	±5.7
42	Cattle fattening/raising	17.3	±2.8
43	Cattle mixed	25.0	±3.1
44	Grazing livestock mixed	2.6	±1.3
50	Granivore	40.1	±4.7
60	Mixed crop	7.5	±1.6
71	Mixed ruminant	29.2	±2.4
72	Mixed granivore	32.7	±4.8
81	Mixed crop/ruminant	65.3	±8.0
82	Mixed crop/livestock	38.6	±3.8

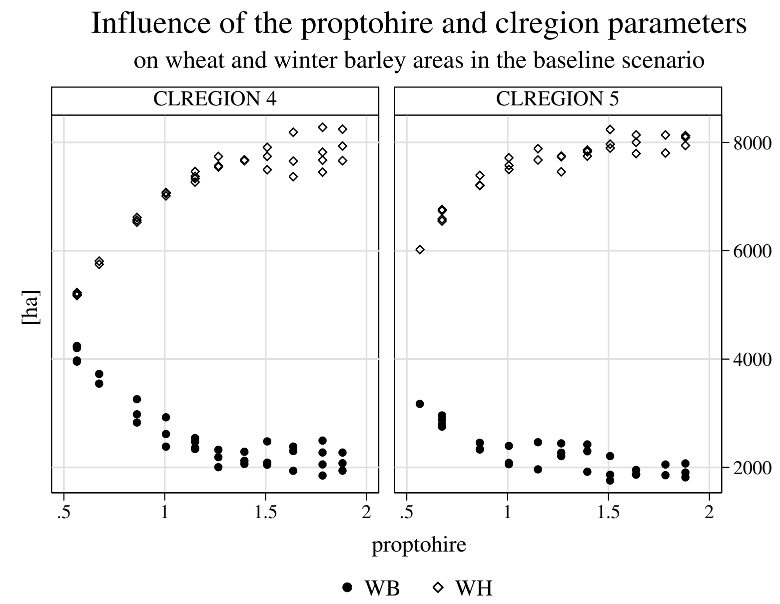

Figure 9.1: *Influence of the proptohire and clregion parameters on winter wheat (WH) and winter barley (WB) areas in the baseline scenario (BS0).*

price scenarios compared to the baseline is calculated. Finally, the contribution of each individual aspect of climate change on the observed future production patterns is analyzed.

9.3.1 Price effects with current climate

Figure 9.2 shows the effect of the two alternative price scenarios ‘O’ and ‘X’ compared to the baseline price scenario ‘B’. Each subgraph in this figure contains one box plot per crop, which illustrates how the simulated change in crop area was distributed over the parameter combinations of the UPCS.

The upper pane in figure 9.2 shows the changes in total crop areas under the two price scenarios compared to the baseline with current long-term average prices (price scenario ‘B’), all simulated for current climate conditions (S0). For scenario ‘O’, the graph shows an increase in winter rape area (+330 to +580 ha), while the winter wheat area (-550 to -80 ha, median -305 ha) and to a minor extent the silage maize area (-260 to +10 ha, median -160 ha) declined. The direction of the effect on other crops was unclear, but its extent was also rather small.

Farmer response in the extreme price scenario ‘X’ was stronger. Most prominently there was a strong decrease of field grass production (-1,300 to -400 ha, median -860 ha) accompanied by an increase in summer barley area (+50 to +920 ha, median +480 ha). The effect on wheat areas could potentially be quite strong (between +1,310 and -70 ha), though under most parameter combinations assessed it was only moderately positive (median: +240 ha). The winter rapeseed area slightly declined (-350 to + 100 ha, median -140 ha) , while the effect on winter barley area was rather ambiguous (-420 to + 650 ha, median: +170 ha).

Calculating first-order sensitivity indices revealed that the *proptohire* parameter again stood out as a major determinant for the variation in the price scenario effects. Its first order effect explained about 26%, resp. 21% of the variance in the effect on winter, resp. summer barley and 17% of the variance in the effect on field grass area under Scenario ‘X’. It also accounted for 37% of the variance of the effect on wheat areas and 24% of the effect on summer barley areas under Scenario ‘O’. In Scenario ‘X’ the *wheat_normal* parameter, scaling the normal wheat yield, was the second major determinant: Its first-order effect was responsible for 17% of the effect on wheat area.

Figure 9.3 shows how the positive effect on barley areas diminished with increasing *proptohire* and *wheat_normal* in Scenario ‘X’. The negative effect on field grass area was less strong for higher *proptohire*. Whether the wheat area increased correspondingly was however dependent on higher values of *wheat_normal*. In Scenario ‘O’, an increasing *proptohire* amplified the positive effect on winter rapeseed and the negative effect on winter barley.

With respect to livestock, both price scenarios showed a decrease in fattening bull (10-20%) and heifer numbers, and a potential though rather minor decrease in heifer, piglet and pig meat production, which could be explained by the relative decrease of meat and milk prices compared to feed and crop prices. This was also reflected in a shift in the farm type distribution from farm types dominated by meat production towards crop and dairy farms. In both cases the effect was stronger for the extreme price (‘X’) scenario than for the OECD-FAO-based scenario (‘O’).

9.3.2 Effects of climate change under different price scenarios

Figure 9.4 shows the pure effect of climate change for each price scenario by comparing each price scenario with climate change (BS3, OS3, XS3) to the corresponding scenario with current climate (BS0, OS0, resp. XS0). Again, each subgraph in this figure contains one box plot per crop, which illustrates how the simulated change in crop area was distributed over the UPCS.

A glance at figure 9.4 shows very similar effects of climate change under each of the three price scenarios. There was a strong increase of winter wheat areas (median: +2,500 ha for ‘B’; +2,750 ha for ‘O’ and +2,770 ha for ‘X’), leading to strong decreases of both winter (median B: -1,390; O: -1,360; X: -2,100 ha) and summer barley areas (median B: -2,130; O: -2,170; X: -1,940 ha). This was accompanied by slight increases in winter rapeseed (median B: +680; O: +465; X: +880 ha) and silage maize areas (median B: +370; O: +440; X: +430 ha). Field grass and fallow areas experienced practically no effect. The direction and order of magnitude of all effects were robust under all parameter settings.

With respect to animals, quite substantial increases in fattening bulls (median B: +740; O: +650; X: +690) and slight increases in heifer numbers (median B: +80; O: +190; X: +170) could be observed compared to current climate. Shifts in the farm type distribution were very minor.

9.3.3 Combined price and climate effects

Figure 9.5 shows the combined effect of climate change and price scenario (BS3, OS3, XS3) compared to the baseline scenario (‘BS0’). (Of course, under price scenario ‘B’, there was no price effect, and the difference was equal to the effect of climate change.) Again, the box plots show the distribution of effects over the 66 repetitions.

When looking at the combined effects of price and climate development on crop areas compared to the baseline (BS0), the climate effects dominated. The increase in wheat area was somewhat dampened by the reduction cause by price scenario ‘O’, and so was the decrease of summer barley area in scenario ‘X’ as consequence of the assumed high prices. The decrease of field grass production in scenario ‘X’ prevailed also under climate change.

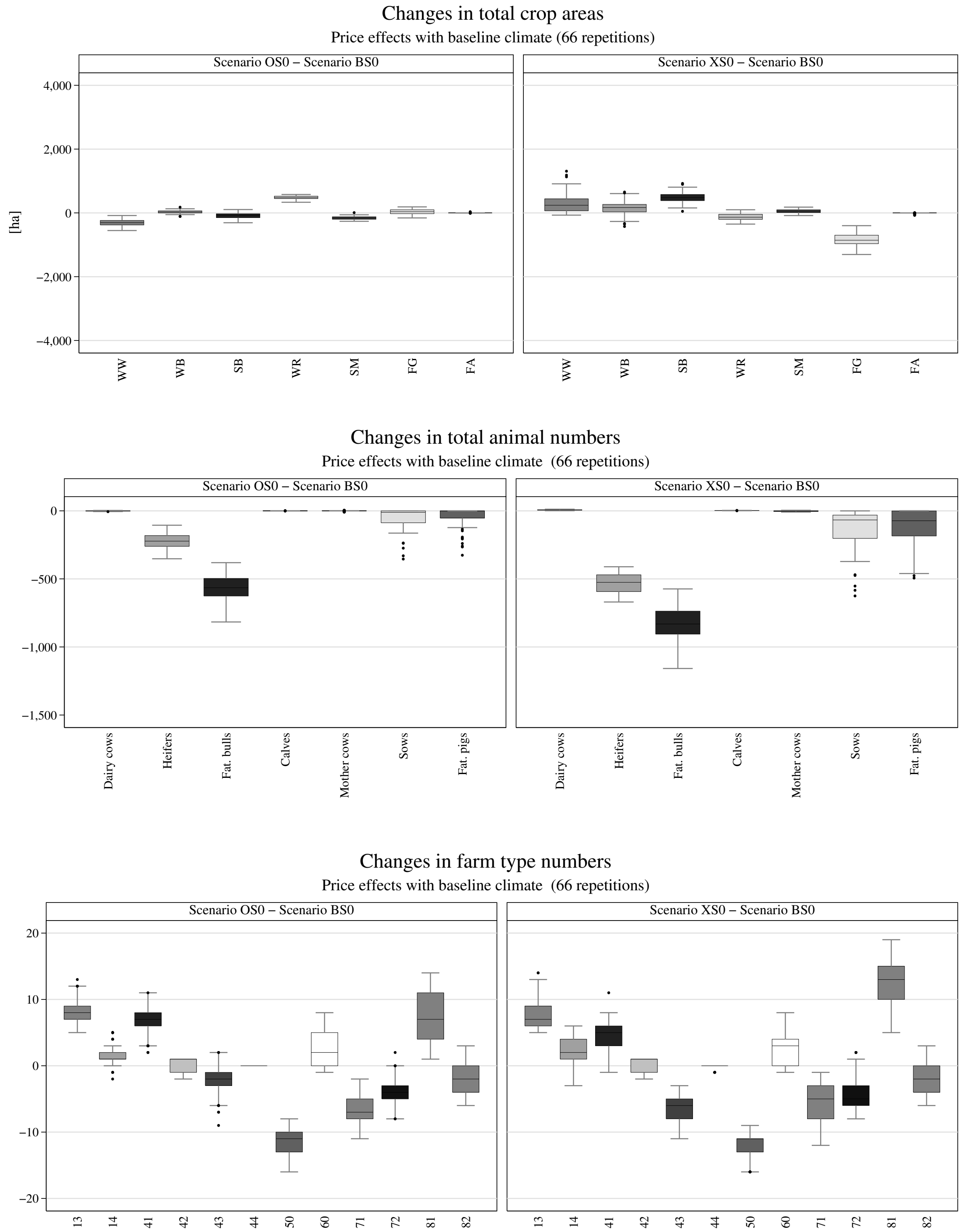

Figure 9.2: *Price effects (Price scenarios compared to baseline) on total crop areas, livestock numbers and farm type distribution under baseline climate. (Crops: WW - winter wheat; WB - winter barley; SB - summer barley; WR - Winter rape; SM - silage maize; FG - field grass; FA - fallow; for farm types cf. tab. 9.4)*

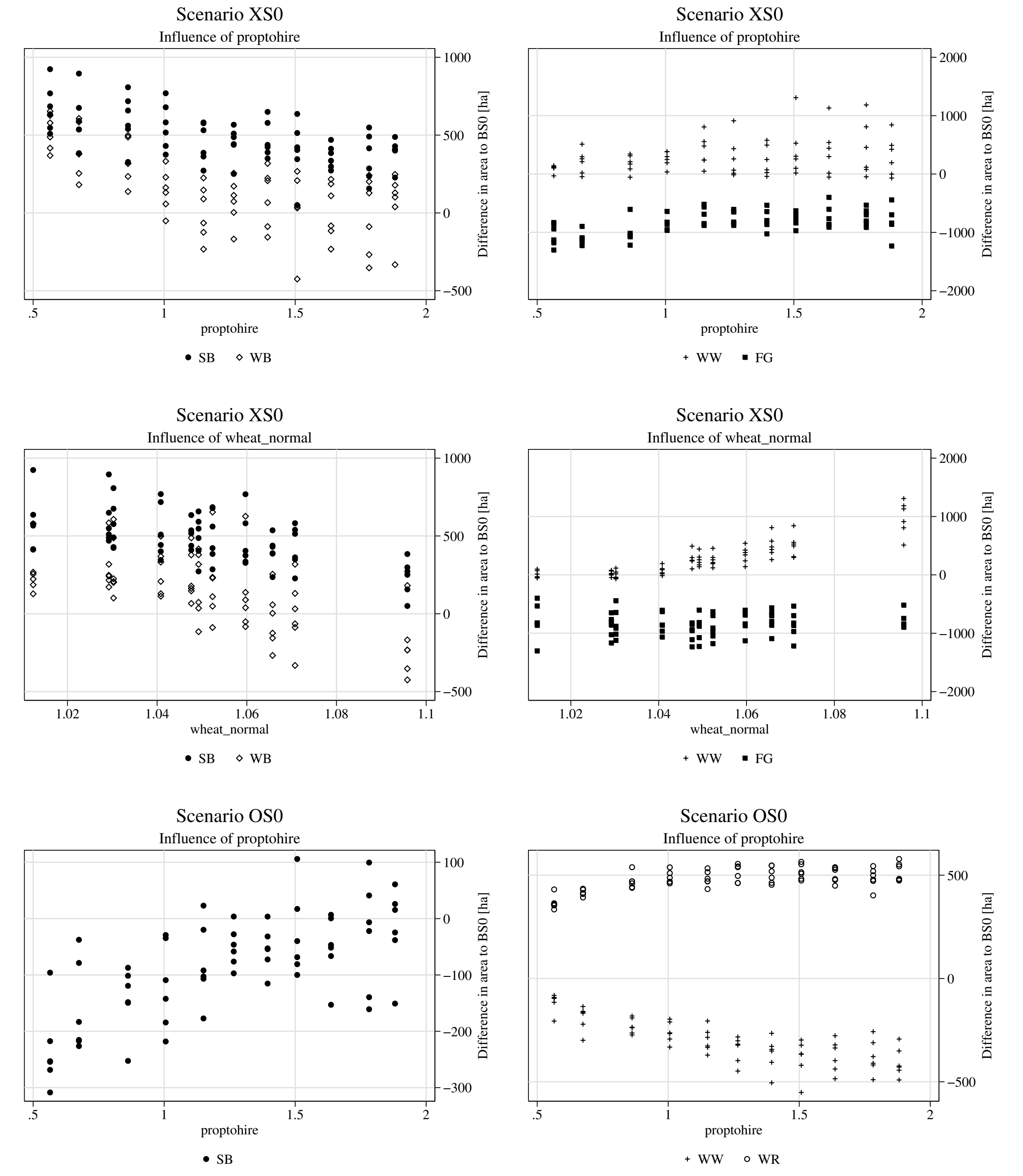

Figure 9.3: *Influence of the* proptohire *and* wheat_normal *parameters on the difference in area between baseline (BS0) and price scenarios (OS0, XS0) for selected crops (for current climate).*

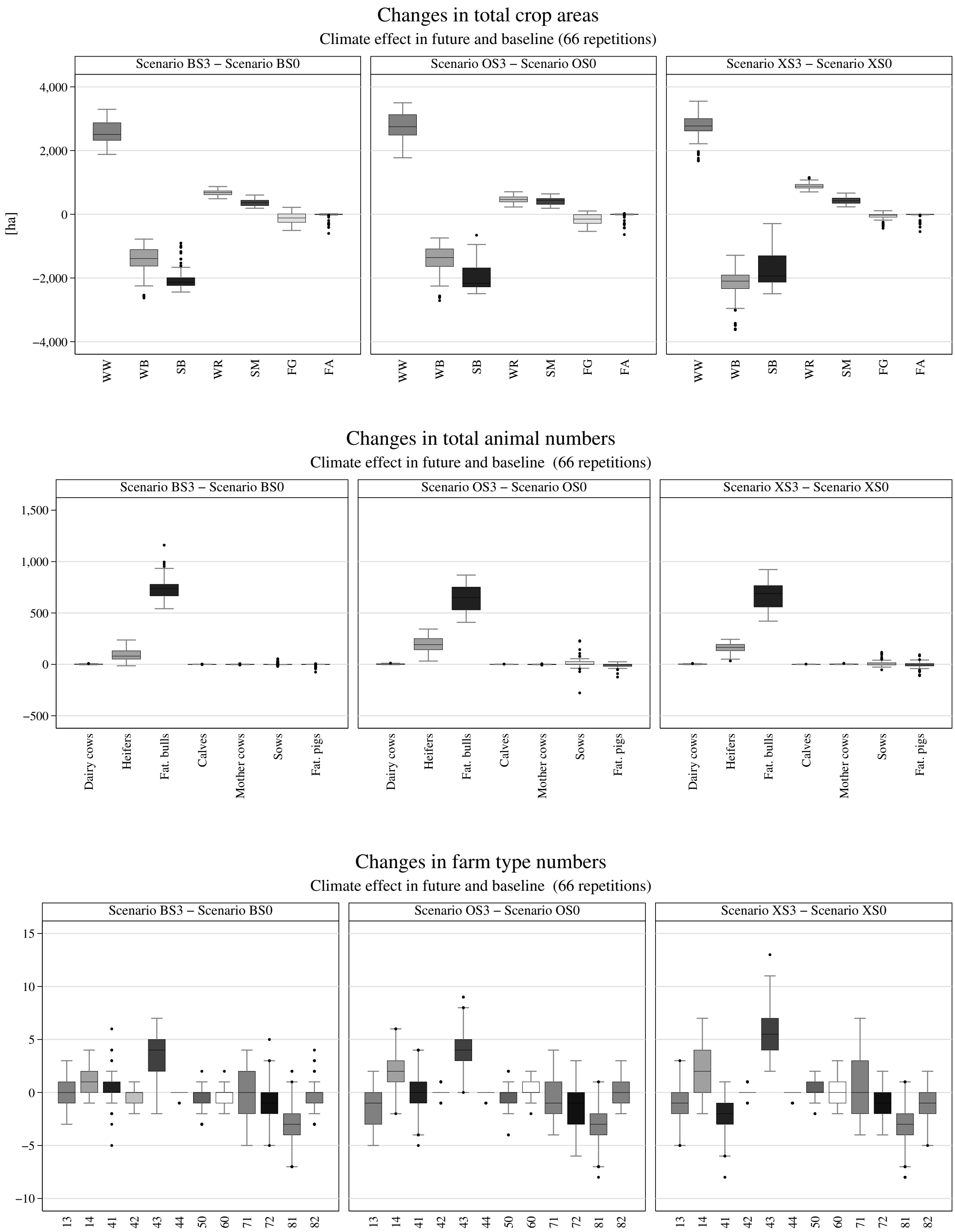

Figure 9.4: *Effect of climate change on total crop areas, animal numbers and farm type distribution under the three different price scenarios. (Crops: WW - winter wheat; WB - winter barley; SB - summer barley; WR - Winter rape; SM - silage maize; FG - field grass; FA - fallow; for farm types cf. e.g. tab. 9.4)*

With respect to livestock, the substantial climate-induced increase in fattening bulls completely compensated the price-induced decrease in scenario 'O' and partly did so in scenario 'X'. As there was little effect of climate change on piglet and pig production, the price effects prevailed and, consequently, there was a minor shift from granivore producers towards crop producers.

9.3.4 Ceteris paribus and omission effects of the individual climatic effects

To disentangle the contribution of climate-induced changes in crop yields, changes in suitable days for field work, and the additional rotation option, three additional climate adaptation scenarios under price scenario B were run, in each of which only one of the impact pathways was considered (BS1_yld, BS1_fwd, BS1_rw). These scenarios were used to calculate *ceteris paribus* (c.p.) effects of the impact pathways by comparison with the baseline scenario (e. g, c.p. yields (B) = BS1_yld - BS0). Further, six additional climate adaptation scenarios were run under price scenarios X and O, in each of which only one of the impact pathways was not considered (OS2_yld_fwd, OS2_yld_rw, OS2_rw_fwd, XS2_yld_fwd, XS2_yld_rw, XS2_rw_fwd). Here, I introduce the term *ceteros assequens*[3] (c.a.) to denote the effect calculated by subtracting the effect observed in these S2 scenarios from the combined effect of all three individual aspects of climate change (e.g., c.a. yields (O) = OS3 - OS2_rw_fwd). A complete global analysis of the sensitivity of adaptation to the individual impact pathways would require assessing the distribution of the effect of activating an individual pathway over all possible combinations of presence and absence of all other pathways. To limit the number of scenarios and for ease of exposition, the c.p. and c.a. effects were selected to represent the extremes of these combinations.

Figure 9.6 shows the simulated c.p. and c.a. effects on crop areas. Again, each box plot represents the distribution of the change in crop area over the 66 repetitions of the experimental design. The figure suggests that the strong increase in wheat areas observed in the climate change scenarios discussed above resulted from the mutually reinforcing effects of all three aspects, the change in field work days causing the strongest shift. The decrease in winter barley areas was mainly due to the yield effect, while the summer barley area decreased because growing rapeseed after winter wheat became possible: Under current climate conditions, farmers in the study area were virtually forced to grow summer barley at some point in the cropping sequence between winter wheat and rapeseed. Under the tested future climate conditions, a more profitable and wheat-intensive crop rotation of wheat-rapeseed-wheat-silage maize was possible. Apart from the wheat expansion, the change in the number of available days for field work increased silage maize areas. Overall, the magnitude of the individual effects of all three aspects of climate change on crop areas seems quite comparable despite showing partly different patterns with respect to different crops.

This looks quite different with respect to livestock, where the effect of an increased availability of field working days clearly bore the (nearly exclusive) responsibility for the positive effect on fattening bull and heifer numbers observed in the climate change scenarios (fig. 9.7). The effect of the available days for field work on bull fattening could be explained through the increase in silage maize production mentioned before and unsurprisingly this also dominated the effect on the farm type distribution (fig. 9.8).

[3] *lat.* catching up with the others

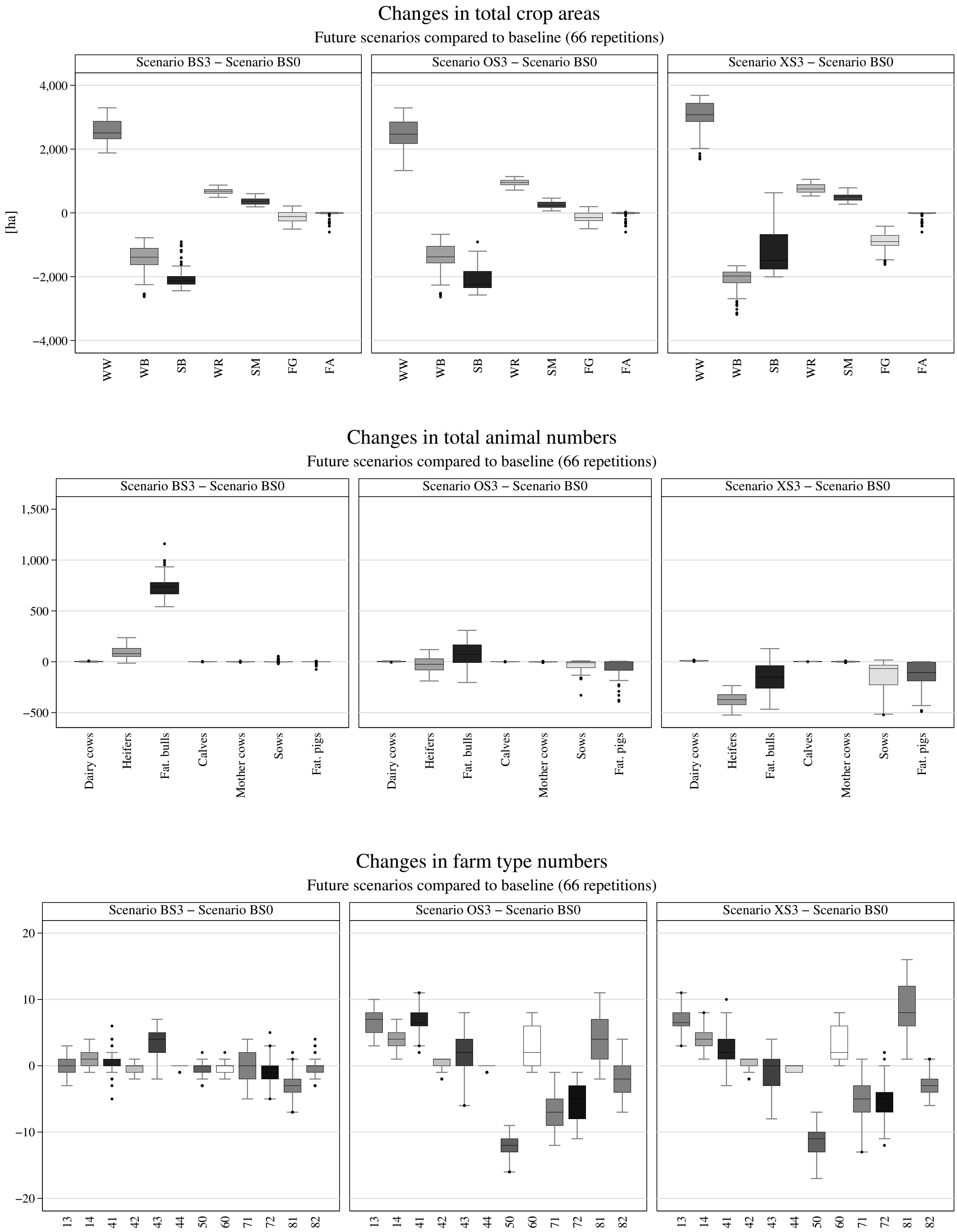

Figure 9.5: *Changes in total crop areas, animal numbers and farm type distribution as predicted in future scenarios compared to baseline. (Crops: WW - winter wheat; WB - winter barley; SB - summer barley; WR - Winter rape; SM - silage maize; FG - field grass; FA - fallow; for farm types cf. e.g. tab. 9.4)*

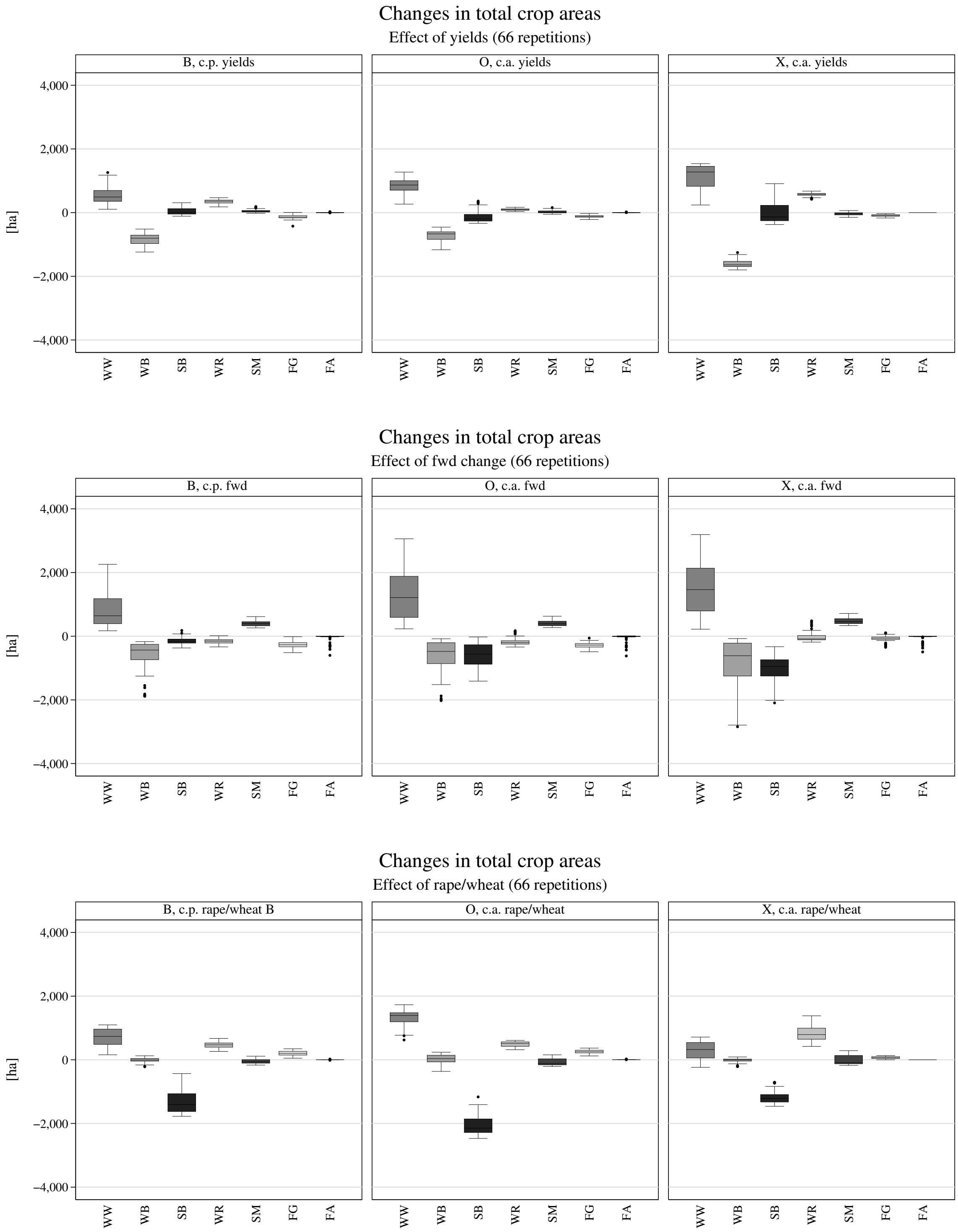

Figure 9.6: *Ceteris paribus and ceteros assequens effects of the individual aspects of climate change on total crop areas. Top: changes in crop yields. Center: changes in available days for field work. Bottom: arpeseed after wheat possible. (Crops: WW - winter wheat; WB - winter barley; SB - summer barley; WR - Winter rape; SM - silage maize; FG - field grass; FA - fallow)*

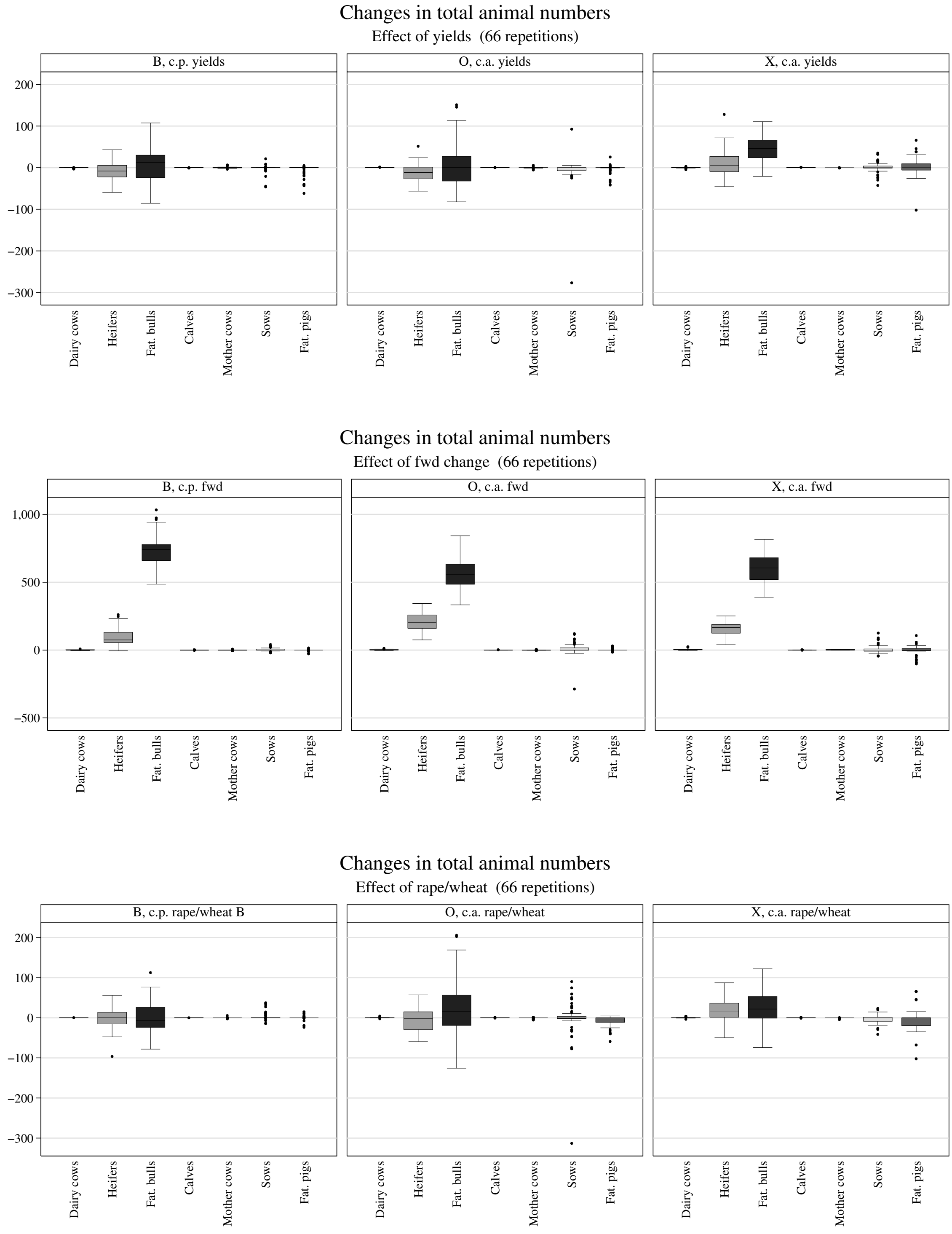

Figure 9.7: *Ceteris paribus and ceteros assequens effects of the individual aspects of climate change on total animal numbers. Top: changes in crop yields. Center: changes in available days for field work. Bottom: rapeseed after wheat possible.*

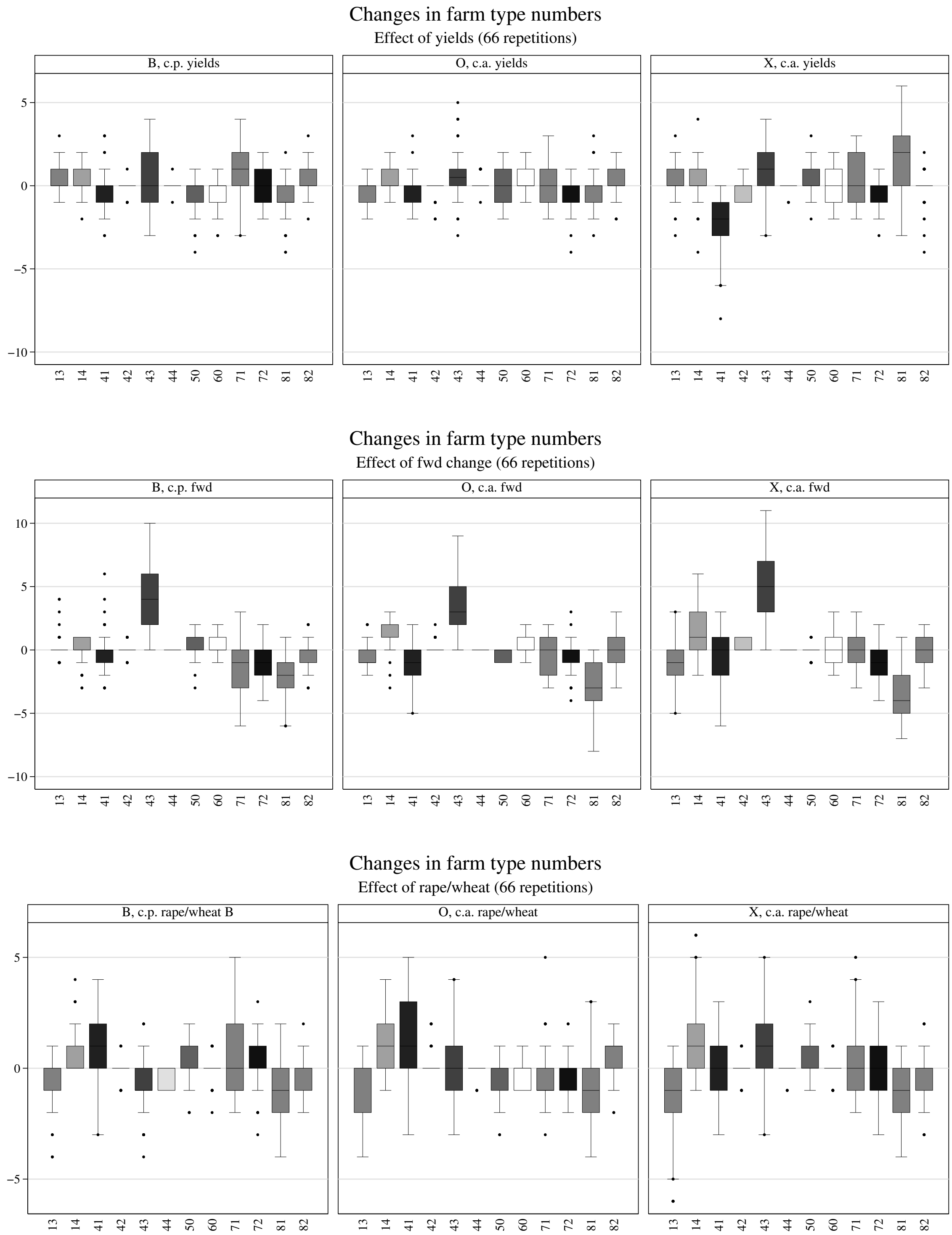

Figure 9.8: *Ceteris paribus and ceteros assequens effects of the individual aspects of climate change on the farm type distribution. Top: changes in crop yields. Center: changes in available days for field work. Bottom: rapeseed after wheat possible. (For farm types cf. e.g. tab. 9.4)*

9.4 Effects on participation in MEKA and investment in biogas plants

How did the adaptations triggered by climate and price changes affect participation in the agri-environmental policy scheme MEKA and investments in biogas plants supported by the EEG?

The model includes four measures of the MEKA program. Measure A2 rewards 20 Euro per hectare for the diversification of crop rotations requiring the production of at least four crops, each with a minimum share of 15% of their total arable area and restricting maize production to a maximum of 40% of the area. The other three measures support extensive grassland use. Commitment to any of these measures requires the farmer to refrain from grassland conversion and the unspecific use of chemical plant protection on grasslands. Under measure B1, farms that restrict livestock density to under 2 livestock units (LSU) per hectare, abstain from grassland conversion and mow 5% of their committed grassland area after the 15th of June receive 50 Euro per hectare of qualifying grassland. Measure B2 imposes maximum livestock densities of 1.4 LSU per hectare of agricultural area and 1.4 grazing livestock units (gLSU) per hectare of fodder area, and a minimum density of 0.3 gLSU per hectare of fodder area, awarding 100 Euro per ha of grassland committed. The third measure, B4, is result-oriented and awards 60 Euro per hectare if at least four out of a catalog of twenty-eight characteristic plant species can be observed on the committed extensive grassland area.

The predicted baseline participation in these MEKA measures is shown in table 9.5, while the effect of the price changes and the combined effect of price and climate changes is shown in figure 9.9. Whereas the grassland related measures B1, B2, and B4 were hardly affected by neither price nor climate change, there was a strong reduction of about a third ('B', 'O'), respectively even two thirds ('X') in both farms and area participating in the A2 measure. This reduction was entirely an effect of climatic changes as the ceteris paribus effect of prices is about zero, respectively even positive in scenario 'X'. As shown in figure 9.10, the different aspects of climate change were mutually reinforcing each other to produce the combined effect.

As mentioned in section 8.5.3, participation in MEKA A2 was generally overestimated in the model, so also the absolute decline will be overestimated. Still, it seems entirely reasonable that the changes in crop rotation leading to increased concentration on wheat, rapeseed and silage maize and decreases in barley areas may lead to more three-part crop rotations, at least on some parts of a farmer's area. This would then also discourage participation in a measure such as A2.

Table 9.5: *Participation in the MEKA agri-environmental scheme as predicted in the baseline.*

	Area [ha]	Participants
A2	15,583 $\pm$716	356 $\pm$ 22
B1	6,319 $\pm$357	272 $\pm$ 9
B2	8,626 $\pm$282	205 $\pm$ 10
B4	5,458 $\pm$153	453 $\pm$ 11

The EEG sets incentives for investments in biogas plants by obliging electricity companies to purchase renewable energy at a guaranteed price. The simulated biogas capacity installed by farm agents amounted to $2,946 \pm 447$ kW in the baseline scenario, with the variation being explained mainly by the choice of the starting population (see fig. 9.11). For the great

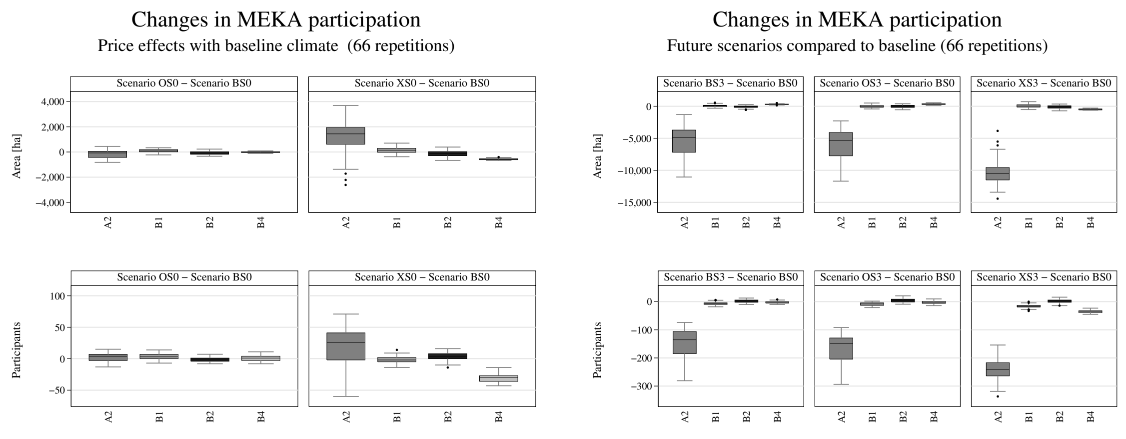

Figure 9.9: *Changes in MEKA participation as predicted in future scenarios (left) and price scenarios with baseline climate (right) compared to baseline.*

majority of repetitions, no effect of the assumed climatic changes on biogas capacity could be observed (fig. 9.12). The price changes assumed for price scenario 'X' led to slightly lower biogas capacities, while the effect was unclear in scenario 'O'.

9.5 Income effects and disaggregation

Figure 9.13 shows the effect of price and climatic changes on farm incomes. For each of the 66 repetitions, the graphs include one box plot. Each box plot depicts the distribution of changes in per-hectare incomes over the full agent population.

Comparing the price effects under current climate (upper pane), there was quite a large range of both income losses and gains for the price scenarios. The price changes of scenario O entailed only minor income effects for the majority of farms, though also substantial losses (up to -400 Euro/ha) and gains (up to 500 Euro/ha) for individual agents could be observed. The extreme price scenario X allowed income gains of 150-500 Euro/ha for more than two thirds of the farms, though it also led to substantial losses for some farms.

The climate change effect on per-ha income was entirely positive for the changes simulated (lower pane): highest gains could be observed in scenario 'X' (median 98-126 €) followed by 'O' (median 72-92 €) and 'B' (median 64-82 €). The lower pane shows the disentangled impacts of the three individual climate effects: The contribution of yield changes and changes in field work days seems about equal, the former having a higher median effect, while the later showed a higher variance over the agent population and stronger effects in the extremes. The contribution of the additional rotation option was rather small (fig. 9.14).

Overall, the distribution of effects did not differ much over the 66 tested parameter combinations confirming the robustness of the estimates. Rather than assessing the contribution of individual parameters on the distribution of income effects, it seemed therefore more interesting to analyze, which agent characteristics determined whether the effect was weak or strong for a specific agent. To do so, standardized regression coefficients (SRCs) were calculated for each scenario and repetition showing the influence of farm size, grassland share, number of cattle and pig stable places, share of specific soil types in arable land and the installed biogas capacity (tab. 9.6).

For the price effects, the picture was quite clear. In scenario 'X', pig fattening places, sow

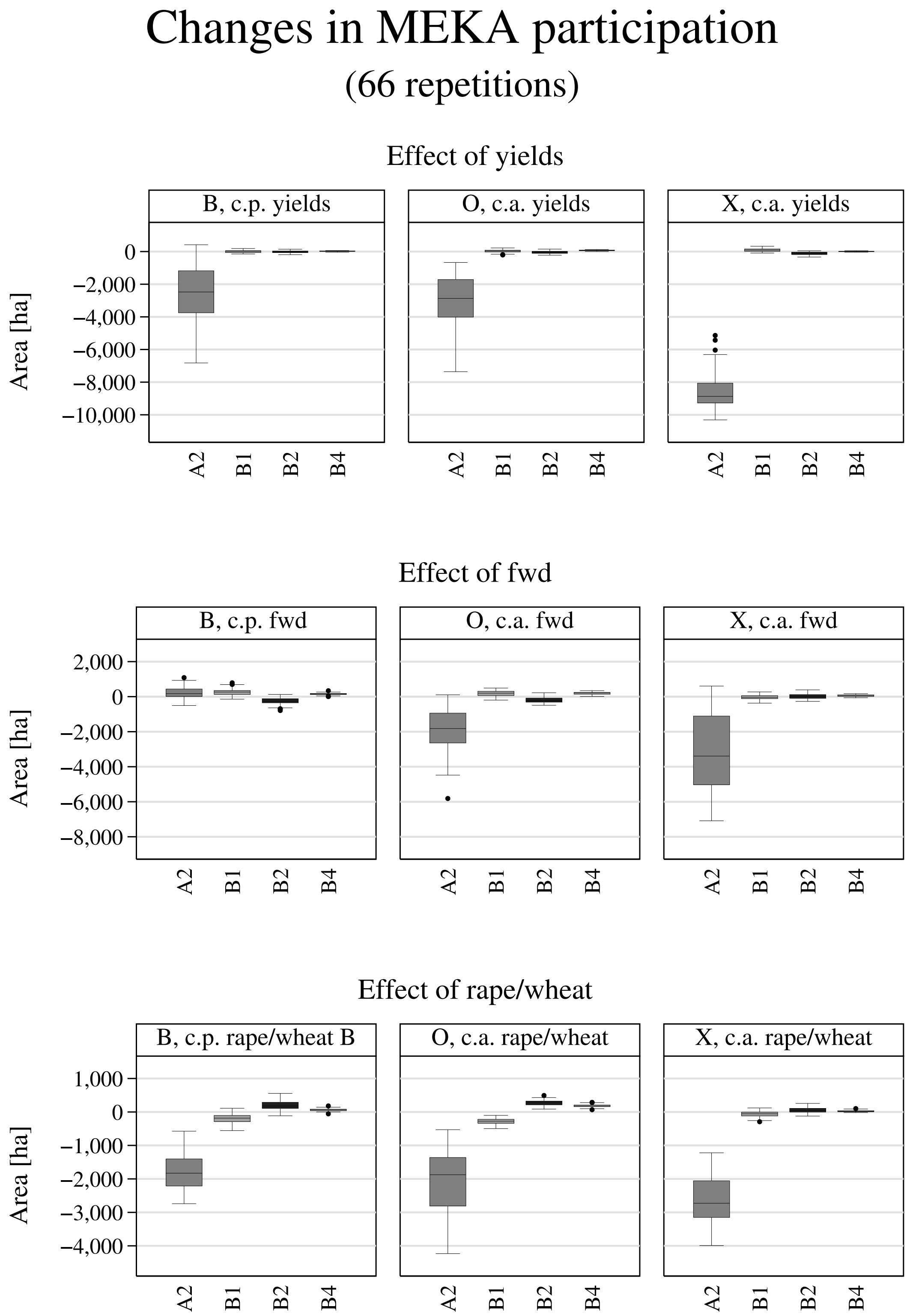

Figure 9.10: *Ceteris paribus (c.p.) and ceteros assequens (c.a.) effects of the three different aspects of climate change on predicted MEKA participation.*

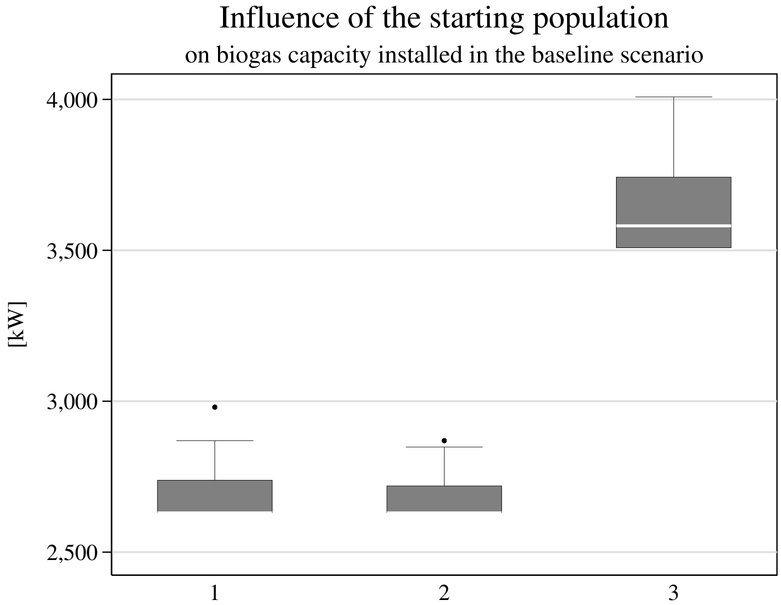

Figure 9.11: *Influence of the choice of starting population on the installed biogas capacity in the baseline scenario.*

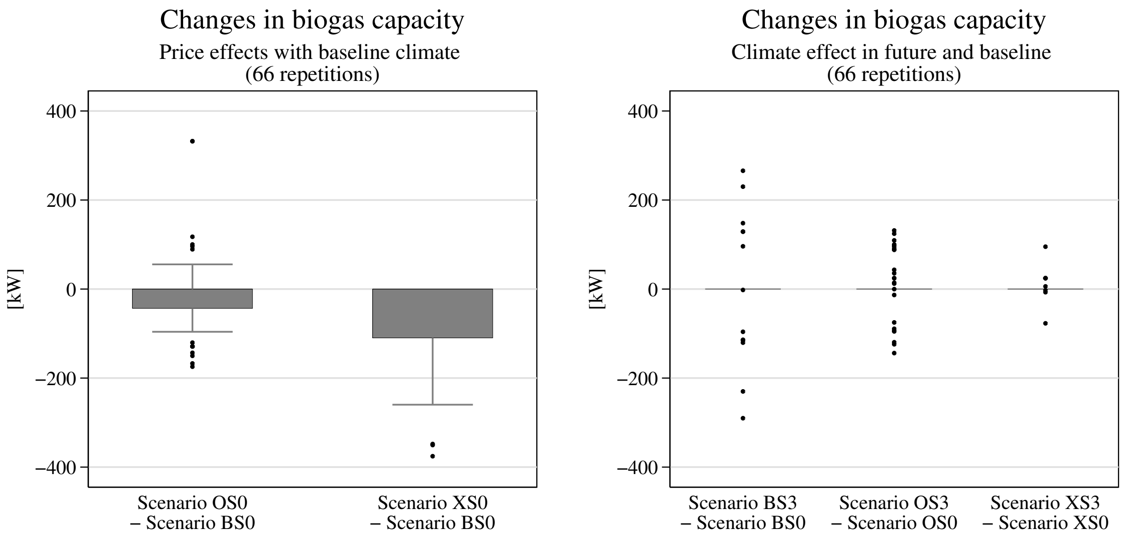

Figure 9.12: *Changes in installed biogas capacity in price scenario without climate change compared to baseline (left) and climate change effect on installed biogas capacity (right).*

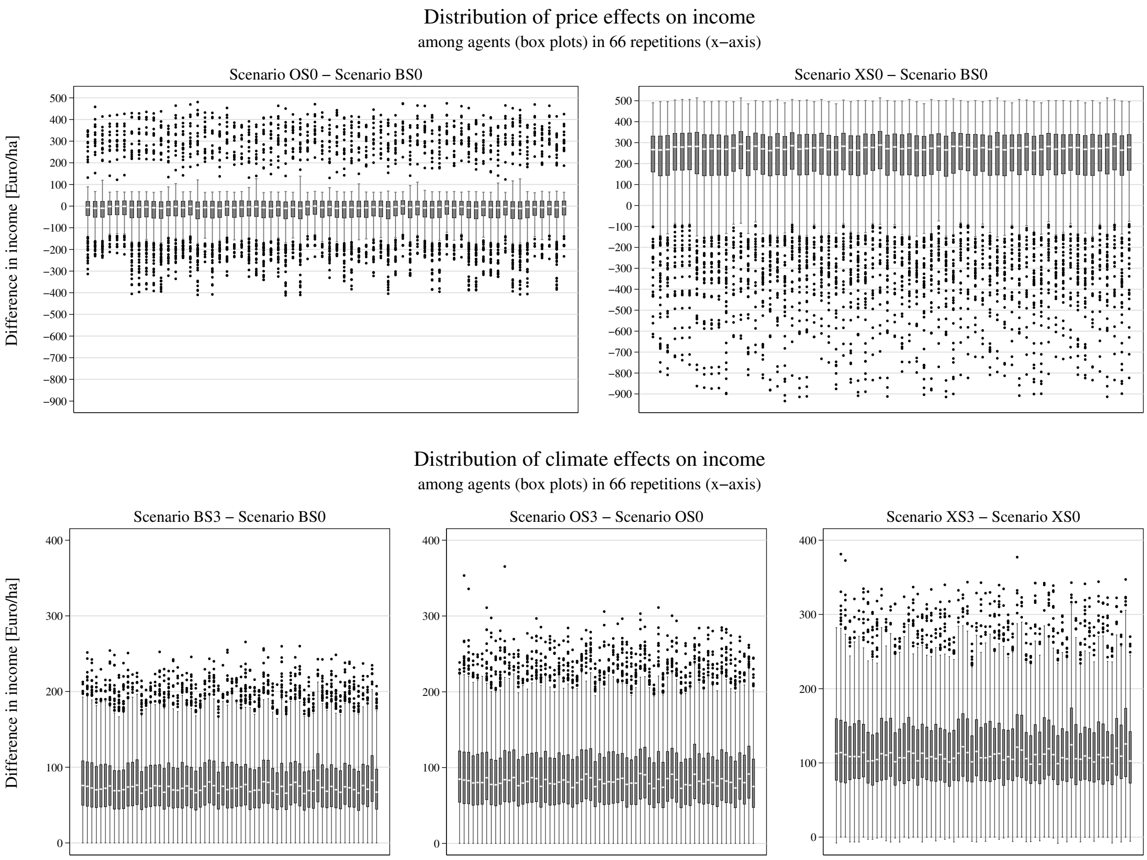

Figure 9.13: *Distribution of the changes in farm income per ha due to price scenarios (upper pane) and climate change (lower pane) over the full agent population (box plot) in 66 repetitions (left-to-right).*

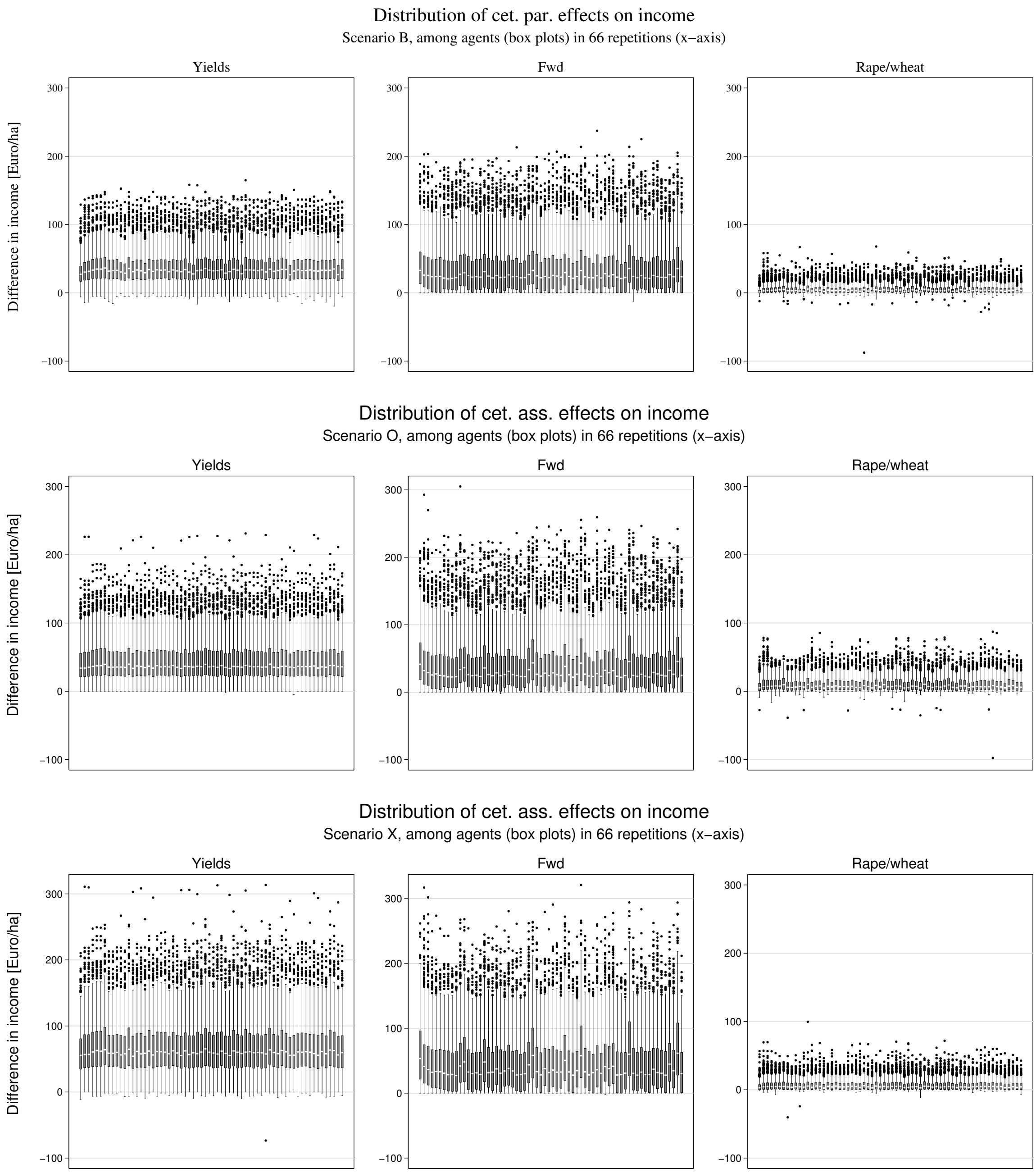

Figure 9.14: *Distribution of the ceteris paribus and ceteros assequens effects of yield change, change in available days for field work and additional rotation options on farm income per ha over the full agent population (box plot) in 66 repetitions (left-to-right).*

places and dairy cow places were associated with losses, while agents with biogas plants benefitted significantly. In scenario 'O', pig fattening, number of sow places and a high share of grassland area were associated with low and negative effects on agent income. The distribution of climate effects, however, could not be captured similarly well by the SRC, pointing to the limits of a linear, additive regression model, such as the one underlying the SRCs for the analysis. The most important indicators for a stronger positive effect on income from climate changes were a higher number of dairy cow stalls per ha, a lower share of grassland, a higher share of soil type 8 (cumulic anthrosols) and a larger farm size. A rather low R^2 of the underlying regression, however, underscored the complexity of the effects.

Table 9.6: *Standardized regression coefficients showing the linear association of simulated price and climate effects on per-ha income to farm attributes. Minima and maxima over 66 repetitions.*

	SRC for price effect				SRC for climate effects					
	Scenario O		Scenario X		Scenario B		Scenario O		Scenario X	
Farm attribute	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Farm size	-0.08	0.00	-0.07	0.00	0.09	0.38	0.11	0.40	0.12	0.42
Sh. of grassland	-0.64	-0.44	-0.18	-0.13	-0.54	-0.40	-0.54	-0.40	-0.61	-0.46
Sh. of soil 1	-0.02	0.00	-0.04	-0.01	0.01	0.05	0.02	0.06	0.00	0.04
Sh. of soil 2	0.01	0.02	-0.03	0.03	0.04	0.11	0.05	0.11	0.05	0.12
Sh. of soil 3	-0.01	0.01	-0.02	0.02	0.01	0.09	0.03	0.10	-0.01	0.09
Sh. of soil 4	0.01	0.02	-0.01	0.02	0.05	0.16	0.06	0.15	0.05	0.18
Sh. of soil 5	-0.01	0.01	-0.03	0.01	0.03	0.10	0.02	0.09	0.04	0.11
Sh. of soil 6	-0.03	-0.02	-0.03	0.01	-0.16	-0.06	-0.14	-0.04	-0.23	-0.13
Sh. of soil 7	0.00	0.02	-0.02	0.01	0.02	0.10	0.02	0.10	0.02	0.10
Sh. of soil 8	0.02	0.05	-0.03	0.00	0.25	0.44	0.30	0.45	0.26	0.45
Cow places	0.07	0.17	-0.43	-0.29	0.60	0.73	0.61	0.72	0.55	0.70
Cattle places	-0.05	-0.01	-0.18	-0.04	0.05	0.24	0.05	0.20	0.05	0.18
Cow places	-0.62	-0.49	-0.50	-0.35	-0.01	0.18	0.00	0.15	0.00	0.13
Piglet places	-0.05	-0.02	-0.05	-0.02	-0.01	0.03	-0.04	0.03	-0.03	0.01
Fat. pig places	-0.83	-0.74	-0.69	-0.40	0.00	0.11	0.02	0.12	0.01	0.13
Biogas capacity	-0.13	-0.08	0.49	0.76	-0.07	0.07	-0.11	0.06	-0.08	0.09
Household labor	-0.04	0.01	-0.03	0.05	-0.16	0.05	-0.16	0.04	-0.15	0.04
Successor labor	-0.01	0.03	-0.01	0.11	-0.13	0.10	-0.14	0.08	-0.08	0.10
R^2	0.96	0.98	0.93	0.97	0.62	0.71	0.62	0.72	0.67	0.76

9.6 Effects on short-term regional supply response to prices

In the simulations presented so far, three specific price scenarios were assessed. Climate-induced adaptation of production – happening all around the world – will, however, trigger price changes on the market that again feed back into the production decisions of farmers. It is hence necessary to characterize the supply behavior of farmers in the study area not only for singular price scenarios, but over the whole range of potential price developments and if possible express it as a function of prices. If done for many regions, a new market equilibrium could then be derived from these functions.

As a first step towards this end, this section presents the results of a second simulation experiment describing the dynamics of land use, livestock production and policy participation with respect to prices, and how these might be affected by climate change. The time-series of prices observed between 2000-2009 served as the basis to determine price ranges for crops, animal products and important inputs that were then expressed as price level coefficients relative to the 2000-2009 price average. To take account of the fact that crop prices are correlated and usually move in similar directions, price coefficients for grain maize, malting and fodder barley, rapeseed and animal feed (ready-mixes, soy and rapeseed meal, etc.) were expressed relative to the development of the wheat price, which consequently served as general crop price level (tab. 9.7). As an example, the price coefficient of summer barley was calculated as

$$dcwh\_sb = \frac{p_{\mathrm{sb},t}}{\bar{p}_{\mathrm{sb}} * pc\_wh_t}$$

and the summer barley price in the simulations was consequently calculated as $p_{\mathrm{sb}} = \bar{p}_{\mathrm{sb}} * pc\_wh * dcwh\_sb$. It is important to notice that $dcwh\_sb$ is not the summer barley to wheat price ratio, but its relative change compared to the 2000-2009 average. The new price ratio can be determined by multiplying $dcwh\_sb$ with the original price ratio $\bar{p}_{\mathrm{sb}}/\bar{p}_{\mathrm{wh}}$.

Table 9.7: *Ranges of price level coefficients*

Price level	Coefficient	Range	Avg. ratio
Beef & young cattle	*pc_beef*	[0.7, 1.3]	
Fertilizer	*pc_fert*	[0.5, 2]	
Fuel & energy	*pc_fuel*	[0.7, 1.5]	
Milk	*pc_milk*	[0.7, 1.2]	
Pork & pigs	*pc_pork*	[0.7, 1.3]	
Wheat	*pc_wh*	[0.5, 2]	
Animal feed (rel. to wheat price development)	*dcwh_fodd*	[0.7, 1.5]	var.
Grain maize (rel. to wheat price development)	*dcwh_mg*	[0.8, 1.2]	1.020
Malting barley (rel. to wheat price development)	*dcwh_sb*	[0.7, 1.3]	1.175
Fodder barley (rel. to wheat price development)	*dcwh_wb*	[0.9, 1.2]	0.908
Rapeseed (rel. to wheat price development)	*dcwh_wr*	[0.7, 1.4]	2.032

These 11 price coefficients were combined with the 11 parameters used already in the previous experiments to form a Latin-hypercube sample with 600 repetitions that were run once under current climate conditions and once using the climate scenario (with all three aspects of climate change applied). The high number of repetitions ensured that the correlation among price coefficients and parameters should be minimal and allow an analysis of the direct influences on prices on outcomes.

Table 9.8 presents a first overview of the results. In the first two columns, it shows the coefficient of variation ($CV = \frac{\sigma}{\mu}$) of each outcome indicator under current (CV_b) and potential future (CV_c) climate indicating the variation of outcomes over the 600 repetitions in each scenario. High values show that prices or parameter settings had a strong influence on these outcome indicators.

The third and fourth column relate the average climate change effect over all repetitions (μ_d) to the mean outcome over all repetitions under current (μ_b), respectively potential future conditions (μ_c). The average climate change effect μ_d was estimated by calculating the difference in response (y) between baseline and climate change scenario at each repetition j and then averaging over all repetitions, i.e $\mu_d = \frac{1}{n}\sum_j^n (y_{c,j} - y_{b,j})$.

The fifth and sixth column then show the ratio between the variance of the climate effect over the repetitions (σ_d^2) and the variance of the outcome indicator in the baseline (σ_b^2), respectively the climate change scenario (σ_c^2). A high value of μ_d/μ_b combined with a low value of σ_d^2/σ_b^2 would indicate that there was a strong effect of climate change on the outcome, but this effect did not vary much over price levels (compared to the baseline variance), and thus one would expect a mere shift in the functions relating prices to the outcome. A high value of σ_d^2/σ_b^2, in contrast, would let us expect a structurally more severe effect that might not only

Table 9.8: *Variation of outcomes (CV_b, CV_c), average climate effect on outcomes (μ_d/μ_b, μ_d/μ_c) and variation of the climate effect on outcomes over price/parameter combinations (σ_d^2/σ_b^2, σ_d^2/σ_c^2)*

Outcome	CV_b	CV_c	μ_d/μ_b	μ_d/μ_c	σ_d^2/σ_b^2	σ_d^2/σ_c^2
Dairy cows	0.02	0.02	0.00	0.00	0.14	0.19
Fattening pigs	0.24	0.23	0.00	0.00	0.01	0.01
Young pigs	0.94	0.93	0.01	0.01	0.00	0.00
Sows	0.29	0.28	0.01	0.01	0.01	0.01
Fattening bulls	0.34	0.31	0.13	0.12	0.03	0.03
Heifers	0.06	0.05	0.02	0.02	0.07	0.10
Summer barley area [ha]	0.24	0.51	-0.22	-0.28	0.79	0.29
Winter wheat area [ha]	0.20	0.18	0.33	0.25	0.37	0.25
Winter barley area [ha]	0.41	0.73	-0.64	-1.80	0.28	0.69
Silage maize area [ha]	0.20	0.19	0.15	0.13	0.15	0.13
Winter rapeseed area [ha]	0.29	0.24	0.15	0.13	0.07	0.08
Fallow area [ha]	2.35	2.95	-0.33	-0.49	0.14	0.20
Field grass area [ha]	0.77	0.79	-0.07	-0.08	0.07	0.07
Meadow and mown pasture [ha]	0.02	0.02	-0.02	-0.03	0.38	0.20
Pasture [ha]	0.36	0.32	0.61	0.38	0.45	0.22
Biogas capacity [kW]	0.20	0.20	0.00	0.00	0.00	0.00
MEKA III A2 area [ha]	0.21	0.48	-0.37	-0.60	1.54	0.73
MEKA III B1 area [ha]	0.11	0.11	0.00	0.00	0.20	0.22
MEKA III B2 area [ha]	0.07	0.07	-0.01	-0.01	0.15	0.18
MEKA III B4 area [ha]	0.12	0.14	0.03	0.03	0.09	0.07

shift but also turn the function. For the combination of a low μ_d/μ_b with a high σ_d^2/σ_b^2, one would expect that the climate effect could be noticed only under specific circumstances (e.g. at high price levels).

Tables 9.9 and 9.10 provide a quick overview over the distance correlations between outcomes and price coefficients in the baseline, respectively the climate scenario. Distance correlations[4] were chosen, because they capture also non-linear and nonmonotonous relationships between variables [Szekely et al., 2007; Szekely and Rizzo, 2013]. As a consequence of capturing also non-monotonous relationships, they cannot indicate the sign of the correlation as Pearson's ρ or Spearman's rank correlation would, but range between 0 (independence) and 1 (perfect correlation). For comparison, the distance correlation between the parameters of the LHS did not surpass 0.1 for any pair of parameters indicating that the sample was largely uncorrelated as desired.

An interesting first observation was that the high σ_d^2/σ_b^2 for the MEKA A2 measure – pointing to a structural change in the relationship to price coefficients – was reflected in an increase of distance correlation between MEKA A2 area and the wheat price rising from 0.29 in the baseline to 0.65 under climate change conditions. Apparently, MEKA A2 participation was much stronger related to the general crop price level under future conditions than under current conditions. The following sections have a closer look at the individual outcome indicators and analyze their relationship with price coefficients and model parameters and whether and how this was affected by climate change.

[4]The algorithm to estimate distance correlations was implemented in stata following the R implementation of Rizzo and Szekely [2013].

Table 9.9: *Distance correlation between outcomes and price level coefficients under current climate conditions.*

	pc_beef	*pc_fert*	*pc_fuel*	*pc_milk*	*pc_pork*	*pc_wh*	*dcwh_fodd*	*dcwh_mg*	*dcwh_sb*	*dcwh_wb*	*dcwh_wr*
Dairy cows	0.06	0.07	0.07	0.21	0.04	0.26	0.08	0.05	0.05	0.04	0.07
Fattening pigs	0.05	0.06	0.09	0.05	**0.30**	**0.45**	0.23	0.05	0.05	0.04	0.06
Young pigs	0.08	0.06	0.10	0.04	**0.93**	0.05	0.05	0.04	0.05	0.06	0.06
Sows	0.05	0.05	0.10	0.04	**0.39**	**0.30**	0.07	0.08	0.04	0.04	0.05
Fattening bulls	**0.59**	0.04	0.16	0.07	0.11	**0.68**	0.11	0.07	0.09	0.06	0.09
Heifers	**0.56**	0.04	0.14	0.08	0.08	**0.62**	0.15	0.07	0.10	0.05	0.08
Summer barley area [ha]	0.13	0.07	0.06	0.06	0.06	0.07	0.11	0.06	**0.80**	0.13	0.13
Winter wheat area [ha]	0.10	0.06	0.06	0.05	0.06	**0.43**	0.09	0.07	**0.35**	**0.40**	0.20
Winter barley area [ha]	0.06	0.06	0.06	0.08	0.08	**0.35**	0.05	0.06	0.22	**0.61**	0.08
Silage maize area [ha]	0.18	0.07	0.23	0.08	0.11	0.20	0.10	0.10	0.14	0.06	**0.36**
Winter rapeseed area [ha]	0.05	0.05	0.06	0.07	0.06	**0.43**	0.09	0.10	0.09	0.06	**0.79**
Fallow area [ha]	0.05	0.05	0.11	0.09	0.07	**0.53**	0.09	0.09	0.06	0.13	0.19
Field grass area [ha]	0.07	0.06	0.14	0.07	0.07	**0.73**	**0.38**	0.05	0.12	0.12	0.18
Meadow …[ha]	0.14	0.06	0.22	0.08	0.06	**0.63**	0.11	0.10	0.07	0.07	0.27
Pasture [ha]	0.13	0.07	0.24	0.06	0.06	**0.72**	0.13	0.11	0.08	0.08	0.27
Biogas capacity [kW]	0.07	0.06	0.06	0.05	0.07	0.08	0.10	0.05	0.05	0.07	0.05
MEKA III A2 area [ha]	0.08	0.08	0.11	0.06	0.08	0.29	0.28	0.08	0.18	**0.58**	0.12
MEKA III B1 area [ha]	**0.41**	0.06	0.07	0.06	0.07	**0.35**	0.29	0.08	0.06	0.08	0.18
MEKA III B2 area [ha]	**0.48**	0.06	0.07	0.13	0.16	**0.43**	0.23	0.09	0.07	0.10	0.25
MEKA III B4 area [ha]	**0.36**	0.05	0.09	0.07	0.11	**0.74**	0.08	0.08	0.07	0.12	0.16

Table 9.10: *Distance correlation between outcomes and price level coefficients under climate change conditions.*

	pc_beef	*pc_fert*	*pc_fuel*	*pc_milk*	*pc_pork*	*pc_wh*	*dcwh_fodd*	*dcwh_mg*	*dcwh_sb*	*dcwh_wb*	*dcwh_wr*
Dairy cows	0.07	0.07	0.05	0.12	0.04	0.16	0.07	0.04	0.04	0.03	0.06
Fattening pigs	0.05	0.07	0.09	0.05	**0.30**	**0.45**	0.23	0.05	0.05	0.04	0.06
Young pigs	0.08	0.06	0.10	0.04	**0.93**	0.05	0.05	0.04	0.05	0.06	0.06
Sows	0.05	0.05	0.10	0.04	**0.38**	0.29	0.07	0.08	0.04	0.04	0.05
Fattening bulls	**0.59**	0.04	0.16	0.06	0.11	**0.67**	0.10	0.07	0.12	0.05	0.09
Heifers	**0.59**	0.04	0.14	0.05	0.07	**0.56**	0.12	0.07	0.10	0.05	0.09
Summer barley area [ha]	0.11	0.05	0.06	0.04	0.07	0.17	0.05	0.06	**0.82**	0.14	0.14
Winter wheat area [ha]	0.10	0.05	0.06	0.05	0.05	**0.42**	0.08	0.06	**0.62**	0.24	0.12
Winter barley area [ha]	0.05	0.05	0.07	0.05	0.07	**0.42**	0.06	0.06	0.29	**0.62**	0.07
Silage maize area [ha]	0.21	0.07	0.20	0.07	0.10	**0.30**	0.07	0.11	0.21	0.06	0.28
Winter rapeseed area [ha]	0.05	0.05	0.06	0.07	0.06	**0.44**	0.08	0.08	0.21	0.05	**0.71**
Fallow area [ha]	0.05	0.04	0.08	0.08	0.08	**0.53**	0.06	0.10	0.08	0.22	0.17
Field grass area [ha]	0.06	0.05	0.15	0.06	0.07	**0.79**	0.26	0.05	0.21	0.15	0.16
Meadow … [ha]	0.13	0.06	0.24	0.05	0.06	**0.70**	0.13	0.13	0.09	0.08	**0.31**
Pasture [ha]	0.13	0.05	0.24	0.05	0.05	**0.76**	0.15	0.13	0.10	0.07	**0.31**
Biogas capacity [kW]	0.06	0.06	0.05	0.05	0.07	0.09	0.10	0.05	0.05	0.07	0.05
MEKA III A2 area [ha]	0.06	0.05	0.07	0.04	0.09	**0.65**	0.07	0.06	0.13	**0.54**	0.05
MEKA III B1 area [ha]	**0.49**	0.05	0.08	0.05	0.08	0.25	0.26	0.08	0.13	0.08	0.10
MEKA III B2 area [ha]	**0.62**	0.05	0.07	0.07	0.14	**0.33**	0.16	0.10	0.14	0.09	0.18
MEKA III B4 area [ha]	0.26	0.05	0.09	0.07	0.10	**0.82**	0.07	0.08	0.16	0.15	0.13

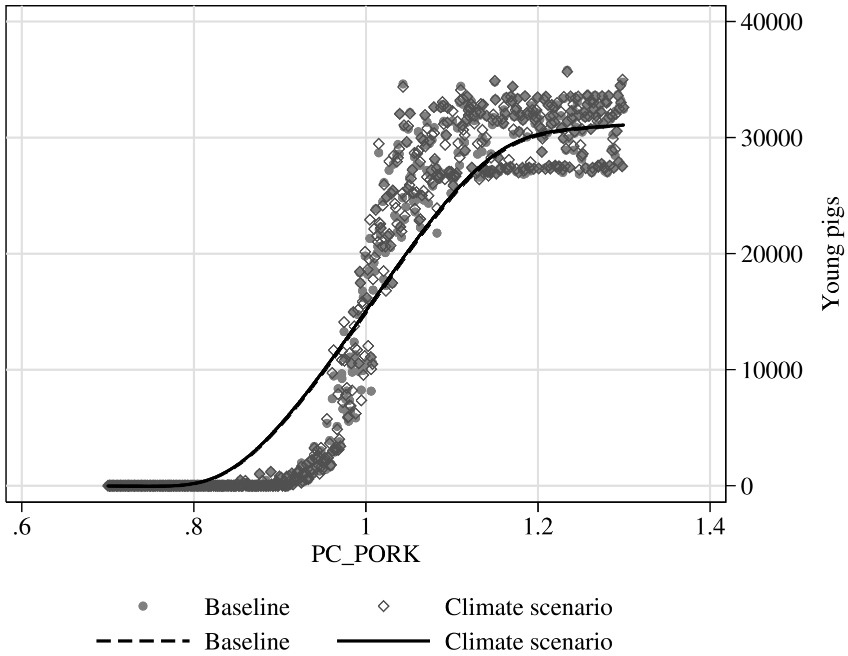

Figure 9.15: *Scatterplot of the number of young pigs being raised as a function of the piglet/pork price level (Lowess smoother: bandwidth=0.8, tricube weighting).*

9.6.1 Pigs

With respect to pig and piglet production, table 9.8 indicates a potentially strong dependence on prices that was hardly affected by the climate change scenario: The coefficients of intra-scenario variation were comparatively high, while the difference between the scenarios was stably small. Unsurprisingly, the distance correlation coefficients in tables 9.9 and 9.10 suggested to have a closer look at the pork and piglet price level as well as the wheat price.

For piglet raising, the correlation with the pork/piglet price level was very strong (0.93). The scatterplot in figure 9.15 reveals a very clear-cut sigmoid relationship. If the price fell below the average price level of 2000-2009 the business became unprofitable and farm agents shut down, while the current stable structure sets an upper limit to expansion in short-term simulations. It seems that the standard price scenario is located in a very sensitive area that warrants increased scrutiny and should receive closer attention in future model development. The lowess smoother that was overlaid over the scatter plot does not differ at all between the climate scenarios.

For piglet production, the number of sows depended on both the wheat and the pork price levels. More specifically as shown in the three dimensional scatter plot in figure 9.16 for current climate conditions, the piglet production broke down when the piglet/pork price was low. In the other cases, the number of sows was determined by the upper limit set by the stable places, which varies with the chosen starting population. This relationship could reasonably well be approximated by a linear regression including the piglet/pork price level, its square, the wheat price level, the interaction of the two price levels and dummies for the starting population chosen. As table 9.11 shows, estimates for the regression coefficients did not change much between the two scenarios. A similar relationship held for the number of fattening pigs with the pork and fodder price levels (fig. 9.16, tab. 9.12).

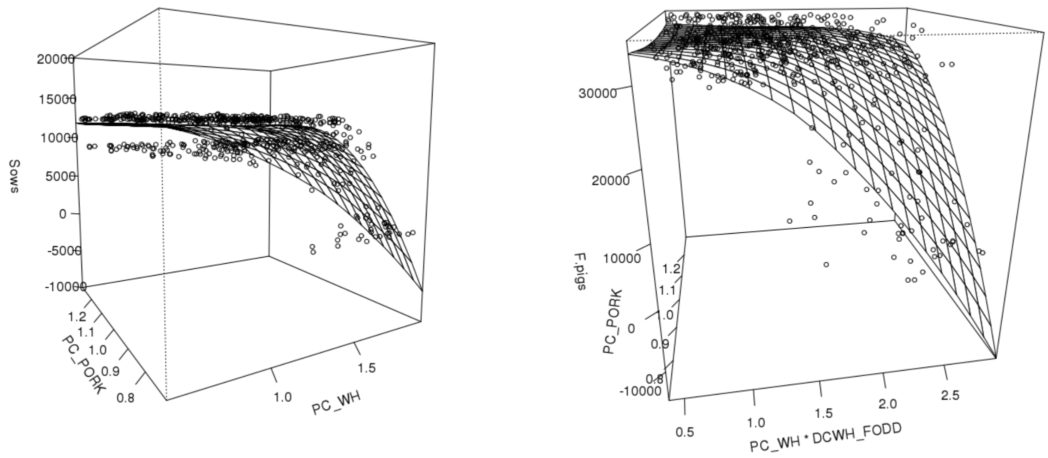

Figure 9.16: *Scatterplots of the number of sows as a function of the wheat and piglet/pork price level under current climate conditions (left) and the number of fattening pigs as a function of the fodder and piglet/pork price level under current climate conditions (right) (Lowess smoother: bandwidth=0.8, tricube weighting).*

Table 9.11: *Linear regression of the number of sows on wheat and piglet/pork price levels (Dummies D_1, D_2 represent starting population.)*

	Baseline			Climate change		
$N = 600$	coef.	s.e.		coef.	s.e.	
pc_pork	51223	5044	***	49685	4857	***
pc_pork sq.	-31669	2466	***	-30653	2374	***
pc_wh	-17716	914	***	-16905	880	***
pc_pork * *pc_wh*	15581	897	***	14897	864	***
D_1	-3054	163	***	-3130	156	***
D_2	-248	162		-280	156	*
constant	-4481	2645	*	-4056	2547	
	$R^2 = 0.71$			$R^2 = 0.72$		

Table 9.12: *Linear regression of the number of fattening pigs on wheat, fodder and piglet/pork price levels (Dummies D_1, D_2 represent starting population.)*

	Baseline			Climate change		
$N = 600$	coef.	s.e.		coef.	s.e.	
pc_pork	35,636	13,187	***	33,560	12,598	***
pc_pork sq.	-30,556	6,536	***	-28,828	6,243	***
pc_wh * *dcwh_fodd*	-20,859	2,319	***	-19,809	2,215	***
pc_wh * *dcwh_fodd* sq.	-4,258	477	***	-4,046	455	***
pc_pork **pc_wh* * *dcwh_fodd*	27,790	1,702	***	26,346	1,626	***
D_1	-3,229	430	***	-3,253	411	***
D_2	-3,029	429	***	-2,979	410	***
constant	27,159	6,754	***	27,604	6,452	***
	$R^2 = 0.64$			$R^2 = 0.64$		

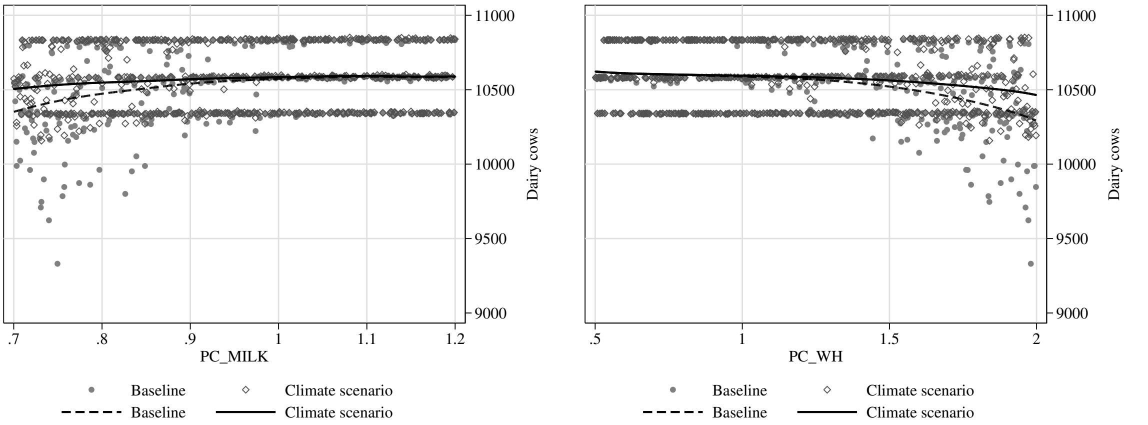

Figure 9.17: *Scatterplots of the number of dairy cows as a function of the milk (left) and wheat (right) price levels (Lowess smoother: bandwidth=0.8, tricube weighting).*

9.6.2 Cattle

Table 9.8 suggests the number of dairy cows to be rather unaffected by price development and the reaction to climate change to be minor. However, there seems to be some variance with respect to the climate change effect worth of closer inspection: Similar to sows and fattening pigs, the number of dairy cows had an upper limit defined by the number of stable places allocated during the initialization of agents, and only a few agents reduced the amount of dairy cows when the milk price was very low and the wheat/fodder price levels were very high at the same time. The climate change scenario dampened the reduction effect observed in the baseline such that the number of dairy cows showed even less reaction to price levels than in the baseline. This effect can be best appreciated in the two dimensional scatter plots shown in figure 9.17.

The relationship between fattening bulls and beef and wheat price levels is a nice example for a function that experienced a mere shift by the climate change scenario. This could already be expected from the statistics in table 9.8, which indicate a noticeable change on average with very little variation over the price coefficient domain and was confirmed by the scatterplots in figure 9.18. In both scenarios, the supply function could be well approximated by a linear regression including the two price levels and the dummies for the stable places initially assigned (table 9.13).

Table 9.13: *Linear regression of the number of fattening bulls on wheat and beef price levels. (Dummies D_1, D_2 represent the choice of starting population.)*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
pc_wheat	-2,566	40	***	-2,587	42	***
pc_beef	5,652	100	***	5,946	105	***
D_1	125	42	***	138	45	***
D_2	315	42	***	507	45	***
constant	1,871	114	***	2,137	120	***
	R^2 = 0.92			R^2 = 0.92		

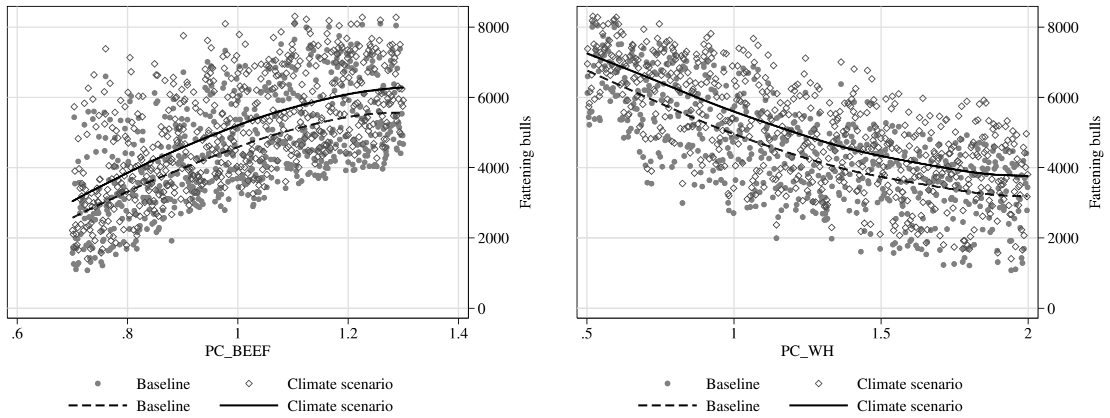

Figure 9.18: *Scatterplots of the number of fattening bulls as a function of the beef (left) and wheat (right) price levels (Lowess smoother: bandwidth=0.8, tricube weighting).*

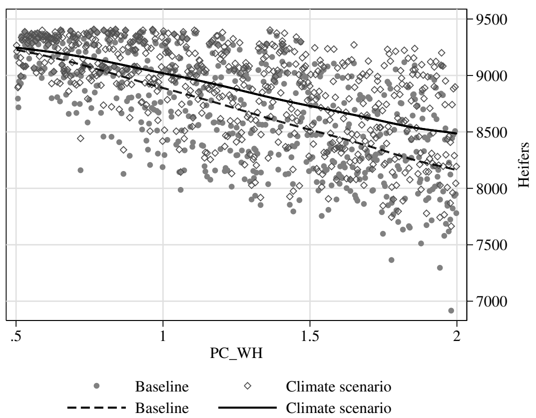

Figure 9.19: *Scatterplot of the number of heifers as a function of the wheat price level (Lowess smoother: bandwidth=0.8, tricube weighting).*

The price sensitivity of the number of heifers was much lower than the one of fattening bulls. The variation in the climate change effects indicated by the last two columns in table 9.8 could mainly be observed in a turn of the partial dependence on the wheat price level: Similar to dairy cows, the reduction caused by high wheat prices diminished in the climate change scenario (fig. 9.19). The relationship to milk and beef/young cattle prices (not shown here) shifted in a similar fashion as for fattening bulls.

9.6.3 Land use

The relationship of the summer barley area to the relative development of the malting barley-to-wheat price ratio (*dcwh_sb*) could be well approximated by a quadratic relationship explaining more than 70% of the variation over the price/parameter domain (tab. 9.14). As shown in figure 9.20, the difference observed between the climate change scenario and the baseline became positive when the summer barley to wheat price ratio increased by more than 20%, i.e. the summer barley-to-wheat price ratio reached more than 1.41 (given a baseline price ratio of 1.175).

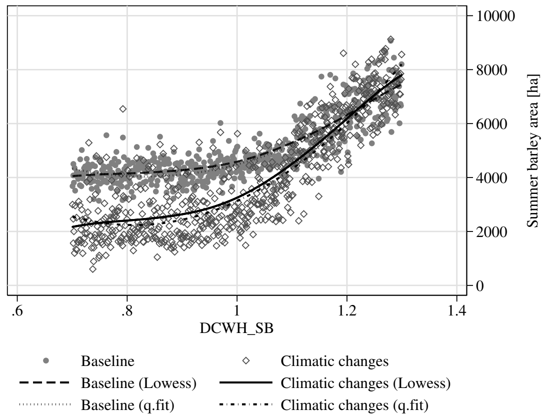

Figure 9.20: *Relationship of the summer barley area to the relative malting barley price development (dcwh_sb) (Lowess smoother: bandwidth=0.8, tricube weighting).*

Table 9.14: *Linear regression of the summer barley area on the relative malting barley price development (dcwh_sb).*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
dcwh_sb	-27,164	1,892	***	-40,704	2,809	***
dcwh_sb sq.	16,283	943	***	25,067	1,400	***
constant	15,370	927	***	18,763	1,376	***
	R^2 = 0.73			R^2 = 0.78		

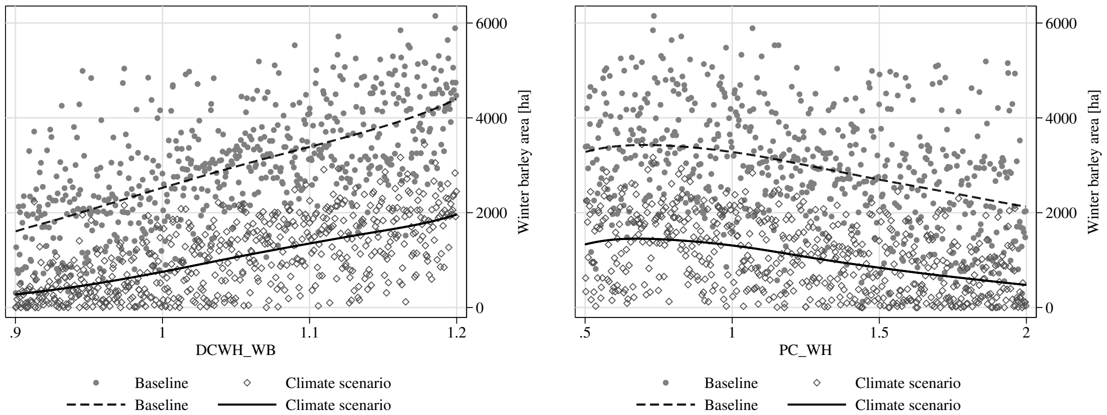

Figure 9.21: *Relationship of the winter barley area to the relative fodder barley price development (dcwh_wb) and the wheat price level (pc_wh) (Lowess smoother: bandwidth=0.8, tricube weighting).*

Winter barley was the crop that experienced the strongest climate effect relative to its baseline area (-0.64, cf. table 9.8). The effect was relatively stable over the whole parameter range, but was strongest when the winter barley price ratio was highest. In other words, the increase in winter barley caused by an increase in the winter-barley to wheat price ratio was lower under climate change conditions than in the baseline (fig. 9.21). The variation of the winter barley area over the price/parameter domain could to a large part be explained by the *dcwh_wb*, *dcwh_sb*, *pc_wh* and *proptohire* parameters (tab. 9.15).

While there was a substitutive relationship of the winter barley area with the wheat area in the baseline (Pearson's ρ = -0.643), the correlation was lower (ρ = -0.296) under climate change conditions due to the strong reduction in winter barley areas. Instead, the negative correlation between wheat and summer barley area became much stronger (Baseline: ρ = -0.354; Climate change: ρ = -0.8761). This was also reflected in the fact that the distance correlation between the winter wheat area and *dcwh_wb* decreased, while the one with *dcwh_sb* increased (tabs. 9.9,9.10). The winter wheat area showed a strong increase on average under climate change. The effect was, however, much lower when the summer barley-to-wheat price ratio was high (fig. 9.22). The effect on the supply curve with respect to the wheat price level was mainly a shift and similar for other price coefficients. The variation of the wheat area over the price/parameter domain is summarized in the regression in table 9.16.

Similar to wheat, the dependence of winter rapeseed areas on the malting barley price ratio increased under climate change conditions (tabs. 9.9, 9.10), while the relationship with respect to the wheat price level and the winter rapeseed price ratio experienced an upward shift (not shown here). The increased importance of malting barley prices could also be observed for silage maize areas (tabs. 9.9, 9.10) and also here the relationships to other price coefficients were predominantly shifted (not shown here). Overall the pattern of variation of the silage maize area over the price/parameter space was quite complex (tab. 9.17). Price coefficients and parameters increasing the profitability of beef and biogas production also increased the silage maize area. There was a substitutive relationship with brewery by-products in feeding as well as with winter rapeseed as an important element in the crop rotation.

The fallow area was inversely related to the wheat price level (fig. 9.23, upper left) and had similar relationships to the the other price coefficients (not shown). Under climate change conditions, the general nature of the curves remained unchanged, though the absolute reduction in fallow area was stronger the larger the fallow area in the baseline. A similar inverse relationship to the wheat price level could be observed for grass production on arable land (fig. 9.23, upper right). Production of roughage was shifted towards an intensification of grassland use (fig. 9.23, lower left). The effect of the rapeseed price ratio on the silage maize area discussed above also transmitted to meadows and mown pastures (fig. 9.23, lower right).

Table 9.15: *Linear regression of the winter barley area on wheat price level (pc_wh), relative fodder and malting barley price developments ($dcwh\_wb$, $dcwh\_sb$) and the proptohire parameter.*

$N = 600$	Baseline coef.	Baseline s.e.		Climate change coef.	Climate change s.e.	
pc_wh	-918	57	***	-739	38	***
dcwh_sb	8,750	286	***	5,621	192	***
dcwh_sb	-1,544	143	***	-1,202	96	***
proptohire	-1,185	57	***	-448	38	***
constant	-2,063	349	***	-2,162	234	***
	$R^2 = 0.75$			$R^2 = 0.73$		

Table 9.16: *Linear regression of the wheat area on selected price coefficients and model parameters.*

$N = 600$	Baseline coef.	Baseline s.e.		Climate change coef.	Climate change s.e.	
pc_wh	1,408	64	***	1,673	67	***
dcwh_wb	-6,319	323	***	-5,119	336	***
dcwh_sb	14,693	2,092	***	30,952	2,172	***
dcwh_sb sq.	-8,795	1,043	***	-18,418	1,083	***
proptohire	1,228	64	***	841	67	***
dcwh_wr	-1,128	138	***	-280	144	*
wheat_normal	9,293	1,363	***	12,010	1,415	***
constant	-4,052	1,775	**	-12,989	1,843	***
	$R^2 = 0.75$			$R^2 = 0.82$		

Table 9.17: *Linear regression of the silage maize area over the price/parameter domain. (Dummies D_1, D_2 represent starting population. Price coefficients cf. table 9.7. Other parameters cf. section 8.4.1.)*

$N = 600$	Baseline coef.	Baseline s.e.		Climate change coef.	Climate change s.e.	
pc_beef	365	56	***	462	58	***
pc_pork	279	56	***	257	58	***
pc_fuel	442	42	***	423	43	***
pc_wh	1,353	145	***	1,548	150	***
pc_wh sq.	-472	57	***	-496	59	***
dcwh_wb	24	112		282	116	**
dcwh_sb	395	720		1,821	744	**
dcwh_wb sq.	-355	359		-1,166	371	***
dcwh_wr	-799	48	***	-703	49	***
D_1	-281	24	***	-275	24	***
D_2	-247	24	***	-200	24	***
biertreber	-354	19	***	-366	20	***
workforhire	156	47	***	163	49	***
wheat_normal	999	470	**	1,096	485	**
kwkyno	215	19	***	225	20	***
constant	-81	621		-1,111	642	*
	$R^2 = 0.68$			$R^2 = 0.70$		

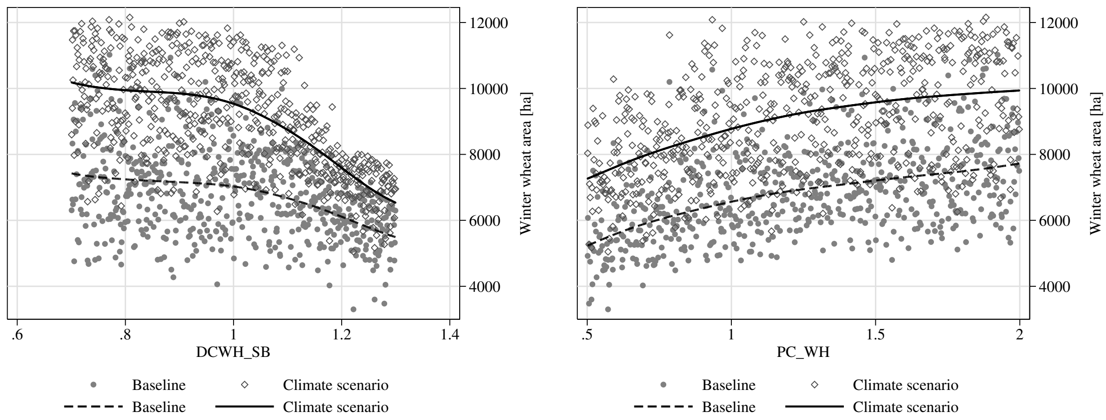

Figure 9.22: *Relationships between winter wheat and the relative malting barley price development (dcwh_sb), resp. the wheat price level (pc_wh).*

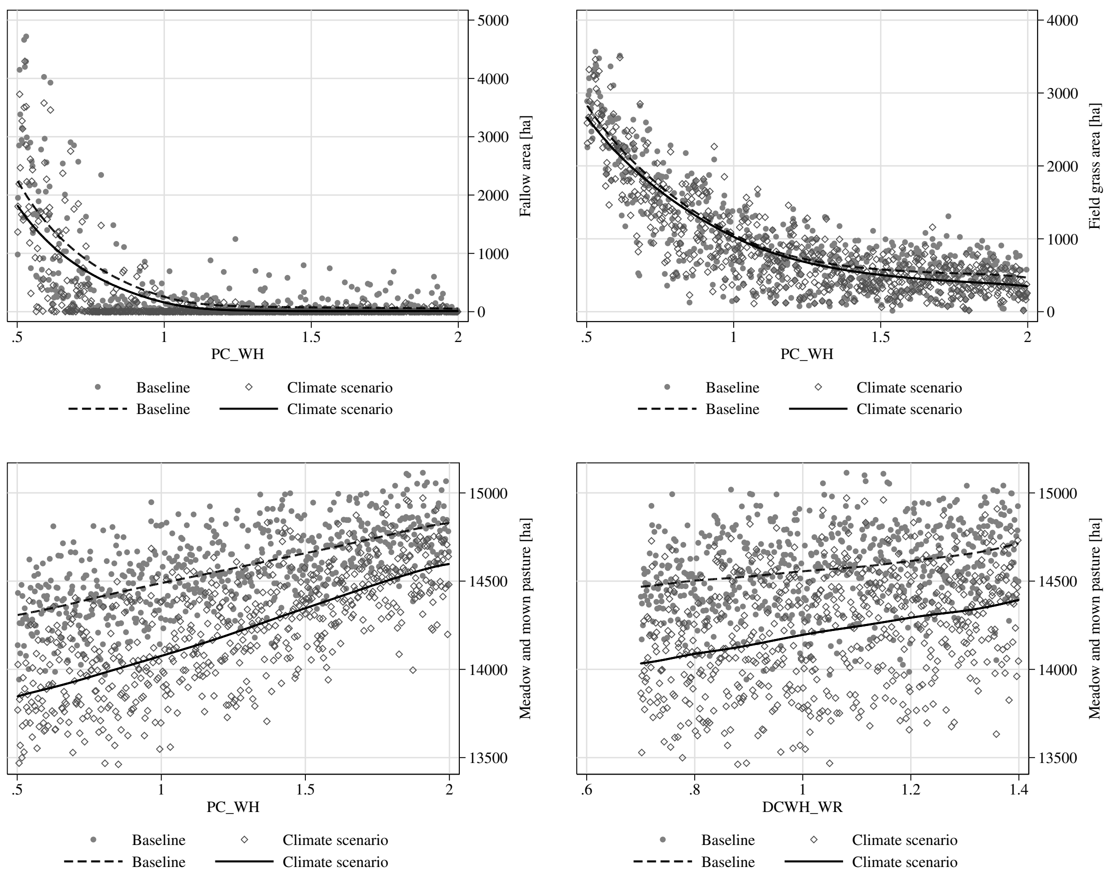

Figure 9.23: *Relationship between fallow, field grass and meadow/mown pasture area and the wheat price level, respectively the relative rapeseed price development (dcwh_wr).*

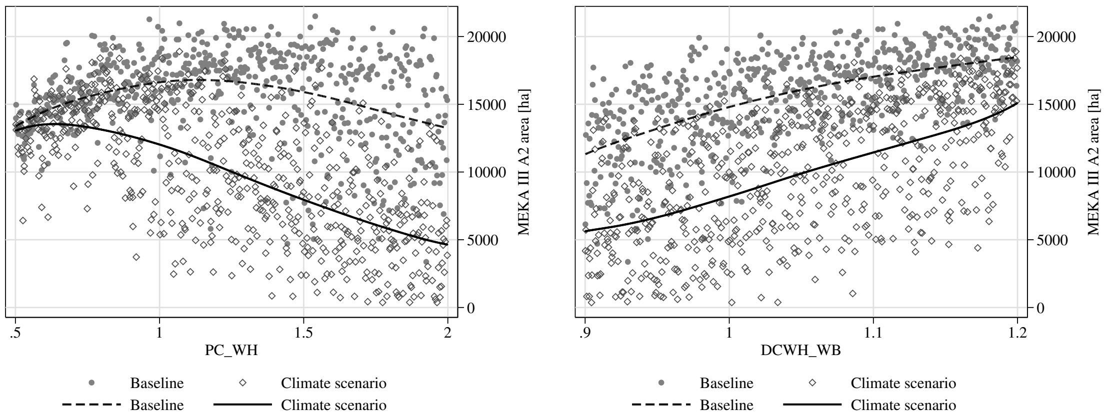

Figure 9.24: *Relationship of MEKA A2 participation [ha] to the wheat and winter barley price levels.*

9.6.4 Biogas & MEKA

The variation of installed biogas capacity was nearly entirely determined by the choice of the initial population and the *kwkyno* parameter controlling whether either everyone or no one has the opportunity to sell heat from the biogas plant. As already noticed above, there was very little effect of climate change at all.

With respect to the MEKA scheme, the effect of the wheat price level on the participation in the diversification measure A2 became stronger in the climate change scenario (tabs. 9.9, 9.10). Figure 9.24 shows that the MEKA A2 area started decreasing with higher wheat price levels much earlier under climate change conditions than in the baseline. On a more general level, table 9.18 illustrates the profound change of the relationship between price coefficients and parameters on the one hand, and the MEKA A2 area, on the other hand.

Participation in MEKA B1 and B2 is mutually exclusive and participation levels were consequently negatively correlated (Pearson's ρ = -0.93 in the baseline, -0.87 under climate change conditions). At low wheat/cereal price levels, agents preferred the higher rewards and tighter restrictions on livestock density of B2 over B1. At medium price levels B1 participation rose, to decrease again for higher wheat price levels, while B2 area remained stable. The climate change scenario slightly shifted the balance towards B1 for lower wheat/cereal price levels (fig. 9.25, upper and middle left). An increasing beef/young cattle price level increased B1 area, while it decreased B2 areas (Fig. 9.25, upper and middle right). Besides these, also fodder barley, rapeseed and feed price levels influenced participation (tabs. 9.19, 9.20).

The MEKA B4 area is a residual area similar to fallow and grassland – agents decide to register their grassland for B4 if they have no better use for it – and consequently the B4 showed a similar inverse relationship to the wheat price level as fallow and field grass (fig. 9.25, lower left). Like for the other MEKA B measures, participation was slightly increased for price levels at which it was low in the baseline.

Table 9.18: *Linear regression of the MEKA A2 area over the price/parameter domain. (Dummies D_1, D_2 represent starting population. For price coefficients cf. table 9.7. For other parameters cf. section 8.4.1.)*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
pc_wh	12,907	1,142	***	-5,991	1,342	***
pc_wh sq.	-16,227	953	***	550	1,120	
dcwh_wb	3,853	1,675	**	32,461	1,969	***
pc_wh * *dcwh_wb*	10,337	803	***	-843	944	
dcwh_fodd	-3,753	332	***	-1,884	391	***
dcwh_wr	902	371	**	-1,082	436	**
dcwh_sb	-4,098	434	***	-796	511	
pc_beef	-847	434	*	-800	511	
kwkyno	-463	151	***	-522	178	***
proptohire	-1,321	174	***	-2,489	204	***
D_1	142	184		-140	217	
D_2	-453	184	**	-770	216	***
constant	15,106	2,074	***	-7,692	2,438	***
	R^2 = 0.69			R^2 = 0.80		

Table 9.19: *Linear regression of the MEKA B1 area over the price/parameter domain. (Dummies D_1, D_2 represent starting population. Price coefficients cf. table 9.7. Other parameters cf. section 8.4.1.)*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
pc_wh	6,152	233	***	4,964	226	***
pc_wh sq.	-2,360	92	***	-2,008	89	***
pc_beef	1,710	89	***	1,971	86	***
dcwh_wr	602	76	***	235	74	***
dcwh_wb	199	89	**	549	86	***
dcwh_fodd	-607	67	***	-454	65	***
D_1	8	38		174	37	***
D_2	-327	38	***	-243	36	***
overmanure	552	107	***	553	103	***
constant	94	267		464	259	*
	R^2 = 0.71			R^2 = 0.69		

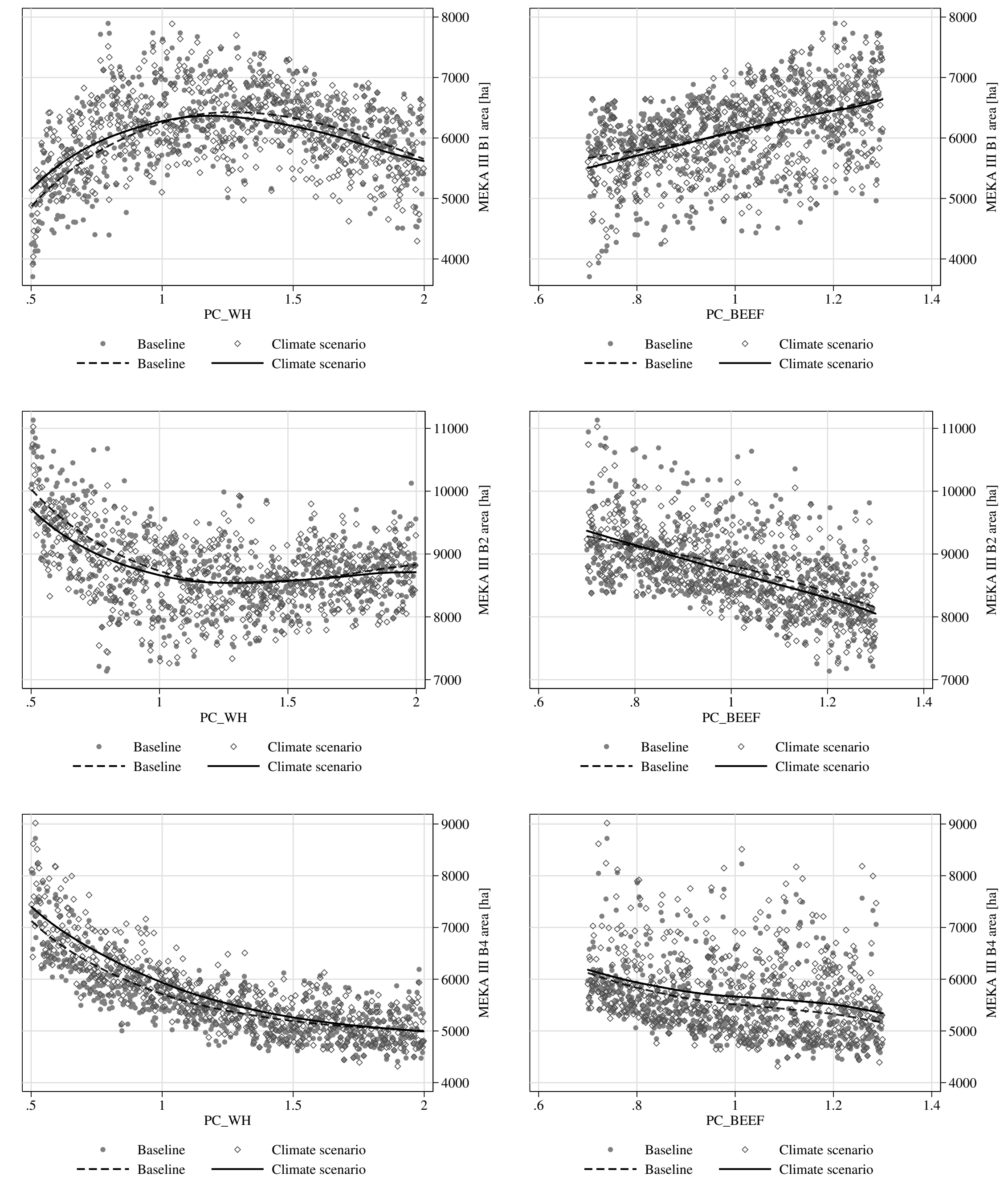

Figure 9.25: *Relationships between MEKA B1 (upper pane), B2 (middle pane) and B4 (lower pane) and the wheat and beef price levels.*

Table 9.20: *Linear regression of the MEKA B2 area over the price/parameter domain. (Price coefficients cf. table 9.7. Other parameters cf. section 8.4.1.)*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
pc_wh	-11,475	987	***	-10,697	827	***
pc_wh sq.	7,375	835	***	7,308	700	***
pc_wh cb.	-1,481	222	***	-1,571	186	***
pc_beef	-1,796	82	***	-2,139	69	***
dcwh_wr	-839	70	***	-599	58	***
dcwh_wb	-348	81	***	-628	68	***
dcwh_fodd	482	62	***	231	52	***
D_1	53	35		-32	29	
D_2	154	35	***	86	29	***
constant	16,580	397	***	16,596	333	***
	R^2 = 0.72			R^2 = 0.76		

Table 9.21: *Linear regression of the MEKA B4 area over the price/parameter domain. (Price coefficients cf. table 9.7. Other parameters cf. section 8.4.1.)*

	Baseline			Climate change		
N = 600	coef.	s.e.		coef.	s.e.	
pc_wh	-10,615	741	***	-9,113	777	***
pc_wh sq.	6,526	627	***	5,036	657	***
pc_wh cb.	-1,366	166	***	-977	174	***
pc_beef	-1,179	62	***	-967	64	***
dcwh_wr	-451	53	***	-432	55	***
dcwh_wb	-809	123	***	-1,354	129	***
dcwh_sb	-312	61	***	-792	64	***
dcwh_fodd	-214	47	***	-236	49	***
overmanure	669	74	***	747	77	***
kwkyno	-235	21	***	-163	22	***
freshgrasslabor	-134	19	***	-99	19	***
constant	13,628	330	***	14,113	346	***
	R^2 = 0.86			R^2 = 0.88		

9.7 Policy analysis: Goal conflicts between EEG and MEKA[5]

Both, the Renewable Energy Act (EEG) and the agri-environmental support scheme MEKA are examples for the development of agricultural policies in the last two decades: The focus shifted from traditional subsidization of agricultural production towards payments for public goods and services, greenhouse gas reduction and environmentally-friendly production. On the one hand, the motivation behind this development can to a certain extent be attributed to the desire of policy makers to maintain a certain level of support for farming. At the same time, they respond to the pressure to phase out coupled support arising in trade negotiations. On the other hand, these policies address growing public concerns for the externalities of food production, climate change and the conservation of traditional rural landscapes and farming systems [Baylis et al., 2008]. The wide array of different objectives addressed by these policies bears the danger that individual policy measures are narrowly targeted at one objective, while inadvertently counteracting another. This danger is even more prevalent if different political departments and scientific communities are targeting different objectives [Poe, 1997].

The EEG aims to contribute to climate change mitigation, a global environmental goal, via the promotion of renewable electricity production e.g. from biogas (see appendix B.1.5). As a consequence, it sets incentives to intensify agricultural production and specialize in silage maize production. High profit margins and guaranteed revenue lead to rising rental prices for farmland and favor large production units. This is at odds with some of the objectives of the MEKA scheme (see appendix B.1.10)that emphasize a reduction of chemical input use, a conservation of biodiversity, extensification, diversification of crop production and an upkeep of traditional agricultural landscapes and production structures. Especially, the EEG-induced expansion of silage maize production has led to growing concerns over a 'maizification' of rural landscapes in Germany. As a reaction to these concerns, the most recent amendment to the EEG in 2012 emphasized limitations to the use of silage maize, a diversification of substrate mixes and co-generation of heat-and-power in order to reduce the environmental side-effects of biogas production.

This section analyzes the counterbalancing effects of EEG and MEKA in determining biogas production, silage maize area, farm incomes and shadow prices of farm land, and examines whether the most recent revision of the EEG can be expected to alleviate potential goal conflicts between the two policy schemes.

9.7.1 Simulation experiments

The analysis was based on a comparison of six scenarios: three EEG scenarios (EEG09, EEG12 and no EEG at all) in combination with two MEKA scenarios (MEKA III or no MEKA at all). All runs assumed price scenario 'B', the EU CAP regulations valid in 2012 and the 2007 agent population. Contrary to previous runs, however, the model was initiated without existing biogas plants. Instead, the simulations show, which of the model agents would invest in a biogas plant under the simulated scenarios, irrespective of whether their real world counterparts did so in the past or not.[6]

[5]The analysis in this section was conducted in cooperation with Teresa Walter, who compiled most of the information and data used for the implementation of biogas production and EEG support in the model.

[6]Since information on existing biogas plants was scarce and it is likely that there was a considerable expansion between 2007 and 2012, the number of existing biogas plants would have been hardly correct. While this was considered a minor problem for climate change simulations, this more general analysis seemed more consistent in this case, where EEG effects were at the center of the analysis.

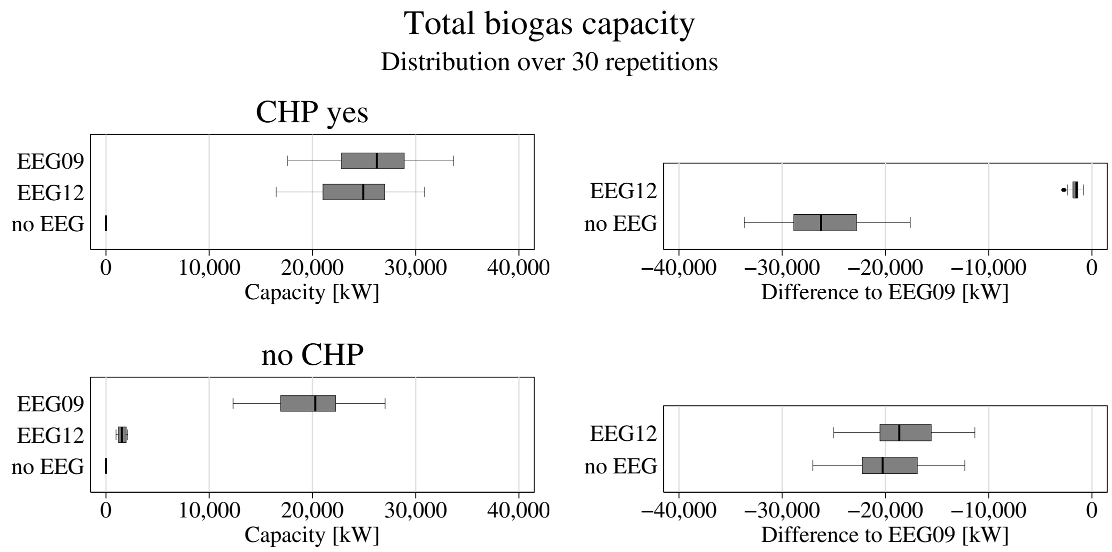

Figure 9.26: *Predicted biogas capacities for the three EEG scenarios and the two heat sales scenarios (with MEKA) over 30 repetitions. Left: predicted capacities. Right: predicted differences to EEG 2009.*

The model was run for one period covering investment and production decisions of agents. Potential investments included investments in biogas plants and farm machinery. Investments in livestock stables were excluded from the analysis. Down payments on biogas plants were disregarded to avoid conditioning the results on current, uncertain levels of farm liquidity. As a consequence, the prediction of biogas plant investment in the simulations has to be understood as representing the upper bound of short-term investment in biogas plants based purely on profitability considerations given no liquidity constraints.

Similar to the previous experiments, the parameter space was first reduced by running an elementary effects screening to identify the parameters with the highest influence on installed biogas capacity, maize areas, land use and MEKA participations (1,200 model runs). The 13 most important parameters were used to construct a Latin hypercube sample (LHS) with 30 repetitions. The *kwkyno* parameter controlling whether either everyone or no one has the opportunity to sell heat from the biogas plant stood out as the most important parameter by far, and it was decided to rerunt the full sample for both options instead of including it in the LHS, i.e. the experimental design comprises a total of 60 repetitions per scenario. With six policy scenarios, this leads to a total of 360 model runs.

9.7.2 Effect of the EEG revision

Figure 9.26 (left) shows the total production capacity for biogas electricity predicted under three different EEG scenarios. Each box plot shows the distribution over the 30 repetitions of the Latin hypercube sample, the upper pane representing the situation, where all farmers can sell biogas heat (CHP yes: $kwkyno = 1$) and the lower pane where no farmer can sell biogas heat (no CHP: $kwkyno = 0$). There was a considerable variation in predicted capacities between repetitions, but differences between EEG scenarios had a consistent direction over repetitions and were thus robust against parameter uncertainty (fig. 9.26, right).

The 2012 revision of the EEG restricts the share of maize in the total feedstock mass to

60%, and requires the combined use of at least 60% of the heat for plants, whose feedstock consists of less than 60% manure. Alternatively, a special higher electricity price is granted for small plants up to 75 kW that use manure for more than 80% of the electricity production (cf. section B.1.5). The model predicted that these policy changes could lead to a reduced investment in biogas capacity of 3-12% compared to the 2009 version of the EEG if all agents had the opportunity to sell the heat from their plant, and to a reduction of 91-94% if no agent could sell the heat. Without the EEG, no agent invested into biogas plants at current energy price levels. The reduced total biogas capacity was mainly caused by a lower number of agents investing in biogas plants (see tab. 9.22), while the average capacity per biogas plant was only lower under EEG 2012 without heat sales.

Table 9.22: *Simulated investments in biogas capacity (Average and standard deviation over 30 runs).*

EEG	CHP	No. investors	Avg. plant size [kW]
2009	yes	254 ± 42	101 ± 3
2009	no	195 ± 38	101 ± 4
2012	yes	223 ± 39	108 ± 4
2012	no	21 ± 5	75 ± 0

The composition of the feedstock used for biogas production is shown in table 9.23. The average plant size and feedstock composition for EEG 2012 in the absence of heat sales indicate that only the ‘small plant (<= 75 kW) with 80% farmyard manure’ category offers a profitable option for agents without CHP potential.

Table 9.23: *Predicted shares of different feedstock in total biogas production in the study area under different EEG scenarios (Mean and standard deviation over 30 repetitions)*

		Feedstock shares		
EEG	CHP	Grass silage	Maize silage	Manure
2009	yes	$33\% \pm 3.6$	$39\% \pm 4.4$	$28\% \pm 1.5$
2009	no	$29\% \pm 3.3$	$41\% \pm 4.5$	$30\% \pm 1.7$
2012	yes	$36\% \pm 3.3$	$39\% \pm 4.1$	$25\% \pm 1.8$
2012	no	$4\% \pm 1.2$	$16\% \pm 1.2$	$80\% \pm 0.0$

9.7.3 EEG and MEKA

Figure 9.27 shows the predicted effect of the different EEG scenarios on the area committed to the four MEKA support schemes included in the model. Box plots show the distribution over the 30 repetitions of the LH sample. The strongest effect could be observed for the A2 measure rewarding crop diversification, i.e. requiring at least four crops, each with an area share of at least 15% in the crop rotation. Without EEG support and under EEG 2012 without heat sales, nearly all the arable area was committed to the measure. Rotation constraints considered in the model mean that most agents who do not grow large amounts of silage maize used a four part crop rotation anyway and participation in MEKA A2 came at no additional cost. Participation was reduced to about a third with EEG support for renewable energy and increased silage maize production.

A similarly strong reduction of committed area in the scenarios with biogas expansion could be observed for measure B4, which rewards conservation of abandoned grasslands. For the support of extensive grassland use, there was a slight decline in area committed to

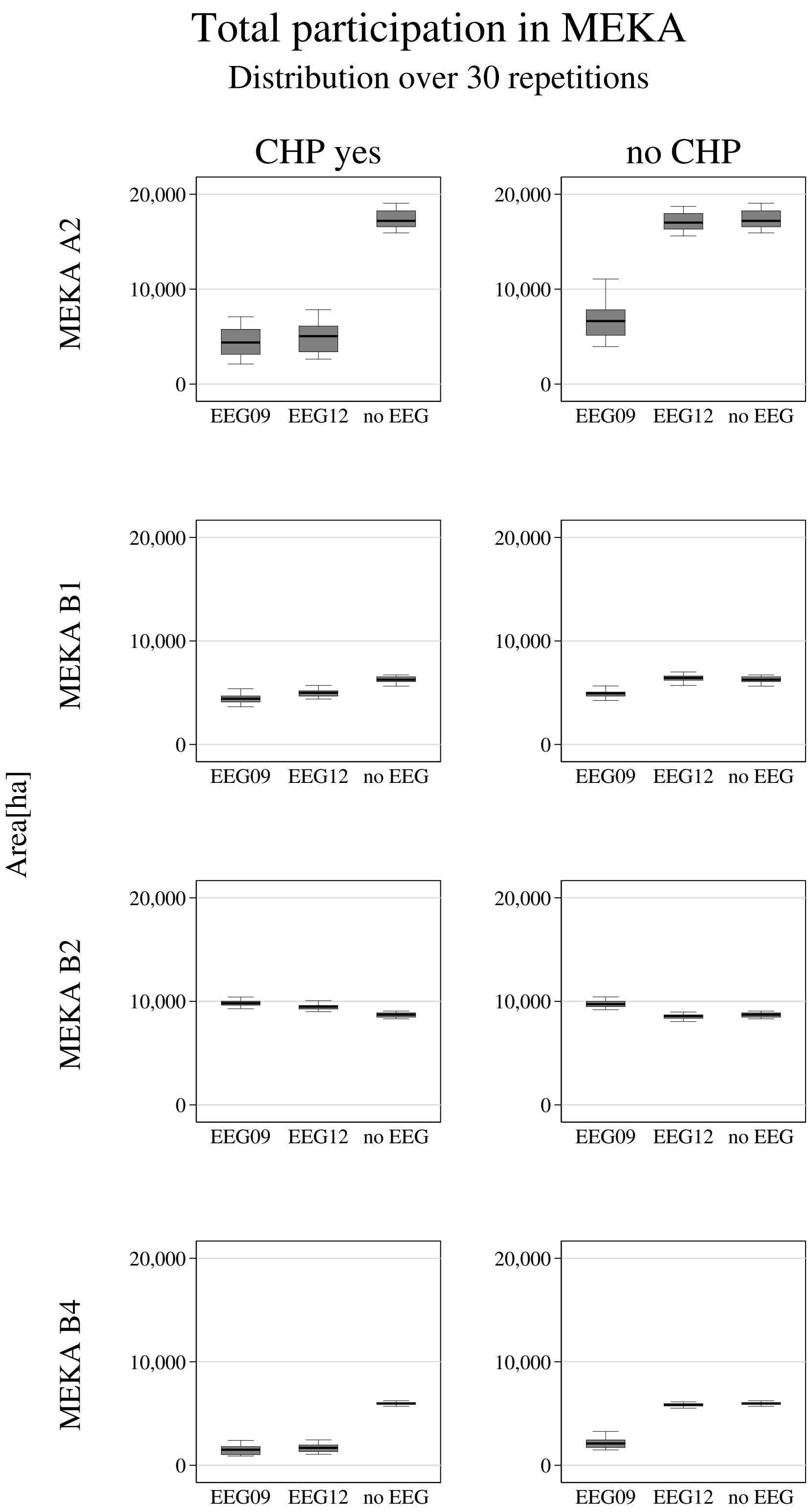

Figure 9.27: *Predicted participation in MEKA measures A2, B1, B2, and B4 in the three EEG and the two heat sales scenarios over 30 repetitions.*

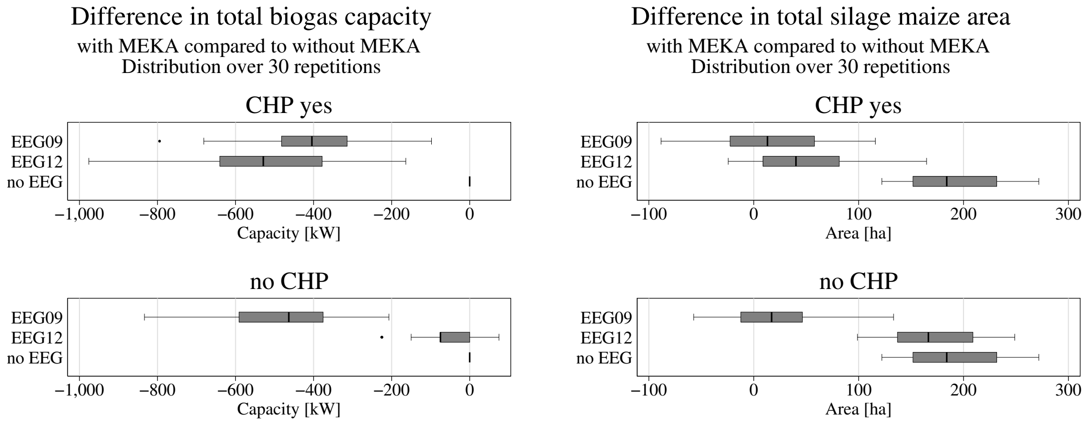

Figure 9.28: *Predicted effect of MEKA on biogas capacities (left) and silage maize areas (right) in three EEG scenarios with and without heat sales over 30 repetitions.*

measure B1, while the area committed to B2 increases slightly.

The effects of the MEKA schemes on biogas investments were smaller as shown in figure 9.28 which shows the difference in predicted total biogas capacity installed with MEKA support compared to a counterfactual situation without MEKA. The relative capacity reduction by MEKA generally ranged between 0 and 5%, with the exception of the EEG 12, no CHP scenario where MEKA caused a reduction of up to 16%, but also an increase of nearly 4% in one repetition.

9.7.4 Effects on silage maize areas

The predicted differences in biogas production had direct consequences for crop production in the model. Silage maize area experienced a three to five-fold increase compared to the situation without EEG support (fig. 9.29). Again predictions for EEG 2009 and 2012 were rather similar with only slight reductions in maize area for EEG 2012 as long as every agent had the opportunity to sell biogas heat, while in the absence of this opportunity silage maize production would be only slightly higher with EEG 2102 than without EEG. The predicted effects of the MEKA scheme on silage maize areas was comparatively small (cf. figure 9.28).

9.7.5 Effect on the value of land

Calculating the shadow prices of arable and grassland soils for each agent gives a first idea on the effects of biogas and MEKA support on long-term dynamics of agricultural structures. The shadow price, i.e. the marginal production value of land for each type of soil, was estimated by solving the production and investment decision of the agents a second time with an additional unit of this type of soil and observing the additional income that the agent obtained from using the additional amount of land. The median shadow price observed among the agents was then compared between scenarios. Table 9.24 summarizes the different relative effects of EEG and MEKA policy scenarios on the marginal production value of the most frequent soil class 0 (mostly rendzic leptosols) and grassland under the two possible assumptions for biogas heat sales.

The median shadow price of the scenario without MEKA and without EEG were set to one hundred for each of the 30 repetitions with and without heat sales, and the values of the

Table 9.24: *Relative effects of EEG and MEKA on the median shadow price of land (Mean and standard deviation over 30 repetitions). Values relative to the scenario without MEKA and without EEG, which was set to 100 for each repetition.*

		Soil class 0		Grassland	
MEKA	EEG	CHP yes	CHP no	CHP yes	CHP no
yes	2009	207± 49	140±18	802±188	577±103
	2012	192± 55	104± 1	762±201	288± 24
	no	104± 1	104± 1	284± 22	284± 22
no	2009	207± 52	137±19	693±203	427±115
	2012	191± 60	100± 0	631±230	104± 3
	no	100± 0	100± 0	100± 0	100± 0

other MEKA and EEG scenarios were expressed relative to this. The table shows mean and standard deviations of relative values over repetitions. The scenarios with EEG 2009 and EEG 2012 with heat sales that show high amounts of biogas production also showed high shadow prices of land with a doubling compared to no EEG support for arable land and even stronger effects for grassland. Without heat sales, the median shadow price remained nearly unaffected by EEG 2012 and showed lower, but still considerable increases for EEG 2009. The effect of MEKA support on the shadow price of arable land was negligible, but strong increases in production value could be observed for grassland areas irrespective of the EEG scenario.

9.7.6 Effects on farm incomes

Income effects of the different EEG versions are shown in figure 9.30. The box plots show the distribution of the absolute income difference with the respective EEG compared to no EEG version over agents and repetitions. A change of income could, of course, only be expected if agents invest in biogas plants, and agents were therefore grouped according to the scenarios in which they invest in a biogas plant: never, only with EEG 09, only with EEG 12, or with both EEG 09 and EEG 12. In general, income gains between 200 and 600 Euro per ha could be observed under EEG 09 and EEG 12 with heat sales. Agents that invest only in one of the two scenarios tended to have lower income gains not reaching 200 Euro per ha in most cases, an observation that also held for investors under EEG 12 without heat sales. Figure 9.31 shows the income effects of the MEKA scheme under the three EEG and two CHP scenarios. Box plots indicate the distribution of per-ha income effects over agents and repetitions. Potential income effect concentrated around 10 to 60 Euro per ha.

9.7.7 Discussion

The simulation results suggest interactions between EEG and MEKA policy schemes that go in both directions. If biogas support is successful in encouraging investments in biogas generation, it simultaneously decreases participation in grassland conservation and crop diversification measures under MEKA considerably.

Whether the EEG revision of 2012 substantially changes the picture depends very much on the assumptions on the demand for excess heat for biogas generation. When demand was too low to satisfy the cogeneration requirement of the new EEG, only few agents opted for the small manure plant and investment in biogas plants and the silage maize area was reduced considerably. Consequently also the effect on MEKA participation was minor. Given the

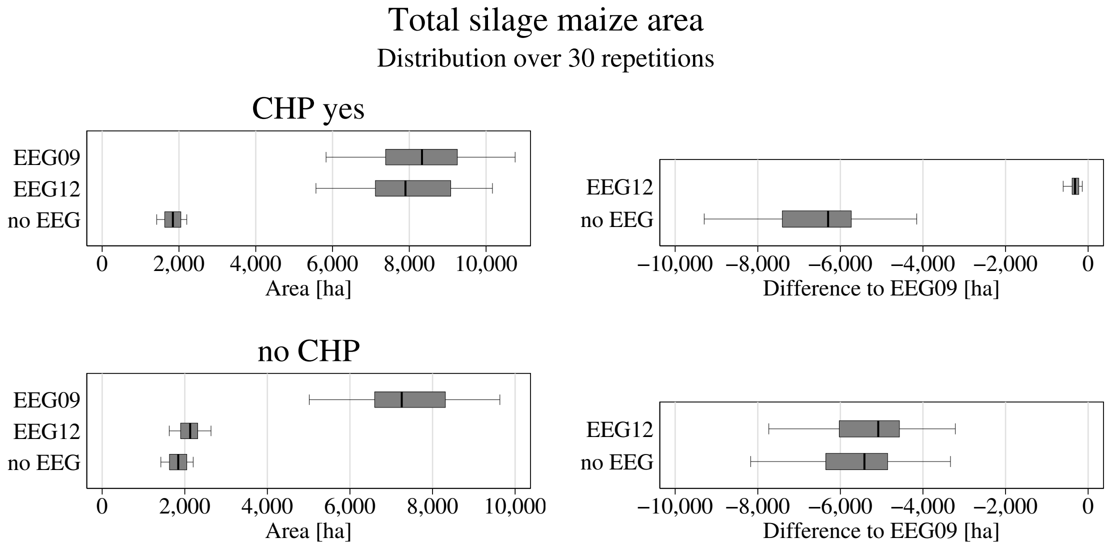

Figure 9.29: *Predicted silage maize areas for the three EEG scenarios and the two heat sales scenarios (with MEKA) over 30 repetitions. Left: predicted capacities. Right: predicted differences to EEG 2009.*

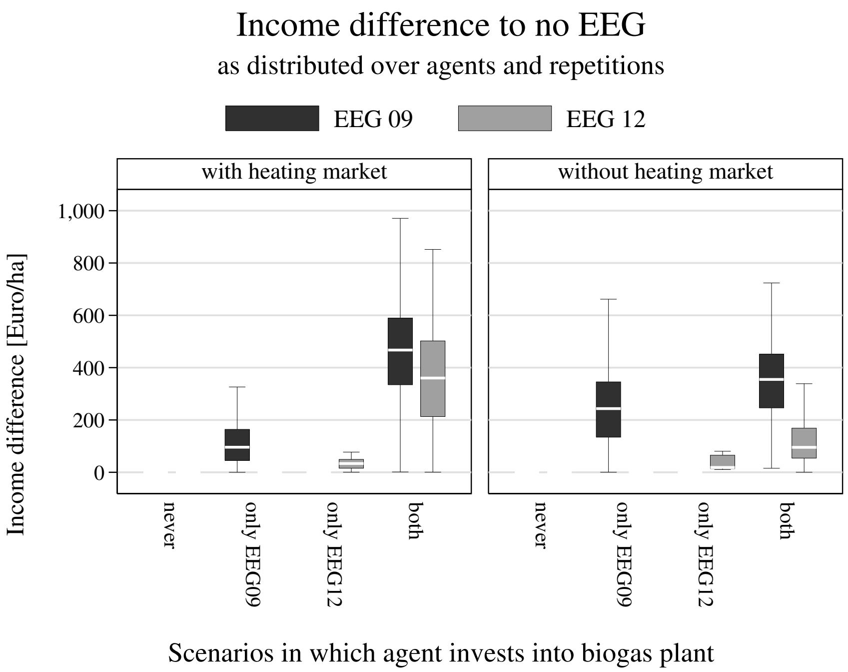

Figure 9.30: *Predicted income effect of EEG 09 and EEG 12 compared to no EEG. Distribution over 30 repetitions and agents grouped by biogas investments.*

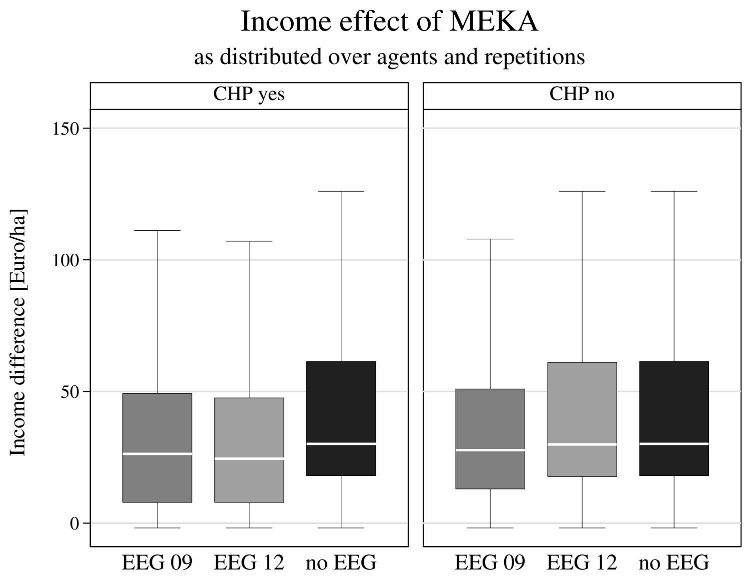

Figure 9.31: *Predicted income effects of MEKA in the three EEG and two CHP scenarios over 30 repetitions.*

importance of grassland in the study area, the newly introduced 60% cap on silage maize as feedstock had little effect: The average feedstock share in the EEG 2009 scenarios was only 40% already and changed little in the EEG 2012 scenario with heat sales.

Maybe a bit more surprising is the minor, but noticeable effect of the MEKA scheme on biogas investments: The income effect of the biogas policy was far greater for those who are able to invest than the income effect of the agri-environmental measures. Still, the comparatively modest compensation for environmental services could apparently keep a few agents from investing into biogas plants. And this leads to another observation: The potential effects on income show great differences between farmers and additionally only about half of the full-time farmer agents actually had the potential to profit from biogas production at all.

This heterogeneity bears the potential to accelerate structural change in agriculture, which is also illustrated by the doubling of the median shadow prices of arable land and the corresponding six- to eightfold increase for grassland. While shadow prices cannot be directly translated to rental prices, they indicate a strong increase in the willingness-to-pay for farmland. It is also interesting to note, how the MEKA scheme increased the value of grassland, albeit starting from rather low levels.

Chapter 10

Recursive-dynamic simulations of climate change effects

Chapter 2 discussed that recursive-dynamic simulations are essential for the analysis of climate change adaptation, especially to assess the potential speed of autonomous adaptation compared to the speed of natural changes. So far, this study analyzed farmer adaptation to climate change, policy implementation and price development using one-period simulations based on long-term averages and assuming agent expectations that coincided with the weather and price conditions applied during the simulations. This chapter takes the model a step further into the direction of fully dynamic models, presenting the result of first recursive-dynamic simulations with the Central Swabian Jura model. These simulations still rely on constant price and weather inputs, representing long-term averages, and an immediate adaptation of price and yield expectations, abstracting from the *adaptation of knowledge* phase, as well as abstracting from weather variability and risk management. Instead, they focus on *long-term adaptation*, i.e. changes to farm structure that cannot necessarily be realized from one season to the other due to constraints by farm liquidity, sunk cost or policy commitments and can hence be observed only to a limited extent in one-period simulations.

10.1 Model enhancements for recursive-dynamic simulations

When moving from static, one-period simulations to recursive-dynamic simulations over several periods, a number of processes gain importance that have only rather small effects in short-term simulations: The full effect of parameters governing the development of farm liquidity, investments, and household composition is realized only in long-term simulations. This section describes a number of model enhancements that were introduced to address problems observed in the results of first test simulations with the recursive dynamic setup. The land market, which also comes into play in dynamic simulations, has already been discussed in chapter 7.

10.1.1 Liquidity and the timing of investments

Investments in machinery usually require at least a certain share of self-financing and can consequently only be realized if agents have enough cash available. For investments in stables and biogas plants, whose size can be freely determined by the agent, this may lead to situations where agents invest in lower sizes than actually optimal for their production setup, because cash reserves are not high enough to cover the required equity share. In reality, these farmers might either negotiate with the bank to reduce the required down payment or wait a year longer to accumulate more liquidity. Implementing this decision in an average year decision problem is not straightforward, and instead, a second investment option without down payments was introduced for the size-dependent part of stables and biogas plants in order to avoid suboptimal decisions. Thus, if farmers have accumulated enough cash to pay the equity share of the size-independent part (the intercept of the linear cost function), they can choose a size optimal for their farm setup and then determine how much of this should be self-financed. More self-financing means lower cost, since the interest rate on foreign capital is higher than the calculatory interest rate on equity.

Reducing the equity share warrants a more careful examination of the repayment capacity of agents. This is also required in order to ensure that agents reserve a part of depreciation for replacement investment. Since the model uses an average year decision problem, it might otherwise be possible that the agent spends cash on a large investment, but is then unable to replace a tractor due to lack of cash the following year rendering the new investment useless. Repayment capacity was added to the model as a constraint that requires the additional debt service (a^{Bdbt}) incurred through investments (x^{iB}) to be smaller or equal to the amount remaining after subtracting the sum ($b^{fixSpend}$) of fix cost (depreciation, interest, rental payments) and minimum consumption from the expected total gross margin.

$$\sum \left(cx^{\neg iB}\right) - \sum_{b} \left(a_b^{Bdbt} x_b^{iB}\right) \geq b^{fixSpend} \tag{10.1}$$

10.1.2 Investments in pig and piglet production

For the short-term production decisions, investments in cattle and pig stables were not enabled. First runs with the recursive dynamic setup showed that allowing investments in stable places leads to a totally unrealistic tenfold increase in pig fattening stalls. Since the validation experiments suggested that the model does not overestimate incomes of granivore specialists, the reason is probably not rooted in a mismatch of pig fattening profitability. Rather, lower-than-expected investments in pig production despite its apparent profitability is a well-known phenomenon: Odening et al. [2005] observe a reluctance of German farmers to invest in hog production despite considerably higher average incomes in livestock production. They cite a much higher production risk in the pig business leading to an investment threshold that lies significantly higher than the break-even point based on net present value (NPV) calculation. They understand this wedge as a product of deferred investment and use the real options theory to explain it. Using the same theoretical framework, Hinrichs et al. [2008] estimate this wedge to lie at 79 Euro on average in Germany.

Reservations against entering into pig production can also be explained by a lack of experience and knowledge for farmers who have never raised pigs before. Such a lack of knowledge can of course be assumed for any production activity not previously done, but would be much more relevant in this case than for a mere switch of crops. Due to the high production risk, farmers may also be reluctant to hire permanent labor and therefore restrict pig production to a size that can still be handled by household labor. Most importantly, pig fattening stables need official approval and the legal process involves considerable transaction cost and legal restrictions and local opposition prevent the construction.

Properly reflecting these different potential explanations in the model will require more in-depth empirical research. For now, two *ad hoc* approaches to limit expansion of the pig branches were introduced into the model: First, an additional cost can be associated to the investment for planning reflecting the top-up on the NPV that e.g. could be explained by the risk considerations cited by Odening et al. [2005]. Second, the amount of stable places an agent can own can be restricted to a defined multiple of the number of stable places owned at the beginning, effectively only allowing those agents that are already pig producers to expand their production.

For the recursive-dynamic simulations presented in the following, three settings for these parameters were selected based on model pre-tests: In the first setting, a wedge of 150 Euro per pig fattening and 100 Euro per sow or piglet stable place were applied. In the second setting, agents cannot expand pig or piglet production, only replacement investments to maintain the capacity owned at simulation start were possible. The third setting restricted the maximum number of stable places that can be owned by an agent to five times the amount initially owned and defines a cost wedge of 100 Euro per stable place for fattening pigs, piglets and sows. While all three settings achieve the goal of limiting the expansion of pig production to more realistic levels, their dynamic validity is a bit questionable: It is uncertain, whether the observed price wedge is constant over time, and although severe restrictions on pig production exist in reality, they are usually less rigid than the ones implemented in the model. As a consequence, it has to be kept in mind that the model is currently not capable to reflect all relevant aspects of investments in pig production when interpreting model results.

10.2 Scenarios

Recursive-dynamic simulations were run for the baseline climate, a climate change scenario considering all three aspects of climate change, and the *ceteris paribus* climate change scenario for yields, all assuming price scenario 'B'. Since climate effects did not differ significantly between the three price scenarios, this setup should allow a first impression on the stability of results under long-term conditions. All three scenarios were run with and without land trading. In all cases, yields and prices were set to long-term averages and not varied during simulation.

Scenarios were repeated over a UPCS design reflecting model uncertainty. Parameters included in the design comprised the eleven parameters of the design used for short-term simulations of climate change and six parameters with potentially important effects on long-term results: The *sconextra* parameter indicates the share of income the agent household will consume on top of its minimum consumption, thereby also controlling the savings rate of the agent. The *sconred* parameter indicates what share of the agent households minimum consumption has to be consumed even if the agent will go bankrupt as a consequence. The *liqcoef* parameter determines the initial per-hectare cash reserves of the agent. The *pigscen* parameter reflects the three scenarios restricting pig and piglet production. The *suc_mincons* parameter represents the minimum income expectation required to make a potential successor take over the farm. It is expressed relative to the agent household minimum consumption. Finally, the two different ways to calculate the investment horizon (based only on the age of the current household head or considering also the age of the potential successor with highest priority) were included (i.e. the third and fourth version of *ihorizon_type*, cf. section 8.4.1). Parameter ranges used are listed in table 10.1.

The 17 parameters lead to a UPCS based on an OA (289, 18, 17, 2) orthogonal array, of which four subarrays (i. e. 68 repetitions) were used for the runs without land markets (204 runs in total). Simulations with land markets were run only for one subarray (17 repetitions) due to time and resource constraints discussed below (51 runs in total).

10.3 Results

In long-term simulations of the agricultural sector, the number of farms is likely to change between periods due to agent exits. As this needs to be taken into account when analyzing land use and income development over time, the following subsection first describes the development of agent exits, before the remainder of this section analyzes policy participation,

Table 10.1: *Ranges of six additional parameters for long-term simulations.*

Parameter	Range/Values
sconextra	[0.25; 0.75]
sconred	[0.5; 1]
liqcoef	[400; 1000]
pigscen	1, 2 ,3
suc_mincons	[1; 1.4]
ihorizon_type	only household head's age; age of potential successor

land use, livestock and income development over the simulation period.

10.3.1 Farm exits

Over the course of the ten simulation periods, farm agents exited due to illiquidity as well as death or retirement of the household head without a successor. The lack of a successor could either be caused by the absence of a potential successor interested in agriculture, or because the economic situation of the farm kept a potential successor from taking over the farm (i.e. the expected income was smaller than *suc_mincons* times the minimum household consumption). As shown in table 10.2, the total number of agents exiting the model in the 10 simulation years ranged between 103 and 201 over the 68 repetitions (median 148) in the baseline scenario. Given 522, resp. 528 agents in the starting population, this amounted to a compound annual loss rate of about 3.6% (2.4 -5.3%) that seems not too unrealistic given the observed reduction from 606 to 533 full-time farms in the study area between 2003 and 2007 corresponding to a compound annual loss rate of 3.2%. The majority of exits were due to illiquidity (33 - 72%) followed by successions failed due to economic reasons (20 - 55%). The wide ranges highlight strong variations making bankruptcies less important under some model parameterizations. Lack of a potential successor (5 - 18%) was the least important reason throughout.

The climate change scenarios showed a slightly lower number of agent exits, predominantly due to a reduction of the number of bankruptcies. Most of the factors governing succession and especially the existence of a potential successor were unaffected by climate change, which is reflected by a median difference of zero between baseline and climate change scenarios. Deviations can, however, be observed for individual parameter combinations due to the differences in illiquidities: In some cases, agents were bankrupt before the potential succession was evaluated. Moreover, the current MPMAS version does not allow full control over the random numbers used to generate stochastic household events (e.g. agent deaths), so different agents may be affected by death in different scenarios.[1]

The number of successful farm successions ranged between 23 and 57 in the baseline (median 40). The climate change scenarios had a rather ambiguous – on average slightly positive – effect on successions (differences to baseline: -11 to +20 for all, resp. -11 to +12 for yield changes only).

Since land markets were not simulated, the land owned or rented by exiting agents could not be used by other agents, leading to a reduction of the model area of 12 to 27% between the beginning and end of the simulation (cf. table 10.2). Moreover, as illustrated in figure 10.1

[1]The current MPMAS version does allow reading in a predetermined sequence of random numbers, but this guarantees exact reproduction of decisions only if all else is equal, otherwise the same random number might be used for a different decision.

Table 10.2: *Statistics on agent exits over 68 repetitions in the three climate change scenarios (without land markets).*

	Baseline			Climate change (all)			C.p. yields		
	med	min	max	med	min	max	med	min	max
Exiting agents	148	103	201	141	103	186	145	104	196
Δ Baseline				-9.5	-23	4	-4	-17	6
Bankruptcies	79	37	139	72	33	125	76	35	134
Δ Baseline				-9	-20	-1	-4	-14	0
Share of exits	55%	33%	72%	52%	29%	67%	54%	31%	69%
No successor	15	9	27	15	8	22	15	7	29
Δ Baseline				0	-14	10	0	-12	9
Share of exits	10%	5%	18%	11%	5%	20%	11%	4%	25%
Succession failed	52	35	73	53	39	72	52	39	72
Δ Baseline				0	-8	7	0	-5	7
Share of exits	35%	20%	55%	38%	23%	56%	36%	23%	57%
Area of exiting agents	7315	4455	10406	6703	4465	8816	6927	4737	9854
Δ Baseline				-662	-2365	381	-321	-1726	638

also nonexiting agents lost area over time: Whenever they were not able to pay land rents, they had to return the land to its owner, and since all land rented in was assumed to belong to outsiders, other model agents could not use these lands in the absence of land markets.

10.3.2 Biogas investments and MEKA participation

Figure 10.2 shows the development of biogas capacity over the simulation periods. To take account of the importance of the *kwkyno* parameter determining whether excess heat can be sold or not (already demonstrated in chapter 9.7), the graph shows two box plots for each period, one showing the distribution over repetitions with heat sales and one showing the distribution over the repetitions without.

For all three scenarios, the graph shows an initial year-to-year increase of installed biogas capacity, which seems to approach saturation in period five, due to investments decreasing from year-to-year. The installed capacity experienced a last sudden upward shift in year six and then remained stable except for a minor reduction in the last years caused by agent exits.

In the recursive dynamic setup, the climate change scenario did have a positive effect on biogas investments if heat sales were assumed. This effect could not be observed in the short-term simulation runs described in chapter 9.4 since it became apparent only from the second period onwards. The effect was less pronounced when only yield changes were considered. Higher biogas capacity with climate change in later years without heat sales were mainly due to a reduction of bankruptcies of biogas producing agents under more favorable climatic conditions.

As expected, the peak biogas capacity reached after period six was considerably lower than the potential capacity estimated under unrestricted conditions in chapter 9.7: The combination of delayed investments due to more realistic liquidity constraints and the yearly reduction of guaranteed prices prevented agents from investing that would have done so for 2012 guaranteed prices if they had not lacked liquidity or repayment capacity.

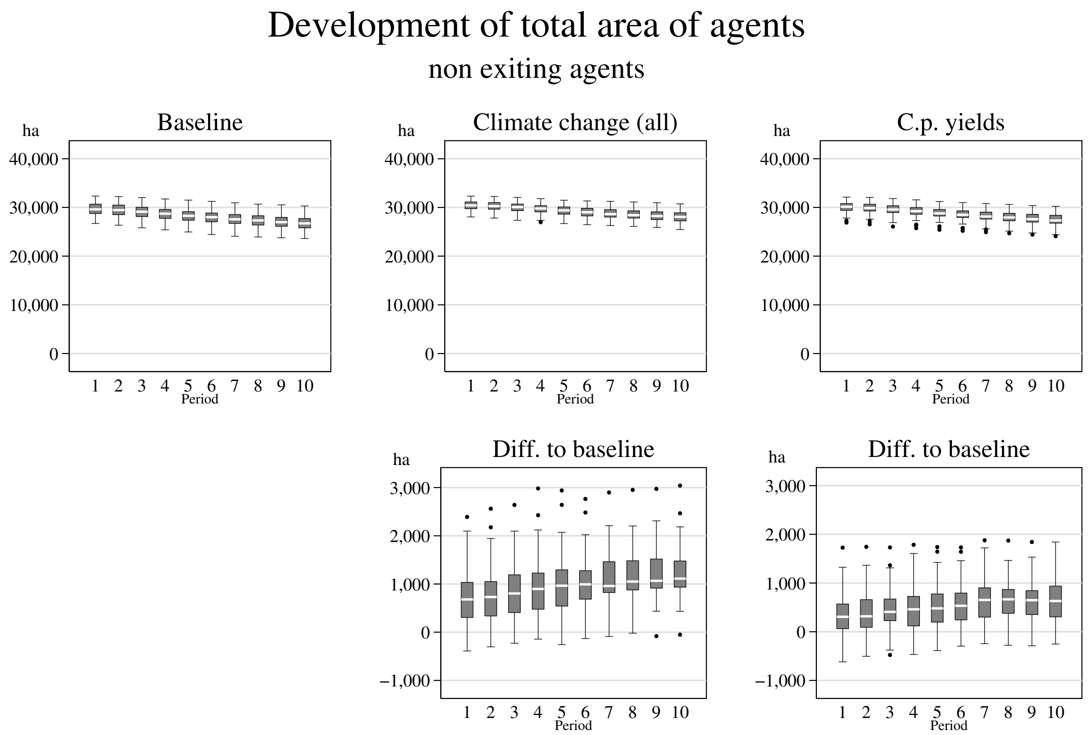

Figure 10.1: *Development of the model area of non-exiting agents over simulation periods. Box plots illustrate the distribution over 68 repetitions.*

Delayed investment is also the reason for the upward shift in biogas capacity from period five to six: Some agents that were not able to invest in biogas plants in the first year committed to MEKA participation for five years, which was apparently not compatible with biogas production even when the later became possible through the accumulation of equity over time. After the commitment had expired by year six, those agents that still found it profitable invested.

As a consequence, shifts can also be observed in the amount of area committed to the MEKA measures (figures 10.3-10.6). MEKA measures A2, B1, and B4 experienced downward shifts, while measure B2 was increased by increased biogas investment, consistent with the pattern observed in chapter 9.7. Due to constant exits of agents, area committed to MEKA also declined over time. To take out this effect, the graphs show MEKA commitments only for those agents that do not exit in a repetition. Still, also these areas declined due to area losses of nonexiting agents.

10.3.3 Land use

Similar to MEKA participation, the area loss through exiting agents lead to generally declining crop areas over simulation periods. To facilitate the detection of other effects, figures 10.7-10.9 show silage maize, field grass, winter wheat and fallow areas only for those agents that did not exit in a given repetition: The development of silage maize area paralleled the development in biogas capacity including a positive effect of climate change that was mainly associated to the changes in rotation options or field work and not so much due to the yield effects. The development of field grass area was linked to biogas production

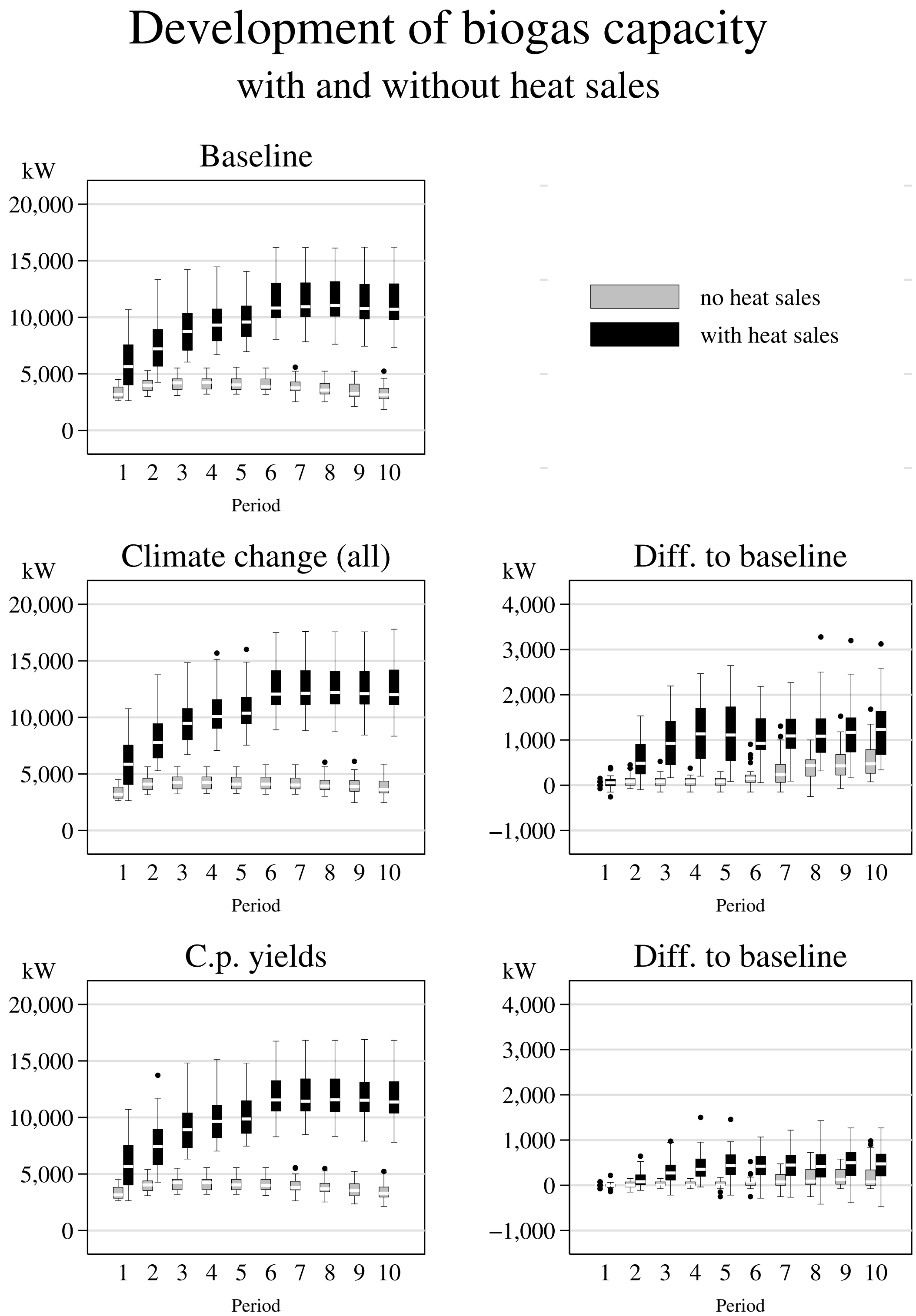

Figure 10.2: *Development of biogas capacity over simulation periods. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

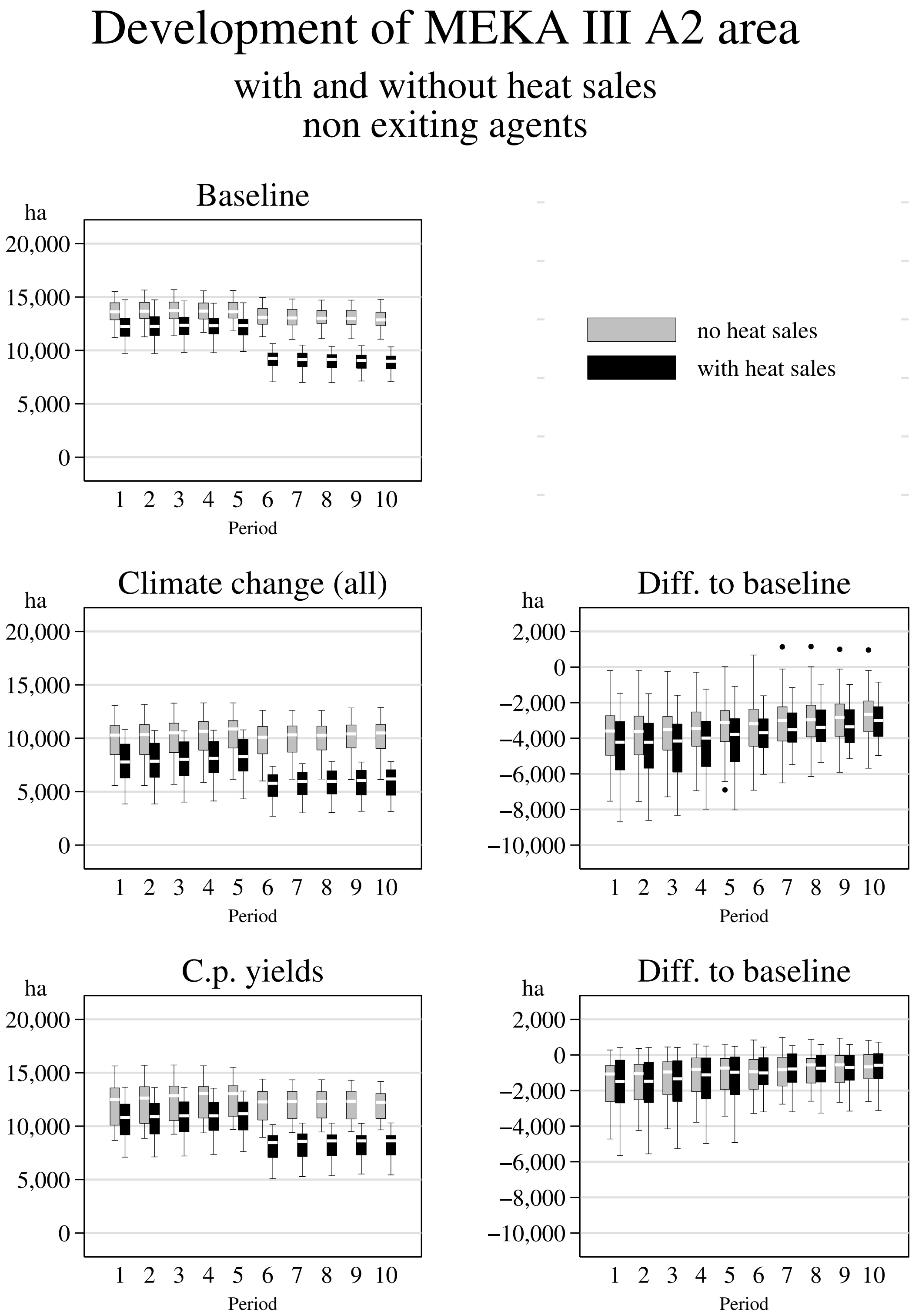

Figure 10.3: *Development of MEKA A2 participation over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

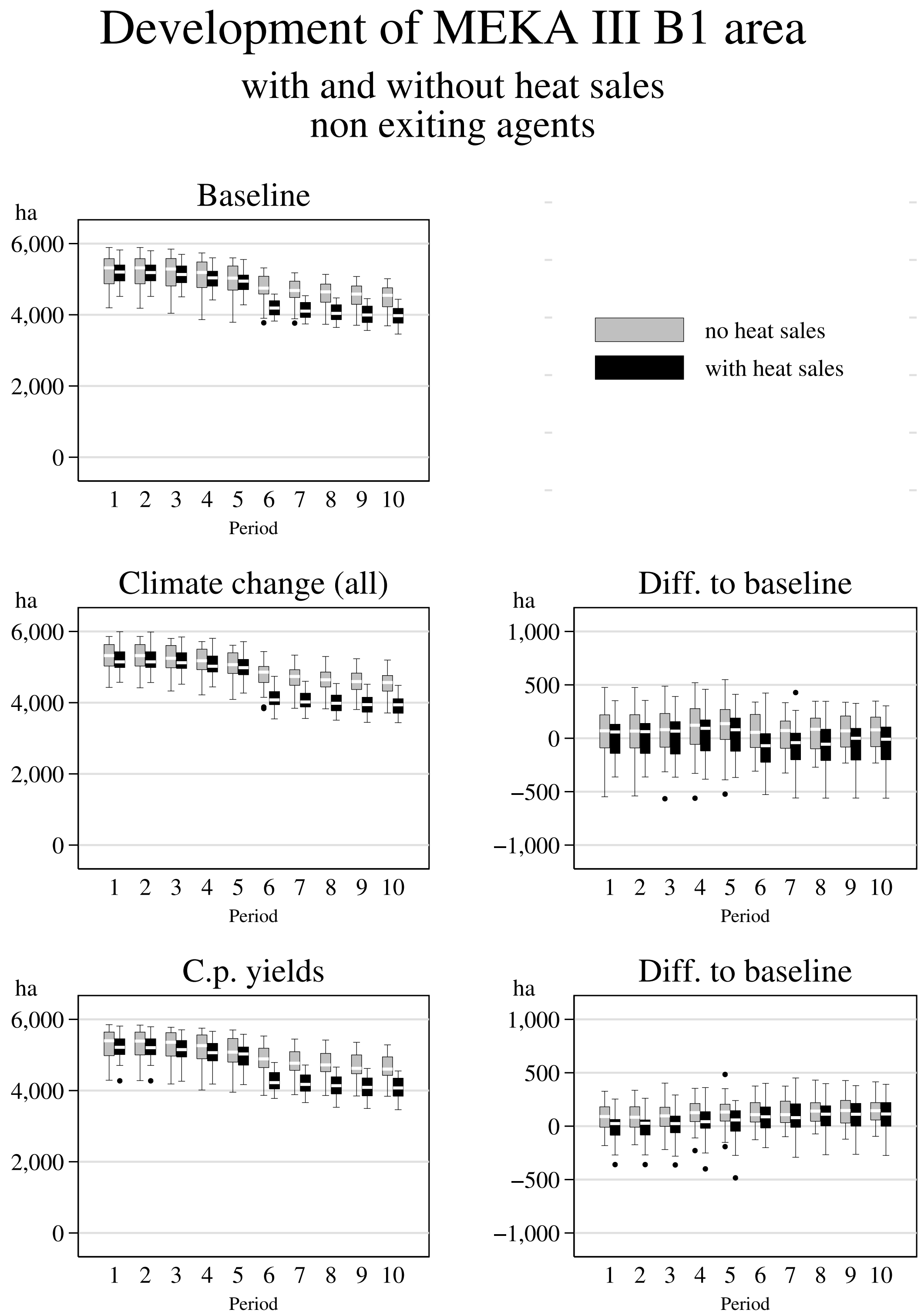

Figure 10.4: *Development of MEKA B1 participation over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

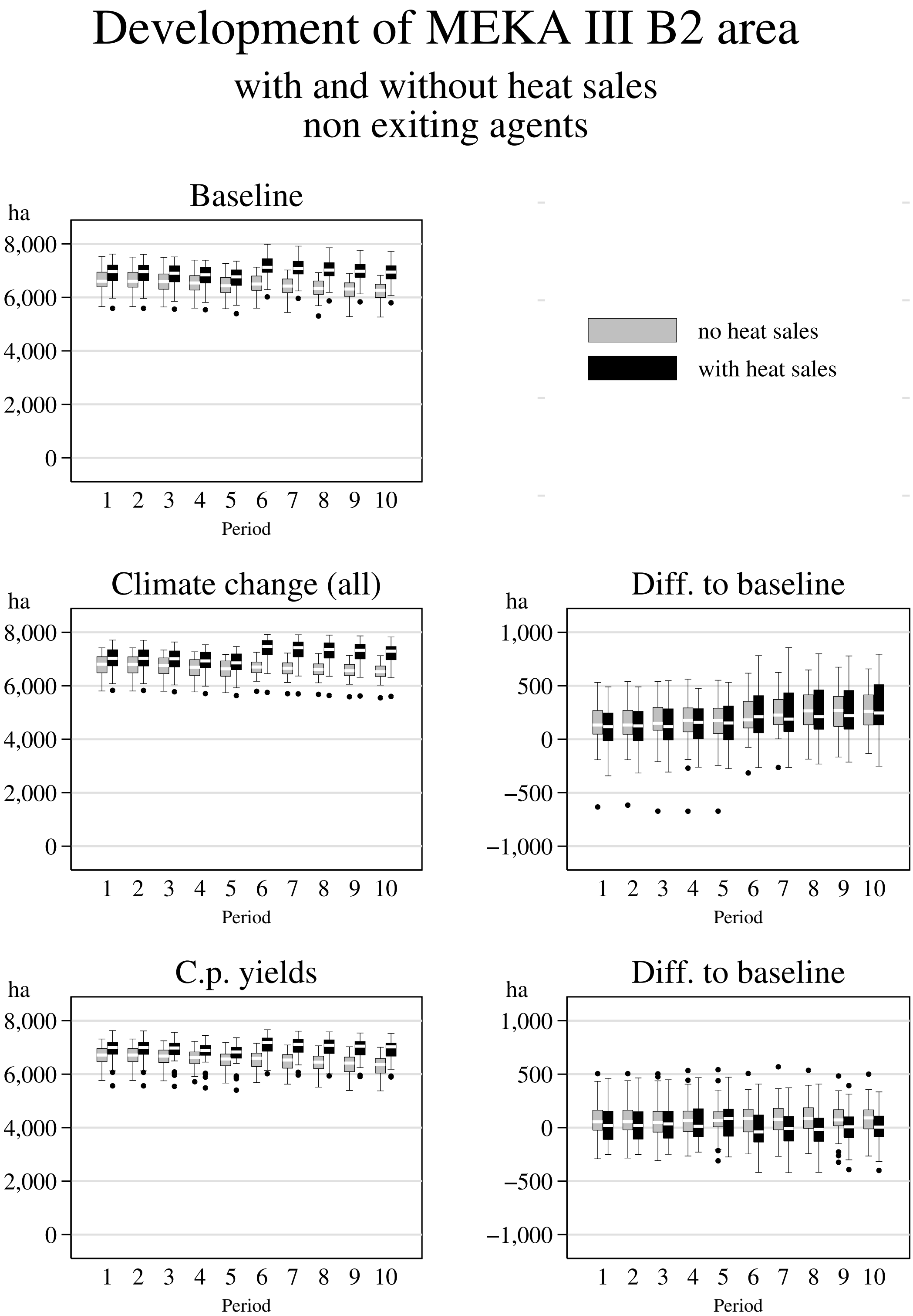

Figure 10.5: *Development of MEKA B2 participation over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

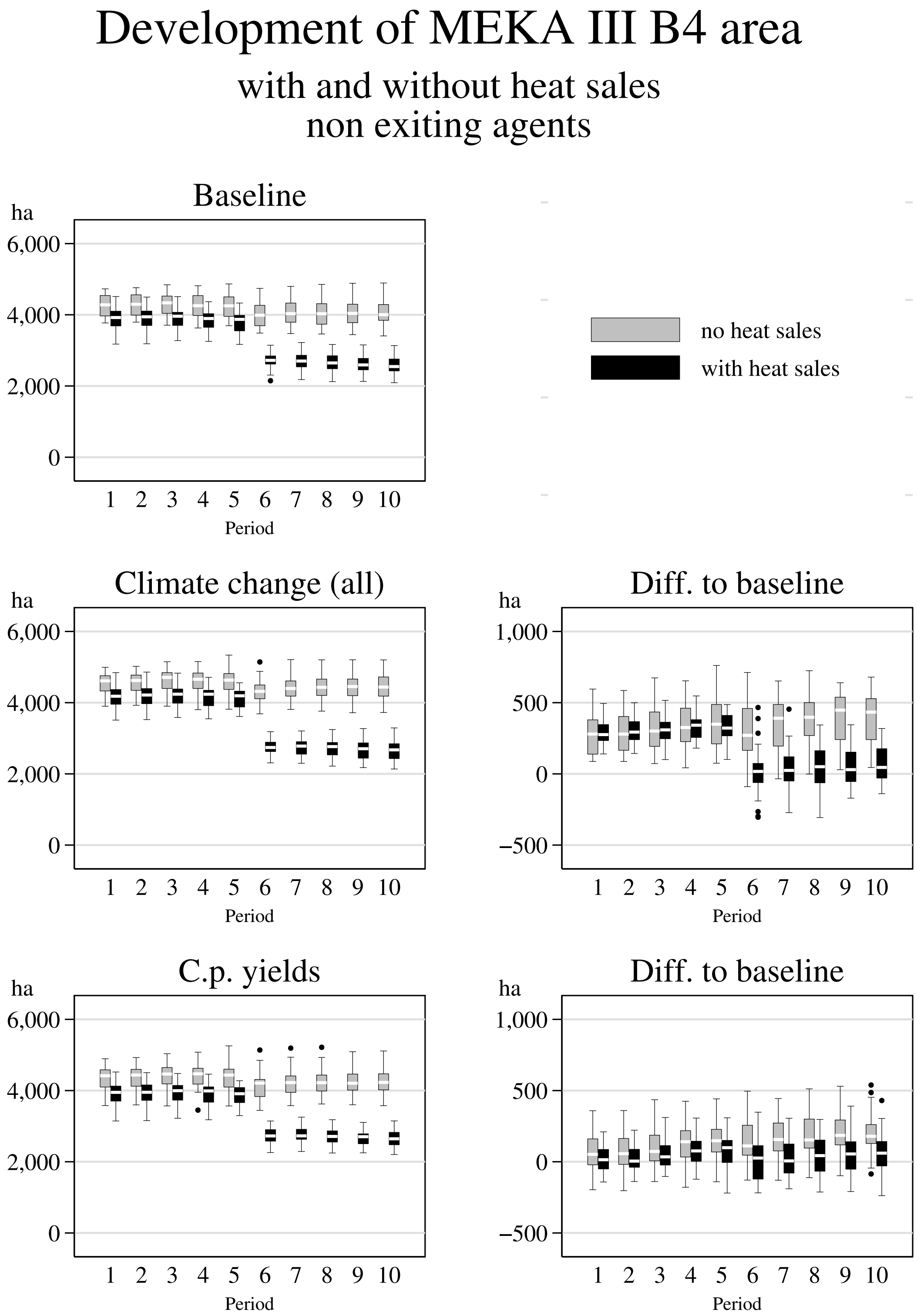

Figure 10.6: *Development of MEKA B4 participation over simulation periodsfor nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

and especially to the shift occurring when the first MEKA commitment period expired after five periods. Interestingly, an increasing effect of climate change on field grass production could only be observed within the first five periods. As a consequence of silage maize and field grass expansion, all other crop areas declined in a similar fashion as shown for winter wheat (fig. 10.9). In the specific case of winter wheat, the increasing effect of climate change predicted in short-term simulations could also be observed in the first periods, but was dampened by silage maize expansion in later periods, when heat sales were assumed.

10.3.4 Livestock

When considering only nonexiting agents, the number of dairy cows remained relatively stable throughout the simulation (fig. 10.10). The effect of climate change was minor and unclear, though probably positive towards the end of simulation time. The number of fattening bulls declined over time (fig. 10.11), while the positive effect of the increased availability of field working days observed in the short-term simulations seems to be confirmed also on the longer term. Due to the restriction by the *pigscen* parameter, fattening pig production remained stable throughout the simulation time (fig. 10.12), when considering only nonexiting agents. Only for some parameter combinations, the lower NPV wedge of 100 Euro in the third *pigscen* setting allowed some agents to invest in pig fattening stalls. Similarly, the number of breeding sows increased with an NPV wedge of 100 Euro and no competition of pig fattening production in the first *pigscen* setting.

10.3.5 Farm types

Figure 10.14 shows the change in the farm type distribution between simulation start and end by indicating the relative difference in the number of agents associated to each principal type of farm (PTOF). The graph shows ambiguous changes for cattle breeders (PTOF 42) and mixed granivore (PTOF 72) agents, while most other types declined, mirroring the overall loss of agents over model run time. Declines were relatively stronger for cereal/oilseed specialists (PTOF 13), mixed cattle (PTOF 43) and field crop-oriented mixed farms (PTOF 60), and less strong for dairy (PTOF 41) and granivore specialists (PTOF 50) as well as ruminant-oriented mixed farms (PTOF 71) and unspecialized farms (PTOF 81, 82). Relative changes for mixed field crop specialists (PTOF 14) and mixed ruminant farms (PTOF 44) can be neglected, since initial agent numbers were very low in these classes and as a consequence small changes showed large impacts on a relative scale. Differences between scenarios were minor.

Tables 10.3-10.5 show the transition matrix between the initial farm type of agents and their farm type at the end of the simulation, expressed as mean and standard deviation over repetitions and distinguished by scenario. Apart from the exit of agents, considerable movement between farm types could be observed. In some cases, such as mixed crop/ruminant farms (PTOF 81) or cattle breeding specialists (PTOF 42) significant additions (partly) compensated for exits and transitions to other types. A large variety of transitions could be observed, although there was a strong tendency to persistence or movement between neighboring classes (e.g. 41 to 42, 43 to 42/71, or 81 to 41/42). Differences between scenarios were minor and mainly associated to a reduction of farm exits in the climate change scenarios.

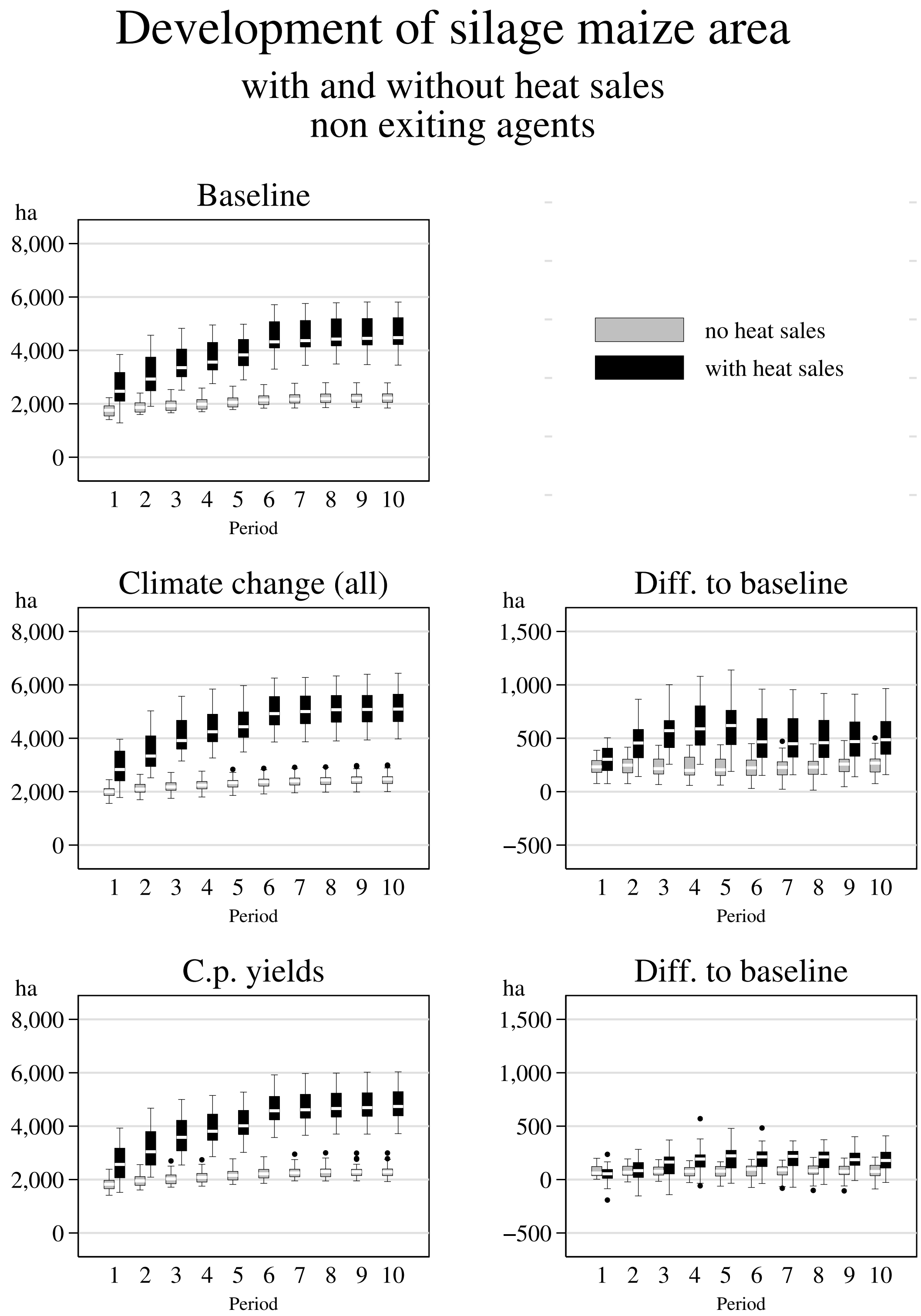

Figure 10.7: *Development of silage maize area over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

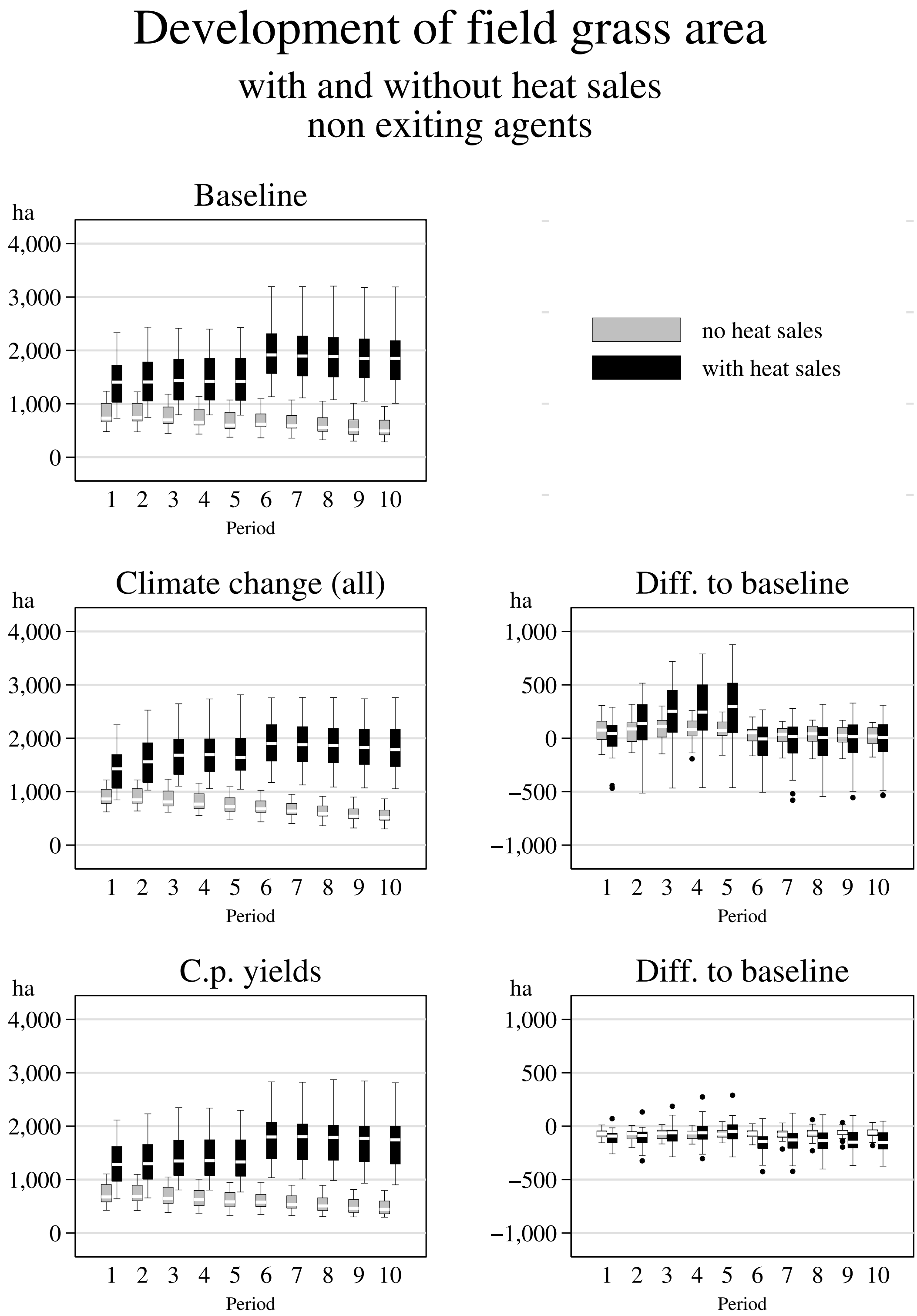

Figure 10.8: *Development of field grass area over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

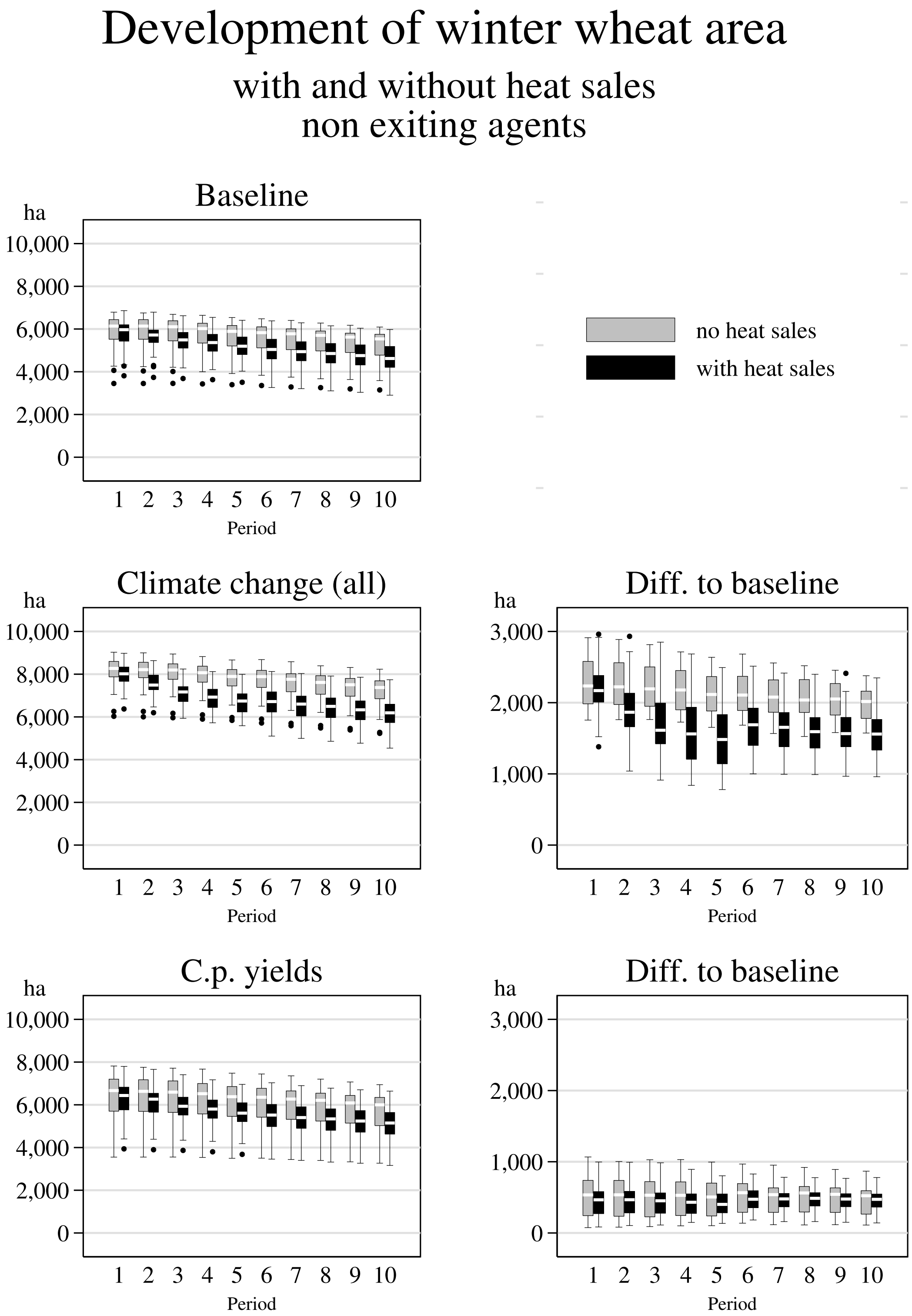

Figure 10.9: *Development of wheat area over simulation periods for nonexiting agents. Box plots illustrate the distribution over repetitions with (black), respectively without heat sales (gray).*

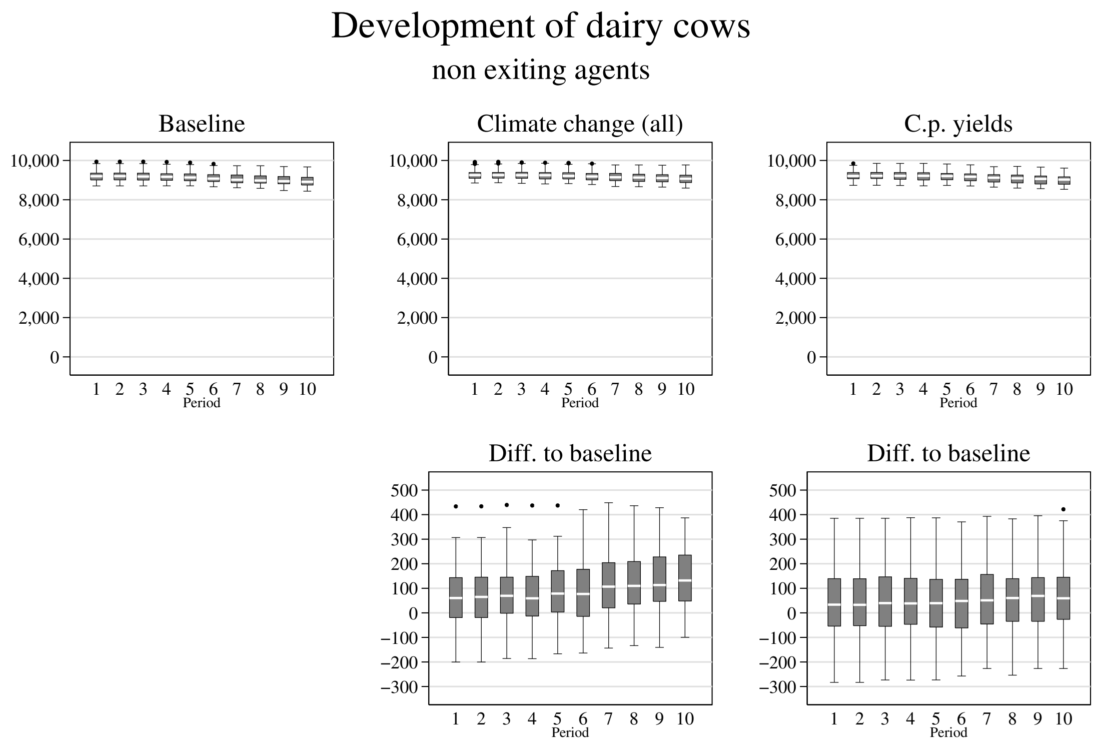

Figure 10.10: *Development of the number of dairy cows over simulation periods for nonexiting agents.*

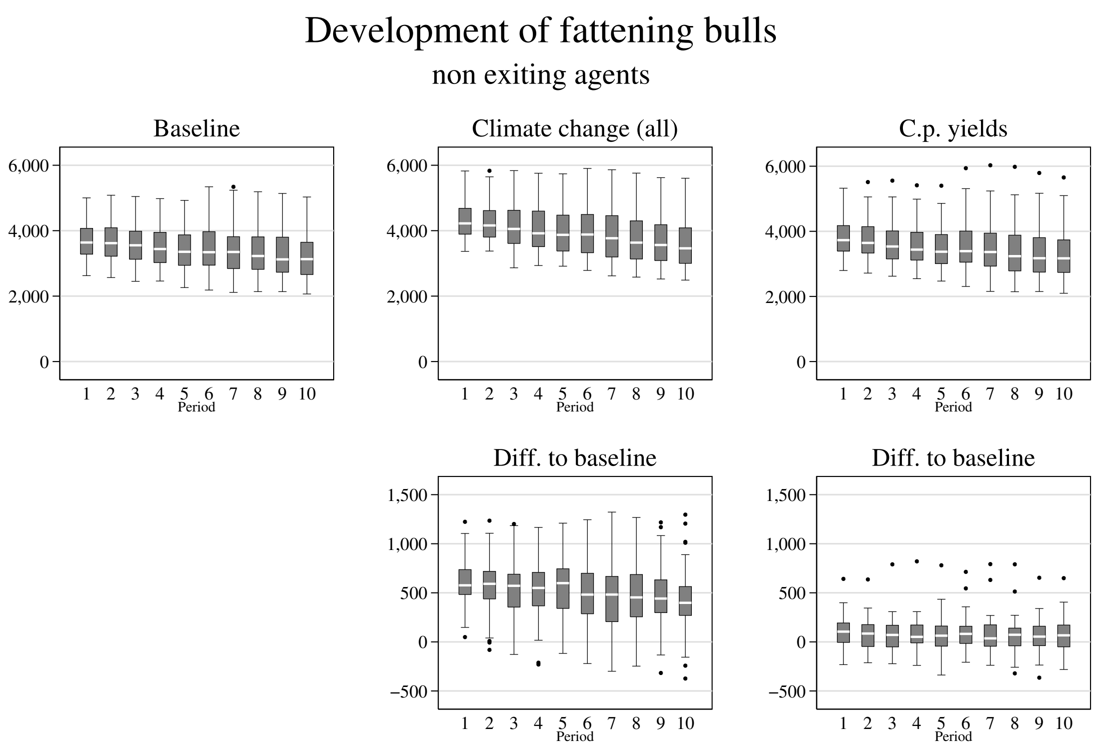

Figure 10.11: *Development of the number of fattening bulls over simulation periods for nonexiting agents.*

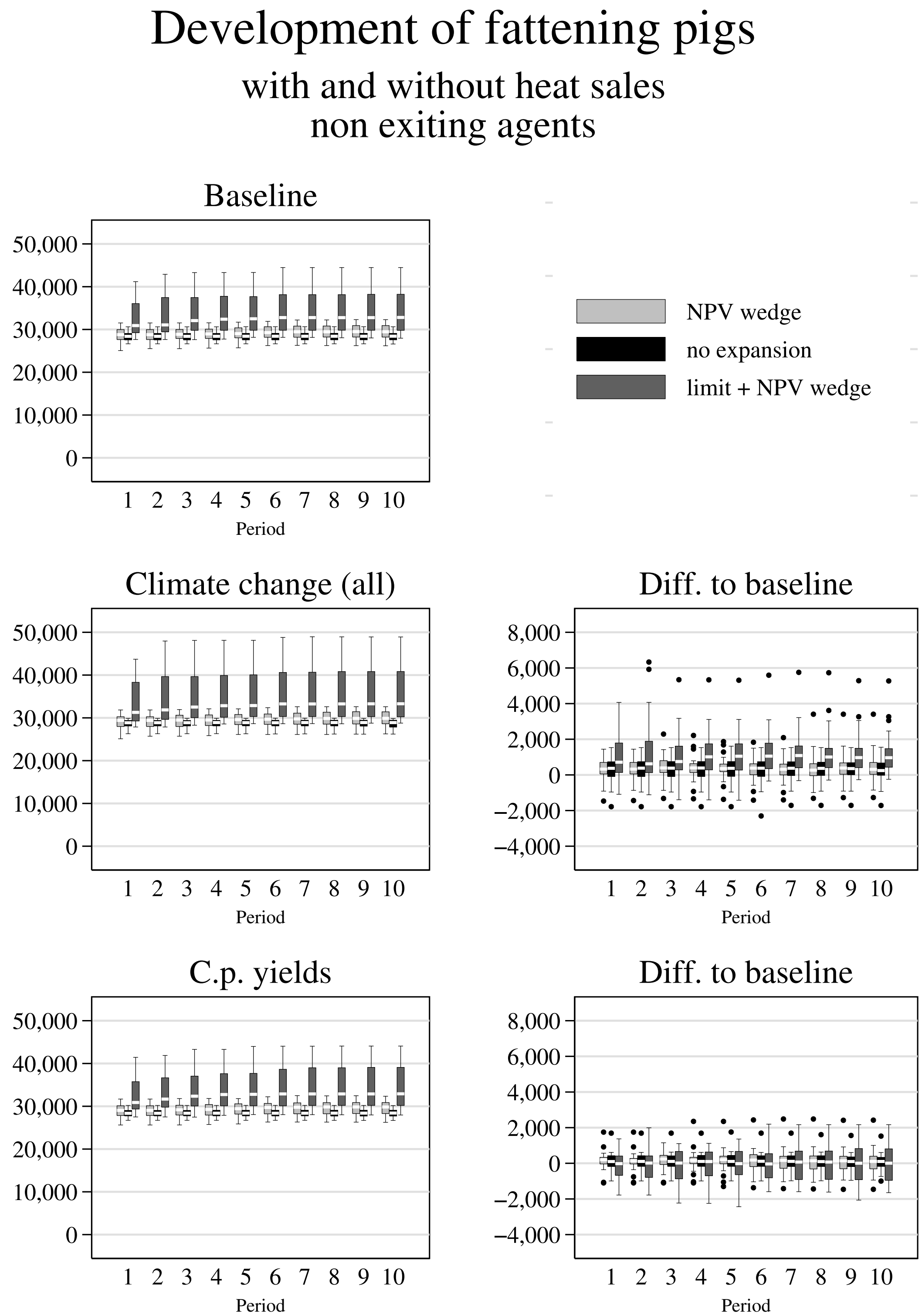

Figure 10.12: *Development of the number of fattening pigs over simulation periods for nonexiting agents depending on parameter pigscen.*

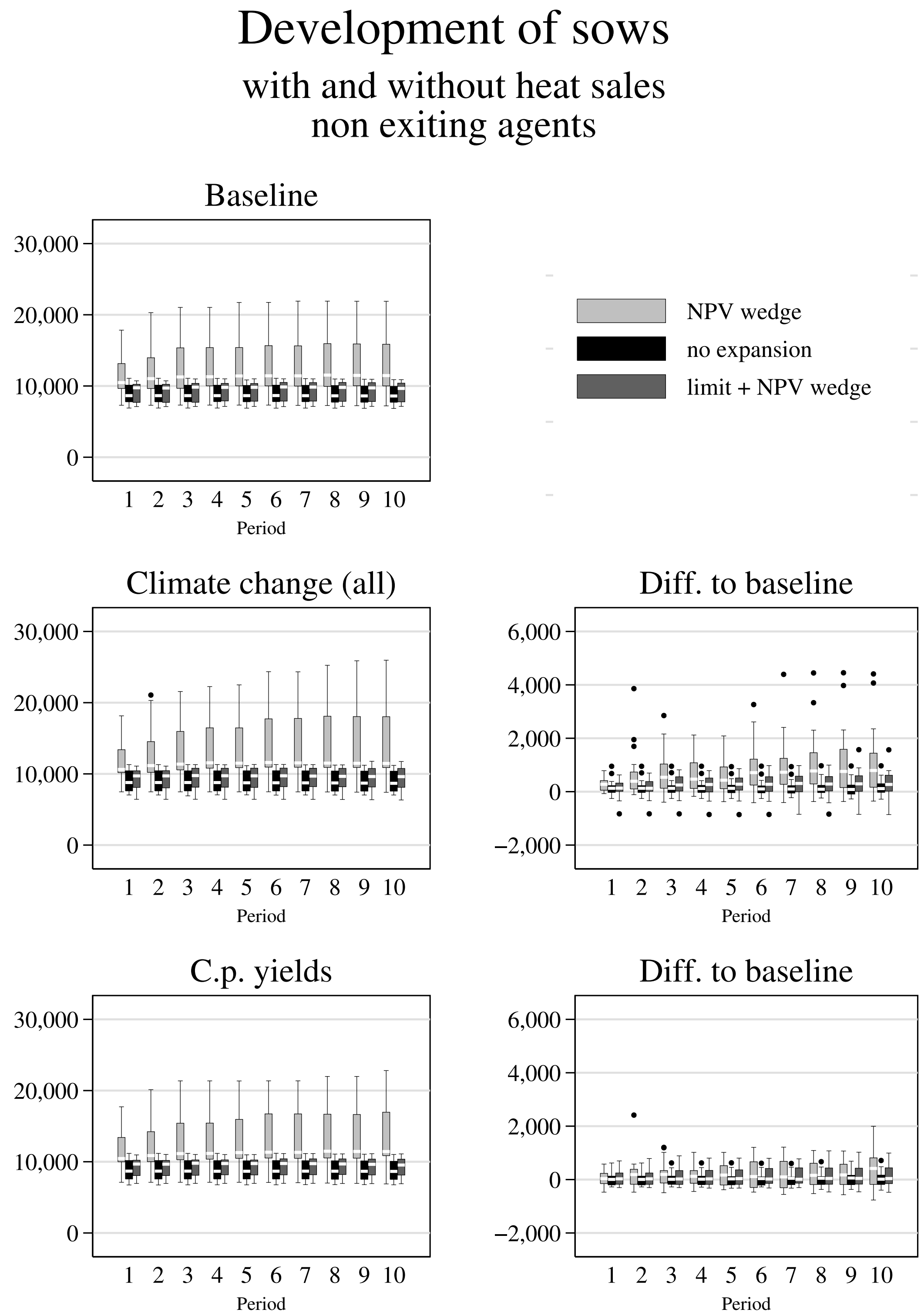

Figure 10.13: *Development of the number of breeding sows over simulation periods for nonexiting agents depending on parameter pigscen.*

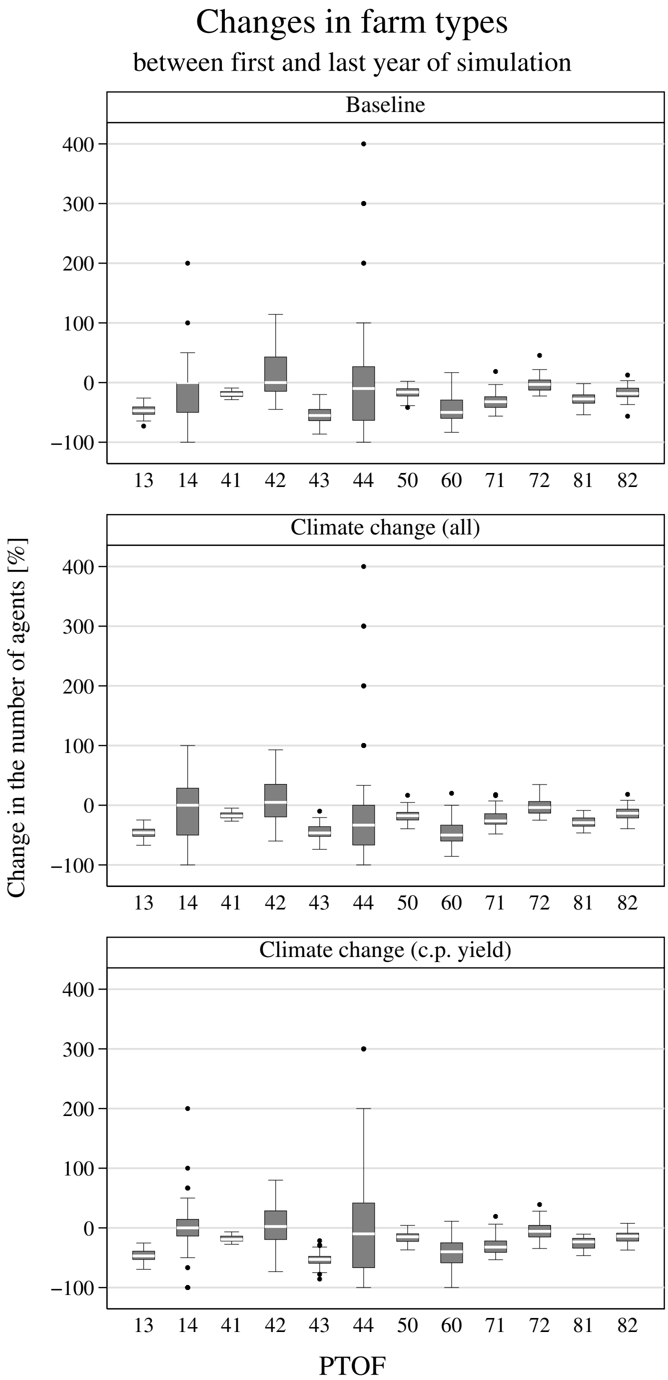

Figure 10.14: *Development of the farm type distribution (PTOF) between start and end of simulation. Box plots over 68 repetitions. (Codes for principal type of farm, PTOF, can be found in tab. 8.11.)*

Table 10.3: *Farm type transition matrix in the baseline. Mean and standard deviation over 68 repetitions. (Codes for principal type of farm, PTOF, can be found in tab. 8.11.)*

Principal type of farm (PTOF)													
	End												
Start	13	14	41	42	43	44	50	60	71	72	81	82	exit
13	**39.4** ± 8	1 ± 1	0 ± 0.2	0.7 ± 0.7		0.4 ± 0.6	0.3 ± 0.5	0.4 ± 0.6	0 ± 0.2	0.2 ± 0.4	3.9 ± 1.8	3.4 ± 1.7	**30.1** ± 8
14	0 ± 0.1	2.6 ± 2.8					0 ± 0.1	0 ± 0.2		0 ± 0.1	0 ± 0.1	0.7 ± 0.8	0.4 ± 0.6
41	0.3 ± 0.5		**142.4** ± 6.4	8.3 ± 3	0.5 ± 0.6	0.7 ± 0.9	0.1 ± 0.3	0.1 ± 0.3	2 ± 1.4	0.7 ± 1	0.9 ± 1	0 ± 0.2	**31.7** ± 6.5
42	0 ± 0.2			2.1 ± 1.8	0 ± 0.2	0.4 ± 0.5			0 ± 0.2		0.3 ± 0.6		**13.7** ± 3.9
43	0.2 ± 0.4		4 ± 2.2	2.5 ± 1.7	8.2 ± 2.9	0.2 ± 0.4		0.1 ± 0.3	2.2 ± 1.1	0.3 ± 0.5	2.3 ± 1.7	0.1 ± 0.3	3.9 ± 2.2
44				0.1 ± 0.3		0.3 ± 0.5					0.1 ± 0.2		2.5 ± 1.2
50		0 ± 0.1				0.1 ± 0.3	**31.1** ± 4.5	0.1 ± 0.3	0 ± 0.2	0.9 ± 0.9	0.1 ± 0.2	2.3 ± 1.6	**10.8** ± 3.8
60	0.2 ± 0.4	0 ± 0.2	0.1 ± 0.3			0 ± 0.1	0.1 ± 0.3	2.7 ± 1.5	0.2 ± 0.5	0.7 ± 0.8	0.4 ± 0.6	1.1 ± 0.7	1.6 ± 1.1
71			2.1 ± 1.9	0.5 ± 0.7	1.3 ± 1.1		0.3 ± 0.5	0.1 ± 0.2	**14.2** ± 3.9	2.8 ± 1.3	1.7 ± 1.4	0.2 ± 0.5	6.8 ± 1.9
72			0.1 ± 0.3			0.2 ± 0.4	2.3 ± 1.7	0 ± 0.2	0.2 ± 0.5	**24.6** ± 5	1.1 ± 0.9	0.6 ± 0.7	5.2 ± 3.1
81	1.7 ± 1.3	0.1 ± 0.3	3.1 ± 3.5	4 ± 2.2	0.9 ± 1	0.3 ± 0.6	0.1 ± 0.3	0.2 ± 0.4	1.7 ± 1.2	0.4 ± 0.6	**33.4** ± 5.6	0 ± 0.1	**15.6** ± 3.9
82	0 ± 0.1	0 ± 0.1					3.7 ± 1.7	0.3 ± 0.5	0 ± 0.3	2.3 ± 1.5	0 ± 0.1	**19.9** ± 2.6	7.7 ± 2.8

Summary

'=' unchanged; '-' to other; '+' from other; 'x' exited

	13	14	41	42	43	44	50	60	71	72	81	82
=	39.4 ± 8	2.6 ± 2.8	142.4 ± 6.4	2.1 ± 1.8	8.2 ± 2.9	0.3 ± 0.5	31.1 ± 4.5	2.7 ± 1.5	14.2 ± 3.9	24.6 ± 5	33.4 ± 5.6	19.9 ± 2.6
-	10 ± 3	1.4 ± 0.6	13.4 ± 3.4	1.5 ± 0.7	11.8 ± 2.9	1 ± 0	3.7 ± 2.2	2.9 ± 1.3	9 ± 2.9	4.6 ± 2.1	12.4 ± 5.9	6.4 ± 2.7
+	2.6 ± 1.5	1.6 ± 0.8	9.5 ± 6.1	16 ± 5.1	2.9 ± 1.5	2.3 ± 1.4	6.9 ± 2.8	1.8 ± 1	6.4 ± 2.4	8.3 ± 2.5	10.8 ± 3.5	8.3 ± 3
x	30.1 ± 8	0.4 ± 0.6	31.7 ± 6.5	13.7 ± 3.9	3.9 ± 2.2	2.5 ± 1.2	10.8 ± 3.8	1.6 ± 1.1	6.8 ± 1.9	5.2 ± 3.1	15.6 ± 3.9	7.7 ± 2.8

Table 10.4: *Farm type transition matrix in the climate change scenario. Mean and standard deviation over 68 repetitions. (Codes for principal type of farm, PTOF, can be found in tab. 8.11.)*

Principal type of farm (PTOF)													
	End												
Start	13	14	41	42	43	44	50	60	71	72	81	82	exit
13	**40.9** ± 8.3	1.9 ± 1.3	0.1 ± 0.3	0.5 ± 0.9		0.4 ± 0.6	0.2 ± 0.6	0.4 ± 0.6	0.1 ± 0.3	0.1 ± 0.3	3.9 ± 2.2	2.4 ± 1.4	**28.1** ± 7.5
14	0.3 ± 0.6	2.3 ± 2.6					0 ± 0.2	0.1 ± 0.2	0 ± 0.1		0 ± 0.2	1 ± 1	0.5 ± 0.5
41	0.1 ± 0.2		**145.9** ± 7	7.1 ± 2.9	0.5 ± 0.7	0.6 ± 0.8	0.1 ± 0.3	0.1 ± 0.3	1.9 ± 1.5	0.7 ± 1.4	0.6 ± 0.7		**30.3** ± 5.8
42	0.1 ± 0.3			2.4 ± 2.2	0.1 ± 0.3	0.2 ± 0.5			0 ± 0.1	0 ± 0.1	0.4 ± 0.7		**13.3** ± 4.1
43	0.1 ± 0.3		3.8 ± 2.2	2.6 ± 2	**12.4** ± 3.2	0.1 ± 0.3		0.1 ± 0.3	2.3 ± 1.5	0.4 ± 0.6	1.4 ± 1.2	0 ± 0.2	3.8 ± 1.8
44	0 ± 0.1			0.1 ± 0.3		0.2 ± 0.5					0.1 ± 0.2		2.4 ± 1.2
50						0.1 ± 0.4	**31.5** ± 4.6	0.1 ± 0.2	0.1 ± 0.3	1.2 ± 1.1	0.1 ± 0.3	2.3 ± 1.6	**10** ± 4
60	0.2 ± 0.4	0.1 ± 0.2	0 ± 0.2	0 ± 0.1	0 ± 0.1		0.1 ± 0.3	2.8 ± 1.5	0.4 ± 0.5	0.5 ± 0.9	0.3 ± 0.5	1.2 ± 0.8	1.4 ± 1.1
71			1.8 ± 1.9	0.7 ± 0.8	1 ± 0.9		0.2 ± 0.5	0 ± 0.2	**15.9** ± 3.3	2.6 ± 1.4	1.1 ± 0.9	0.1 ± 0.4	6.1 ± 1.9
72			0.1 ± 0.3		0 ± 0.2	0.2 ± 0.4	2.1 ± 1.7		0.2 ± 0.5	**24.4** ± 5.3	1.1 ± 1	0.7 ± 1	4.8 ± 2.5
81	1.2 ± 1.2	0 ± 0.2	3.4 ± 3.8	3.9 ± 2.1	1.1 ± 1	0.3 ± 0.7	0.1 ± 0.3	0.2 ± 0.4	1.8 ± 1.2	0.4 ± 0.7	**33.3** ± 5.1	0 ± 0.2	**14** ± 3.6
82			0 ± 0.1				3.2 ± 1.6	0.1 ± 0.2	0.1 ± 0.3	2.1 ± 1.4		**21.1** ± 2.6	7.2 ± 2.7

Summary

'=' unchanged; '-' to other; '+' from other; 'x' exited

	13	14	41	42	43	44	50	60	71	72	81	82
=	40.9 ± 8.3	2.3 ± 2.6	145.9 ± 7	2.4 ± 2.2	12.4 ± 3.2	0.2 ± 0.5	31.5 ± 4.6	2.8 ± 1.5	15.9 ± 3.3	24.4 ± 5.3	33.3 ± 5.1	21.1 ± 2.6
-	9.9 ± 3.3	1.8 ± 0.9	11.6 ± 3.6	1.5 ± 0.8	10.9 ± 2.9	1.1 ± 0.3	4.2 ± 2.2	2.8 ± 1	7.6 ± 2.5	4.4 ± 2.2	12.3 ± 5.7	5.4 ± 2
+	2.4 ± 1.2	2.2 ± 1.3	9.2 ± 6.9	15 ± 4.8	2.9 ± 1.5	2.1 ± 1.2	6 ± 2.9	1.5 ± 0.8	6.9 ± 2.8	8 ± 2.8	9 ± 3.5	7.8 ± 3.1
x	28.1 ± 7.5	0.5 ± 0.5	30.3 ± 5.8	13.3 ± 4.1	3.8 ± 1.8	2.4 ± 1.2	10 ± 4	1.4 ± 1.1	6.1 ± 1.9	4.8 ± 2.5	14 ± 3.6	7.2 ± 2.7

Table 10.5: *Farm type transition matrix in the scenarios with* ceteris paribus *yield effects. Mean and standard deviation over 68 repetitions. (Codes for principal type of farm, PTOF, can be found in tab. 8.11.)*

Principal type of farm (PTOF)													
	End												
Start	13	14	41	42	43	44	50	60	71	72	81	82	exit
13	**41.3** ± 8.6	1.5 ± 0.9	0 ± 0.2	0.6 ± 0.9		0.4 ± 0.7	0.2 ± 0.5	0.5 ± 0.6	0 ± 0.2	0.1 ± 0.4	3.7 ± 1.7	2.6 ± 1.5	**28.7** ± 7.3
14	0 ± 0.2	2 ± 2.8						0 ± 0.2			0 ± 0.2	0.8 ± 0.8	0.5 ± 0.6
41	0.2 ± 0.5		**143.6** ± 6.3	7.2 ± 2.7	0.4 ± 0.6	0.8 ± 0.9	0.2 ± 0.4	0.1 ± 0.3	2.1 ± 1.5	0.7 ± 1.1	1.2 ± 0.9	0.1 ± 0.3	**30.9** ± 6.1
42	0.1 ± 0.3		0 ± 0.1	2.1 ± 1.7	0.1 ± 0.3	0.4 ± 0.6			0 ± 0.2		0.4 ± 0.6		**13.5** ± 4
43	0.1 ± 0.4		4 ± 2.1	2.4 ± 1.7	8.9 ± 3.1	0.1 ± 0.3	0 ± 0.1	0.1 ± 0.3	2.1 ± 1.1	0.2 ± 0.4	2.6 ± 2	0 ± 0.2	3.7 ± 2.1
44				0.1 ± 0.2		0.2 ± 0.4					0 ± 0.2		2.5 ± 1.2
50		0 ± 0.1				0.1 ± 0.3	**30.9** ± 4.6	0.1 ± 0.3	0 ± 0.2	1 ± 1	0.1 ± 0.3	2.3 ± 1.5	**10.4** ± 3.8
60	0.1 ± 0.3	0 ± 0.2	0.1 ± 0.2	0 ± 0.1			0.1 ± 0.4	2.8 ± 1.4	0.3 ± 0.5	0.6 ± 0.9	0.4 ± 0.6	1 ± 0.8	1.4 ± 0.9
71			2.3 ± 1.9	0.6 ± 0.7	1.1 ± 1.1		0.3 ± 0.5	0.1 ± 0.2	**15** ± 3.8	2.6 ± 1.5	1.7 ± 1.4	0.3 ± 0.6	7 ± 1.9
72			0.1 ± 0.3			0.3 ± 0.4	2.4 ± 1.8	0.1 ± 0.3	0.2 ± 0.6	**24.7** ± 5.2	1 ± 0.9	0.7 ± 0.9	5.1 ± 2.8
81	1.4 ± 1.2	0.1 ± 0.3	3.4 ± 3.8	3.9 ± 2	1 ± 1	0.3 ± 0.5	0.1 ± 0.3	0.2 ± 0.5	1.8 ± 1.3	0.4 ± 0.6	**33.5** ± 5.8		**14.9** ± 4
82		0 ± 0.1				0 ± 0.1	3.5 ± 1.6	0.1 ± 0.3		2.1 ± 1.7	0 ± 0.2	**21.2** ± 2.8	7.4 ± 2.6

Summary

'=' unchanged; '-' to other; '+' from other; 'x' exited

	13	14	41	42	43	44	50	60	71	72	81	82
=	41.3 ± 8.6	2 ± 2.8	143.6 ± 6.3	2.1 ± 1.7	8.9 ± 3.1	0.2 ± 0.4	30.9 ± 4.6	2.8 ± 1.4	15 ± 3.8	24.7 ± 5.2	33.5 ± 5.8	21.2 ± 2.8
-	9.6 ± 3.1	1.5 ± 0.7	12.8 ± 3.6	1.7 ± 1.3	11.8 ± 3.1	1 ± 0	4 ± 2	2.6 ± 1.2	8.9 ± 2.8	4.8 ± 2.4	12.5 ± 6.2	5.7 ± 2.1
+	2.3 ± 1.3	1.8 ± 1	9.9 ± 6.6	14.7 ± 5.1	2.8 ± 1.4	2.3 ± 1.3	6.8 ± 2.8	1.7 ± 0.9	6.5 ± 2.7	7.7 ± 2.9	11.3 ± 3.7	7.9 ± 3
x	28.7 ± 7.3	0.5 ± 0.6	30.9 ± 6.1	13.5 ± 4	3.7 ± 2.1	2.5 ± 1.2	10.4 ± 3.8	1.4 ± 0.9	7 ± 1.9	5.1 ± 2.8	14.9 ± 4	7.4 ± 2.6

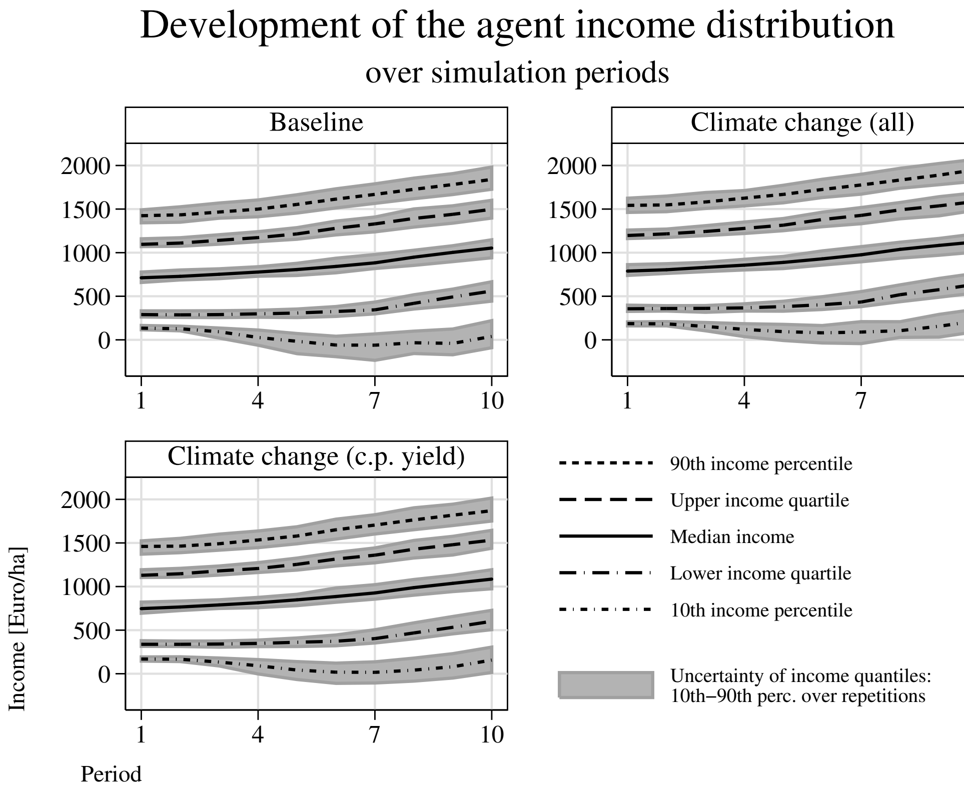

Figure 10.15: *Development of the per-ha income distribution among all agents over simulation periods. Median (lines) and 10th and 90th percentile (shaded area) of income quantiles over 68 repetitions.*

10.3.6 Income

Figure 10.15 shows the development of the income distribution over the course of the simulation, illustrated by the development of income quartiles and first and last deciles. Except for the lowest decile, which declined initially , a largely parallel development of rising quantiles can be observed. Thus, at first glance, the graph seems to indicate an overall income increase affecting at least the upper three income quartiles rather equally. It should not be forgotten, however, that agent bankruptcies and failed successions affect lower-income agents overproportionally: Agent exits can therefore be expected to lead to an increase of income quantiles even if incomes of remaining agents do not increase at all.

To control for this effect, figure 10.16 shows the development of income quantiles for nonexiting agents only: Again, except for the lowest decile, a common increase could be observed. Absolute increases tended to be higher for higher quantiles (P25: +42 to +261 Euro; P75: +236 to +564 Euro in the baseline), while relative increases tended to be higher for lower quantiles, which, however, also showed a higher variance over repetitions than higher quantiles (P25: +10 to +57%; P75: +20 to +39%). The lowest decile was stable on average, although both increases of up to 160 Euro and decreases of up to 250 Euro were observed in some repetitions and scenarios.

Quantiles only show the development of the overall income distribution, which results from the aggregate perspective on an individually much more heterogeneous income development: An increase in, for example, the lower quartile does not mean the agents that

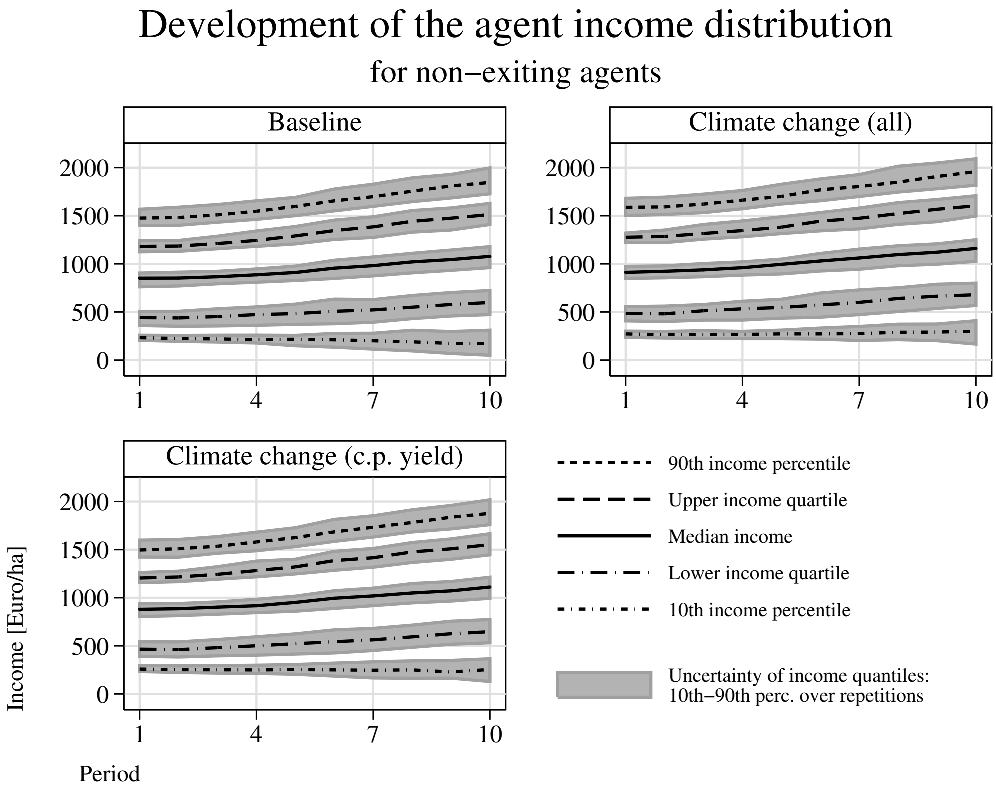

Figure 10.16: *Development of the per-ha income distribution among nonexiting agents over simulation periods. Median (lines) and 10th and 90th percentile (shaded area) of income quantiles over 68 repetitions.*

represented the lower quartile in the beginning increased their income. It may as well have been stable or declining, while agents with initially lower incomes overtook them. Depending on the repetition, only between 45-65% of the non-exiting agents showed a change of less than $\pm 5\%$ in their per-ha income rank relative to the other non-exiting agents. The average income difference between first and last simulation period over all nonexiting agents and all repetitions was 172 Euro/ha ± 347 in the baseline, 185 Euro/ha ± 336 in the climate change, and 183 Euro/ha ± 345 in the *ceteris paribus* yields scenario. Individual agents, however, had positive and negative differences of up to 2,500 Euro/ha in individual repetitions and scenarios.

Table 10.6 identifies the agent characteristics that determine the difference between an agent's per-ha income in the first and the last period of the simulations based on an OLS regression over all nonexiting agents and all repetitions of the experimental design. Strong income differences were mainly associated with the expansion of dairy cow, fattening pig and biogas production. Agents with a high grassland share and high income at simulation start tended to show lower or more negative income differences, while agents with a higher livestock intensity at simulation start showed more positive income differences. Larger farm size lead to a higher income difference, but this effect was mainly observable for smaller farms and less important among large farms. (This is why the reciprocal of farm size was used in the regression.) As a special case, farm size reductions resulting from returns of leased lands to owners in case of insolvency had an nonnegligible positive effect on per-ha income development, since they reduced the total farm area relatively more than the income.

Apart from raising the overall income difference as evidenced by the higher constant term, the climate change scenarios made the grassland share more important, which is clearly a consequence of the omission of climate effects on grass yields compared to mostly positive impacts on crop production.

10.3.7 Land markets

In reality only a part of the land of closed farms is put to nonagricultural uses, while the rest is rented by farms that want to expand their area. The final simulations described in this thesis were designed to model this process using an adaptation of the MPMAS land rental market.

Based on the observation that German full-time farmers tend to rent out all or most of their land, or even their whole farm, when they are deciding to exit the farming business, but hardly rent out small portions of land, the simulations considered only land offers by the 'other land owner' agent, which 'administrates' land of exited agents (cf. chapter 7). Agents could expand their farm area by renting land, but could not offer land themselves. The overall maximum distance for considering renting a cell was set to 20 km, while the *markdown* parameter was set to 0.7. Land market simulations were run for all three scenarios and each of the 17 repetitions of the first subarray of the UPCS used in the recursive-dynamic simulations.

Due to the necessity to calculate a shadow price for every agent and each of the ten soil types, the investment decision of every agent has to be evaluated at least 11 times in addition to the normal investment and production decision plus an update for both owner and buyer after a successful contract has been established. These additional evaluations boost computation times considerably and unfortunately meant that none of the 51 runs finished a full ten-year cycle within 3 days, the time limit set for the longest standard queue used on bwGriD. One repetition did not even finish the third period.

Comparing the first five periods over the 15 repetitions that reached the fifth period to the simulations without land markets, a reduction of land loss from 9.9% to 5.5% can be observed for the baseline. Also for the climate change scenarios land loss is approximately halved. Given the incomplete results, further evaluation of simulation results must, however, wait until a technical or methodological solution allows a complete and reliable evaluation of land markets and is left for future research.

Table 10.6: *Determinants of the income difference between first and last period. OLS regressions over all nonexiting agents and repetitions distinguished by scenario (SRC: standardized regression coefficient)*

Baseline				
	coef.	s.e.		SRC
Sh. of grassland	-0.07	0.04		-0.005
Income/ha t_1	-0.06	0.01	***	-0.077
1 / farm size t_0	-3442.58	134.96	***	-0.096
Δ *farm size*	4.81	0.06	***	0.254
Biogas cap./ha t_1	15.78	4.69	***	0.009
Δ *biogas cap/ha.*	311.20	4.68	***	0.187
Fat. pigs/ha t_1	23.54	0.59	***	0.182
Δ *fat. pigs/ha*	126.70	2.74	***	0.124
Dairy cows/ha t_1	376.42	7.78	***	0.333
Δ *dairy cows/ha*	2201.13	9.59	***	0.703
constant	181.51	4.37	***	
Dep.var.: Δ *income/ha*	$R = 0.801$			

Climate change (all)				
	coef.	s.e.		SRC
Sh. of grassland	-0.16	0.04	***	-0.013
Income/ha t_1	-0.07	0.01	***	-0.096
1 / farm size t_0	-3190.27	133.72	***	-0.094
Δ *farm size*	5.76	0.07	***	0.273
Biogas cap./ha t_1	15.62	4.63	***	0.010
Δ *biogas cap./ha*	320.92	4.46	***	0.212
Fat. pigs/ha t_1	23.77	0.59	***	0.190
Δ *fat. pigs/ha*	134.32	2.47	***	0.152
Dairy cows/ha t_1	351.62	8.09	***	0.321
Δ *dairy cows/ha*	2232.14	10.37	***	0.681
constant	202.24	4.54	***	
Dep.var.: Δ *income/ha*	$R = 0.785$			

C. p. yields				
	coef.	s.e.		SRC
Sh. of grassland	-0.12	0.04	***	-0.009
Income/ha t_1	-0.06	0.01	***	-0.078
1 / farm size t_0	-3738.91	135.58	***	-0.106
Δ *farm size/ha*	5.00	0.06	***	0.251
Biogas cap./ha t_1	16.04	4.77	***	0.010
Δ *biogas cap./ha*	324.73	4.64	***	0.201
Fat. pigs/ha t_1	24.05	0.59	***	0.186
Δ *fat. pigs/ha*	128.84	2.77	***	0.128
Dairy cows/ha t_1	383.14	7.69	***	0.340
Δ *dairy cows/ha*	2192.03	9.89	***	0.692
constant	190.37	4.49	***	
Dep.var.: Δ *income/ha*	$R = 0.797$			

Chapter 11

Discussion

The previous three chapters described simulations that approached the task of assessing climate change adaptation and policy analysis in the Central Swabian Jura from different angles including one-period and recursive-dynamic simulations, contrasting climate and policy scenarios, and observing price responses over a full domain of potential price developments. In all this, techniques and technical infrastructure identified or developed in the first part of this thesis were employed. The subsequent discussion will hence not only focus on the results of climate change and policy analysis, but also recapitulate on the experience that could be gained with the use of these approaches in a practical application.

11.1 The model

Considerable effort went into the construction of a model that is capable of exploring the impact of climate change on farm production decisions in the study area. Information was gathered from a large variety of secondary datasets and complemented with area-specific information from expert interviews and a farm survey. The farm decision model constructed for the analysis shows a size, comprehensiveness and complexity that is unprecedented for MPMAS models and to my knowledge for agricultural economic agent-based models in general. The decision problem faced by German farmers today is quite complex and the model still reflects only a part of it.

11.1.1 Comprehensiveness and complexity

Comprehensiveness was prescribed by the objective to build a generic model that encompasses all individual situations any farm in a heterogeneous population might find itself in now and in the future and allows farm agents to – gradually – evolve over time without unduly restricting their path of development to a fixed composition of labor force, machinery, stables and farm size.

To create this flexibility, many restricting capacities that usually would be exogenously defined in representative farm modeling are endogenously calculated in the model. For example, while often crop production activities are defined in terms of a pre-defined mechanization level (see e.g Berger 2001, Troost 2009, Freeman et al. 2009, Freeman 2005, Lobianco and Esposti 2010), the model used in this study treats the choice of machinery separately from crop management choice and allows also the use of different types of machinery for different works of the same production activity. Only due to this separation, the model is able to capture options to react to climate-induced changes of the field work capacity of the farm by investing in machinery.

The complexity of the model is a – somewhat inevitable – consequence of this separation as well as of the ambition to model economies of scale and hysteresis, to capture the set of incentives and restrictions defined by agricultural policy, and a sophisticated crop rotation system that does not only consider upper limits, but also the compatibility between preceding and following crops and thereby allows to model the impact of the additional rotation option farmers expect for the future. On top of the inherent complexity, discretization triggered by the need to formulate the model as a mixed integer programming problem further increased the size of the problem. Still, given the large number of simulations, increasing the problem size to be able to use a mixed integer linear solver seemed to be the better choice compared with the lower efficiency of solvers capable of solving problems with nonlinearity in constraints (though this was not empirically tested here).

The high level of complexity certainly does not make the model straightforward to understand, but an increased effort has been made to make the modeling process transparent and understandable: An extensive documentation of the model and especially all the equations of the decision module following the ODD protocol is provided in chapter 7. The calibration procedure as well as the construction of the validation and calibration dataset is comprehensively documented in chapter 8. Consequences of parameter uncertainty are consistently highlighted throughout the result analysis in chapters 9 and 10.

11.1.2 Achievements and limitations

The simulation results showed that the model succeeds in capturing a considerable degree of the heterogeneity of study area farms and their behavior: First, the validation experiments (section 8.5) showed that the model is able to reproduce the distribution of farm types pretty well. To a certain extent, this has to be attributed to the high quality of the statistical estimates and data used for the initialization of the farming population, but these alone would not produce a good fit if the model was not able to translate heterogeneity in starting assets into behavioral heterogeneity.

A second indication is the smoothness of the supply response curves in section 9.6. Albeit subject to model uncertainty, the aggregate land use, animal stock and policy participation curves showed gradual, partly nonlinear price responses and confirmed the postulate that the problems of overspecialization and jumpy behavior exhibited by linear programming models (cf. section 4.2.1) can be smoothened out by incorporating a high degree of heterogeneity into an agent-based model.

The example of the price response curve of piglet raising – also in section 9.6 – highlights that the degree of smoothness is much lower for production activities that are largely isolated from the rest of the production setup and thus much more governed by processes not incorporated into the model (e.g. vertical integration, fragmentation of markets, farmer experience and value) than by the specific agro-economic setup of the farm. To a somewhat lower degree, this also holds for piglet production and fattening pigs, although the integration with other production activities is higher here due to the larger amount of manure and the importance of on-farm production of feed. The challenge of modeling pig production is also reflected in the recursive-dynamic simulations (chapter 10) when investments in stable places were allowed and the gaps in the modeling of restrictions to pig production required the inclusion of – empirically justified, but theoretically somewhat unsatisfactory – restrictions in order to avoid unrealistic model behavior. Here, a more in-depth analysis of pig production and investment behavior is necessary as one of the the next steps of model development.

The case of biogas investments provides an especially clear example of the benefits of recursive-dynamic simulations: The potential effect of climate change is not detectable in one-period simulations if realistic liquidity restrictions are applied (cf. section 9.4). If instead the potential capacity is estimated abstracting from such restrictions as in section 9.7, the biogas investment is far greater than the one that is ultimately achieved in recursive-dynamic simulations that do consider realistic restrictions (cf. section 10.3.2).

On a more general level, a more detailed and disaggregated analysis of one process raises the demand for a corresponding level of detail representations of other processes: The model currently pairs a sophisticated description of the relationship of field work capacity to machinery ownership, labor availability and weather conditions with a rather *ad hoc* representation of the possibility to hire field work services. Since also the comparison with FADN data suggest a general underestimation of labor cost (section 8.5), the whole topic of labor and service provision and hiring needs to receive specific scrutiny in order to improve the reliability of simulations and is probably best dealt with by making it a topic in interactive modeling and validation sessions with farmers.

The provision and hiring of field work services is also one example of a longer list of topics concerning inter-farm cooperation and trade that could only receive rudimentary representation in the current model version: Trade of silage and manure (especially in connection with biogas production), trade of young animals, demand for excess heat, and vertical integration (especially, with respect to meat production) will most probably play important roles in shaping the future structural development of the agricultural sector. Research into these types of agent-agent interactions needs to be a major focus of future research.

The simulation of land markets is still hampered by computational limits that can probably be overcome by an internal parallelization of MPMAS. So far, the use of the parallel computing facilities has been limited to trivial parallelization, i.e. running many repetitions in parallel, but each using one processor only. To fully exploit the advantages of distributed computing, agent decisions of the same run could be evaluated in parallel on several processors shortening the duration of a single run and allowing to stay within the time limits set by the computing facilities. Being able to store the current model state, interrupt it and continue it at a later point of time would be an alternative or additional option.

Apart from technical solutions, a methodological revision of the land market could contribute to solving efficiency, but also make it empirically more accurate: Trade of land hectare-by-hectare takes a lot of computation time and is also unrealistic as plots are usually based on natural limits, comprise larger units, and owners will try to interact with a few counterparts only instead of having to deal with a separate one for each hectare. Basing land trade on cell bundles instead of individual cells would be a step in the right direction, but would require a conceptual revision of the current land market implementation.

With the stage of validation and the thorough and detailed description of model behavior achieved in this thesis, the model provides the opportunity to test improved or newly developed representations for these and other processes, e.g. risk and learning processes, against an empirically tested baseline and well explored model behavior. In this way, the effort necessary to construct, parameterize and validate the model is counterbalanced by the possibility to use it as the basis for future theory development and many types of agricultural and environmental policy analysis. The comprehensiveness of the model facilitates its use for various purposes besides analyzing climate change adaptation as the applications for policy analysis have demonstrated. Each application will probably require a few enhancements or adaptations of the model setup, but that also means that knowledge about the model can be improved and more experience on the strength and weakness of different process representations is gained with every application. Continuous maintenance of the model will, however, be necessary in order to keep it up to date and a suitable basis for further research. The fact that the model has been built around a normalized database will hopefully help in its long-term maintenance.

11.2 Calibration, validation and uncertainty analysis

The empirical application in the Central Swabian Jura showed that the suggestions developed in chapter 4 present a viable way to deal with the challenges of model uncertainty in agro-economic agent-based models. The high degree of detail in the model brought along a high degree of parameter uncertainty. Despite a comparatively good and comprehensive database, conditions for empirical parameter estimation were not ideal: On the one hand, datasets for initialization and validation were available only at a relatively coarse resolution and did not contain all information necessary to parameterize the model. On the other hand, unobservables and a number of omissions (e.g. risk, local markets) led to an unknown, but probably biased, nonnormal error distribution.

In this situation, a conservative calibration criterion was chosen in order to avoid overfitting the model, and as a consequence, the number of parameters that could be fixed is relatively modest. The result obtained in the calibration procedure is intuitively appealing: The calibrated model assumes zero supply of young female cattle, and zero demand for silage maize and contracted energy crops. In reality, there certainly is some trade in these items, but it is not (yet) widespread and assuming unlimited demand, resp. supply will certainly lead to a much greater error than assuming no trade at all. For the future, this may change especially with respect to silage markets and this simplification needs to be kept in mind and addressed by further model development, as already mentioned above. Likewise, decreasing the productivity of pasturing compared to theoretical optima is certainly in line with empirical observations showing little evidence of pastured cattle except for extensive production activities. In this case, the process representation could certainly be improved by a differentiation between different types of grassland and plot distance to the farm. Calibration also suggests using a high probability of occurrence when calculating the days with suitable weather for field work. Though also experience suggests that farmers tend to be risk averse and plan conservatively, this later calibration result might also be another hint that other constraints relating to labor use are not tight enough.

The model predictions held up well to expert scrutiny in a Turing test and the calibrated model reproduced observed land use data with a satisfactory level of accuracy. The observed biases were consistent with model simplifications, with the possible exception of labor cost and use. It has to be noted that the comparison with state averages of FADN data was of limited use, and one next step in model improvement could be to use the individual accounts. The example of the FDZ data shows that anonymization requirements need not preclude a fruitful use.

Further validation can be achieved by implementing a model version for Kraichgau, the second study area of the FOR 1695 project. The model is sufficiently general to be applied to other study areas with modest modifications: Certain parameters are area-specific (e.g. crop yields, management plans, crop rotations coefficients, and soil types) and crops not grown and machinery not used in the first study area will have to be added, but the general structure of the model will remain the same and even a lot of data and parameters (prices, labor demands, technical coefficients of machinery) are not region specific and valid for other regions in Germany, too. When implementing a model for another area, only observable or theoretically justified differences between the areas should be adapted, while model structure and parameters should largely be reused. The conservative calibration criterion should then be expanded requiring observable improvements in both areas if a parameter is to be fixed or its range reduced.

In light of the theoretical considerations discussed in chapters 3 and 4, a stronger reduction of parameter uncertainty by calibration would not be justified. Rather the remaining uncertainty had to be considered in further simulation experiments. The results presented show that this needs not preclude robust conclusions with respect to the simulated impacts of climate change. Of course, the compromise between feasible computation time and completely covering all possible parameter combinations meant that some extreme values may have escaped the analysis, however, the experimental designs ensured an optimal representation of global parameter uncertainty for the given number of runs.

Any substantial future change to the model setup will require re-running the calibration procedures and most model applications for policy analysis will require specific screening experiments and experimental designs. The examples given in this thesis show, however, that this is perfectly feasible with the available infrastructure and should not be an argument for incomplete or inconsistent uncertainty analysis.

11.3 The technical infrastructure

The creation of a new preprocessing interface for MPMAS allowed the creation of a complex, comprehensive model and, together with the use of computational resources provided by the bwGRID initiative, made running the necessary number of simulations for uncertainty analysis and estimating price response curves possible. Experienced MPMAS users will agree that constructing a farm decision model of the size and complexity used in the Central Swabian Jura study and running the number of simulations necessary would have hardly been feasible with in the *MpmasExcel* setup, although a proof cannot currently be provided, because so far nobody has made the attempt to do so.

On a 3.33 GHz desktop PC, `mpmasql` takes roughly 1.5 minutes to create the input files of the Central Swabian Jura for one simulation run (without creating commented files and using the compact matrix format) plus about one minute of overhead for loading the data from the database independent of the number of runs created. On a dual core processor two instances of `mpmasql` can be run in parallel. Although preprocessing time is low compared to a simulation time of about 30 minutes per period, preprocessing is currently still a bottle neck, because the number of available desktop computers is much lower than the number of processors accessible on the computing grid. Migration of the preprocessing routines to bwGRID should be attempted as one of the next steps of technical development. It is currently mainly hampered by the lack of an SQL Server accessible from bwGRID.

The `mpmasql` package has been designed as a generic application usable for any MPMAS application, although time constraints lead to a focus on implementing interfaces for those features of MPMAS that were required for this thesis. Now gradually being enhanced to include the features demanded by other MPMAS applications, it has already been taken up by other projects including applications in Chile, Ethiopia, and Brazil.

Maximum run-times on bwGRID clusters currently present a constraining factor for long-term simulations including land markets as reported in chapter 10. Solutions to overcome this current obstacle include (i) reducing the computation time of a single run by internal parallelization of the `mpmas` executable, i.e. distributing the production decisions of different agents over several processors, (ii) improving the solution efficiency of individual MIPs used for shadow price calculation, e.g. by starting with the previous solution instead of starting from scratch or using the parametric analysis tool provided by the OSL, (iii) a more efficient design of the land market itself that reduces the number of MIPs which need to be evaluated based on theoretical or empirical knowledge on potential bidders and increasing the unit size of traded terrains.

11.4 Simulation results

In light of the uncertainties and gaps in the agricultural economic model discussed in the previous sections, the uncertainty in AOGCM predictions discussed in chapter 1, the potential inconsistency and nonstationarity of weather relationships in the statistical downscaling procedure, the uncertainty in the crop growth model and considering the fact that only one specific climate scenario was used in the analysis, the simulation results should not be interpreted as forecasts of future agricultural production and land use in the study area. Apart from improvements of the model, such a forecast would require more reliable meteorological and agro-climatic forecasts, and also incorporating climate effects on grassland. Especially, the development of suitable days for field work used in the scenarios can only be understood as a first assumption.

Nevertheless, the results provide valuable insights for the analysis of climate change impacts and adaptation and the assessment of climate-related policies:

11.4.1 Relevance of different climate-induced changes

The simulation results show that besides effects on yields also other climate change-induced effects on the conditions of agricultural production may have important impacts on land use decisions of farmers and deserve more attention in climate change impact analysis. In the case of the Central Swabian Jura, the model suggests that potential impacts of changes in the time slots suitable for field work and of an additional rotation option are comparable to the – partly substantial – impact of the changes in yields predicted by a crop growth model. This finding is robust over the range of candidate parameterization representing epistemic model uncertainty and stable over long-term simulations.

The effects of these different aspects of climate change may reinforce or counterbalance each other. In the case of winter wheat and winter barley in the study area, short-term simulation results suggest a mutual reinforcement leading to considerable increases in wheat area (cf. section 9.3) and decreases in winter barley area. On the long run, the wheat expansion may be somewhat dampened by a higher silage maize area for biogas production under climate change conditions (cf. section 10.3). Climate effects on silage maize and summer barley are each dominated by a different simulated aspect of climate change.

Taken together the simulated climatic changes show stronger impacts on land use and agricultural production than currently expected price developments for the next decade or even permanently elevated prices at the levels observed during the 2007/2008 cereal price peak. It has to be taken into account, however, that the price projections expect a largely parallel development of crop prices. Changes in the price of a specific crop relative to the prices of other crops, e.g. caused by a climate change-induced decrease in supply outside the study area, are likely to cause stronger impacts in the study area as the assessment of price response curves in chapter 9.6 suggests.

The analysis of price response curves highlights some more dynamic aspects of climate change adaptation: The partial substitution of summer barley by winter wheat production observed in the short-term simulations (section 9.3) creates a much stronger dependence of winter wheat areas on the development of summer barley prices relative to wheat prices, i.e. it increases the cross-price elasticity of winter wheat on summer barley prices (cf. section 9.6.3).

Farm incomes are generally positively affected, which is not very surprising given largely positive yield changes and relaxed restrictions in the climate change scenarios. Here the contribution of the additional rotation option was low, while the effects of yield changes and changes in field work days were comparable. Again observed effects were robust against parameter uncertainty, but very heterogeneous over the agent population. Part of the heterogeneity can be explained by modeling choices, specifically the lower effects on agents with high grassland share is a logical consequence of the neglect of climate effects on grassland productivity. In general, however, the observed heterogeneity underlines the importance of a disaggregate analysis of climate change impacts.

11.4.2 Policies

The simulations suggest potential climate change impacts on participation in the agro-environmental policy scheme MEKA. Especially simulated participation in crop rotation diversification is strongly reduced in the climate change scenarios. The analysis of price response curves in section 9.6.3 underlines the dependence of participation levels – and even the preference of specific grassland diversification measures – on expected price levels. Again the model is capable of revealing potential climate change impacts on dynamic relationships exemplified by the shift of the peak of participation in MEKA A2 towards lower crop price levels (cf. fig. 9.24).

Climate change effects on biogas investments supported by the Renewable Energy Act become apparent only over time and are consequently missed when only looking at short-term effects underlining the important role of a thorough representation of liquidity and credit constraints in the model. The climate change-induced increase in biogas capacity lies around 1000 kW, roughly 10-15% of the baseline capacity, when assessed after six years. Given the latest EEG revision, the future expansion of biogas capacity crucially depends on the opportunity to use or sell excess process heat, since a significant heat use was established as a condition for EEG price guarantees. The newly established small manure plant remuneration tier, which is exempted from this requirement, cannot be expected to attract a large quantity of farmers.

The simulations suggest a noteworthy interplay between the two policy schemes that is observable in both, the simulations specifically designed to analyze this interaction (chapter 9.7) and the recursive-dynamic simulations (chapter 10):

Increased investments in biogas plants increase the demand for silage maize production and intensive grassland areas and consequently decrease the attractiveness of participation in MEKA crop diversification and grassland extensification measures. Although the observed switch from B1 to B2 participation seems to contradict this statement, it actually confirms it: B2 has tighter restrictions on livestock density than B1, but B1 requires 5% of the grassland area to be cut late. B2 is preferred since it does not include this requirement and a restriction of livestock density is not necessarily relevant for biogas farmers. The short-term simulations show a minor dampening effect of the MEKA scheme on biogas investments, while long-term runs suggest a delaying effect of MEKA commitments on biogas investments.

All these results are subject to the reservation that the informal comparison with observed MEKA participation rates suggest a general overestimation of MEKA participation in the model. To a certain extent, this bias is expected, i.e. it would rather be a reason to worry if the model did not overestimate participation: The decision module assesses participation only based on the profitability of the measure itself. It does not consider transaction costs, the reluctance of farmers to commit for the required 5 years, and potential incongruence with farmer values [Burton et al., 2008]. The model also uses an average year investment problem, which might be one reason for the delaying effect of MEKA commitments. In reality, farmers might well decide to build the biogas plant already earlier to secure the higher EEG guaranteed price, while the full potential of biogas production is only realized after the MEKA commitment expired.

Further, grassland conservation support outside the MEKA scheme, e.g. individually tendered contracts, have not been included in the model and the B4 area also acts as a proxy for these types of rewarded grassland conservation. It should not be forgotten either that the hay and grass silage yields and the dedication of grassland area to measure B4 are purely based on the management decision of the farmer, while natural conditions affecting grass yields and species diversity are not considered due to lack of data. Parts of the grassland in the study area may not support two or more cuts per year even under best management. Given the important share of grass silage in the biogas feedstock, this may also affect investments in biogas plants and the estimated effects on land shadow prices.

As a consequence, the simulated effects cannot be interpreted as precise forecasts of policy effects. Nevertheless, they highlight the general pattern of price and climate change impacts and the potential magnitudes of interaction between the two groups of policy measures and provide a sound basis upon which to refine future analysis.

Conclusions

The scientific work presented in this thesis started from two hypotheses: On the one hand, agent-based models can provide valuable contributions to the analysis of climate change impacts on and climate change adaptation in agriculture. On the other hand, there is a number of challenges associated with the empirical use of agent-based models and related to process uncertainty, data requirements, validation and calibration, and necessary computational resources that need to be overcome before they can be fruitfully employed for climate change analysis.

Contributions to climate change adaptation analysis

Summarizing knowledge about climate change effects on agricultural production and reviewing the literature on agent-based models in agriculture, chapters 1 and 2 concluded that agent-based models can be expected to make the most valuable contributions when applied at the regional level for an explicit simulation of climate change impacts and climate-related policies that are not easily captured in other types of models. Specifically, this includes the consequences of farm heterogeneity, changes in opportunity costs that have to be understood in a whole-farm context, the speed of adaptation given learning, sunk costs, path dependency and household demographics, as well as the consequences of interactions between farmers through local markets and cooperation. Generally, the research interest in climate change adaptation can be associated to three types of analysis: (i) predicting differences in production, resource use and structure of the agricultural sector, (ii) assessing the adaptive capacity and resilience of farms, and (iii) *ex ante* policy analysis.

Models that combine innovative approaches to represent agent-agent interactions and learning processes with a comprehensive, process-based understanding of agent decision-making seem best-suited for these types of analysis. The potential for dynamic analysis of long-term changes in farm structure and composition of the agricultural sector, the incorporation of variability and risk management, learning and innovation, taking account of the many different pathways through which climate change may affect agricultural production as well as considering the potential feedbacks of adaptation to climate were identified as important model elements requiring further development. Empirical parameterization, data availability and technical efficiency are preconditions for their usefulness.

The model developed for the Central Swabian Jura in the second part of this thesis focuses on advancing experience on the simulation of various climatic effects, empirical parameterization and and the simulation of long-term structural changes of a heterogeneous population. Simulation results demonstrate the suitability and potential of the model to contribute to all three types of analysis mentioned above:

Analyzing expected differences in the agricultural sector

Potential differences in production, land use and farm type structure were assessed considering detailed representations of several types of effects of climate change on production conditions. The simulations with the model underlined the importance of such a widening of the focus of analysis over merely looking at the potential effects of climate change on crop yields when analyzing agricultural adaptation to climate change: In the case of the study area, a shift of sowing and harvest dates enabling farmers to follow winter wheat with winter rape production and an increased number of days suitable for field work were predicted to have impacts similar in magnitude as yield changes predicted by crop growth models.

Simulated impacts for the study area pointed to an expansion of wheat and silage maize areas at the expense of barley areas. The partial crowding out of summer barley by wheat area held for current price relations and is less strong at higher relative prices for summer barley. The price response analysis indicated that the winter wheat production enters a substitutive relationship with summer barley production under climate change conditions, while competition with winter barley area diminishes leading also to a higher elasticity of the wheat area with respect to relative summer barley prices. Detecting such changes in structural relationships is one of the motivations for using process-based models in the analysis of climate change impacts and this result exemplifies the models capability of simulating these effects.

Further, the simulation of price response functions highlights how agent-based models can be made useful also for larger-scale integrated modeling endeavors that e.g. analyze climate change effects on world or national markets or the feedbacks between land use decisions and climate. The agent-based model (ABM) can be run in a designed experiment to estimate production or land use response functions that can then be used in large scale models.

To improve the representation of structural change processes, the model application extended the MPMAS framework by a demographic component capable of representing the reality of German farms, whose long run survival and investment behavior depends crucially on the existence of a potential successor. It also incorporated a consideration of repayment capacity to avoid overinvestment and ensure replacement of assets. A comparison of the recursive-dynamic results with the potential biogas capacity expansion simulated in unrestricted short-term runs underlines the importance of taking household and liquidity dynamics into account. Simulated biogas expansion, farm type transitions and realistic farm exit rates illustrate that agent-based models will in the future be able to complement currently favored statistical Markov chain approaches [Zimmermann et al., 2009] with a process-based empirical analysis that, with further development, is suitable to reflect the interaction effects emphasized by Huettel and Margarian [2009].

Certainly, a full exploitation of the potential of agent-based models requires further theoretical, empirical, and technical advancements. The model developed provides a well-tested and comprehensive foundation for the development and testing of those processes that have been out of the scope of the applied analysis conducted in this thesis, such as agent-agent interactions, learning processes, variability and risk assessment, and also land markets.

Assessing the adaptive capacity of heterogeneous farms

With respect to assessments of the adaptive capacity and resilience of farms, the model emphasized a comprehensive representation of agent heterogeneity based on observable agent characteristics and the use of functional relationships that translate initial heterogeneity in assets into heterogeneous behavior over time. The simulations confirmed that such a representation is able to endogenously reproduce the observed farm type distribution and land use pattern to a satisfactory degree. This farm type heterogeneity did not suddenly collapse over long-term simulations and was at the same time not unduly restricted, but experienced a gradual development that could be judged realistic, at least if in this specific application the expansion of pig production was exogenously restricted or correctly modeled. The simulations also confirmed that observable agent heterogeneity translates into heterogeneous income effects of climate change and heterogeneous farmer response to climate, policy and price developments – again, underlining the importance of taking farm heterogeneity into account. Since the simulated climatic effects were overwhelmingly positive, only a rudimentary analysis of vulnerability was conducted.

Policy analysis

The model demonstrates its strengths for policy analysis at a disaggregate level, enabling the modeler to analyze the development of investments in biogas production and participation in agri-environmental policy schemes including the competition between biogas support and extensification support. To my knowledge, such an interaction has not been assessed so far in the context of German agricultural policies. Existing analysis of German EEG biogas support [e.g. Sorda et al., 2013; Delzeit, Britz and Holm-Müller, 2012; Delzeit, Holm-Müller and Britz, 2012] have taken a regional, aggregate perspective that cannot easily be adapted to capture this competition. Such a centralized optimization of biogas number and spatial distribution is theoretically equivalent to the aggregate outcomes of individual farmers decision-making are when perfectly functioning regional markets and inter-farm cooperation are assumed [Hazell and Norton, 1986]. However, cooperation between farms is limited in reality and especially participation in agri-environmental schemes is often an activity pursued on area that depends very much on farm-specific circumstances. It is seldom the main source of farm income and the main driver of agricultural production decisions, but rather taken up if it fits into the general production setup of the farm.

The conditions established by the latest EEG revision mean that further development of biogas capacity will crucially depend on the existence of demand for excess process heat, because the alternative option of using high manure shares seems to be unattractive for farmers in the area according to the simulation results. Since biogas investments are a major driver of land use development in the area, a better understanding of existing opportunities to sell excess heat is important.

The recursive simulations demonstrate how the gradual, asymptotic expansion of biogas technology can be explained by the interplay of the development of cash reserves, household composition, annually decreasing guaranteed prices and the duration of existing policy commitments, not even considering innovation diffusion.

Overcoming the challenges of agent-based modeling

In the introduction of this thesis, data requirements, process uncertainty, model validity and computational requirements were identified as the major challenges that need to be overcome for a successful empirical application of agent-based models.

Data requirements

The model application in the Central Swabian Jura showed that at least in Germany a comprehensive dataset can be constructed by combining available secondary datasets like farm censuses, price and demographic statistics, soil and land cover maps and information provided by agricultural extension services. Expert interviews and farm surveys are essential tools to discover regionally specific farming conditions. For the Central Swabian Jura model application, a nonparametric estimation of marginal and joint probability distributions has been developed that was able to comply with privacy restrictions of the FDZ [2010] dataset and at the same time provide a suitable basis to reflect agent heterogeneity in the model. This approach can certainly be refined and needs to be adapted for other cases and regions, but it provides an example showing that privacy restrictions attached to microdatasets need not hamper the analysis and can be dealt with by detailed analysis of the joint statistical distributions observed in the datasets.

Process uncertainty

Process uncertainty can ultimately only be overcome by empirical research and the comparison of different process representations in models for different model applications. It has also become clear that disaggregating one process into invariant partial processes will often require reaching a corresponding level of detail of connected processes.

Incomplete knowledge of processes or parameters is often dealt with by choosing plausible values and implementations *ad hoc*. The Central Swabian Jura model is no exception, but it shows that *ad hoc* parameter choices still allow for a meaningful analysis if their implications are properly assessed in model validation and uncertainty analysis.

Validation and uncertainty analysis

The methodological discussion of model validation in the first part of this thesis emphasized that validity is not an absolute characteristic of a model, but can only be understood relative to a given purpose, and ultimately is not a characteristic of the model itself, but of the conclusions drawn using the model.

It concluded that to ensure the validity of conclusions drawn from modeling studies, a transparent documentation, and an assessment whether the invariant elements of the model can really be expected to be invariant between scenarios assessed is required. It further recommends employing empirical calibration of the model only to the extent warranted by available observation and knowledge about the expected error distribution as well as evaluating and communicating the effect of process uncertainty on the conclusions. Various loss functions and measures of goodness-of-fit were discussed in chapter 3. Not all were applied or applicable in the Central Swabian Jura application, but the discussion of their implications hopefully helps other MPMAS modelers to select a measure suitable for their application.

For the case of agricultural agent-based models that employ mathematical programming, there is little knowledge of error distributions and suitably structured observations are often scarce. In this situation, modelers are well advised to rely on robust calibration criteria and refrain from identifying one single best parameter combination. Rather scenarios can be run for a suitably designed sample of the remaining parameter space and results can then be presented as ranges or distributions over all repetitions of the sample communicating the influence of uncertainty on outcomes.

In the Central Swabian Jura model application, the uncertain parameter space was reduced only when goodness-of-fit improved across the structural breaks incorporated in all three observation years of the calibration dataset. The simulation results show that despite a rather modest reduction of parameter uncertainty the conclusions drawn with respect to climate change effects are robust with respect to the model uncertainty assessed and ambiguous effects can be traced to the effect of individual parameters if a suitable experimental design is used.

Computational requirements

Agent-based models usually need to evaluate one or several decision models for each agent, leading to generally higher computation times than those of aggregate models. Extensive uncertainty analysis requires many repeated model evaluations and hence computational resources multiply when used with agent-based models.

In the Central Swabian Jura model, the computational challenge was addressed by the combination of three essential elements: the use of the computation power of computing grids, the use of efficient experimental designs to keep the necessary model repetitions low, and the development of efficient pre- and postprocessing tools that allow a flexible creation of scenarios and repetitions based on a basic complete model version. Of course, also an efficient software implementation of the MPMAS executable and a fast, reliable solver algorithm were necessary preconditions.

Though land market simulations have for the moment been hampered by computational limits, considerable potential to reduce run-time through parallelization of agent decisions and a more realistic and efficient implementation of land markets have not yet been exploited. The simulations showed that under these preconditions climate change and policy analysis are perfectly possible even with a complex agent-based model.

Future research

The methodological discussion and the practical experience in this thesis suggest a number of research topics that should receive heightened attention in the future to advance the understanding of climate change adaptation and the usability of agent-based models.

Methodologically, the validation and calibration of economic agent-based models should be further developed. While currently robust and conservative criteria for parameter estimation need to be the method of choice, more experience as to their reliability should be gathered. On the long run, analytical and numerical analysis of the properties of error distributions associated to aggregated and disaggregated economic decision models can make a valuable contribution to both, the refinement of robust methods and the establishment of formal Bayesian approaches to model validation. Some attention should be paid to analyzing the sensitivity of model outcomes to different process implementations and the level of detail in order to evaluate potentials for simplification for specific integrated analysis.

An important aspect of validation is the transparency of the model to reviewers and stakeholders. A lot of effort went into a comprehensive documentation of the Central Swabian Jura model and a transparent communication of the effects of *ad hoc* modeling choices and model uncertainty on simulation results. The newly developed *MpmasMySQL* interface has the potential to increase transparency as it allows for a more concise representation of the model than a full-scale Excel MIP matrix setup, a documentation of transformation rules and storage and exposition of the data in their original format. Since it is able to produce commented input files similar to the Excel input files of the *MpmasExcel* setup, it is not falling behind in transparency compared with the traditional preprocessing procedure. Still, the logic and syntax employed in control files is not necessarily intuitive for users unfamiliar with macro- or programming languages and the control files for the Central Swabian Jura application have not been written with a focus on transparency, but rather efficiency. Some future effort should be dedicated to evaluating and improving the understandability of the model files in order to increase the potential of model review. Despite being less prone to simple errors than the *MpmasExcel* setup, the complexity of the model required a long process of verification that could potentially be shortened if more people could get involved and familiar with the model increasing the reliability of the implementation.

The Central Swabian Jura application showed that the disaggregate nature and the use of many repetitions to capture uncertainty create another challenge for the presentation and analysis of the results: It needs to depict outcomes spanning over several dimensions, i.e. agents, scenarios, repetitions and time. The present work underlined that uncertainty reduction can be achieved by focusing on the distribution of differences between scenarios at each repetition rather than directly comparing the distributions of outcomes and it explored different forms of graphical representations including box plots of differences, repeated box plots, regression analysis and scatter plot smoothing. Further expansion and refinement of these methods was out of the scope of this thesis, but is certainly desirable.

From a technical point of view, further improvements in efficiency both in the model and pre- and post processing tools would certainly increase the potential to simulate larger regions or more complex decisions. Alternatively, the option to store, alter and restart the model would help to cope with computational limits. A more pressing essential requirement is the implementation of full control over random numbers used for aleatory model processes in order to ensure a consistent analysis of recursive-dynamic scenarios.

Empirically and theoretically, the effects of climate change on production conditions that go beyond direct effects on crop yields need to receive more attention: On the one hand, models predicting the availability of time slots for field work need to be incorporated into integrated models of climate change analysis in agriculture. Further research needs to show whether and when a change of climate can be expected to consistently allow the additional rotation option in the study area. On the other hand, the current difficulty of growing winter rapeseed after winter wheat identified for the Central Swabian Jura is a regionally specific production constraint that might hold for some other mountainous areas in Germany, but cannot be generalized for all of Germany. Such regionally specific production constraints that might be affected by climate change need to be identified and gathered in a database to allow also modelers working at supra-regional scales to take these changes into account and potentially identify generalizable patterns. Moreover, the three effects considered in the application are only a part of the effects discussed in section 1.2. For example, the yield effect considered still abstracts from CO_2 fertilization and associated yield quantity and quality effects.

The calibration and validation process showed the importance of parameters that were introduced to determine the existence of demand, respectively supply of resources and intermediate products in the absence of fully modeled local markets and suggests that these markets should receive heightened attention in further model development. Following the development of a working and empirically validated land market, attention should then shift to the analysis and modeling of farm cooperation and local markets, especially for silage maize, manure and young animals. Besides land markets, also the analysis of labor and machinery hiring and the hiring of field work services will be crucial to simulate long term structural developments in agriculture.

The inclusion of existing learning and risk management models into the Central Swabian Jura model, which was identified as a desirable model feature in chapter 2, still remains as a major task. It will however be greatly facilitated with the working, validated model created in this work as a basis. Still, the restriction on including only observable agent characteristics will be stretched by the inclusion of individual risk aversion parameters and will likely require either the identification of correlations between risk aversion and observable agent characteristics, or a further development of uncertainty analysis to take account of the many possible joint distributions of risk aversion and other agent characteristics. Psychological and behavioral economic research has shown that there are more consistent, likely invariant patterns of human behavior than merely the choice of the objectively preferable of two alternatives [Rabin, 1998], and learning and risk management will provide the interesting opportunity to incorporate these research results into the model.

Ultimately and fortunately, human behavior will always retain its degree of unpredictability and predictions of economic behavior will maybe never get to the (still limited) reliability of climate projections. Nevertheless, the patterns of human behavior that are generalizable have not yet been exploited as far as possible in order to anticipate relevant developments in the agricultural sector in the future. The present thesis has shown that it is worthwhile to consider the use of agent-based models for this analysis and that the notorious challenges associated to these types of models can be overcome.

References

Acosta-Michlik, L. and Espaldon, V. [2008], 'Assessing vulnerability of selected farming communities in the Philippines based on a behavioural model of agent's adaptation to global environmental change', *Global Environmental Change* **18**(4), 554 – 563.

Alder, D. and Silva, J. [2000], 'An empirical cohort model for terra firme forests in Brazilian Amazon.', *Forest Ecology and Management* **130**, 141–157.

Allen, R., Pereira, L., Raes, D. and Smith, M. [1998], Crop evapotranspiration: Guidelines for computing crop water requirements, Irrigation and Drainage Paper 56, FAO.
http://www.fao.org/docrep/X0490E/X0490E00.htm

An, L., Linderman, M., Qi, J., Shortridge, A. and Liu, J. [2005], 'Exploring complexity in a human–environment system: An agent-based spatial model for multidisciplinary and multiscale integration', *Annals of the Association of American Geographers* **95**(1), 54–79.

Angus, S., Parris, B. and Hassani-M., B. [2009], Climate change impacts and adaptation in Bangladesh: An agent-based approach, *in* '18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009'.
http://mssanz.org.au/modsim09

Antle, J. M. and Capalbo, S. M. [2001], 'Econometric-process models for integrated assessment of agricultural production systems', *American Journal of Agricultural Economics* **83**(2), 389–401.

Antle, J. M. and Capalbo, S. M. [2010], 'Adaptation of agricultural and food systems to climate change: An economic and policy perspective', *Applied Economic Perspectives and Policy* **32**(3), 386–416.

Arnold, T. [2010], Linking farm economics and hydrology: Model integration for watershed-level irrigation management, applied to Chile, Diss., University of Hohenheim, Stuttgart.
http://nbn-resolving.de/urn:nbn:de:bsz:100-opus-4992

Arnold, T., Uribe, H., Troost, C. and Berger, T. [2010], Irrigation management in chile: Integrated modeling of access to water., *in* D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli and T. Filatova, eds, 'International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software, Ottawa, Canada.'.

Aune, J. and Lal, R. [1995], The Tropical Soil Productivity Calculator: A model for assessing effects of soil management on productivity., *in* R. Lal and B. Stewart, eds, 'Soil Management. Experimental Basis for Sustainability and Environmental Quality.', Lewis Publishers, London, pp. 499–520.

Aurbacher, J., Parker, P. S., Calberto Sánchez, G. A., Steinbach, J., Reinmuth, E., Ingwersen, J. and Dabbert, S. [2013], 'Influence of climate change on short term management of field crops – a modelling approach', *Agricultural Systems* **119**, 44 – 57.

Balbi, S. and Giupponi, C. [2009], Reviewing agent-based modelling of socio-ecosystems: a methodology for the analysis of climate change adaptation and sustainability, Working Paper 15/WP/2009, Department of Economics, Ca' Foscari University of Venice.

Balmann, A. [1997], 'Farm-based modelling of regional structural change: a cellular automata approach', *European Review of Agricultural Economics* **24**(1), 85–108.

Balmann, A., Odening, M., Weikard, H.-P. and Brandes, W. [1996], 'Path-dependence without increasing returns to scale and network externalities', *Journal of Economic Behavior & Organization* **29**(1), 159 – 172.

Barlas, Y. [1996], 'Formal aspects of model validity and validation in system dynamics', *System Dynamics Review* **12**(3), 183–210.

Barreteau, O., Bousquet, F. and Attonaty, J.-M. [2001], 'Role-playing games for opening the black box of multi-agent systems: Method and lessons of its application to Senegal River Valley irrigated systems', *Journal of Artificial Societies and Social Simulation* **4**, 2.
http://www.soc.surrey.ac.uk/JASSS/4/2/5.html

Baylis, K., Peplow, S., Rausser, G. and Simon, L. [2008], 'Agri-environmental policies in the EU and United States: A comparison', *Ecological Economics* **65**(4), 753 – 764.

Beck, M., Ravetz, J., Mulkey, L. and Barnwell, T. [1997], 'On the problem of model validation for predictive exposure assessments', *Stochastic Hydrology and Hydraulics* **11**(3), 229–254.

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

Bellman, R. and Åström, K. [1970], 'On structural identifiability', *Mathematical Biosciences* **7**(3-4), 329 – 339.

BenDor, T., Scheffran, J. and Hannon, B. [2009], 'Ecological and economic sustainability in fishery management: A multi-agent model for understanding competition and cooperation', *Ecological Economics* **68**(4), 1061 – 1073.

Berendonk, C. [2011], *Standortgerechte Weide- und Mähweidenutzung des Dauergrünlands - Grünlandanlage, Pflege, Düngung, Nutzung, Weidemanagement*, Landwirtschaftskammer Nordrhein-Westfalen, Kleve.
http://www.riswick.de/pdf/standortgerechte_weide-_und_maehweidenutzung.pdf

Berger, T. [2001], 'Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis.', *Agricultural Economics* **25**(2/3), 245–260.

Berger, T. [2005], Innovation as an alternative to migration? Exemplary results from a multiple-agent programming model applied to Chile, *in* J. Unruh, M. Krol and N. Kliot, eds, 'Environmental Change and its Implications for Population Mitigation. Advances in Global Change Research.', Kluwer Academic Publishers.

Berger, T., Birner, R., Díaz, J., McCarthy, N. and Wittmer, H. [2007], 'Capturing the complexity of water uses and water users within a multi-agent framework', *Water Resources Management* **21**, 129–148.

Berger, T., Schilling, C., Troost, C. and Latynskiy, E. [2010], Knowledge-brokering with agent-based models: Some experiences from irrigation-related research in Chile., *in* D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli and T. Filatova, eds, 'International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software, Ottawa, Canada.'.

Berger, T. and Schreinemachers, P. [2006], 'Creating agents and landscapes for multiagent systems from random samples', *Ecology & Society* **11**(2), 19.

Berger, T., Schreinemachers, P. and Woelcke, J. [2006], 'Multi-agent simulation for the targeting of development policies in less-favored areas', *Agricultural Systems* **88**, 28–43.

Berger, T. and Troost, C. [2012], Agent-based modelling in the agricultural economics tradition of recursive farm modelling and adaptive micro-systems, *in* R. Seppelt, A. Voinov, S. Lange and D. Bankamp, eds, 'International Environmental Modelling and Software Society (iEMSs): 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany'.
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Berger, T. and Troost, C. [2014], 'Agent-based modelling of climate adaptation and mitigation options in agriculture', *Journal of Agricultural Economics* **65**(2), 323–348.

Berkhout, F., Hertin, J. and Gann, D. [2006], 'Learning to adapt: Organisational adaptation to climate change impacts', *Climatic Change* **78**, 135–156.

Betzholz, T. [2011], Der landwirtschaftliche Grundstücksmarkt in Baden-Württemberg, Statistisches Monatsheft, Statistisches Landesamt Baden-Württemberg.

Beven, K. and Binley, A. [1992], 'The future of distributed models: Model calibration and uncertainty prediction', *Hydrological Processes* **6**, 279–298.

Beven, K. and Freer, J. [2001], 'Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology', *Journal of Hydrology* **249**, 11–49.

Beven, K. J., Smith, P. J. and Freer, J. E. [2008], 'So just why would a modeller choose to be incoherent?', *Journal of Hydrology* **354**, 15–32.

Bharwani, S., Bithell, M., Downing, T. E., New, M., Washington, R. and Ziervogel, G. [2005], 'Multi-agent modelling of climate outlooks and food security on a community garden scheme in Limpopo, South Africa', *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1463), 2183–2194.

Bithell, M. and Brasington, J. [2009], 'Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution', *Environmental Modelling & Software* **24**(2), 173 – 190.

BMELV [2006], *Die EU-Agrarreform – Umsetzung in Deutschland. Ausgabe 2006.*, Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz.

BMELV [2013], 'Archiv Testbetriebsnetz/Buchführungsergebnisse'.
http://www.bmelv-statistik.de/de/service/publikationen/archiv-testbetriebsnetz-buchfuehrungsergebnisse/

Boland, L. A. [1994], Scientific thinking without scientific method: Two views of Popper, *in* R. Backhouse, ed., 'New Directions in Economic Methodology', Routledge, London, pp. 154–172.

Boote, K., Jones, J., Hoogenboom, G. and Pickering, N. [1998], The CROPGRO model for grain legumes., *in* G. Tsuji, G. Hoogenboom and P. Thornton, eds, 'Understanding Options for Agricultural Production. Systems Approaches for Sustainable Agricultural Development.', Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 99–128.

Bousquet, F., Barreteau, O., Le Page, C., Mullon, C. and Weber, J. [1999], An environmental modelling approach. The use of multi-agent simulations, *in* F. Blasco and A. Weill, eds, 'Advances in Environmental and Ecological Modelling', Elsevier, Amsterdam, pp. 113–121.

Brander, K. M. [2007], 'Global fish production and climate change', *Proceedings of the National Academy of Sciences* **104**(50), 19709–19714.

Brenner, T. and Werker, C. [2007], 'A taxonomy of inference in simulation models', *Computational Economics* **30**, 227–244.

Breustedt, G. and Glauben, T. [2007], 'Driving forces behind exiting from farming in Western Europe', *Journal of Agricultural Economics,* **58**(1), 115–127.

Britz, W., Heckelei, T. and Wolff, H. [2003], 'Symmetric positive equilibrium problem: A framework for rationalizing economic behavior with limited information: Comment', *American Journal of Agricultural Economics* **85**(4), 1078–1081.

Brown, B. W. and Walker, M. B. [1989], 'The Random Utility Hypothesis and inference in demand systems', *Econometrica* **57**(4), 815–829.

Brown, D. G., Page, S. E., Riolo, R. and Rand, W. [2004], 'Agent-based and analytical modeling to evaluate the effectiveness of greenbelts', *Environmental Modelling & Software* **19**(12), 1097 – 1109.

Browne, M. W. [2000], 'Cross-validation methods', *Journal of Mathematical Psychology* **44**(1), 108–132.

Burnham, K. P. and Anderson, D. [2004], 'Multimodel inference: understanding AIC and BIC in model selection', *Sociological Methods & Research* **33**, 261–304.

Burton, R. J., Kuczera, C. and Schwarz, G. [2008], 'Exploring farmers' cultural resistance to voluntary agri-environmental schemes', *Sociologia Ruralis* **48**(1), 16 – 37.

Busemeyer, J. R. and Wang, Y.-M. [2000], 'Model comparisons and model selections based on generalization criterion methodology', *Journal of Mathematical Psychology* **44**(1), 171 – 189.

Buysse, J., Huylenbroeck, G. V. and Lauwers, L. [2007], 'Normative, positive and econometric mathematical programming as tools for incorporation of multifunctionality in agricultural policy modelling', *Agriculture, Ecosystems & Environment* **120**(1), 70 – 81.

bwGRiD [2013], (http://www.bw-grid.de), member of the German D-Grid initiative, funded by the Ministry for Education and Research (Bundesministerium fuer Bildung und Forschung) and the Ministry for Science, Research and Arts Baden-Wuerttemberg (Ministerium fuer Wissenschaft, Forschung und Kunst Baden-Wuerttemberg).

Calberto Sanchez, G. [2015], Translating plant growth variability due to climate change into farm embedded risk calculation, PhD thesis, Universität Hoheneim, Stuttgart, Germany.

Caldwell, B. J. [1991], 'Clarifying Popper', *Journal of Economic Literature* **29**(1), 1–33.

Campo, P. C., Mendoza, G. A., Guizol, P., Villanueva, T. R. and Bousquet, F. [2009], 'Exploring management strategies for community-based forests using multi-agent systems: A case study in Palawan, Philippines', *Journal of Environmental Management* **90**(11), 3607 – 3615.

Campolongo, F., Cariboni, J. and Saltelli, A. [2007], 'An effective screening design for sensitivity analysis of large models', *Environmental Modelling & Software* **22**(10), 1509–1518.

Castella, J.-C., Boissau, S., Trung, T. N. and Quang, D. D. [2005], 'Agrarian transition and lowland-upland interactions in mountain areas in Northern Vietnam: Application of a multi-agent simulation model', *Agricultural Systems* **86**(3), 312 – 332.

Castella, J. C., Trung, T. N. and Boissau, S. [2005], 'Participatory simulation of land-use changes in the northern mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system.', *Ecology & Society* **10**(1), 27.
http://www.ecologyandsociety.org/vol10/iss1/art27/

Christensen, J., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R., Kwon, W.-T., Laprise, R., Rueda, V. M., Mearns, L., Menéndez, C., Räisänen, J., Rinke, A., Sarr, A. and Whetton, P. [2007], Regional climate projections., *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M.Tignor and H. Miller, eds, 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., chapter 11, pp. 847–940.

CLC2000 [2004], 'CORINE Land Cover (CLC2000); Federal Environment Agency, DLR-DFD.'.

CLC2006 [2009], 'CORINE Land Cover (CLC2006); Federal Environment Agency, DLR-DFD'.

Cobelli, C. and DiStefano, J. J. [1980], 'Parameter and structural identifiability concepts and ambiguities: A critical review and analysis', *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **239**(1), R7–R24.

Davis, G. C. [2004], 'The structure of models: Understanding theory reduction and testing with a production example.', *Journal of Agricultural and Resource Economics* **29**(1), 65 – 78.

Day, R. [2008], 'Micro foundations for meso and macro economic theory', *Journal of Evolutionary Economics* **18**, 261–273. 10.1007/s00191-007-0084-2.

de Brabander, D., de Boever, J., Vanacker, J., Boucque, C. and Bottermann, S. [1999], Evaluation of physical structure in dairy cattle nutrition., *in* P. Garnsworthy and J. Wiseman, eds, 'Recent Advances in Animal Nutrition', Nottingham Univ. Press, pp. 111–145.

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

Dean, J. S., Gumerman, G., Epstein, J. M., Axtell, R. L., Swedlund, A. C., Parket, M. T. and McCarroll, S. [2000], Understanding Anasazi cultural change through agent-based modelling., *in* T. A. Kohler and G. J. Gumerman, eds, 'Dynamics in human and primate studies: Agent-based modeling of social and spatial processes.', Oxford University Press, New York, Oxford, p. 179–206.

Dederer, M. and Messner, J. [2011], *Biogaserzeugung in Baden-Württemberg*, Bildungs- und Wissenszentrum Boxberg (Landesanstalt für Schweinezucht - LSZ), Boxberg.

Deffuant, G., Huet, S. and Amblard, F. [2005], 'An individual-based model of innovation diffusion mixing social value and individual benefit', *American Journal of Sociology* **110**(4), 1041–1069.

Deichsel, S. and Pyka, A. [2009], 'A pragmatic reading of Friedman's methodological essay and what it tells us for the discussion of ABMs', *Journal of Artificial Societies and Social Simulation* **12**(4), 6.
http://jasss.soc.surrey.ac.uk/12/4/6.html

Delzeit, R., Britz, W. and Holm-Müller, K. [2012], 'Modelling regional input markets with numerous processing plants: The case of green maize for biogas production in Germany', *Environmental Modelling & Software* **32**, 74 – 84.

Delzeit, R., Holm-Müller, K. and Britz, W. [2012], 'Ökonomische Bewertung des Erneuerbare Energien Gesetzes zur Förderung von Biogas', *Perspektiven der Wirtschaftspolitik* **13**(3), 251–265.

destatis [2011], Arbeitskräfte - Landwirtschaftszählung 2010, Land- und Forstwirtschaft, Fischerei - Fachserie 3 Heft 2, Statistisches Bundesamt, Wiesbaden.

destatis [2012*a*], 'GENESIS online - Tabelle 12411-0008: Bevölkerung Deutschland'.
https://www-genesis.destatis.de/genesis/online

destatis [2012*b*], 'GENESIS online - Tabelle 12612-0008: Geburtenziffern Deutschland'.
https://www-genesis.destatis.de/genesis/online

destatis [2012*c*], 'GENESIS online - Tabelle 12613-0003: Gestorbene Deutschland'.
https://www-genesis.destatis.de/genesis/online

destatis [2012*d*], 'GENESIS online - Tabelle 41122-0001: Landwirtschaftliche Betriebe, Landwirtsch. genutzte Fläche (Rep. ASE): Deutschland, Jahre (bis 2007), Rechtsform, Größenklassen der landwirtschaftlich genutzten Fläche'.
https://www-genesis.destatis.de/genesis/online

destatis [2012*e*], 'GENESIS online - Tabelle 61221-0002: Index der Einkaufspreise landwirtschaftl. Betriebsmittel: Deutschland, Wirtschaftsjahr, Messzahlen mit/ohne Umsatzsteuer, Landwirtschaftliche Betriebsmittel'.
https://www-genesis.destatis.de/genesis/online

Doherty, J. [2005], *PEST: Model Independent Parameter Estimation. Fifth edition of User Manual.*, Watermark Numerical Computing, Brisbane, Australia.

Döll, P. [2002], 'Impact of climate change and variability on irrigation requirements: A global perspective', *Climatic Change* **54**(3), 269–293.

Duan, Q., Gupta, V. and Sorooshian, S. [1993], 'A shuffled complex evolution approach for effective and efficient global minimization', *Journal of Optimization Theory and Applications* **76**(3), 501–521.

DWD [2011], 'Climate data - historical time series - daily values - Stötten'.
http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU2/KU21/klimadaten/german/download/tageswerte/kl__10836__hist__txt,templateId=raw,property=publicationFile.zip/kl_10836_hist_txt.zip

Easterling, W., Aggarwal, P., Batima, P., Brander, K., Erda, L., Howden, S., Kirilenko, A., Morton, J., Soussana, J.-F., Schmidhuber, J. and Tubiello, F. [2007], Food, fibre and forest products., *in* M. Parry, O. Canziani, J. Palutikof, P. van der Linden and C. Hanson, eds, 'Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, Cambridge, UK, pp. 273–313.

Eboli, F., Parrado, R. and Roson, R. [2010], 'Climate-change feedback on economic growth: Explorations with a dynamic general equilibrium model', *Environment and Development Economics* **15**(05), 515–533.

Efron, B. and Tibshirani, R. [1997], 'Improvements on cross-validation: The .632+ bootstrap method', *Statistical Science* **92**(438), 548–560.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J. [2012], 'HESS Opinions "Should we apply bias correction to global and regional climate model data?"', *Hydrology and Earth System Sciences* **16**(9), 3391–3404.

Etienne, M. [2003], 'SYLVOPAST: a multiple target role-playing game to assess negotiation processes in sylvopastoral management planning', *Journal of Artificial Societies and Social Simulation* **6**(2), 5.
http://jasss.soc.surrey.ac.uk/6/2/5.html

Evans, T. P. and Kelley, H. [2004], 'Multi-scale analysis of a household level agent-based model of landcover change', *Journal of Environmental Management* **72**, 57–72.

Fachagentur nachwachsende Rohstoffe e.V. [n.d.], 'Abb. 27 Entwicklung Biogasanlagen'.
http://mediathek.fnr.de/grafiken/daten-und-fakten/bioenergie/biogas/entwicklung-biogasanlagen.html

Fagiolo, G., Moneta, A. and Windrum, P. [2007], 'A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems.', *Computational Economics* **30**(3), 195 – 226.

Falloon, P. and Betts, R. [2010], 'Climate impacts on European agriculture and water management in the context of adaptation and mitigation—the importance of an integrated approach', *Science of the Total Environment* **408**(23), 5667–5687.

FDZ [2010], *AFiD-Panel Agrarstruktur 1999, 2003, 2007*, Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder, Kiel.

Feil, J.-H., Mußhoff, O. and Balmann, A. [2011], The effects of different political schemes on the willingness to invest, firm profitability and economic efficiency in the dairy sector - an agent based real options approach, *in* 'Paper prepared for the 122nd EAAE Seminar "Evidence-Based Agricultural And Rural Policy Making: Methodological And Empirical Challenges Of Policy Evaluation", Ancona, February 17-18, 2011'.

Filatova, T., Van Der Veen, A. and Parker, D. C. [2009], 'Land market interactions between heterogeneous agents in a heterogeneous landscape - Tracing the macro-scale effects of individual trade-offs between environmental amenities and disamenities', *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* **57**(4), 431–457.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N. and Snyder, P. K. [2005], 'Global consequences of land use', *Science* **309**(5734), 570–574.

Forrester, J. W. and Senge, P. M. [1980], Tests for building confidence in system dynamics models, *in* A. A. Legasto, Jr., J. W. Forrester and J. M. Lyneis, eds, 'System Dynamics', Vol. 14 of *TIMS Studies in the Management Sciences*, North-Holland, New York, Amsterdam, pp. 209–228.

Forster, M. R. [2000], 'Key concepts in model selection: Performance and generalizability', *Journal of Mathematical Psychology* **44**(1), 205 – 231.

Freeman, T., Nolan, J. and Schoney, R. [2009], 'An agent-based simulation model of structural change in Canadian prairie agriculture, 1960–2000', *Canadian Journal of Agricultural Economics* **57**(4), 537–555.

Freeman, T. R. [2005], From the ground up: an agent-based model of regional structural change, Master's thesis, University of Saskatchewan.

Frisch, R. [1933], 'Editorial', *Econometrica* **1**, 1–5.

Gass, S. I. [1983], 'Decision-aiding models: Validation, assessment, and related issues for policy analysis', *Operations Research* **31**(4), 603–631.

Gaube, V., Kaiser, C., Wildenberg, M., Adensam, H., Fleissner, P., Kobler, J., Lutz, J., Schaumberger, A., Schaumberger, J., Smetschka, B., Wolf, A., Richter, A. and Haberl, H. [2009], 'Combining agent-based and stock-flow modelling approaches in a participative analysis of the integrated land system in Reichraming, Austria', *Landscape Ecology* **24**, 1149–1165.

Gibbons, J., Cox, G., Wood, A., Craigon, J., Ramsden, S., Tarsitano, D. and Crout, N. [2008], 'Applying Bayesian model averaging to mechanistic models: An example and comparison of methods', *Environmental Modelling & Software* **23**(8), 973 – 985.

Gigerenzer, G. and Goldstein, D. G. [1996], 'Reasoning the fast and frugal way: Models of bounded rationality', *Psychological Review* **103**(4), 650–669.

Godwin, D. and Jones, C. [1991], Nitrogen dynamics in soil-plant systems, *in* J. Hanks and J. Ritchie, eds, 'Modeling Soil and Plant Systems', Agronomy Society of America, Madison, p. 287–322.

Gotts, N. M. and Polhill, J. G. [2009], 'When and how to imitate your neighbours: Lessons from and for FEARLUS', *Journal of Artificial Societies and Social Simulation* **12**(3), 2. *http://jasss.soc.surrey.ac.uk/12/3/2.html*

Gotts, N., Polhill, J. and Law, A. [2003], 'Agent-based simulation in the study of social dilemmas', *Artificial Intelligence Review* **19**, 3–92. 10.1023/A:1022120928602.

Granovetter, M. [1978], 'Threshold models of collective behavior', *The American Journal of Sociology* **83**(6), 1420–1443.

Gregory, P. J., Johnson, S. N., Newton, A. C. and Ingram, J. S. I. [2009], 'Integrating pests and pathogens into the climate change/food security debate', *Journal of Experimental Botany* **60**(10), 2827–2838.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U. and DeAngelis, D. L. [2006], 'A standard protocol for describing individual-based and agent-based models', *Ecological Modeling* **198**(1-2), 115 – 126.

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

Gross, J., McAllister, R., Abel, N., Smith, D. S. and Maru, Y. [2006], 'Australian rangelands as complex adaptive systems: A conceptual model and preliminary results', *Environmental Modelling & Software* **21**(9), 1264 – 1272.

Grothmann, T. and Patt, A. [2005], 'Adaptive capacity and human cognition: The process of individual adaptation to climate change', *Global Environmental Change* **15**(3), 199 – 213.

Guzy, M. R., Smith, C. L., Bolte, J. P., Hulse, D. W. and Gregory, S. V. [2008], 'Policy research using agent- based modeling to assess future impacts of urban expansion into farmlands and forests.', *Ecology & Society* **13**(1), 37.
http://www.ecologyandsociety.org/vol13/iss1/art37/

Haas, A. and Jaeger, C. [2005], 'Agents, Bayes, and climatic risks - a modular modelling approach', *Advances in Geosciences* **4**, 3–7.

Hailegiorgis, A. B., Kennedy, W. G., Rouleau, M., Bassett, J. K., Coletti, M., Balan, G. C. and Gulden, T. [2010], An agent based model of climate change and conflict among pastoralists in East Africa, *in* D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli and T. Filatova, eds, 'International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada'.
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Hall, C., McVittie, A. and Moran, D. [2004], 'What does the public want from agriculture and the countryside? A review of evidence and methods', *Journal of Rural Studies* **20**(2), 211 – 225.

Hands, D. W. [2001], *Reflection without Rules - Economic Methodology and Contemporary Science Theory*, Cambridge: Cambridge University Press.

Hansen, L. P. and Heckman, J. J. [1996], 'The empirical foundations of calibration', *Journal of Economic Perspectives* **10**(1), 87–104.

Happe, K., Balmann, A., Kellermann, K. and Sahrbacher, C. [2008], 'Does structure matter? The impact of switching the agricultural policy regime on farm structures', *Journal of Economic Behavior and Organization* **67**(2), 431–444.

Happe, K., Schnicke, H., Sahrbacher, C. and Kellermann, K. [2009], 'Will they stay or will they go? Simulating the dynamics of single-holder farms in a dualistic farm structure in Slovakia', *Canadian Journal of Agricultural Economics* **57**(4), 497–511.

Harms, J. and Wendl, G. [2009], Automatisierung in der Milchviehhaltung – Stand der Technik und Entwicklungstendenzen, *in* G. Wendl, ed., 'Strategien für zukunftsorientierte Milchviehbetriebe in Bayern. Landtechnisch-bauliche Jahrestagung am 25. November 2009 in Triesdorf', Schriftenreihe, Bayerische Landesanstalt für Landwirtschaft.

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T. and Huth, A. [2011], 'Statistical inference for stochastic simulation models – theory and application.', *Ecology Letters* **14**, 816–827.

Hartmann, A. [2008], Wie viel Fläche wird für Biogas benötigt?, Statistisches Monatsheft Baden-Württemberg 7/2008, Statistisches Landesamt Baden-Württemberg.

Hazell, P. and Norton, R. [1986], *Mathematical programming for economic analysis in agriculture*, Macmillan, New York.

Heckelei, T. and Wolff, H. [2003], 'Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy', *European Review of Agricultural Economics* **30**(1), 27–50.

Helton, J., Johnson, J., Sallaberry, C. and Storlie, C. [2006], 'Survey of sampling-based methods for uncertainty and sensitivity analysis', *Reliability Engineering & System Safety* **91**, 1175–1209.

Hill, M. C., Cooley, R. L. and Pollock, D. W. [1998], 'A controlled experiment in ground water flow model calibration', *Ground Water* **36**(3), 520–535.

Hinrichs, J., Mußhoff, O. and Odening, M. [2008], 'Economic hysteresis in hog production', *Applied Economics* **40**(3), 333–340.

Hobbs, N. T. and Hilborn, R. [2006], 'Alternatives to statistical hypothesis testing in ecology: A guide to self teaching', *Ecological Applications* **16**, 5–19.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. [1999], 'Bayesian model averaging: A tutorial', *Statistical Science* **14**(4), 382–417.

Hoffmann, M., Kelley, H. and Evans, T. [2002], Simulating land-cover change in South-Central Indiana: An agent-based model of deforestation and afforestation., *in* M. A. Janssen, ed., 'Complexity and ecosystem management: the theory and practice of multi-agent systems.', Edward Elgar, Cheltenham, UK, p. 218–247.

Högy, P. and Fangmeier, A. [2008], 'Effects of elevated atmospheric CO_2 on grain quality of wheat', *Journal of Cereal Science* **48**(3), 580 – 591.

Hohenegger, C., Brockhaus, P. and Schär, C. [2008], 'Towards climate simulations at cloud-resolving scales', *Meteorologische Zeitschrift* **17**, 383–394.

Holtz, G. and Pahl-Wostl, C. [2012], 'An agent-based model of groundwater over-exploitation in the Upper Guadiana, Spain', *Regional Environmental Change* **12**, 95–121.

Howitt, R. E. [1995], 'Positive mathematical programming', *American Journal of Agricultural Economics* **77**(2), 329–342.

Huettel, S. and Margarian, A. [2009], 'Structural change in the West German agricultural sector', *Agricultural Economics* **40**, 759–772.

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.

Hunt, R. J., Doherty, J. and Tonkin, M. J. [2007], 'Are models too simple? Arguments for increased parameterization', *Ground Water* **45**(3), 254–262.

Hyndman, R. J. and Koehler, A. B. [2006], 'Another look at measures of forecast accuracy', *International Journal of Forecasting* **22**, 679–688.

Iglesias, A. and Rosenzweig, C. [2010], *Effects of Climate Change on Global Food Production under Special Report on Emissions Scenarios (SRES) Emissions and Socioeconomic Scenarios: Data from a Crop Modeling Study.*, Socioeconomic Data and Applications Center (SEDAC), Columbia University., Palisades, NY. (Accessed Aug 3, 2012).
http://sedac.ciesin.columbia.edu/mva/cropclimate/

Ines, A. V. and Hansen, J. W. [2006], 'Bias correction of daily GCM rainfall for crop simulation studies', *Agricultural and Forest Meteorology* **138**(1–4), 44 – 53.

Izquierdo, L., Gotts, N. and Polhill, J. [2003], FEARLUS-W: An agent-based model of river basin land use and water management., *in* M. Dijst, P. Schot and K. de Jong, eds, 'Framing land use dynamics: integrating knowledge on spatial dynamics in socio-economic and environmental systems for spatial planning in Western urbanized countries.', Faculty of Geographical Sciences, Utrecht University, Utrecht, The Netherlands, p. 163–165.

Jakeman, A., Letcher, R. and Norton, J. [2006], 'Ten iterative steps in development and evaluation of environmental models', *Environmental Modelling & Software* **21**, 602–614.

Janmaat, J. and Anputhas, M. [2010], Development of an agent based coupled economic and hydrological model of the Deep Creek watershed in British Columbia, Canada., *in* D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli and T. Filatova, eds, 'International Environmental Modelling and Software Society (iEMSs). 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada'.
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Janssen, M. A. [2007], 'Coordination in irrigation systems: An analysis of the Lansing-Kremer model of Bali', *Agricultural Systems* **93**(1-3), 170 – 190.

Janssen, M. A., Walker, B. H., Langridge, J. and Abel, N. [2000], 'An adaptive agent model for analysing co-evolution of management and policies in a complex rangeland system', *Ecological Modelling* **131**(2-3), 249 – 268.

Janssen, M. and de Vries, B. [1998], 'The battle of perspectives: A multi-agent model with adaptive responses to climate change', *Ecological Economics* **26**(1), 43 – 65.

Janssen, M. and Ostrom, E. [2006], Adoption of a new regulation for the governance of common-pool resources by a heterogeneous population., *in* J. Baland, P. Bardhan and S. Bowles, eds, 'Inequality, Cooperation and Environmental Sustainability.', Princeton University Press, Princeton, NJ, chapter 4, pp. 60–96.

Janssen, S. and van Ittersum, M. K. [2007], 'Assessing farm innovations and responses to policies: A review of bio-economic farm models', *Agricultural Systems* **94**(3), 622 – 636.

Jansson, T. and Heckelei, T. [2011], 'Estimating a primal model of regional crop supply in the European Union', *Journal of Agricultural Economics* **62**(1), 137–152.

Jepsen, M. R., Leisz, S., Rasmussen, K., Jakobsen, J., Møller-Jensen, L. and Christiansen, L. [2006], 'Agent-based modelling of shifting cultivation field patterns, Vietnam', *International Journal of Geographical Information Science* **20**(9), 1067–1085.

Juliá, R. and Duchin, F. [2007], 'World trade as the adjustment mechanism of agriculture to climate change', *Climatic Change* **82**, 393–409.

Just, R. E. [1993], 'Discovering production and supply relationships: Present status and future opportunities', *Review of Marketing and Agricultural Economics* **61**(1), 11–14.

Kaufmann, P., Stagl, S. and Franks, D. W. [2009], 'Simulating the diffusion of organic farming practices in two New EU Member States', *Ecological Economics* **68**(10), 2580 – 2593.

Kaye-Blake, W., Li, F. Y., Martin, A. M., McDermott, A., Rains, S., Sinclair, S. and Kira, A. [2010], Multi-agent simulation models in agriculture: A review of their construction and uses, Research report 318, Lincoln University. Agribusiness and Economics Research Unit.
http://hdl.handle.net/10182/1928

Kerr, J. T. [2001], 'Butterfly species richness patterns in Canada: energy, heterogeneity, and the potential consequences of climate change.', *Conservation Ecology* **5**(1), 10.
http://www.consecol.org/vol5/iss1/art10/

Kirschbaum, M. [1999], 'CenW, a forest growth model with linked carbon, energy, nutrient and water cycles.', *Ecological Modelling* **118**, 17–59.

Klappholz, K. and Agassi, J. [1959], 'Methodological prescriptions in economics', *Economica* **26**(101), 60–74.

Kleijnen, J. P., Bettonvil, B. and van Groenendaal, W. [1998], 'Validation of trace-driven simulation models: A novel regression test', *Management Science* **44**, 812–819.

Koertge, N. [1975], 'Popper's metaphysical research program for the human sciences', *Inquiry* **18**(4), 437–462.

Koertge, N. [1979], 'The methodological status of Popper's rationality principle', *Theory and Decision* **10**, 83–95. 10.1007/BF00126332.

Kohler, T., Kresl, J., West, C. V., Carr, E. and Wilshusen, R. H. [2000], Be there then: A modelling approach to settlement determinants and spatial efficiency among late ancestral populations of the Mesa Verde region, US Southwest., *in* T. A. Kohler and G. J. Gumerman, eds, 'Dynamics in human and primate studies: agent-based modeling of social and spatial processes', Oxford University Press, New York, Oxford, p. 145–178.

Krause, P., Boyle, D. and Bäse, F. [2005], 'Comparison of different efficiency criteria for hydrological model assessment', *Advances in Geosciences* **5**, 89–97.

Kreienkamp, F., Spekat, A. and Enke, W. [2013], 'The weather generator used in the empirical statistical downscaling method, WETTREG', *Atmosphere* **4**(2), 169–197.

KTBL [2008], *KTBL-Datensammlung. - Betriebsplanung Landwirtschaft 2008/09*, Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V, Darmstadt.

KTBL [2010], *KTBL-Datensammlung. - Betriebsplanung Landwirtschaft 2010/11*, Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V, Darmstadt.

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. [2006], 'Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters', *Journal of Hydrology* **331**, 161–177.

Kuhn, T. S. [1962], *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

Kuhn, T. S. [1977], Objectivity, value judgement and theory choice, *in* 'The Essential Tension: Selected Studies in Scientific Tradition and Change.', University of Chicago Press, pp. 320–339.

Kumbhakar, S. C. and Tsionas, E. G. [2011], 'Stochastic error specification in primal and dual production systems', *Journal of Applied Econometrics* **26**(2), 270–297.

Kydland, F. E. and Prescott, E. C. [1996], 'The computational experiment: An econometric tool', *Journal of Economic Perspectives* **10**(1), 69–85.

Lambin, E., Rounsevell, M. and Geist, H. [2000], 'Are agricultural land-use models able to predict changes in land-use intensity?', *Agriculture, Ecosystems & Environment* **82**(1–3), 321 – 331.

Lansing, J. S. and Kremer, J. N. [1993], 'Emergent properties of Balinese water temple networks: Coadaptation on a rugged fitness landscape', *American Anthropologist* **95**(1), 97–112.

Latynskiy, E., Troost, C. and Berger, T. [2010], Impacts of climate variability and climate change on agriculture and rural poverty in Central Chile, *in* 'Contributed Paper at the 117th Seminar of the EAAE: Climate Change, Food Security and Resilience of Food and Agricultural Systems in Developing Countries: Mitigation and Adaptation Options, November 25-27, 2010.Stuttgart-Hohenheim.'.

Law, A. M. [2007], *Simulation Modeling & Analysis - International Edition*, 4th edition edn, McGraw-Hill, Boston.

Le, Q. B., Park, S. and Vlek, P. [2008], 'Land use dynamic simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. 1. Structure and theoretical specification', *Ecological Informatics* **3**, 135–153.

Legates, D. R. and McCabe, G. [1999], 'Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation', *Water Resources Research* **35**(1), 233–241.

Lei, Z., Pijanowski, B. C., Alexandridis, K. T. and Olson, J. [2005], 'Distributed modeling architecture of a Multi-Agent-based Behavioral Economic Landscape (MABEL) model', *SIMULATION* **81**(7), 503–515.

LEL [2010], *Material aus der Ernährungswirtschaft des Landes Baden-Württemberg: Getreide und Futtermittel 2008/2009*, Landesanstalt für Entwicklung der Landwirtschaft und der ländlichen Räume (LEL), Schwäbisch Gmünd.

LEL [2011*a*], *Material aus der Ernährungswirtschaft des Landes Baden-Württemberg: Milch 2010*, Landesanstalt für Entwicklung der Landwirtschaft und der ländlichen Räume (LEL), Schwäbisch Gmünd.

LEL [2011*b*], *Material aus der Ernährungswirtschaft des Landes Baden-Württemberg: Vieh und Fleisch 2010*, Landesanstalt für Entwicklung der Landwirtschaft und der ländlichen Räume (LEL), Schwäbisch Gmünd.

LEL [2012], 'Datensammlung Kalkulationsdaten Marktfruchtbau Ernte 2012 in Baden-Württemberg; Version 01/2012'.
https://www.landwirtschaft-bw.info/servlet/PB/show/1178224_l1/LEL_Kalkulationsdaten_Marktfruechte.zip

Lewbel, A. [2001], 'Demand systems with and without errors', *The American Economic Review* **91**(3), 611–618.

LfL [2012], 'Fachbeiträge aus dem Institut für Pflanzenbau und Pflanzenzüchtung'.
http://www.lfl.bayern.de/ipz/

LfL, ed. [2010], *Gruber Tabelle zur Fütterung der Milchkühe Zuchtrinder Schafe Ziegen*, 32nd edn, Bayerische Landesanstalt für Landwirtschaft, Freising-Weihenstephan.

LfL, ed. [2011], *Gruber Tabelle zur Fütterung in der Rindermast: Fresser, Bullen, Ochsen, Mastfärsen, Mastkühe*, 16th edn, Bayerische Landesanstalt für Landwirtschaft, Freising-Weihenstephan.

Linder, S. [2013], Räumliche Diffusion von Photovoltaik-Anlagen in Baden-Württemberg, PhD thesis, Universität Würzburg.
http://opus.bibliothek.uni-wuerzburg.de/volltexte/2013/7778/

Lippert, C., Krimly, T. and Aurbacher, J. [2009], 'A Ricardian analysis of the impact of climate change on agriculture in Germany', *Climatic Change* **97**(3-4), 593–610.

Lobell, D. [2010], Crop responses to climate: Time-series models, *in* D. Lobell and M. Burke, eds, 'Climate Change and Food Security', Vol. 37 of *Advances in Global Change Research*, Springer, chapter 5, pp. 85–98.

Lobianco, A. and Esposti, R. [2010], 'The Regional Multi-Agent Simulator (RegMAS): An open-source spatially explicit model to assess the impact of agricultural policies', *Computers and Electronics in Agriculture* **72**(1), 14 – 26.

Loibl, W. and Toetzer, T. [2003], 'Modeling growth and densification processes in suburban regions - Simulation of landscape transition with spatial agents', *Environmental Modelling & Software* **18**(6), 553 – 563.

Longino, H. [1992], Essential tensions - Phase two: Feminist, philosophical and social studies of science, *in* E. McMullin, ed., 'The Social Dimensions of Science', University of Notre Dame Press, Notre Dame, IN, pp. 198–216.

Lotze-Campen, H. and Schellnhuber, H. [2009], 'Climate impacts and adaptation options in agriculture: what we know and what we don't know', *Journal für Verbraucherschutz und Lebensmittelsicherheit* **4**, 145–150.

LUBW [2007], *Wasser- und Bodenatlas Baden-Württemberg*, Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg.

Lucas, Jr., R. E. [1976], Econometric policy evaluation: A critique, *in* K. Brunner and A. Meltzer, eds, 'The Philips Curve and Labor Markets', Vol. 1 of *Carnegie-Rochester Conference Series on Public Policy*, American Elsevier, New York, pp. 19–46.

Luck, J., Spackman, M., Freeman, A., Trębicki, P., Griffiths, W., Finlay, K. and Chakraborty, S. [2011], 'Climate change and diseases of food crops', *Plant Pathology* **60**(1), 113–121.

Mann, S. [2007], 'Wie entstehen HofnachfolgerInnen?', *Agrarwirtschaft/German Journal for Agricultural Economics* **56**(3), 161–165.

Manson, S. M. [2005], 'Agent-based modeling and genetic programming for modeling land change in the Southern Yucatán Peninsular Region of Mexico', *Agriculture, Ecosystems & Environment* **111**(1-4), 47 – 62.

Marohn, C., Schreinemachers, P., Quang, D. V., Berger, T., Siripalangkanont, P., Nguyen, T. T. and Cadisch, G. [2013], 'A software coupling approach to assess low-cost soil conservation strategies for highland agriculture in Vietnam', *Environmental Modelling & Software* **45**, 116 – 128.

Marohn, C., Siripalangkanont, P., Berger, T., Lusiana, B. and Cadisch, G. [2010], Integrated land use modeling in an interdisciplinary project: The LUCIA model., *in* 'Proceedings of the LANDMOD 2010 Conference, Montpellier, 3-5 February 2010.'.

Matthews, R. [2006], 'The People and Landscape Model (PALM): Towards full integration of human decision-making and biophysical simulation models', *Ecological Modelling* **194**(4), 329 – 343.

Matthews, R., Gilbert, N., Roach, A., Polhill, J. and Gotts, N. [2007], 'Agent-based land-use models: A review of applications', *Landscape Ecology* **22**, 1447–1459. 10.1007/s10980-007-9135-1.

McCarl, B. and Apland, J. [1986], 'Validation of linear programming models', *Southern Journal of Agricultural Economics* **18**(2), 155–164.

McCloskey, D. N. [1983], 'The rhetoric of economics', *Journal of Economic Literature* **21**(2), 481–517.

McCloskey, D. N. and Ziliak, S. [1996], 'The standard error of regressions', *Journal of Economic Literature* **34**, 97–114.

McElroy, M. B. [1987], 'Additive general error models for production, cost, and derived demand or share systems', *Journal of Political Economy* **95**(4), 737–757.

McKay, M. D., Beckman, R. J. and Conover, W. J. [1979], 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', *Technometrics* **21**(2), 239–245.

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., Kitoh, A., Knutti, R., Murphy, J., Noda, A., Raper, S., Watterson, I., Weaver, A. and Zhao, Z.-C. [2007], Global climate projections., *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M.Tignor and H. Miller, eds, 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., chapter 10, pp. 747–845.

Milner-Gulland, E., Kerven, C., Behnke, R., Wright, I. and Smailov, A. [2006], 'A multi-agent system model of pastoralist behaviour in Kazakhstan', *Ecological Complexity* **3**(1), 23 – 36.

MLREV [2011], 'Verwaltungsvorschrift des Ministeriums für Ländlichen Raum, Ernährung und Verbraucherschutz zur Förderung der Erhaltung und Pflege der Kulturlandschaft und von Erzeugungspraktiken, die der Marktentlastung dienen (Marktenlastungs- und Kulturlandschaftsausgleich - MEKA III)', https://www.landwirtschaft-bw.info/servlet/PB/show/1318342/28.02.2011%20-%20Verwaltungsvorschrift%20MEKA%20III%20-%20MLR%2025-8872.53.pdf.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L. [2007], 'Model evaluation guidelines for systematic quantification of accuracy in watershed simulations', *Transactions of the ASABE* **50**(3), 885–900.

Morris, M. D. [1991], 'Factorial sampling plans for preliminary computational experiments', *Technometrics* **33**(2), 161–174.

Morris, M. D., Moore, L. M. and McKay, M. D. [2008], 'Using orthogonal arrays in the sensitivity analysis of computer models', *Technometrics* **50**(2), 205–215.

Morton, J. F. [2007], 'The impact of climate change on smallholder and subsistence agriculture', *Proceedings of the National Academy of Sciences* **104**(50), 19680–19685.

Moss, S. [2008], 'Alternative approaches to the empirical validation of agent-based models', *Journal of Artificial Societies and Social Simulation* **11**(1), 5.
http://jasss.soc.surrey.ac.uk/11/1/5.html

Moss, S., Pahl-Wostl, C. and Downing, T. [2001], 'Agent-based integrated assessment modelling: The example of climate change', *Integrated Assessment* **2**, 17–30. 10.1023/A:1011527523183.

Müller, B., F., A., Drees, R., Dressler, G., J., G., Klassert, C., M., S., Schulze, J., Weise, H. and Schwarz, N. [2013], 'Describing human decisions in agent-based models - ODD+D, an extension of the ODD protocol.', *Environmental Modeling and Software* **48**, 37–48.

Müller, T., Magid, J. and Jensen, L. [2003], 'Decomposition of plant residues of different quality in soil – DAISY model calibration and simulation based on experimental data.', *Ecological Modelling* **166**, 3–18.

Munkres, J. [1957], 'Algorithms for the assignment and transportation problems', *Journal of the Society for Industrial and Applied Mathematics* **5**(1), 32–38.

Nakićenović, N. and Swart, R., eds [2000], *Special Report on Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Nash, S. and Sutcliffe, J. [1970], 'River flow forecasting through conceptual models, 1, a discussion of principles', *Journal of Hydrology* **10**, 282–290.

Nerlove, M. [1958], 'Distributed lags and estimation of long-run supply and demand elasticities: Theoretical considerations', *Journal of Farm Economics* **40**(2), 301–310.

NIST/SEMATECH, ed. [2012], *e-Handbook of Statistical Methods.*
http://www.itl.nist.gov/div898/handbook/

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V. and Alvin, K. F. [2002], 'Error and uncertainty in modeling and simulation', *Reliability Engineering & System Safety* **75**(3), 333 – 357.

Odening, M., Mußhoff, O. and Balmann, A. [2005], 'Investment decisions in hog finishing: an application of the real options approach', *Agricultural Economics* **32**(1), 47–60.

OECD/FAO, ed. [2012], *OECD-FAO Agricultural Outlook 2012*, OECD Publishing.

Olesen, J. E. and Bindi, M. [2002], 'Consequences of climate change for European agricultural productivity, land use and policy', *European Journal of Agronomy* **16**(4), 239–262.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. [1994], 'Verification, validation, and confirmation of numerical models in the earth sciences', *Science* **263**(5147), 641–646.

Pacala, S., Canham, C., Saponara, J., Silander, J., Kobe, R. and Ribbens, E. [1996], 'Forest models defined by field measurements: Estimation, error analysis and dynamics', *Ecological Monographs* **66**, 1–43.

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

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

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M. and Fischer, G. [2004], 'Effects of climate change on global food production under SRES emissions and socio-economic scenarios', *Global Environmental Change* **14**(1), 53 – 67. Climate Change.

Parry, M., Rosenzweig, C., Iglesias, A., Fischer, G. and Livermore, M. [1999], 'Climate change and world food security: a new assessment', *Global Environmental Change* **9**(Supplement 1), 51 – 67.

Parton, W., Stewart, J. and Cole, C. [1988], 'Dynamics of C, N, P and S in grassland soils: A model', *Biogeochemistry* **5**, 109–131.

Patt, A. and Siebenhüner, B. [2005], 'Agent based modeling and adaptation to climate change', *Vierteljahreshefte zur Wirtschaftsforschung* **74**(2), 310–320.

Perez, P., Ardlie, N., Kuneepong, P., Dietrich, C. and Merritt, W. S. [2002], 'CATCHCROP: Modeling crop yield and water demand for integrated catchment assessment in Northern Thailand', *Environmental Modelling & Software* **17**(3), 251 – 259.

Petritsch, R., Hasenauer, H. and Pietsch, S. [2007], 'Incorporating forest growth response to thinning within Biome-BGC.', *Forest Ecology and Management* **242**, 324–336.

Phillips, T. J. and Gleckler, P. J. [2006], 'Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics', *Water Resources Research* **42**(3).

Pielke, Sr., R. A., Adegoke, J. O., Chase, T. N., Marshall, C. H., Matsui, T. and Niyogi, D. [2007], 'A new paradigm for assessing the role of agriculture in the climate system and in climate change', *Agricultural and Forest Meteorology* **142**, 234–254.

Poe, G. L. [1997], 'Extra-market values and conflicting agriculture environmental policies', *Choices* **12**(3), 4–8.

Poeter, E. P. and Hill, M. C. [1997], 'Inverse models: A necessary next step in ground-water modeling', *Ground Water* **35**(2), 250–260.

Pope, R. D. and Just, R. E. [1991], 'On testing the structure of risk preferences in agricultural supply analysis', *American Journal of Agricultural Economics* **73**(3), 743–748.

Pope, R. D. and Just, R. E. [2002], 'Random profits and duality', *American Journal of Agricultural Economics* **84**(1), 1–7.

Popper, K. [1963], *The Open Society and its Enemies*, Princeton University Press, Princeton.

Priesack, E., Achatz, S. and Stenger, R. [2001], Parametrisation of soil nitrogen transport models by use of laboratory and field data., *in* M. Shaffer, M. Liwang and S. Hansen, eds, 'Modeling Carbon and Nitrogen Dynamics for Soil Management', CRC Press Inc., Boca Raton, Florida, pp. 461–484.

Priesack, E., Gayler, S. and Hartmann, H. [2006], 'The impact of crop growth sub-model choice on simulated water and nitrogen balances.', *Nutrient Cycling in Agroecosystems* **75**, 1–13.

Pujol, G., Iooss, B. and Janon, A. [2012], *R Package 'sensitivity'*.
http://cran.r-project.org/web/packages/sensitivity/index.html

Quang, D. V., Schreinemachers, P. and Berger, T. [2014], 'Ex-ante assessment of soil conservation methods in the uplands of Vietnam: An agent-based modeling approach', *Agricultural Systems* **123**, 108 – 119.

Quine, W. V. O. [1951], 'Two dogmas of empiricism', *The Philosophical Review* **60**, 20–43.

Rabin, M. [1998], 'Psychology and economics', *Journal of Economic Literature* **36**(1), 11–46.

Randall, D., Wood, R., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R., Sumi, A. and Taylor, K. [2007], Climate models and their evaluation., *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M.Tignor and H. Miller, eds, 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., chapter 8, pp. 589–662.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L. and Vanrolleghem, P. A. [2007], 'Uncertainty in the environmental modelling process - a framework and guidance', *Environmental Modelling & Software* **22**(11), 1543 – 1556.

Reichert, P. and Omlin, M. [1997], 'On the usefulness of overparameterized ecological models', *Ecological Modelling* **95**(2-3), 289–299.

Reidsma, P., Ewert, F., Lansink, A. O. and Leemans, R. [2010], 'Adaptation to climate change and climate variability in European agriculture: The importance of farm level responses', *European Journal of Agronomy* **32**, 91–102.

Reynolds, R. G., Kohler, T. A. and Kobti, Z. [2003], 'The effects of generalized reciprocal exchange on the resilience of social networks: An example from the Prehispanic Mesa Verde Region', *Computational & Mathematical Organization Theory* **9**, 227–254.

Ritchie, J. [1985], A user-oriented model of the soil water balance of wheat, *in* A. R. Day, W., ed., 'Wheat Growth and Modelling.', NATO ASI Series A: Life Sciences, Plenum Press, New York, p. 293–305.

Ritchie, J., Singh, U., Godwin, D. and Bowen, W. [1998], Cereal growth, development and yield, *in* G. Tsuji, G. Hoogenboom and P. Thornton, eds, 'Understanding Options for Agricultural Production. Systems Approaches for Sustainable Agricultural Development.', Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 79–98.

Ritchie, J. T. [1991], Wheat phasic development., *in* J. Hanks and J. Ritchie, eds, 'Modeling Plant and Soil Systems', Vol. 31 of *Agronomy Monograph*, ASA-CSSA-SSSA, Madison, WI, USA, pp. 31–54.

Rizzo, M. L. and Szekely, G. [2013], 'energy : E-statistics (energy statistics). R package version 1.6'. *http://cran.r-project.org/web/packages/energy/index.html*

Robinson, D. T., Brown, D. G., Parker, D. C., Schreinemachers, P., Janssen, M. A., Huigen, M., Wittmer, H., Gotts, N., Promburon, P., Irwin, E., Berger, T., Gatzweiler, F. and Barnaud, C. [2007], 'Comparison of empirical methods for building agent-based models in land use science', *Journal of Land Use Science* **2**(1), 31–55.

Rosenzweig, C. and Parry, M. [1994], 'Potential impact of climate change on world food supply', *Nature* **367**, 133–138.

Rosenzweig, C. and Tubiello, F. [2007], 'Adaptation and mitigation strategies in agriculture: An analysis of potential synergies', *Mitigation and Adaptation Strategies for Global Change* **12**, 855–873.

Rötter, R. and van de Geijn, S. [1999], 'Climate change effects on plant growth, crop yield and livestock', *Climatic Change* **43**, 651–681. 10.1023/A:1005541132734.

Rykiel, E. J. [1996], 'Testing ecological models: The meaning of validation', *Ecological Modelling* **90**, 229–244.

Saltelli, A. and Annoni, P. [2010], 'How to avoid a perfunctory sensitivity analysis', *Environmental Modelling & Software* **25**, 1508–1517.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. [2004], *Sensitivity analysis in practice. - A guide to assessing scientific models.*, Chichester:Wiley.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F. and Bura, S. [1997], 'SIMPOP: A multiagent system for the study of urbanism', *Environment and Planning B: Planning and Design* **24**(2), 287–305.

Saqalli, M., Gérard, B., Bielders, C. and Defourny, P. [2011], 'Targeting rural development interventions: Empirical agent-based modeling in Nigerien villages', *Agricultural Systems* **104**(4), 354 – 364.

Sauer, N. and Hardeweg, B. [2006], *Standarddeckungsbeiträge (SDB) - Kalkulation der Rechenwerte zur Betriebsklassifizierung nach der EU-Typologie*, Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (KTBL).

Schaeffli, B. and Gupta, H. V. [2007], 'Do Nash values have value?', *Hydrological Processes* **21**, 2075–2080.

Schlenker, W. [2010], Crop response to climate and weather: Cross-section and panel models, *in* D. Lobell and M. Burke, eds, 'Climate Change and Food Security', Vol. 37 of *Advances in Global Change Research*, Springer, chapter 6, pp. 99–108.

Schoups, G. and Vrugt, J. A. [2010], 'A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors', *Water Resources Research* **46**(W10531), 1–17.

Schreinemachers, P. and Berger, T. [2006], 'Land-use decisions in developing countries and their representation in multi-agent systems', *Journal of Land Use Sciences* **1**, 29–44.

Schreinemachers, P. and Berger, T. [2011], 'MP-MAS: An agent-based simulation model of human-environment interaction in agricultural systems', *Environmental Modelling & Software* **26**, 845–859.

Schreinemachers, P., Berger, T. and Aune, J. [2007], 'Simulating soil fertility and poverty dynamics in Uganda: A bio-economic multi-agent systems approach.', *Ecological Economics* **64**(2), 387–401.

Schreinemachers, P., Berger, T., Sirijinda, A. and Praneetvatakul, S. [2009], 'The diffusion of greenhouse agriculture in Northern Thailand: Combining econometrics and agent-based modeling', *Canadian Journal of Agricultural Economics* **57**(4), 513–536.

Schreinemachers, P., Potchanasin, C., Berger, T. and Roygrong, S. [2010], 'Agent-based modeling for ex-ante assessment of tree crop innovations: Litchis in northern Thailand', *Agricultural Economics* **41**(6), 519–536.

Schulla, J. and Jasper, K. [2000], Model description WaSiM-ETH, Technical report, IAC ETH Zurich.

Schweizer, B. and Sklar, A. [2011], Copula, *in* 'Encyclopedia of Mathematics.', Springer. *http://www.encyclopediaofmath.org/index.php?title=Copula&oldid=19072*

Sengupta, R., Lant, C., Kraft, S., Beaulieu, J., Peterson, W. and Loftus, T. [2005], 'Modeling enrollment in the conservation reserve program by using agents within spatial decision support systems: An example from Southern Illinois', *Environment and Planning B: Planning and Design* **32**(6), 821–834.

Shiffrin, R. M., Lee, M. D., Kim, W. and Wagenmakers, E.-J. [2008], 'A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods', *Cognitive Science* **32**(8), 1248–1284.

Simon, C. and Etienne, M. [2010], 'A companion modelling approach applied to forest management planning', *Environmental Modelling & Software* **25**(11), 1371 – 1384.

Smit, B., Burton, I., Klein, R. J. and Wandel, J. [2000], 'An anatomy of adaptation to climate change and variability', *Climatic Change* **45**, 223–251.

Smit, B. and Skinner, M. W. [2002], 'Adaptation options in agriculture to climate change: A typology', *Mitigation and Adaptation Strategies for Global Change* **7**, 85–114. 10.1023/A:1015862228270.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., O'Mara, F., Rice, C., Scholes, B., Sirotenko, O., Howden, M., McAllister, T., Pan, G., Romanenkov, V., Schneider, U., Towprayoon, S., Wattenbach, M. and Smith, J. [2008], 'Greenhouse gas mitigation in agriculture', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1492), 789–813.

Sorda, G., Sunak, Y. and Madlener, R. [2013], 'An agent-based spatial simulation to evaluate the promotion of electricity from agricultural biogas plants in Germany', *Ecological Economics* **89**, 43 – 60.

Spear, R. and Hornberger, G. [1980], 'Eutrophication in Peel inlet—II. identification of critical uncertainties via generalized sensitivity analysis', *Water Research* **14**(1), 43 – 49.

Stanley, T. D. [1998], 'Empirical economics? An econometric dilemma with only a methodological solution', *Journal of Economic Issues* **32**(1), 191–218.

Statististisches Landesamt Baden-Württemberg [2012], 'Struktur- und Regionaldatenbank', http://www.statistik.baden-wuerttemberg.de/SRDB/home.asp.

Statististisches Landesamt Baden-Württemberg [2013], 'Besitz- und Pachtverhältnisse landwirtschaftlicher Betriebe'.
http://www.statistik.baden-wuerttemberg.de/Landwirtschaft/Landesdaten/BesitzPacht.asp

Stedinger, J. R., Vogel, R. M., Lee, S. U. and Batchelder, R. [2008], 'Appraisal of the generalized likelihood uncertainty estimation (GLUE) method', *Water Resources Research* **44**(W00B06), 1–17.

Stenger, R., Priesack, E., Barkle, G. and Sperr, C. [1999], Expert-N. A tool for simulating nitrogen and carbon dynamics in the soil-plant-atmosphere system, *in* M. Tomer, M. Robinson and G. Gielen, eds, 'NZ Land Treatment Collective. Proceedings Technical Session 20: Modelling of Land Treatment Systems.', New Plymouth, pp. 19–28.

Szekely, G. J. and Rizzo, M. L. [2013], 'Energy statistics: A class of statistics based on distances', *Journal of Statistical Planning and Inference* **143**(8), 1249 – 1272.

Szekely, G., Rizzo, M. L. and Bakirov, N. K. [2007], 'Measuring and testing dependence by correlation of distances', *The Annals of Statistics* **35**(6), 2769–2794.

Taub, D. R., Miller, B. and Allen, H. [2008], 'Effects of elevated CO_2 on the protein concentration of food crops: A meta-analysis', *Global Change Biology* **14**(3), 565–575.

Tebaldi, C. and Knutti, R. [2010], Climate models and their projections of future changes, *in* D. Lobell and M. Burke, eds, 'Climate Change and Food Security - Adapting agriculture to a warmer world', Vol. 37 of *Advances in Global Change Research*, Springer, Dordrecht, NL, chapter 3.

Tesfatsion, L. [2006], Agent-based computational economics: A constructive approach to economic theory, *in* L. Tesfatsion and K. Judd, eds, 'Handbook of Computational Economics', Vol. 2, Elsevier, chapter 16, pp. 831 – 880.

Troost, C. [2009], Applying multi-agent modeling for policy analysis: Public subsidies for advanced irrigation technologies in the Province of Linares, Chile, M.Sc. thesis, Universität Hohenheim.

Troost, C. and Berger, T. [2014], 'Dealing with uncertainty in agent-based simulation: Farm-level modeling of adaptation to climate change in southwest Germany', *American Journal of Agricultural Economics* . (in press).
http://dx.doi.org/10.1093/ajae/aau076

Troost, C., Berger, T. and Latynskiy, E. [2010], Building integrated assessment tools for climate change adaptation in agriculture - A case for coupled land model systems at regional scale with fine resolution, *in* '117th Seminar of the EAAE: Climate Change, Food Security and Resilience of Food and Agricultural Systems in Developing Countries: Mitigation and Adaptation Options, November 25-27, 2010. Stuttgart-Hohenheim'.

Troost, C., Calberto, G., Berger, T., Ingwersen, J., Priesack, E., Warrach-Sagi, K. and Walter, T. [2012], Agent-based modeling of agricultural adaptation to climate change in a mountainous area of South West Germany, *in* R. Seppelt, A. Voinov, S. Lange and D. Bankamp, eds, 'International Environmental Modelling and Software Society (iEMSs): 2012 International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany'.
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Valbuena, D., Verburg, P., Bregt, A. and Ligtenberg, A. [2010], 'An agent-based approach to model land-use change at a regional scale', *Landscape Ecology* **25**, 185–199.

van Oel, P. R., Krol, M. S., Hoekstra, A. Y. and Taddei, R. R. [2010], 'Feedback mechanisms between water availability and water use in a semi-arid river basin: A spatially explicit multi-agent simulation approach', *Environmental Modelling & Software* **25**(4), 433–443.

Vanclay, J. [1994], *Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests.*, CAB International, Wallingford.

Voas, D. and Williamson, P. [2001], 'Evaluating goodness-of-fit measures for synthetic microdata', *Geographical & Environmental Modelling* **5**(2), 177–200.

Wagenmakers, E.-J., Grünwald, P. and Steyvers, M. [2006], 'Accumulative prediction error and the selection of time series models', *Journal of Mathematical Psychology* **50**(2), 149 – 166.

Walker, W., Harremoes, P., Rotmans, J., van der Sluijs, J., van Asselt, M. and Janssen, P. and, K. v. K. M. [2003], 'Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support', *Integrated Assessment* **4**(1), 5–17.

Weisbuch, G. and Boudjema, G. [1999], 'Dynamical aspects in the adoption of agri-environmental measures.', *Advances in Complex Systems* **2**, 11–36.

WETTREG [2010], 'WR2010_EH5_1_A1B: UBA-WETTREG ECHAM5/OM 20C + A1B Lauf 1: 1961-2100', http://cera-www.dkrz.de/WDCC/ui/Project.jsp?proj=WETTREG2010.

White, J. W. and Hoogenboom, G. [2010], Crop response to climate: Ecophysiological models, *in* D. Lobell and M. Burke, eds, 'Climate Change and Food Security', Vol. 37 of *Advances in Global Change Research*, Springer, chapter 4, pp. 59–83.

Whittaker, G., Confesor, R., Di Luzio, M. and Arnold, J. G. [2010], 'Detection of overparameterization and overfitting in an automatic calibration of SWAT', *Transactions of the American Society of Agricultural and Biological Engineers* **53**(5), 1487–1499.

Wilby, R. L. [1997], 'Non-stationarity in daily precipitation series: Implications for GCM down-scaling using atmospheric circulation indices', *International Journal of Climatology* **17**(4), 439–454.

Wilby, R. and Wigley, T. [1997], 'Downscaling general circulation model output: A review of methods and limitations', *Progress in Physical Geography* **21**(4), 530–548.

Willmott, C. [1981], 'On the validation of models', *Physische Geographie* **2**, 184–194.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J. and Rowe, C. M. [1985], 'Statistics for the evaluation and comparison of models,', *Journal of Geophysical Research* **90**(C5), 8995–9005.

Willmott, C. J. and Matsuura, K. [2005], 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', *Climate Research* **30**, 79–82.

Windrum, P., Fagiolo, G. and Moneta, A. [2007], 'Empirical validation of agent-based models: Alternatives and prospects', *Journal of Artificial Societies and Social Simulation* **10**(2).
http://jasss.soc.surrey.ac.uk/10/2/8.html

Wreford, A., Moran, D. and Adger, N. [2010], *Climate change and agriculture: impacts, adaptation, and mitigation.*, OECD, Paris.

Yapo, P. O., Gupta, H. V. and Sorooshian, S. [1998], 'Multi-objective global optimization for hydrologic models', *Journal of Hydrology* **204**(1-4), 83 – 97.

Yin, X. and van Laar, H. [2005], *Crop Systems Dynamics: An ecophysiological model for genotype-by-*

environment interactions, Wageningen Academic Publishers, Wageningen, The Netherlands.

Ziervogel, G., Bithell, M., Washington, R. and Downing, T. [2005], 'Agent-based social simulation: A method for assessing the impact of seasonal climate forecast applications among smallholder farmers', *Agricultural Systems* **83**(1), 1 – 26.

Zimmermann, A., Heckelei, T. and Pérez Domínguez, I. [2009], 'Modelling farm structural change for integrated ex-ante assessment: Review of methods and determinants', *Environmental Science & Policy* **12**(5), 601–618.

Ziska, L. H., Blumenthal, D. M., Runion, G. B., Hunt, E. R. and Diaz-Soltero, H. [2011], 'Invasive species and climate change: An agronomic perspective', *Climatic Change* **105**(1-2), 13–42.

Zucchini, W. [2000], 'An introduction to model selection', *Journal of Mathematical Psychology* **44**(1), 41–61.

Appendix

Appendix A

Symbology used in mathematical programming equations

A.1 General remarks

For the description of the equations of the mathematical programming problems, the following symbols are used: The letter x is used for decision variables, the letter b for capacities, the letter a for the coefficients of decision variables in a constraint equation and the letter c for the coefficients of the decision variables in the objective function. M stands for a high value representing infinity, which is chosen high enough to completely relax the associated restriction, but low enough not to cause any problems for the branch-and-bound algorithm.

The type of decision variable is indicated by the superscript, which usually refers to a set, denoted by a capital letter. The specific member of the set to which a decision variable is associated is denoted by a lowercase letter in the subscript. E.g. L denotes the set of land use activities. The $\mathbf{x}^L$ is the vector of areas of all land use activities and x_l^L is the area of land use activity $l \in L$.

Subsets of sets are denoted by adding lowercase letters to the capital letter denoting the superset, e.g. Lg denotes the set of all grassland and field grass production activities, which is a subset of L.

In some cases, a decision variable is associated with several sets, e.g. $x_{e,t,k}^{E,T,K}$ is the number of hours equipment e is used in work season t for work with weather sensitivity k . In other cases, several decision variables are associated to the same set and a small letter is added to the superscript in order to allow a distinction, e.g. x_g^{sG} and x_g^{bG} denote sales and purchase of good g, respectively.

A similar convention is used for coefficients and capacities. E.g. b_e^E denotes the number of equipments of type e owned, $a_{l,g}^{LG}$ the amount of good g yielded by land use activity l, and c_g^{sG} the sales price of good g.

The symbol $\forall$ to the right of a displayed constraint equation is used to indicate that this type of relation is repeated for each member of the indicated set(s).

Symbols marked with a tilde˜denote expected, rather than actual values.

A.2 List of symbols

Table A.1: *Overview of sets in the CSA decision model*

Symbol		Description
A		types of animal production
B		investment
C		cash
	Cf	fix costs, debt payments, depreciation,rental payments
	Cmc	minimum household consumption
D		feeding season
E		equipment
G		consumable goods
	Gb	goods that can be used for biogas production
	Ge	biogas electricity
	Gc	crops
	Gg	fresh grass (cut or pastured)
	Ggp	fresh grass (pastured)
	Gi	pure inputs
	Gm	maize
	Gn	untraded intermediates
	Go	manure
	Gs	pure products
	Gt	traded intermediates
H		labor
J		
	Ja	animal groups for feeding
	Jf	crop groups for following position in rotation
	Jp	crop groups for preceding position in rotation
	Jr	crop groups for overall rotational limit
	Js	crop groups for rotational limit on self-following
	Jym	crop groups for MEKA diversification measures
K		weather sensitivity level
L		land use activity
	$Lb4$	grassland, with one conservation cut per season
	Lg	grassland and field grass production
	Lgp	grassland and field grass production used for pasture only
	Lgg	grassland, cross-compliance conformant
	$Lgg1$	grassland, one use per season
	$Lgg2$	grassland, two uses per season
	$Lmai$	maize
	Lmf	land use counted as main forage area
	$Loil$	oilseeds
M		infrastructure and machinery for animals
N		nutrients
	N_b	basic nutrients
	N_l	nutrients with a lower limit on dry matter share
	N_r	nutrients with a two-sided limit

Table A.1: *Overview of sets in the CSA decision model [cont.]*

Symbol		Description
	N_u	nutrients with an upper limit on dry matter share
O		services for animals
P		Tractive power class
S		soil types
T		work season
U		biogas production
V		types of services for animal production
W		types of field work
	Wh	types of work that can be contracted
Y		subsidy & policy related
	Ya	single farm payment entitlements, arable, EU MTR
	Yb	special male cattle premium, EU Agenda 2000
	Yc	crop premium, EU Agenda 2000
	Yd	milk quota
	Ye	extensification premium, EU Agenda 2000
	Yg	single farm payment entitlements, set-aside, EU MTR
	Yg	single farm payment entitlements, grassland, EU MTR
	Yk	slaughter premium cattle, EU Agenda 2000
	Ym	commitments eligible under MEKA
	Yo	mother cow premium, EU Agenda 2000
	Ys	set aside premium, EU Agenda 2000
	Yu	size limits for rewarding biogas electricity through EEG
	Yx	small manure biogas plant, EEG 2012
	Yy	year of establishment of biogas plant
	Yz	relationship of remuneration classes for biogas electricity, EEG 2012
Z		Tractor class

Appendix B

Model Documentation: Submodels

B.1 The farm decision model

In the following subsections, the elements of the mixed integer programming (MIP) problem (decision variables, objective function and constraint equations) which was developed to represent production, investment, harvest and rental decisions are described in detail.

B.1.1 Objective function

The objective function that agents maximize results from subtracting the sum of all planning-dependent cost from the sum of all revenue. Revenue can be created by selling goods ($\mathbf{x}^{sG}$) , receiving interest on deposits ($\mathbf{x}^{dC}$), receiving premiums awarded by different policy schemes ($\mathbf{x}^{Y}$), and selling biogas electricity ($\mathbf{x}^{sYyYu}$, $\mathbf{x}^{osYyYu}$). Costs result from the purchase of goods ($\mathbf{x}^{bG}$), the use of machinery and buildings ($\mathbf{x}^{ZTK}$, $\mathbf{x}^{\alpha O}$, $\mathbf{x}^{\beta O}$, $\mathbf{x}^{\beta tM}$, $x^{\beta mUe}$), hiring permanent ($\mathbf{x}^{bpH}$) and temporary labor ($\mathbf{x}^{btHTK}$), payment of interest on short-term credit ($\mathbf{x}^{bC}$) and the direct cost of land use ($\mathbf{x}^{L}$) and animal production activities ($\mathbf{x}^{A}$).

Debt payments on assets bought in the past are omitted from the objective function, which thus represents expected total farm gross margin ($\tilde{\pi}_{tgm}$) rather than expected farm income ($\tilde{\pi}$). However, as debt payments are considered planning independent fix cost, maximizing the total gross margin function is equivalent to maximizing total income.

The complete objective function is shown in equation B.1. Explanation of the individual decision variables and the associated objective function coefficients are given throughout the subsequent sections. A comprehensive overview and explanation of symbols used in the MIP equations is also given in appendix section A.

$$
\begin{aligned}
\max!\quad \tilde{\pi}_{tgm} = \\
&\sum_{g} \tilde{c}_g^{sG} \tilde{x}_g^{sG} + \sum_{y} c_y^{Y} x_y^{Y} + c^{dC} x^{dC} + \sum_{y_u,y_y} \left(c_{y_u,y_y}^{sYyYu} x_{y_u,y_y}^{sYyYu} \right) + \sum_{y_u,y_y} \left(c_{y_u,y_y}^{osYyYu} x_{y_u,y_y}^{osYyYu} \right) \\
&+ c^{sYx} x^{sYx} + \sum_{y_u,y_y,y_z} \left(c_{y_u,y_y,y_z+1}^{slYyYuYz} x_{y_u,y_y,y_z+1}^{slYyYuYz} + c_{y_u,y_y,y_z}^{slYyYuYz} x_{y_u,y_y,y_z}^{suYyYuYz} \right) \\
&- \sum_{g} c_g^{bG} x_g^{bG} - \sum_{l} \left(c_l^{L} x_l^{L} \right) - \sum_{a} \left(c_a^{A} x_a^{A} \right) - c^{pH} x^{bpH} - \sum_{t,k} c_{t,k}^{tH} x_{t,k}^{btHTK} - c^{bC} x^{bC} \\
&- \sum_{z,t,k} \left(c_z^{Z} x_{z,t,k}^{ZTK} \right) - \sum_{o} \left(c_o^{\alpha O} x_o^{\alpha O} \right) - \sum_{o} \left(c_o^{\beta O} x_o^{\beta O} \right) - \sum_{m} \left(c_m^{\beta M} x_m^{\beta tM} \right) - c^{\beta mUe} x^{\beta mUe}
\end{aligned}
\tag{B.1}
$$

When considering the employment of a potential successor, the labor cost for employing the family member is part of the total gross margin, but is counterbalanced by the utility of employing the successor as described in section B.2.3. The objective function then differs from the total gross margin function by the wage paid to the potential successor.

B.1.2 Market interaction & goods balances

The agent interacts with goods markets by selling (x_g^{sG}) or buying goods (x_g^{bG}). The goods balances ensure that the farm agent cannot sell or use more of a good $g \in G$ than he/she bought or produced him/herself. Not all goods can be sold or bought, and only a limited range of goods can be produced by the farm agent itself. Since goods have different potential uses, e.g. as fodder, fertilizer or fuel, many terms in the following general balance equation are actually omitted or their corresponding coefficients are zero for an individual good.

$$x_g^{sG} - x_g^{bG} \pm \sum_l \left(a_{l,g}^{LG} x_l^L\right) \pm \sum_a \left(a_{a,g}^{AG} x_a^A\right) + \sum_{j_a,d} x_{g,j_a,d}^{fGJaD} + x_g^{uG} + \sum_a \left(a_{z,g}^{ZG} x_{z,t,k}^{ZTK}\right) \leq 0 \quad \forall g \tag{B.2}$$

Specifically, the model distinguishes

Pure products (Gs**)** Goods that are only sold by the farm agent, but not bought, e.g. malting barley, rapeseed, milk, meat.

Pure inputs (Gi**)** Goods that are only bought by the farm agent, but not sold, e.g. fuel, soybean meal and other industrial fodder.

Traded intermediates (Gt**)** Goods that can be both sold and bought by the farm, e.g. fodder barley, fodder wheat, young animals.

Non-traded intermediates (Gn**)** Goods that are produced by one process and used as an input for another process on the farm, but not traded, e.g. hay, grass silage.

Manure (Go**)** Non-traded intermediates with specific treatment due to their potential use in biogas plants (see section B.1.6).

Fresh grass (Gg**)** Non-traded intermediates with specific treatment due to their only seasonal availability (see B.1.3).

For some cases, e.g. silage maize, the group a good falls into is varied according to the assumptions embodied in a specific parameter combination selected during calibration (see Ch. 8).

B.1.3 Land use: crop production and grassland cultivation

Crop production and grassland cultivation are the major land uses considered in the model. Each element l of the vector of land use activities ($\mathbf{x}^L$) represents a combination of a crop g_c, a soil type s, and a management plan. Grassland cultivation and grass/clover cultivation on arable land can have several products as the same area can be used up to four times a year for silage, hay, pasture and cutting of fresh grass.

The model does not explicitly account for the perennial nature of grasslands at the moment, but rather distinguish between arable land and grassland. Conversion of grassland to arable land or vice versa is not considered for simplification. The land use statistics of the area show no significant changes in overall grassland area in the study area between 1999 and 2007 (see ch. 6) and grassland conversion has effectively been forbidden in the state of Baden-Württemberg as of 2011, so this simplification seems justified.

Arable crops

The model includes winter wheat, winter wheat silage, winter rapeseed, winter fodder barley, summer fodder barley, summer malting barley, silage maize and fallow as potential production activities for arable land. Cultivating field grass is also possible on arable land, but due to the several potential harvests the associated management plan is structurally more similar to grassland, and thus described in the next subsection.

Management plans were derived from standard recommendations of German extension services [KTBL, 2010, 2008; LEL, 2012; LfL, 2012] and cross-checked and updated in expert interviews, survey results and observations on the field measurement sites. In general, three fertilization schemes (only mineral fertilizer, with pig manure and with cattle manure) and two tillage regimes (plough tillage and low tillage using rotary tillers) are distinguished in the model. For the two summer crops, summer barley and silage maize, management plans with and without winter cover crops (field mustard) are included. Different levels of pesticide use are not distinguished, rather a standard plant protection practice for each crop is assumed, as it was not possible to simulate or estimate the yield effect of pesticide use. For some EU Agenda 2000 support schemes, it is necessary to ensure a certain use of the product (see B.1.9). Whenever this applies, this commitment is also considered part of the management plan.

Management plans determine the quantity of physical inputs required, the necessary field work (tillage, sowing, fertilization, plant protection, harvest) and its timing. Except for animal manure, physical inputs are multiplied by prices and aggregated to direct cost, which enter the MIP as the objective function coefficient (c_l^L) of the corresponding land use activity (x_l^L).

Management plans do affect expected ($\tilde{a}_{l,g}^{LG}$) and actual yields ($a_{l,g}^{LG}$), which are discussed in detail in section 8.1.1.

Grassland & field grass production

For grassland cultivation, four intensity levels of production are distinguished in the model:

Level	Description	Use	Production [t dm /a]
0	Abandoned	Not even minimum requirements of cross-compliance fulfilled	0
1	Very extensive use	late cuts, extensive pasture	25
2	Extensive grassland use	a maximum of two cuts per year	62
3	Intensive grassland use	a maximum of three cuts per year	83

Potential uses of grassland are grazing (G), cutting fresh grass for direct feeding (C), production of grass silage (S) or production of hay (H). Combinations of one type of fresh and one type of conserved fodder production on a single grassland plot are possible. In this case, the harvest of conserved fodder always precedes harvests of fresh fodder. The maximum number of uses is determined by the intensity level. Thus, SG, SS, SC, HH, HC, HG, G, C[1]

[1]For simplification, letters C and G are not repeated and always refer to all further potential uses till the end of the season

are the potential combinations of uses for a plot managed at intensity level 2, and SSS, SSG, SSC, SG, SC, HHH, HHC, HC, HHG, HG, G, C are potential use combinations for intensity level 3. Additional grazing at the end of the season is always possible for both intensity levels. For very extensive grassland (level 1), three use options were included: year-round pasture, late cut (beginning of July, every two years) and very late cut (beginning of October, every two years). Abandoned grassland is not used by agents at all.

The implementation of field grass production is structurally similar to grassland, the only difference being that it is not restricted to grassland plots, but rather to arable plots. Only one intensity level is considered for field grass, which allows up to four uses a year.

Grass yields of the individual uses are determined using simple regrowth parameters that relate daily regrowth during a specific half month to total expected dry matter production in the year. These parameters were calculated from data given in Berendonk [2011]. The yield obtained by a specific use is then the total regrowth between the date of the harvest and the date of the previous harvest, respectively the beginning of the growing season. Total annual dry matter production depends on the intensity level. The specific dates for the individual uses depends on both, the intensity level and the combination of uses (use profile) of a grassland plot.

Intensity level and use profile also determine the amount of fertilization and cultivation work. Similar to the implementation of arable crops, three fertilization schemes are included: one with mineral fertilizer only, one with pig manure and one with cattle manure. Again, only manure is treated as an explicit input in the MIP (see B.1.6) and other physical inputs are aggregated to direct costs and form the objective function coefficient (c_l^L) of the corresponding land use activity (x_l^L).

Due to the different potential uses, several different products ($g \in G$) can be obtained from one grassland plot. Additionally, for both, conserved and fresh grass, the model distinguishes fodder obtained from the first cut of the year, and fodder obtained from latter cuts, due to their different nutritional composition.

For fresh grass products ($g \in G_g$), separate balances for different feeding seasons ($d \in D$) are distinguished to take account of the fact that these cannot be stored and are only available at the certain point of time, when they are mature and have been harvested (see B.1.4 for more detail).

The part of the MIP that links grassland and field grass activities to the balances of its products and the objective function, thus looks like this:

$$\begin{array}{lllllcl}
\sum\limits_g c_g^{sG} x_g^{sG} & -\sum\limits_g c_g^{bG} x_g^{bG} & -\sum\limits_l \left(c_l^L x_l^L\right) & +\dots & & \rightarrow & \text{obj} \\
x_g^{sG} & -x_g^{bG} & -\sum\limits_l \left(a_{l,g}^{LG} x_l^L\right) & +\sum\limits_{j_a,d} x_{g,j_a,d}^{fGJaD} & +x_g^{uG} & \leq & 0 \quad \forall g \notin Gg \\
 & & -\sum\limits_l \left(a_{l,g,d}^{LG} x_l^L\right) & +\sum\limits_{j_a} x_{g,j_a,d}^{fGJaD} & & \leq & 0 \quad \forall (g \in Gg, d)
\end{array} \tag{B.3}$$

While all of the grassland areas become available for grazing at a certain point of the season, most farmers in the study area often do not make use of this option as pasturing can be quite labor intensive. For this reason, work for cutting grass is accounted for with the land use activity, but work for pasturing is accounted for at the respective feeding activities and thus only required if the area is actually used for pasturing. This implementation, however, would allow a farm agent to declare a grassland plot to be pasture only to fulfill cross compliance requirements, but then not use it at all avoiding both the work for pasturing and conservation cuts. To avoid this, an additional constraint is introduced requiring that at least

50% of the grass available in the second feeding period is actually pastured, whenever a plot is declared pasture only ($l \in Lgp$).

$$0.5 \sum_{l \in Lp} \left(a^{LG}_{l,g,d} x^{L}_{l} \right) - \sum_{j_a} x^{fGJaD}_{g,j_a,d} \leq 0 \quad \forall (g_g \in Ggp, d = 2) \tag{B.4}$$

Crop rotation

Land use activities are obviously constrained by the area of a certain soil type available to each individual agent (b^S_s). More precisely, crops do either require a corresponding part of the soil to be incorporated into crop rotation (x^{rS}_s), or kept out of the crop rotation (x^{nS}_s), depending on whether they form part of the crop rotation or not.

$$\begin{array}{rrrcl} & x^{rS}_s & +x^{nS}_s & = & b^S_s \quad \forall s \\ \sum_l \left(1^{LS}_{s,l} 1^{LrS}_{s,l} x^L_l \right) & -x^{rS}_s & & = & 0 \quad \forall s \\ \sum_l \left(1^{LS}_{s,l} 1^{LnS}_{s,l} x^L_l \right) & & -x^{nS}_s & = & 0 \quad \forall s \end{array} \tag{B.5}$$

Cultivation of arable crops has to respect crop rotation rules. Following good agricultural practice and as observed in the study area, it is assumed that farmers tend to implement production plans that can – at least in theory – be upheld for several years without violating crop rotational rules. Two types of rotation rules are distinguished in the model:

First, there are maximum limits on the share of a crop or crop group in the rotation, e.g. if a crop should be grown maximum once every three years a maximum of 33% of the arable area should be cultivated with this crop. Crops under consideration were classified into rotation groups (Jr), for each of which a specific limit a_{j_r} applies. This limit is multiplied with the part of the soil that is included into the crop rotation x^{rS}_s to give the maximum area of crops of this group that can be grown by the agent.

$$\sum_l \left(1^{LS}_{s,l} 1^{LJr}_{l,j_r} x^L_l \right) - a^{Jr}_{j_r} x^{rS}_s \leq 0 \quad \forall (s, j_r) \tag{B.6}$$

Second, it may not be recommendable to grow crop A after crop B for plant health reasons or it may even be impossible due to incompatible timing of sowing A and harvesting B in the study area. Two classifications of land use activities were created, which group together land use activities with similar characteristics as a preceding land use (Jp), respectively as a following land use (Jf). This classification is not only crop-, but also management-specific, as different management plans for the same crop may entail different timings and thus affect compatibility with other crops. Compatibility coefficients $1^{Jpf}_{j_p,j_f}$ for each combination of ($j_p \in Jp, j_f \in Jf$) were then created, which may take a value of 1 indicating compatibility or 0 meaning noncompatibility.

$$\begin{array}{rrcl} -\sum_l \left(1^{LS}_{s,l} 1^{LJp}_{l,j_p} x^L_l \right) & +\sum_{j_f} x^{SJpJf}_{j_p,j_f,s} & \leq & 0 \quad \forall (s, j_p) \\ \sum_l \left(1^{LS}_{s,l} 1^{LJf}_{l,j_f} x^L_l \right) & -\sum_{j_p} \left(a^{Jpf}_{j_p,j_f} x^{SJpJf}_{j_p,j_f,s} \right) & \leq & 0 \quad \forall (s, j_f) \end{array} \tag{B.7}$$

Some crops are compatible with themselves, but should not be repeated on the same plot more than a certain number of times (n) in a row. In this case, all sequences containing only crop Js require the inclusion of sufficient other preceding-following crop relations. E.g. if a crop is to follow itself maximum once, each hectare where it is grown after itself has to be complemented by another hectare, where it is grown after another crop.

The corresponding coefficient $a^{JsJpJf}_{j_p,j_f,j_s}$ is -1 for all relations including crop j_s only as a following crop, and $\frac{n+1}{n} - 1$ for all relations, where crop j_s follows itself.

$$\sum_{j_p,j_f,j_s} \left(a^{JsJpJf}_{j_s,j_p,j_f} x^{SJpJf}_{j_p,j_f,s} \right) \leq 0 \quad \forall (s, j_s) \tag{B.8}$$

Field work & weather dependency

Every land use activity x^L_l requires certain types of field work ($w \in W$) to be executed at certain points of time. The model distinguishes nine work seasons ($t \in T$) comprising between one and seven half-months (with fine resolution in summer and coarse resolution in winter). Each type of field work requires a different amount of time and tractor power depending on the equipment used and the resistance of the soil. Field work activities ($x^{WTSqE}_{w,t,s_q,e}$ [h]) are therefore combinations of a type of work w, the equipment e used, the work season t and the soil resistance class s_q .

Farm agents can do field work using own machinery or by contracting external machinery and workers (x^{bWTSq}_{w,t,s_q}). The amount of field work has to be balanced for each type of work, work season and soil resistance class as shown in equation B.9,

$$\sum_{l} \left(a^{LWT}_{l,w,t} 1^{LSq}_{l,s_q} x^L_l \right) - \sum_{w,e} \left(a^{WTSqE}_{w,e} x^{WTSqE}_{w,t,s_q,e} \right) - x^{bWTSq}_{w,t,s_q} \leq 0 \quad \forall (w, t, s_q) \tag{B.9}$$

with $a^{LWT}_{l,w,t}$ being the amount of field work w required in work season t for land use activity x^L_l, $a^{WTSqE}_{w,e}$[ha $\cdot$ h^{-1}] being the area that can be worked when using equipment e for work type w for an hour, and $1^{LSq}_{l,s_q}$ being an indicator function that links land use activities to soil resistance classes.

The amount of work with own machinery that can be done in a work season is limited by the number of equipments and tractors owned by the agent, the amount of labor available and the number of days with suitable weather for the type of work to be done. KTBL [2010] provides a division of Germany into climatic regions and an estimate of expected days for field work of different weather sensitivity levels (k) for each region and half-month of the growing season. Following this approach, five levels of weather sensitivity are distinguished: (i) cereal harvest; (ii) hay harvest (soil dried); [(iii) hay harvest (shed dried)] [2]; (iv) harvest of grass silage; (v) medium sensitive activities such as harvest of silage maize, mineral fertilization, and sowing; and finally (vi) less sensitive activities such as organic fertilization and incorporation of crop residues into the soil. Based on this, the coefficient $a^{TK}_{t,k}$ representing the available hours for field work of level k in work season t was calculated by assuming nine work hours per day.

The indicator function $1^{WK}_{w,k}$ links every type of field work to the corresponding weather sensitivity level, but also to all levels representing lower weather sensitivity: A day that is suitable for field work of level (v) is also suitable for work of level (vi), and if a person performs level (v) work it cannot simultaneously perform level (vi) work.

[2]Currently not used in the model.

Equipment capacity The capacity constraints for equipment e are then represented by the following linear equations (B.10), with b_e^E being the number of equipments of type e owned.

$$\begin{array}{rrcl} \sum_{w,s_q}\left(1_{w,k}^{WK}x_{w,t,s_q,e}^{WTSqE}\right) - a_{t,k}^{TK}x_e^{tE} & \leq & 0 & \forall(k,t,e) \\ x_e^{tE} & \leq & b_e^E & \forall e \end{array} \tag{B.10}$$

Tractor capacity In a similar fashion, field work is constrained by available tractor capacity. The major difference is that a 83 kW tractor can, of course, also be used for work, which requires only 45 kW of tractor power, although fuel consumption and variable cost will be higher than for a less powerful tractor. To take account of this fact the model contains tractor capacity balances and tractor power balances. The capacity of a tractor type is calculated in a similar way as equipment capacities

$$\begin{array}{rrcl} x_{z,t,k}^{ZTK} - a_{t,k}^{TK}x_z^{tZ} & \leq & 0 & \forall(k,t,z) \\ x_z^{tZ} & \leq & b_z^Z & \forall z \end{array} \tag{B.11}$$

and feeds into the corresponding tractor power balance for each combination of work season and weather sensitivity. The transfer activities x^{PPTK} allow the use of higher tractor power for less power-demanding work, too.

$$\sum_{w,s_q,e}\left(1_{w,k}^{WK}a_{w,s_q,e,p}^{WSqEP}x_{w,t,s_q,e}^{WTSqE}\right) - \sum_{z}\left(1_{z,p}^{ZP}x_{z,t,k}^{ZTK}\right) - x_{p+1,t,k}^{PPTK} + x_{p,t,k}^{PPTK} \leq 0 \quad \forall(k,t,p) \tag{B.12}$$

In any case, fuel consumption ($a_{z,g}^{ZG}$) and maintenance cost (c_z^Z) per hour of tractor use ($x_{z,t,k}^{ZTK}$) depend only on the type of tractor used, not on the type of work done with it. So, whenever available, a smaller tractor will be preferred over a heavier one if both can do the work in question.

Labor capacity As a third restriction, the farm needs to be able to provide the necessary amount of labor during the suitable days. This labor capacity is calculated similarly to equipment and tractor capacity (see section B.1.7).

Restrictions on hiring labor and machinery It is debatable and therefore left to sensitivity analysis and open discussion at this point, in how far contracting of field work is restricted by days with suitable weather. Also contracted work can only be done with suitable weather, however, how much work can actually be done depends on whether the farmer is able to find a provider and the available capacity of the provider.

$$\begin{array}{rcll} x_{t,k}^{btHTK} & \leq & b_{t,k}^{btHTK} & \forall(t,k) \\ x_{w,t,s_q}^{bWTSq} & \leq & b_{w,t,s_q}^{bWTSq} & \forall(w,t,s_q) \end{array} \tag{B.13}$$

The capacity b_{w,t,s_q}^{bWTSq} is calculated as follows:

$$b_{w,t,s_q}^{bWTSq} = \zeta_{\text{proptohire}} * \zeta_{\text{hoursaday}} * a_{t,k}^{TK} * 1_{w,k}^{WK} \tag{B.14}$$

The parameter *proptohire* is subject to calibration.

B.1.4 Animal husbandry

The model includes cattle and pig related animal production activities. For cattle, dairy production, calf raising (0-3 months, male and female), heifer raising (3 - 30 months), bull fattening (3 - 18 months) and suckler cows are included. Except for suckler cows, where specific suckler cow races are considered, each of these activities have been included once for Fleckvieh and Holstein-Friesian races.[3] For pigs, piglet production (< 8 kg), piglet raising (8-28 kg) and pig fattening (28-118 kg) have been included.

Each animal production activity ($a \in A$) is associated with a decision variable (x_a^A) in the MIP, which indicates the number of stable places used. For dairy cows, two different production levels (3000 kg milk /a and 8000 kg/a) are considered. Due to the linear nature of the MIP, any production level between the two levels included is (theoretically) achievable by a linear combination of the two. For all other animal production activities, just one standard specification has been included. The standard specification defines the duration of a turnover and the quantity produced, and correspondingly nutrition requirements, input use, manure production, and work and infrastructure requirements.

Animal production activities usually produce several products (including live animals) that form an input for other animal production activities. E.g. besides milk, dairy production supplies male and female calves. Live animals, nutritional inputs and heating are treated in explicit balances in the MIP, while all other inputs (e.g. water, straw, veterinary costs, insurance) are multiplied by prices and aggregated to direct costs that enter the MIP as the objective function coefficient (c_a^A) of the corresponding animal production activity (x_a^A).

The part of the MIP which links animal production activities to the balances of its products and the objective function, thus looks like this:

$$\begin{array}{llllll} \sum_g c_g^{sG} x_g^{sG} & -\sum_g c_g^{bG} x_g^{bG} & -\sum_a \left(c_a^A x_a^A\right) & +\dots & \rightarrow & \text{obj} \\ x_g^{sG} & -x_g^{bG} & -\sum_l \left(a_{a,g}^{AG} x_a^A\right) & & \leq & 0 \quad \forall g \in Ga \end{array} \tag{B.15}$$

Nutrition

Each animal production activity ($a \in A$) requires the provision of certain quantities ($a_{a,n,d}^{AND}$) of selected basic nutrients ($n_m \in N_m$) in each of six feeding seasons ($d \in D$). Nutrients considered are metabolizable energy (ME), raw protein (XP) and lysine for pigs, net energy lactation (NEL) and usable raw protein (nXP) for dairy cows, and metabolizable energy and usable raw protein for other cattle. For cattle, it is also necessary that the raw fibre content of the fodder ration is high enough, and for dairy cows standard limits on structure value (SV, de Brabander et al. 1999), sugar and starch content, raw fat content and ruminal nitrogen balance (RNB) are applied. Nutrition demand of animals was taken from LfL [2010, 2011] and KTBL [2010]. Balances for these nutrients are distinguished for feeding groups ($j_a \in Ja$), each of which comprises several animal production activities.

The agent is free to choose any suitable combination of bought or self-produced fodder in order to satisfy the nutrient demand of its animals. The feeding decision is represented in the MIP by the vector of feeding activities ($x_{g,j_a,d}^{fGJaD}$), which indicate the quantity of a product g fed to a feeding group j_a in feeding season d. Feeding activities, obviously, form part of the product balance:

[3] However, using both in parallel caused problems, and led to the exclusion of HF activities for the moment.

$$-x_g^{bG} \quad -\sum_l \left(a_{l,g}^{LG} x_l^L\right) \quad + \sum_{j_a,d} x_{g,j_a,d}^{fGJaD} \quad \leq \quad 0 \quad \forall g \tag{B.16}$$

Each feeding activity is associated with coefficients (a_{g,j_a,n_b}^{GJaNb}) that determine the quantity of the respective nutrients (n) in each unit of the good g fed to animal group j_a. Values were taken from KTBL [2010] and LfL [2010]. In order to ensure a healthy diet, the nutrient demand and supply should be balanced for each animal group in each feeding seasons, allowing for sufficient fodder with the right mixture of nutrients. To allow for some flexibility in solving the model, a minimum and a maximum constraint was included instead of an equality, allowing a 1% oversupply per nutrient:

$$\begin{aligned}
\sum_a \left(1_{a,j_a}^{AJa} a_{a,n_b,d}^{AND} x_a^A\right) \quad & -\sum_g \left(a_{g,j_a,n_b}^{GJaNb} x_{g,j_a,d}^{fGJaD}\right) \quad \leq \quad 0 \quad \forall (n_b, j_a, d) \\
-1.01 \sum_a \left(1_{a,j_a}^{AJa} a_{a,n_b,d}^{AND} x_a^A\right) \quad & + \sum_g \left(a_{g,j_a,n_b}^{GJaNb} x_{g,j_a,d}^{fGJaD}\right) \quad \leq \quad 0 \quad \forall (n_b, j_a, d)
\end{aligned} \tag{B.17}$$

For cattle (with the exception of calves), a healthy diet requires a minimum raw fibre content of 18% of dry matter fed. For dairy cows, sugar and starch content should not surpass 28% of dry matter and raw fat content should lie below 4% of dry matter, while the structure value (SV, de Brabander et al. 1999) should at least reach an average of 1.2 per kg dry matter following recommendations of LfL [2010].

These restrictions have been implemented in the model using the following system of equations, distinguishing between those nutrients with upper limits (N_u) on dry matter share and those with a lower limit (N_l). The coefficients a_{g,j_a,n_l}^{GJaNl}, resp. a_{g,j_a,n_u}^{GJaNu} indicate the nutrient content (% of dm) of feedstock g, while a_{g,j_a}^{GJaNd} indicates its dry matter content. Values were obtained from LfL [2010]. x^{tNd} are transfer activities used to close the equations, and the coefficients $a_{n_l}^{Nl}$ and $a_{n_u}^{Nu}$ represent the lower, respectively upper limits imposed on dry matter share for each nutrient.

$$\begin{aligned}
-\sum_g \left(a_{g,j_a,n_l}^{GJaNl} x_{g,j_a,d}^{fGJaD}\right) \quad & +a_{n_l}^{Nl} x_{j_a,d}^{tNd} \quad \leq \quad 0 \quad \forall (j_a, d, n_l) \\
\sum_g \left(a_{g,j_a,n_u}^{GJaNu} x_{g,j_a,d}^{fGJaD}\right) \quad & -a_{n_u}^{Nu} x_{j_a,d}^{tNd} \quad \leq \quad 0 \quad \forall (j_a, d, n_u) \\
\sum_g \left(a_{g,j_a}^{GJaNd} x_{g,j_a,d}^{fGJaD}\right) \quad & -x_{j_a,d}^{tNd} \quad \leq \quad 0 \quad \forall (j_a, d)
\end{aligned} \tag{B.18}$$

The ruminal nitrogen balance for dairy cows is restricted to lie between 0 and 30 g per day, by the following constraints, in which $a_{n_r,d}^{NrDl}$ and $a_{n_r,d}^{NrDu}$ represent the lower, respectively upper limit to the ruminal balance in each feeding period d.

$$\begin{aligned}
a_{n_r,d}^{NrDl} \sum_a \left(1_{a,j_a}^{AJa} a_{a,n,d}^{AND} x_a^A\right) \quad & -\sum_g \left(a_{g,j_a,n}^{GJaN} x_{g,j_a,d}^{fGJaD}\right) \quad \leq \quad 0 \quad \forall (n_{\neg s}, j_a, d) \\
-a_{n_r,d}^{NrDu} \sum_a \left(1_{a,j_a}^{AJa} a_{a,n,d}^{AND} x_a^A\right) \quad & + \sum_g \left(a_{g,j_a,n}^{GJaN} x_{g,j_a,d}^{fGJaD}\right) \quad \leq \quad 0 \quad \forall (n_{\neg s}, j_a, d)
\end{aligned} \tag{B.19}$$

Work, machinery and infrastructure: services for animal production

Animal production is usually subject to economies of size, because work time required per stable place and cost of infrastructure per animal decline with an increasing number of stable places. In the model, labor, infrastructure and machinery use of animals are subsumed under the term services. Each animal production activity (a) requires certain quantities of different types of services. The required type of service ($v \in V$) is given as a scale independent coefficient $a^{AV}_{a,v}$ per stable place, e.g. for dairy cows it just indicates that every stable place used requires the capacity to milk one cow.

These service requirements can be satisfied by the provision of services ($o \in O$), which may require labor, machinery or infrastructure capacity ($m \in M$), cash or again other service types. E.g. milking with a herringbone milking parlor requires labor and the capacity of a corresponding parlor. Economies of size are represented in the model by assuming that for (most) services, the quantity of labor, cash or other inputs can be represented by a linear equation with a binary fixed ($x^{\alpha O}$) and a proportional ($x^{\beta O}$) component, resulting in a decreasing specific input demand function with increasing production volume. A similar assumption holds for investments in infrastructure ($x^{\alpha iM}, x^{\beta iM}$), which also necessitates the inclusion of fixed and proportional capacities ($b^{\alpha iM}, b^{\beta iM}$). Services related to feeding are represented by separate balances for each feeding season d.

The corresponding system of equations is shown in block B.20.[4]

$$\begin{array}{lllllll}
\sum_a \left(a^{AV}_{a,v} x^A_a\right) & & -\sum_o \left(a^{OV}_{o,v} x^{\beta O}_o\right) & -\sum_m \left(a^{\beta MV}_{m,v} x^{\beta tM}_m\right) & \leq & 0 & \forall v \\
 & -M x^{\alpha O}_o & +x^{\beta O}_o & & \leq & 0 & \forall o \\
 & & & x^{\beta tM}_m & \leq & b^{\beta M}_m & \forall m
\end{array} \tag{B.20}$$

Some services are specific to the type of fodder fed to the animals – e.g. feeding of silage requires a totally different type of work than pasturing – and are thus associated with the feeding rather than the animal production activities. This also makes it necessary to disaggregate related services and service types by feeding season (d), and requires the distinction between time-specific (Od, Vd) and non-time-specific (On, Vn) services and service types. Cutting of fresh grass requires field work, and pasturing requires labor in certain field work seasons. (These details have been omitted from equation B.20.)

B.1.5 Biogas production[5]

Maize, wheat and grass silage, as well as manure can be used in fermenters to produce biogas, which is then transformed to heat and electricity in generators. The production of biogas electricity from specific goods is represented by the decision variables x^{uG}_g. Electricity yields a^{uG}_g are specific to the feedstock used.

Production of biogas from a certain feedstock is obviously constrained by production or purchase of this feedstock,

$$x^{uG}_g - \sum_l \left(a^{LG}_{l,g} x^L_l\right) - \sum_a \left(a^{AG}_a x^A_a\right) - x^{bG} \leq 0 \quad \forall g, \tag{B.21}$$

as well as the total electric (b^{Ue}) capacities installed. Similar to other infrastructure, biogas plants are split into a fixed ($b^{\alpha Ue}$) and a variable ($b^{\beta Ue}$) part to reflect economies of scale in investment size in the model. Use of capacity requires maintenance (x^{mUe}), with

[4]Note: Some services/infrastructure items have no independent part, while others have a fixed size. Equations B.20 include terms for service provision of the fixed part, too, which have been omitted for clearer exposition here.

[5]The implementation of biogas production in the model was developed in cooperation with Teresa Walter, who also compiled most of the information and data used for the implementation.

size-dependent monetary maintenance cost ($c^{\beta mUe}$) and size-independent daily maintenance work ($a^{H\alpha mUe}$).

$$\begin{array}{llcr} \sum_g \left(a_g^{GUe} x_g^{uG}\right) & -x^{\beta mUe} & \leq & 0 \\ & x^{\beta mUe} & = & b^{\beta Ue} \\ & x^{\beta mUe} \quad -Mx^{\alpha mUe} & \leq & 0 \\ & x^{\alpha mUe} & = & b^{\alpha Ue} \end{array} \tag{B.22}$$

Biogas production (x_g^{uG}) further requires constant daily labor and process electricity, which is reflected by including corresponding coefficients (a_g^{uGH}, a_g^{GbG}) for x_g^{uG} in the daily labor, respectively the product balance of conventional electricity.

The German Renewable Energy Act (*Erneuerbare Energien Gesetz*, EEG) obliges electricity companies to purchase electricity from renewable sources like biogas plants at a fixed price. Prices are guaranteed to the electricity producer for twenty years from the start of electricity production. The individual price paid for a kWh of a certain biogas plant depends on the year the plant first entered production ($y_y \in Yy$) and is tiered by volume. The EEG was first established in 2000 and has been subject to revisions in 2004, 2009 and 2012. If an agent is not willing to comply with the EEG requirements for receiving the guaranteed prices, it can still sell the electricity at market prices (x^{sGe}). (The model assumes this price to lie at 30% of the consumer electricity price following long-term spot market price development observed at the European electricity exchange in Leipzig.)

EEG 2000-2009

The general mode of tiered payments has not changed between the 2000, 2004 and 2009 versions of the EEG. The EEG 2004 introduced additional boni for the use of energy plants and farm manure (NaWaRo bonus), as well as combined heat and power generation (KWK), while the EEG 2009 mainly changed the amount of the guaranteed prices and added a specific manure bonus on top of the NaWaRo bonus.

In the model, an individual decision variable x_{y_u,y_y}^{sYyYu} reflects the sale of a quantity of electrical energy for the price c^{sYyYu} valid under tier $y_u \in Yu$ for plants established in year $y_y \in Yy$.

$$x^{sGe} + \sum_{y_u,y_y} x_{y_u,y_y}^{sYyYu} - \sum_{g \in Gb} \left(a_g^{uG} x_g^{uG}\right) \leq 0. \tag{B.23}$$

Further, it requires an EEG application corresponding to biogas plant capacity established in the given year ($b_{y_y}^{Yy}$),

$$\sum_{y_u,y_y} x_{y_u,y_y}^{sYyYu} \leq b_{y_y}^{Yy} \quad \forall y_y \tag{B.24}$$

and that the volume allowed under the corresponding tier ($b_{y_u}^{YyYu}$) has not yet been exhausted.

$$x_{y_u,y_y}^{sYyYu} \leq b_{y_u}^{Yu} \quad \forall (y_u, y_y) \tag{B.25}$$

Since all feedstock categories considered in the model fulfill the requirements for the NaWaRo bonus, it is automatically added to the biogas sales price. The manure bonus of EEG 2009 requires a minimum of 30% manure (($g \in Go$)) in the total mass of the feedstock. This condition has been implemented using a binary decision of either accepting the condition and receiving the bonus (x^{yuo}), or relaxing the condition on minimum manure use (x^{nuo}) and forgo the bonus (eq. B.26).

$$
\begin{array}{llllllll}
\sum\limits_{y_u,y_y} x^{osYyYu}_{y_u,y_y} & -\sum\limits_{g\in Go}\left(a^{uG}_g x^{uG}_g\right) & & & & \leq & 0 & \\
 & \sum\limits_{g} x^{uG}_g & -x^{tuGo} & & -Mx^{nuo} & \leq & 0 & \\
 & -\sum\limits_{g\in Go}\left(x^{uG}_g\right) & 0.3x^{tuGo} & & & \leq & 0 & \\
\sum\limits_{y_u,y_y} x^{osYyYu}_{y_u,y_y} & & & -Mx^{yuo} & & \leq & 0 & \\
 & & & x^{yuo} & +x^{nuo} & \leq & 1 & \\
x^{osYyYu}_{y_u,y_y} & & & & & \leq & b^{oYyYu}_{y_u,y_y} & \forall(y_u,y_y)
\end{array}
\quad (B.26)
$$

Apart from the electricity also the heat produced during the burning of biogas, can potentially be sold or used as input for animal production on the farm. The combined use of heat and electricity was rewarded with an additional KWK bonus under *Erneuerbare Energien Gesetz* (EEG) 2004 and 2009 (^{sYuh}).

$$
\begin{array}{lllllllll}
 & & -\sum\limits_{g\in Gb}\left(a^{uGh}_g x^{uG}_g\right) & & +x^{tuh} & & \leq & 0 & \\
x^{suh} & & & & -x^{tuh} & +x^{tuh2} & \leq & 0 & \\
 & -x^{bG}_g & & +\sum\limits_{a}\left(a^{AG}_g x^{A}\right) & & -x^{tuh2} & \leq & 0, & g = heat \\
x^{yuh} & & & & -a^{tueh}x^{tuh} & & \leq & 0 &
\end{array}
\quad (B.27)
$$

Both, manure and KWK bonus can only be rewarded for electricity for which also the base rate is awarded:

$$
\begin{array}{lllll}
\sum\limits_{y_u,y_y} x^{osYyYu}_{y_u,y_y} & & -\sum\limits_{y_u,y_y} x^{sYyYu}_{y_u,y_y} & \leq & 0 \\
 & x^{suh} & -\sum\limits_{y_u,y_y} x^{sYyYu}_{y_u,y_y} & \leq & 0
\end{array}
\quad (B.28)
$$

EEG 2012

The newest revision of the EEG replaced the old system of a base price and boni by introducing two remuneration classes, into which biogas feedstock was classified. The remuneration is granted according to the share of the feedstock classes in the total methane produced. As the remuneration remains tiered (Yu), ensuring a remuneration of the same feedstock composition in all tiers requires introducing a quadratic relationship into the model constraints. In a mixed integer linear model, this could only be resolved using discretization: Remuneration activities (Yz) were defined each associated to a fixed relationships between the two remuneration classes ranging from 100% remuneration class I to 100% remuneration class II in steps of 10%. Except for the extremes, two activities were introduced at each step, one ($x^{slYyYuYz}$) serving as the lower bound of a 10% interval and the other as the upper bound ($x^{suYyYuYz}$).

The remuneration activities are complemented by mutually exclusive binary activities ($x^{ynoYyYz}$) that make sure the boundary activities of only one interval within a tier can be used. This ensures that the relationship between the remuneration classes is (at least approximately) equal in all tiers. Otherwise the optimization might lead to the remuneration of electricity of one class in the lower tier and of the other one in a higher tier (the relationship between rewards granted for each remuneration class is not the same in all tiers).

$$\begin{array}{lllll} x^{slYyYuYz}_{y_u,y_y,y_z+1} & +x^{suYyYuYz}_{y_u,y_y,y_z} & -a^{Yu}x^{ynoYyYz}_{y_y,y_z} & \leq & 0 \quad \forall y_z, y_u, y_y \\ & & \sum_{y_z}\left(x^{ynoYyYz}_{y_y,y_z}\right) & \leq & 1 \quad \forall y_y \end{array} \tag{B.29}$$

The EEG 2012 further restricts the share of maize in the total feedstock mass to 60%,

$$\begin{array}{llllll} & -\sum_g x^{uG}_g & +x^{tuGm} & & & \leq 0 \\ & \sum_{g\in Gm}\left(x^{uG}_g\right) & -0.6x^{tuGm} & & -Mx^{n12m} & \leq 0 \\ \sum_{y_u,y_y,y_z}\left(x^{slYyYuYz}_{y_u,y_y,y_z+1}+x^{suYyYuYz}_{y_u,y_y,y_z}\right) & & & -Mx^{y12m} & & \leq 0 \\ & & & x^{y12m} & +x^{n12m} & \leq 1 \end{array} \tag{B.30}$$

and requires the combined use of at least 60% of the heat for plants whose feedstock consists of less than 60% manure. A special unitary premium (x^{sYx}) is granted for small plants up to 75 kW that use manure for more than 80% of the electricity production. In the model, together with the manure bonus of EEG 2009, these are combined into a mutually exclusive set of reward options using binary activities,

$$x^{yYx} + x^{y60o} + x^{yuo} + x^{nuo} \leq 1 \tag{B.31}$$

which are used to apply different manure share requirements,

$$\begin{array}{lllllll} \sum_g x^{uG}_g & -x^{tYx} & -x^{tu60o} & -x^{tuGo} & -Mx^{nuo} & \leq & 0 \\ -\sum_{g\in Go}\left(x^{uG}_g\right) & +0.8x^{tYx} & +0.6x^{tu60o} & +0.3x^{tuGo} & & \leq & 0 \end{array} \tag{B.32}$$

with

$$\begin{array}{lllll} x^{tu60o} & & -Mx^{y60o} & & \leq 0 \\ & x^{tuGo} & & -Mx^{yuo} & \leq 0 \end{array} \tag{B.33}$$

and then allow the use of the respective schemes:

$$x^{sYx} - Mx^{yYx} \leq 0 \tag{B.34}$$

$$a^{rKWK}\sum_{y_u,y_y,y_z}\left(x^{slYyYuYz}_{y_u,y_y,y_z+1}+x^{suYyYuYz}_{y_u,y_y,y_z}\right) - Mx^{y60o} \leq 0 \tag{B.35}$$

For lack of data on potential heat uses, only the two extreme scenarios that either all or no agents are able to sell all of the available heat were considered. The requirement to use the heat is therefore not explicitly implemented in the model. Under the assumption that all agents have the potential of external heat use, the coefficient a^{rKWK} is set to zero as the condition is fulfilled per se, while in the other case no biogas plant with less than 60% of manure can be rewarded according to EEG 2012.

Further the remuneration activities are subject to the same constraints regarding biogas production and the establishment of an EEG contract as the EEG 2000-2009 activities.

$$\begin{array}{llllll} x^{sYx} & +x^{sGe} & +\sum_{y_u,y_y,y_z}\left(x^{slYyYuYz}_{y_u,y_y,y_z+1}+x^{suYyYuYz}_{y_u,y_y,y_z}\right) & -\sum_{g\in Gb}\left(a^{uG}_g x^{uG}_g\right) & \leq & 0 \\ x^{sYx} & & +\sum_{y_u,y_y,y_z}\left(x^{slYyYuYz}_{y_u,y_y,y_z+1}+x^{suYyYuYz}_{y_u,y_y,y_z}\right) & & \leq & b^{Yy}_{y_y} \quad \forall y_y \end{array} \tag{B.36}$$

B.1.6 Manure

The manure balance links land use, animal production and biogas production. Manure produced by animals can be either used in a biogas plant ($x^{uG}_{g_o}$) or directly spread on the field or grassland ($x^{tGo}_{g_o}$).

$$x^{uG}_{g_o} + x^{tGo}_{g_o} - \sum_{a} \left(a^{AGo}_{a,g_o} x^{A}_{a} \right) \leq 0 \quad \forall g_o \tag{B.37}$$

Currently, only two types of manure ($g_o \in Go$), cattle and pig manure are distinguished. For simplification, it is assumed that the residue from biogas production from manure is equivalent to the manure input with respect to fertilization (which seems justified at least with respect to total nitrogen amounts). Residue from biogas production with silage feedstock is transformed into pig and cattle manure equivalents based on nitrogen content, so that the balance for organic fertilization can be formulated as:

$$\sum_{l} \left(a^{LGo}_{l,g_o} x^{L}_{l} \right) - x^{uG}_{g_o} - \sum_{g_b \notin Go} \left(a^{uGGo}_{g_o,g_b} x^{uG}_{g_b} \right) - x^{tGo}_{g_o} \leq 0 \quad \forall g_o \tag{B.38}$$

At the same time, all the manure produced also has to be spread on the field. For computational reasons, a certain slack is allowed here in order to give some flexibility to the solver; the corresponding coefficient ζ_{manure} is subject to calibration.

$$-\zeta_{\text{manure}} \sum_{l} \left(a^{LGo}_{l,g_o} x^{L}_{l} \right) + \sum_{a} \left(a^{AGo}_{a,g_o} x^{A}_{a} \right) + \sum_{g_b \notin Go} \left(a^{uGGo}_{g_o,g_b} x^{uG}_{g_b} \right) \leq 0 \quad \forall g_o \tag{B.39}$$

Second, the farm needs to have storage capacity for all manure produced, which is implemented as an infrastructure service (v_{go}) as described in section B.1.4.

$$\sum_{a,g_o} \left(a^{AGo}_{a,g_o} x^{A}_{a} \right) + \sum_{g_o,g_b \notin Go} \left(a^{uGGo}_{g_o,g_b} x^{uG}_{g_b} \right) - \sum_{m} \left(a^{\beta MV}_{m,v} x^{\beta tM}_{m} \right) \leq 0 \quad , v = v_{go} \tag{B.40}$$

B.1.7 Labor

Labor capacity depends on the number of household members working on the farm ($x^{H1} + x^{H2}$, see section B.2.3) and hired permanent employees (x^{bpH}). This labor can be either used for the seasonal field work (x^{tHw}), or for constant daily tasks (x^{tHd}), as they typically are required for animal and biogas production.

$$x^{tHw} + x^{tHd} - x^{bpH} - x^{H1} - 0.25x^{H2} \leq 0 \tag{B.41}$$

Labor capacity in each field work season is calculated like equipment and tractor power capacity, using the available number of field working days expected in each work season ($a^{TK}_{t,k}$). Additionally, temporary labor can be hired (x^{btHTK}_{t,k_i}) on an hourly basis for each work season.

$$\sum_{w,s_q,e} \left(1^{WK}_{w,k} x^{WTSqE}_{w,t,s_q,e} \right) - a^{TK}_{t,k} x^{tHw} - \sum_{k_i \leq k} \left(x^{btHTK}_{t,k_i} \right) \leq 0 \quad \forall (k,t) \tag{B.42}$$

Labor reserved for constant daily labor is multiplied by the assumed amount of daily working hours (a^{Hd}), and is available for animal production and related services as well as biogas production, assuming the same tasks have to be realized everyday.

$$- a^{Hd} x^{tHd} + \sum_a \left(a_a^{AH} x_a^A\right) + \sum_o \left(a_o^{\alpha OH} x_o^{\alpha O}\right) + \sum_o \left(a_o^{\beta OH} x_o^{\beta O}\right) + \sum_m \left(a_m^{\beta MH} x_m^{\beta tM}\right) + \sum_g \left(a_g^{uGH} x_g^{uG}\right) + a^{H\alpha mUe} x^{\alpha mUe} \leq 0 \quad \text{(B.43)}$$

B.1.8 Financial activities & liquidity

While the objective function reflects the expected total farm gross margin, which could be negative, the financial balances ensure that the agent cannot use more liquid means than actually available. Cash available to the agent at the start of the season (b^C) can either be deposited on the bank to earn interest (x^{dC}) or used in the production process (x^{tC}).

$$x^{dC} + x^{tC} \leq b^C \quad \text{(B.44)}$$

This mainly concerns expenses for inputs of crop production x^L that have to be pre-financed. If cash reserves are insufficient for the later, they can be extended by short-term credit x^{bC}.

$$\sum_l a_{l,c}^{LC} x_l^L - x^{tC} - x^{bC} \leq 0 \quad \text{(B.45)}$$

Usually, the standing crop can be used as a collateral and extends the credit limit of the farm.

$$x^{bC} - \sum_l a_{l,c}^{LC} x_l^L \leq 0 \quad \text{(B.46)}$$

B.1.9 EU CAP premiums

During the time covered in validation simulations, the CAP regulations changed several times. The regulations applicable to the first season simulated (1998/99) still date from the 1992 MacSharry reforms. From 1999/2000 on, the changes under the Agenda 2000 applied. The CAP MTR of 2003 enacted regulations applying from seasons 2004/05 on, which were only slightly adapted under the CAP Health Check in 2008.

Milk quota

Throughout the whole period considered, milk sales are restricted by the milk quota b^{Yd}.

$$x^s G_{g=\text{milk}} \leq b^{Yd} \quad \text{(B.47)}$$

MacSharry reforms & Agenda 2000

Regulations under MacSharry and Agenda 2000 are structurally similar and differ mainly in parameters and can therefore be implemented similarly.

Cereal, oilseed and protein crop premium Farmers could apply for area premiums ($y_c \in Yc$) for cereals (wheat, barley, maize), oilseed and protein crops, which were paid per area of crops grown.

$$x_{y_c}^{Yc} - \sum_l \left(1_{l,y_c}^{\mathrm{LYc}} x_l^L\right) \leq 0 \quad \forall y_c \tag{B.48}$$

A certain percentage (10%) of the area to be subsidized had to be set aside (x^{Ys}), with the exemption of farmers applying for premiums on an area equivalent to less than 92 t reference yield ($a_{y_c}^{Yc}$). (The binary variables x^{yYs} and x^{nYs} represent the decision to set-aside land at all.)

$$\begin{array}{rrlrrcr}
 & x^{Ys} & -\sum_l \left(1_l^{\mathrm{LYs}} x_l^L\right) & & & \leq & 0 \\
0.1 \sum_{y_c} x_{y_c}^{Yc} & -0.9x^{Ys} & & -Mx^{nYs} & & \leq & 0 \\
\sum_{y_c} \left(a_{y_c}^{Yc} x_{y_c}^{Yc}\right) & & & & -Mx^{yYs} & \leq & 92 \\
 & & & x^{nYs} & x^{yYs} & \leq & 1
\end{array} \tag{B.49}$$

As a premium (c^{Ys}) was also paid for set-aside areas, farmers had an incentive to set-aside more land than required. This voluntary set-aside could be extended up to 33% of the total subsidized area.

$$\begin{array}{rrrcr}
-0.33 \sum_{y_c} x_{y_c}^{Yc} & -0.33x^{Ys} & x^{tYs} & \leq & 0 \\
 & x^{Ys} & -x^{tYs} & \leq & 0
\end{array} \tag{B.50}$$

Table B.1: *EU area premiums and reference yields 1999-2004*

		Year (of harvest)			
Coefficient	Land use	1999	2000	2001	2002-2004
c^{Yc}[€]	cereals	279	303	324	324
	maize	396	429	459	459
	oilseeds	545	474	421	324
	protein crops	403	384	384	384
c^{Ys} [€]	set-aside	363	310	333	333
a^{Yc} [t/ha]	cereals	5.29	5.14	5.14	5.14
	maize	5.29	7.28	7.28	7.28
	oilseeds	2.97	5.70	5.79	5.79
	protein crops	5.29	5.29	5.29	5.29

Energy crops (NaWaRo) could be grown on set-aside land, if their use for nonfood and nonfeeding purpose was ensured. To capture this, separate NaWaRo product balances for relevant crops were introduced into the model. Biogas production is based on the NaWaRo product balances, while selling and feeding is based on the normal product balances. For all concerned production activities, a duplicate was introduced and marked as "production destined for NaWaRo". The yield of these activities is transferred to the NaWaRo product balance. Feedstock can be transferred from the normal product balance to the NaWaRo balance, but not vice versa.

Suckler cow premium and special premium for male cattle A special premium for male cattle (Yb) was granted by the EU for each bull once in its lifetime, and for each ox twice in its lifetime. As a turnover time of 15 months is assumed for bull fattening in the model, this results in 0.8 potential premium applications (a_a^{AYb}) per stable place and year in the model.

$$x^{Yb} - \sum_a \left(a_a^{AYb} x_a^A \right) \leq 0 \tag{B.51}$$

Suckler cow premium (Yo) was granted for each suckler cow every year (a_a^{AYo} = 1).

$$x^{Yo} - \sum_a \left(a_a^{AYo} x_a^A \right) \leq 0 \tag{B.52}$$

The maximum amount of suckler cows to be subsidized is limited by the suckler cow quota (b^{Yo}) owned by the agent.

$$x^{Yo} \leq b^{Yo} \tag{B.53}$$

Further, a combined upper limit for suckler cow and special cattle premiums was given by the available area used for feeding these animals (x^{tYbo}) after subtracting the area used for feeding any dairy cows ($a \in Am$). Or, alternatively, under a small producer scheme (x^{yYbs}), subsidies for up to 15 LU could be granted irrespective of feeding area. The amount of livestock unit which could be subsidized per hectare of feeding area (a^{tYbo}) was 2 in 1999, and later reduced to 1.9 in 2002 and to 1.8 in 2003. Suckler and dairy cows were counted as 1 LU, fattening bulls as 0.6 LU per stable place.

$$1.0x^{Yo} + 0.6x^{Yb} + 1.0 \sum_{a \in Am} \left(x_a^A \right) - a^{tYbo} x^{tYbo} - 15x^{yYbs} - Mx^{nYob} \leq 0 \tag{B.54}$$

Due to the involvement of dairy cows in equation B.54, the constraint needs to be relaxed completely in case the agent chooses neither to apply for cattle nor suckler cow premiums (x^{nYob}). This is reflected in the following equations, which also incorporates the choice between the small and regular producer conditions ($x^{nYob}, x^{yYob}, x^{nYbs}, x^{yYbs}$ are binary integer variables).

$$\begin{array}{lllllllll}
x^{Yo} & +x^{Yb} & & & -Mx^{yYob} & & & \leq & 0 \\
 & & & x^{nYob} & +x^{yYob} & & & \leq & 1 \\
 & & & & & x^{nYbs} & +x^{yYbs} & \leq & 1 \\
 & & x^{tYbo} & & & -Mx^{nYbs} & & \leq & 0
\end{array} \tag{B.55}$$

Total forage area is calculated as the sum of all land uses suitable for feeding ruminants (1_l^{LYbo}).

$$x^{tYbo} - \sum_l \left(1_l^{LYbo} x_l^L \right) \leq 0 \tag{B.56}$$

Under the EU regulations of MacSharry and Agenda 2000, cereal area could be counted as forage area, but could then not be used to apply for the crop premium and its yield could not be used for feeding other animals (e.g. horses or pigs). To account for this in the model, product balances have been split up and growing activities duplicated in a similar fashion as for the NaWaRo rule of the set-aside scheme of the crop premium. Agents can use yields from feeding areas only for feeding their cattle (or other ruminants), while yields of other areas can be used for selling, biogas production and feeding alike.

Extensification premium In addition to every suckler cow or special male cattle premium granted, an extensification bonus (Ye) could be awarded

$$x^{Ye} - x^{Yo} - x^{Yb} \leq 0 \tag{B.57}$$

if the ratio of livestock unit (a_a^{AYe}, counting only dairy cows, mother cows, heifers, male cattle and sheep) to forage area (x^{tYe}) is less or equal to 1.4,

$$\sum_a \left(a_a^{AYe} x_a^A\right) - 1.4x^{tYe} - Mx^{nYe} \leq 0 \tag{B.58}$$

and at least 50% of this area is pasture .

$$0.5x^{tYe} - \sum_{l \in Lgp} x_l^L \leq 0 \tag{B.59}$$

Again, equation B.58 has to be relaxed in case the agent does not opt for the extensification premium (x^{nYe}), requiring the following additional restrictions (with x^{yYe} and x^{nYe} being binary integers):

$$\begin{array}{llllll} x^{Ye} & -Mx^{yYe} & & \leq & 0 \\ & x^{yYe} & +x^{nYe} & \leq & 1 \end{array} \tag{B.60}$$

The potential feeding area is calculated from suitable land use activities (1_l^{LYe}), which does not include cereals and oilseeds in this case.

$$x^{tYe} - \sum_l \left(1_l^{LYe} x_l^L\right) \leq 0 \tag{B.61}$$

Slaughter premium for cattle For cattle over eight months of age to be slaughtered or exported outside the European Union a slaughter premium (Yk) was granted under MacSharry and Agenda 2000. This general premium could be topped-up by EU member states according to a fixed budget, which was distributed among all applicants. The corresponding model coefficient a_a^{AYk}, indicating the number of potential applications per stable place and year, is calculated based on the turnover time assumed for the respective animal production activity a.

$$x^{Yk} - \sum_a \left(a_a^{AYk} x_a^A\right) \leq 0 \tag{B.62}$$

A similar premium was granted for slaughtering calves (without top-up), though this was not considered in the present version of the model, as the model does not distinguish whether calves sold by farm agents are directly slaughtered or raised. For similar considerations, the cattle slaughter premium is only considered for fattening bulls as well as the replacement of dairy and mother cows.

EU Transition from Agenda 2000 to MTR

In the course of the MTR reforms, the crop-specific area premiums were transformed into payment entitlements, which now allow receiving the new farm premium irrespective of what is grown on the plots as long as the area is kept in 'good' conditions according to cross compliance regulations [BMELV, 2006].

Table B.2: *EU animal premiums 1999-2004*

		Year (of harvest)			
Coefficient	Type	1999	2000	2001	2002-2004
c^{Yo}[€]	suckler cow premium	145	163	182	200
c^{Yb}[€]	special premium male cattle	135	160	185	210
c^{Ye}[€]	extensification premium	51.65	100	100	100
c^{Yk}[€]	slaughter premium incl. top-up	0	34	66	100

In 2005, farmers in the EU received four categories of payment entitlements: for grassland (Yg), arable areas (Ya), set-aside (Yf) and special entitlements. Special entitlements could be awarded to producers without any land (e.g. sheep herders), but are currently not considered in the model.

Until 2013, the amount payable for each entitlement was specific for each farmer as it was partly determined based on the crop area premiums, cattle, extensification and suckler cow premiums received between 2000 and 2002. Representing this in a linear model, requires discretization, i.e. including a separate entitlement for each potential value a grassland or arable entitlement could take on (set-aside entitlements have fixed values). Discretization steps of 25 € were chosen, leading to the inclusion of 198 levels of grassland entitlements ($y_g \in [72, 107, \ldots, 5000]$) and 188 levels of arable entitlements ($y_g \in [303, 328, \ldots, 5000]$).

The transition itself has not been implemented in the model, as I have not found a feasible MIP implementation for this problem yet. (This currently bars us from running dynamic simulation from 2003 to 2007).

EU MTR and Health Check

Beginning with season 2004/2005, agents can receive the single farm payment ($x^{Ya}_{y_a}, x^{Yg}_{y_g}, x^{Yf}_{y_f}$) according to the respective entitlements owned (b^{Ya}, b^{Yg}, b^{Yf}),

$$x^{Yg}_{y_g} \leq b^{Yg}_{y_g} \quad \forall y_g \tag{B.63}$$

$$x^{Ya}_{y_a} \leq b^{Ya}_{y_a} \quad \forall y_a \tag{B.64}$$

Until the EU Health Check in 2007/2008, receiving premiums was bound to the condition that all set-aside entitlements were activated by setting a corresponding amount of area to set-aside.

This condition is enforced in the model by the following equality constraint,

$$x^{Yf}_{y_f} + x^{tnYf} = b^{Yf} \tag{B.65}$$

which is only relaxed (x^{tnYf}) if the agent chooses not to receive single farm payments. x^{yYf} and x^{nYf} are binary integer variables reflecting the decision for, respectively against receiving payments.

$$\begin{array}{llllll} \sum_{y_g} x^{Yg}_{y_g} & +\sum_{y_a} x^{Ya}_{y_a} & & -Mx^{yYf} & & \leq 0 \\ & & x^{tnYf} & & -Mx^{tnYf} & \leq 0 \\ & & & x^{yYf} & +x^{nYf} & \leq 1 \end{array} \tag{B.66}$$

Set-aside entitlements can only be activated using fallow or NaWaRo land use activities (1_l^{LYf}).

$$x_{y_f}^{Yf} - \sum_l \left(1_l^{LYf} x_l^L \right) + x^{tYf} \leq 0 \quad \text{(B.67)}$$

Grassland and arable entitlements can be activated using any land use that fulfills cross compliance requirements, including set-aside land (x^{tYf}):

$$\sum_{y_g} x_{y_g}^{Yg} + \sum_{y_a} x_{y_a}^{Ya} - \sum_l \left(1_l^{LY\neg f} x_l^L \right) - x^{tYf} \leq 0 \quad \text{(B.68)}$$

Following the EU regulations, from 2005 till 2009 the payment per entitlement in the model corresponds to the nominal amount of the agent entitlement (e.g. $c_{y_a}^{Ya} = y_a$). After that the payments are gradually adapted such that in 2013 all entitlements of every agent in the region have the same value (*regional target value, rtv*). Specifically, the difference between the individual value of an entitlement in 2009 and the regional target value is calculated, and in the next years the difference is gradually reduced such that agent entitlements worth less than the regional target value gradually increase in value, and higher valued entitlements decrease in value. The value in a given year between 2010 and 2013 is calculated according to the following formula

$$c_{y.,year}^{Y.} = rtv + \psi_{year}(c_{y.,2009}^{Y.} - rtv) \quad \text{(B.69)}$$

with ψ_{year} according to the following table [BMELV, 2006]:

Year	2010	2011	2012	2013
ψ_{year}	0.9	0.7	0.4	0

B.1.10 MEKA

The Compensation Scheme for Market Easing and Landscape Protection (*Marktentlastungs- und Kulturlandschaftsausgleich*, MEKA) program rewards farmers with payments for agricultural practices, which contribute to extensification, landscape conservation and environmentally friendly production [MLREV, 2011]. The MEKA catalog – as of 2011 – contains about 30 different measures ($y_m \in Ym$). In the model, only a few selected ones that can reasonably be modeled in the current model setup and that are related to grassland extensification and crop rotation diversification have been considered.

So far, there have been three phases of the MEKA program: MEKA I from 1994 to 1999, MEKA II from 2000 till 2006, and MEKA III from 2007-2013. In each phase, the measures and associated conditions and rewards were revised substantially, and require a separate implementation in our model.

A constant feature throughout all phases has been the general principal of awarding a measure-specific number of points ($a_{y_m}^{Ym}$) per unit ($x_{y_m}^{Ym}$, e.g. ha, animal, farm) included under a certain measure y_m. For each point received (x^{sYm}), the agent is rewarded with c^{sYm} Euro.

$$x^{sYm} - \sum_{y_m} \left(a_{y_m}^{Ym} x_{y_m}^{Ym} \right) \leq 0 \quad \text{(B.70)}$$

A minimum amount rewarded (a^{llYm}, in Euro) is required for participation (x^{yYm}, integer) and a maximum of b^{ulYm} Euro can be awarded per agent.

$$\begin{array}{rrcr} c^{sYm}x^{sYm} & & \leq & b^{ulYm} \\ x^{sYm} & -Mx^{ylYm} & \leq & 0 \\ -c^{sYm}x^{sYm} & +a^{llYm}x^{ylYm} & \leq & 0 \end{array} \tag{B.71}$$

A second constant feature of the program has been the requirement to commit to the application of a measure for five years.[6] Agent participation in a measure is therefore bound to a corresponding commitment ($b^{Ym}_{y_m}$);

$$x^{yYm}_{y_m} = b^{Ym}_{y_m} \quad \forall y_m \tag{B.72}$$

or, where the commitment is not a yes or no decision, but covers a specified area:

$$x^{Ym}_{y_m} = b^{Ym}_{y_m} \quad \forall y_m \tag{B.73}$$

MEKA I

During the first MEKA phase, 20 DEM (i.e. $c^{sYm} \approx 10.22$ €) were awarded per point, the upper limit was 40,000 DEM ($b^{ulYm} \approx 20,452$ €) and the lower limit 100 DEM ($a^{llYm} \approx 51.13$ €).

For the first MEKA phase, only the extensive grassland measures listed under chapters 3.1 (use of grassland) and 3.2.2 (limits on the number of grassland cuts) of the MEKA I catalog were included in the model. For MEKA I, the state of Baden-Württemberg was subdivided into three grassland support areas ("Förderkulissen"), where support was focused on either (i) groundwater protection, (ii) erosion prevention, or (iii) landscape value. For measure 3.1, differences between groundwater protection areas (3.1A) and the other two areas (3.1) applied. This differentiation is due to the fact that grassland conversion was (and is) not allowed in groundwater protection areas.

Extensive grassland (3.1) In groundwater protection areas, participation in measure 3.1A required maintaining an animal-to-land ratio between 0.3 and 1.4 roughage-consuming livestock unit (RLU) per ha of main forage area (MF), and rewarded 8 points per ha of grassland of the farmer. In other areas, merely maintaining grassland was rewarded with 2 points per ha, respectively 3 points if an animal-to-land ration of less than 1.8 RLU per ha MF was maintained, or 5 points if it was below 1.2 RLU per ha MF. In the model, these regulations have been implemented accordingly and the decision to participate in the measures is represented by the binary integers x^{yYm}_{131A}, x^{yYm}_{131h}, x^{yYm}_{131m} and x^{yYm}_{131l}, whereas x^{nYm}_{131} denotes nonparticipation. These decision alternatives are mutually exclusive,

$$x^{yYm}_{131A} + x^{yYm}_{131h} + x^{yYm}_{131m} + x^{yYm}_{131l} + x^{nYm}_{131} \leq 1 \tag{B.74}$$

and only possible if the agent is part of the corresponding support focus area:

$$x^{yYm}_{131A} \leq Mb^{y31A} \tag{B.75}$$

$$x^{yYm}_{131h} + x^{yYm}_{131m} + x^{yYm}_{131l} \leq Mb^{y31A} \tag{B.76}$$

[6] On completion of the five years, usually a one or two year extension until the end of the phase was offered, if applicable.

Depending on the choice, different restrictions on the total farm RLU apply:

$$\begin{array}{rllllll} \sum_a \left(a_a^{Arlu} x_a^A\right) & -1.4x_{131A}^{clru} & -1.2x_{131l}^{clru} & -1.8x_{131m}^{clru} & -Mx_{131hn}^{clru} & \leq & 0 \\ -\sum_a \left(a_a^{Arlu} x_a^A\right) & +0.3x_{131A}^{clru} & & & & \leq & 0 \end{array} \tag{B.77}$$

with

$$x_{y_a}^{clru} - Mx_{y_a}^{yYm} \leq 0 \quad \forall\, y_a \in \{131A, 131l, 131m\} \tag{B.78}$$

$$x_{131hn}^{clru} - Mx_{131h}^{yYm} - Mx_{131}^{nYm} \leq 0. \tag{B.79}$$

and the transfer variables (x^{clru}) required to equalize the main forage area:

$$\sum_{y_a \in Ym131} \left(x_{y_a}^{clru}\right) - \sum_{l \in Lmf} x_l^L = 0 \tag{B.80}$$

Fulfillment of these conditions allows agents to retrieve the associated bonuses

$$x_{y_a}^{Ym} - Mx_{y_a}^{yYm} \leq 0 \quad \forall\, y_a \in \{131A, 131l, 131m, 131h\} \tag{B.81}$$

based on total grassland area of the agent:

$$\sum_{y_a \in Ym131} \left(x_{y_a}^{Ym}\right) - \sum_{l \in Lgg} \left(x_l^L\right) \leq 0 \tag{B.82}$$

Limitation of grassland cutting (3.2.2) Under measure 3.2.2, 1 point was awarded per ha of grassland whose use had to be restricted to two cuts per year (x_{13222}^{Ym}) ; 2 points for maximum one cut (x_{13221}^{Ym}).

The model implementation of these measures is straightforward:

$$\begin{array}{rrrll} x_{13221}^{Ym} & & +x^{tYm1322} & -\sum\limits_{l \in Lgg1} x_l^L & \leq 0 \\ & x_{13222}^{Ym} & -x^{tYm1322} & -\sum\limits_{l \in Lgg2} x_l^L & \leq 0 \end{array} \tag{B.83}$$

MEKA II

During the second MEKA phase, $c^{sYm} = 10$ € were awarded per point, the upper limit (b^{ulYm}) was 40,000 € and the lower limit (a^{llYm}) was 100 €.

Participation in any MEKA II measure required maintaining an animal-to-land ratio of 2.5 LU per ha of agricultural area (AA) as an overall condition. This is implemented using two binary integer variables x^{yYm2} and x^{nYm2}.

$$\begin{array}{rrrrrll} x^{yYm2} & +x^{nYm2} & & & & \leq & 1 \\ & -Mx^{nYm2} & +\sum_a \left(a_a^{Alu} x_a^A\right) & -2.5x_2^{clu} & & \leq & 0 \\ -Mx^{yYm2} & & & +x_2^{clu} & & \leq & 0 \\ & & & x_2^{clu} & -\sum_l x_l^L & \leq & 0 \end{array} \tag{B.84}$$

Diversification of crop rotation (A7) Under measure A7 of the MEKA II catalog, diversity in crop production was awarded, requiring the cultivation of at least 4 different crops, each with at least 15% of the total arable area of the farm and a restriction of maize area to 40% of the total arable area. Oilseeds could be counted as crops to fulfill diversification requirements, but no points were awarded for oilseed areas.

The corresponding model implementation therefore requires the inclusion of several binary integer variables: two variables to represent the decision whether to participate (x_{2A7}^{yYm}) or not (x_{2A7}^{nYm}), which are of course mutually exclusive,

$$x_{2A7}^{yYm} + x_{2A7}^{nYm} \leq 1 \tag{B.85}$$

x_{2A7}^{yYm} requires participation in MEKA II in general:

$$x_{2A7}^{yYm} - Mx^{yYm2} \leq 0 \tag{B.86}$$

Then for each group crop (Jym) potentially included in the agent crop rotation and counted for diversification, two binary integer variables indicate whether it has been included (x_{jym}^{yJym}) or not (x_{jym}^{nJym}). The condition of requiring at least four crops with a minimum share of 15% is enforced in the model by the following system of equations (using the soil in rotation variables x_s^{Sr} – see section B.1.3 – to sum up all arable land):

$$\begin{array}{llllllll}
4x_{2A7}^{yYm} & -\sum\limits_{jym} x_{jym}^{yJym} & & & & & \leq & 0 \\
& x_{jym}^{yJym} & +x_{jym}^{nJym} & & & & \leq & 1 \quad \forall j_{ym} \\
& & & \sum\limits_{s} x_s^{Sr} & -x_{2A7}^{t1Ym} & -x_{2A7}^{t2Ym} & \leq & 0 \\
& & -Mx_{jym}^{nJym} & & +0.15x_{2A7}^{t1Ym} & +0.15x_{2A7}^{t2Ym} \quad -\sum\limits_{l}\left(1_{l,jym}^{LJym} x_l^L\right) & \leq & 0 \quad \forall j_{ym}
\end{array} \tag{B.87}$$

Two more variables are needed to distinguish between oilseed area (x_{2A7}^{t1Ym}) and nonoilseed area (x_{2A7}^{t2Ym}), only the later can be counted to achieve the point:

$$\begin{array}{llllll}
x^{sYm} & & & -1x_{2A7}^{t2Ym} & & \leq 0 \\
& -Mx_{2A7}^{yYm} & +x_{2A7}^{t1Ym} & +x_{2A7}^{t2Ym} & & \leq 0 \\
& & x_{2A7}^{t1Ym} & & -\sum\limits_{l \in Loil} x_l^L & \leq 0 \\
& & & x_{2A7}^{t2Ym} & -\sum\limits_{l \notin Loil, Lgg} x_l^L & \leq 0
\end{array} \tag{B.88}$$

Further, the MEKA restriction on maize cultivation is implemented as follows:

$$\sum_{l \in Lmai} -0.4x_{2A7}^{t1Ym} - 0.4x_{2A7}^{t2Ym} - Mx_{2A7}^{nYm} \leq 0 \tag{B.89}$$

Extensive grassland (B1, B2, B4) For MEKA II, the distinction of different support focus areas used for MEKA I was dropped and support for extensive grassland use unified in the whole area of Baden-Württemberg. Measure B1 awarded nine points for maintaining grassland use, i.e. abstaining from grassland conversion and maintaining a minimum level of use on all grassland plots, while restricting the animal-to-land ratio below 2 RLU / ha MF. Measure B2 awarded an additional 4 points for maintaining an animal-to-land ratio between 0.5 and 1.4 RLU / ha MF.

These restrictions are represented in the model by the different limit calculation activities (x^{crlu}) in the following two equations,

$$\begin{array}{llllll} \sum_a \left(a_a^{Arlu} x_a^A\right) & -2.5x_{2B1}^{crlu} & -1.4x_{2B2}^{crlu} & -Mx_{n2B}^{crlu} & & \leq 0 \\ -\sum_a \left(a_a^{Arlu} x_a^A\right) & & +0.5x_{2B2}^{crlu} & & & \leq 0 \\ & x_{2B1}^{crlu} & +x_{2B2}^{crlu} & +x_{n2B}^{crlu} & -\sum\limits_{l \in Lmf} x_l^L & = 0 \end{array} \tag{B.90}$$

while the choice of the correct calculation activity is a function of the choice of participating or not participating in B1 and B2 (x_{2B1}^{yYm}, x_{2B1}^{nYm}, x_{2B2}^{yYm}, x_{2B2}^{nYm}, all binary integer variables),

$$\begin{array}{llllll} x_{2B1}^{crlu} & +x_{n2B}^{crlu} & -Mx_{2B2}^{nYm} & & & \leq 0 \\ & x_{n2B}^{crlu} & & -Mx_{2B1}^{nYm} & & \leq 0 \\ & & x_{2B2}^{nYm} & +x_{2B2}^{yYm} & & \leq 1 \\ & & & x_{2B1}^{nYm} & +x_{2B1}^{yYm} & \leq 1 \end{array} \tag{B.91}$$

which then also allows receiving corresponding points depending on the total grassland area owned by the agent:

$$\begin{array}{llllll} -Mx_{2B1}^{yYm} & & +x_{2B1}^{Ym} & & & \leq 0 \\ & -Mx_{2B2}^{yYm} & & +x_{2B2}^{Ym} & & \leq 0 \\ & & x_{2B1}^{Ym} & & -\sum\limits_{l \in Lgg} \left(x_l^L\right) & \leq 0 \\ & & & x_{2B2}^{Ym} & -\sum\limits_{l \in Lgg} \left(x_l^L\right) & \leq 0 \end{array} \tag{B.92}$$

Measure B4 rewarded very extensive grassland use with five points per ha (replacing MEKA I 3.2.2). The existence of at least four out of a catalog of 28 characteristic species was used as an indicator for low intensity grassland use. For simplification, it is assumed in the model that grassland activities of intensity level 1 with conservation cuts ($Lb4$) fulfill these requirements.

$$x_{2B4}^{Ym} - \sum_{l \in Lb4} \left(x_l^L\right) \leq 0 \tag{B.93}$$

Participation in B1 was a prerequisite for participation in measures B2 and B4, and itself required participation in MEKA II in general. Further, farmers could not participate in MEKA B2 and at the same time receive the EU Agenda 2000 extensification premium (section B.1.9).

$$\begin{array}{llllll} x_{2B1}^{yYm} & & & -Mx^{yYm2} & & \leq 0 \\ -Mx_{2B1}^{yYm} & +x_{2B2}^{yYm} & +x_{2B4}^{Ym} & & & \leq 0 \\ & x_{2B2}^{yYm} & & & +x^{yYe} & \leq 1 \end{array} \tag{B.94}$$

MEKA III

As in the second MEKA phase, also the third MEKA phase $c^{sYm} = 10$ € were awarded per point. The upper limit (b^{ulYm}) remained at 40,000 €, while the lower limit (a^{llYm}) was raised to 250 €. The general requirement of restricting the animal-to-land ratio to 2.5 LU /ha AA to participate in the MEKA program was dropped.

Diversification of crop rotation (A2) With measure A2, the MEKA III catalog contained a diversification support similar to measure A7 of MEKA II. In contrast to phase II, points were also awarded for oilseed areas, while fallow/set-aside areas counted as element of the rotation, but no points were awarded for these areas. Compensation was increased to two points for each hectare of arable land of the farmer. In the model, the measure could thus be implemented analogous to the implementation of measure A7 of MEKA II (see section B.1.10).

The MEKA III catalog included additional support for a five-part crop rotation (measure A3), which has not been implemented in the model as it required at least 5% legumes in the rotation and currently the model does not include any legume among the crops eligible to agents. Measure A3 has been implemented in the model for future use, and can be activated once legumes are included in the model.

Extensive grassland (B1, B2, B4) The measures supporting extensive use of grassland were again revised for the third phase (and implemented in the model accordingly): Under measure B1, five points were awarded per ha grassland included under B1, if

- a maximum animal-to-land ratio of 2.0 LU per ha AA was not surpassed;
- at least 5% of the area was cut the first time after 15th July;
- no grassland conversion was performed;
- a number of other restrictions (e.g. no use of chemical plant protection on grassland at the farm, documentation of organic fertilization and use, pasture care) were respected (These are not relevant in the model, however).

Under measure B2, ten points were awarded to farmers (and agents in the model) for each hectare of grassland, if

- the animal-to-land ratio was 1.4 LU/ ha AA and between 0.3 and 1.4 RLU/ ha MF
- no grassland conversion was performed;
- and a number of other restrictions (e.g. no use of chemical plant protection on grassland at the farm, no sprinkler irrigation of grassland, no amelioration on grassland, pasture care) were respected (which again are not relevant in the model).

As in reality, agent participation in B1 or B2 is mutually exclusive (x^{yYm}_{3B1}, x^{yYm}_{3B2}, and x^{nYm}_{3B} are binary integers).

$$x^{yYm}_{3B1} + x^{yYm}_{3B2} + x^{nYm}_{3B} \leq 1 \tag{B.95}$$

The restriction on the animal-to-land ratio is implemented in the model using different calculation activities (x^{clu}, x^{crlu}). For the LU to AA ratio the equation system is as follows:

$$\begin{array}{lllllll} \sum\limits_a \left(a^{Alu}_a x^A_a\right) & -2.0x^{clu}_{3B1} & -1.4x^{clu}_{3B2} & & & -Mx^{nYm}_{3B} & \leq 0 \\ & x^{clu}_{3B1} & +x^{clu}_{3B2} & -\sum\limits_l x^L_l & & & \leq 0 \\ & x^{clu}_{3B1} & & & -Mx^{yYm}_{3B1} & & \leq 0 \end{array} \tag{B.96}$$

, and for the RLU to MF ratio, the equation system is as follows:

$$\begin{array}{rrrrrrl}
\sum\limits_a \left(a_a^{Alu} x_a^A\right) & -1.4x_{3B2}^{crlu} & -Mx_{n3B2}^{crlu} & & & \leq & 0 \\
-\sum\limits_a \left(a_a^{Alu} x_a^A\right) & +0.3x_{3B2}^{crlu} & & & & \leq & 0 \\
& x_{3B2}^{crlu} & +x_{n3B2}^{crlu} & -\sum\limits_{l \in Lmf} x_l^L & & = & 0 \\
& & x_{n3B2}^{crlu} & -Mx_{3B1}^{yYm} & -Mx_{3B}^{nYm} & \leq & 0
\end{array} \tag{B.97}$$

Receiving points requires fulfillment of the conditions and is restricted by the available grassland area, and in the case of B1 on the additional condition of cutting 5% of the area the first time after the fifteenth of July.

$$\begin{array}{rrrrrrl}
x_{3B1}^{Ym} & +x_{3B2}^{Ym} & -\sum\limits_{l \in Lgg} x_l^L & & & \leq & 0 \\
0.05x_{3B1}^{Ym} & & -\sum\limits_{l \in Lb4} x_l^L & & & \leq & 0 \\
x_{3B1}^{Ym} & & & -Mx_{3B1}^{yYm} & & \leq & 0 \\
& x_{3B2}^{Ym} & & & -Mx_{3B2}^{yYm} & \leq & 0 \\
-x_{3B1}^{Ym} & & & +x_{3B1}^{yYm} & & \leq & 0 \\
& -x_{3B2}^{Ym} & & & +x_{3B2}^{yYm} & \leq & 0
\end{array} \tag{B.98}$$

In MEKA III, measure B4 is continued similar as in MEKA II, though six points were awarded per hectare of late-cut grassland and participation was not conditioned on participating in measure B1 anymore. The implementation in the model is analogous to the implementation of measure B4 of MEKA II (see section B.1.10) and not repeated here.

B.1.11 Investments

For the agent investment decision, the production problem described in the previous subsections is augmented by investment activities (x^{iB}). These include investments in tractors (x^{iZ}) and equipments (x^{iE}), the fixed and size-dependent part of biogas plants ($x^{i\alpha Ue}, x^{i\beta Ue}$) and infrastructure ($x^{i\alpha M}, x^{i\beta M}$), but also other decisions with effects that last longer than one season: for example, the 5-year commitment to a MEKA measure (x^{iYm}), and obtaining the right to sell biogas electricity at the guaranteed prices of the current year for the next twenty years (x^{iYy}).

In general, every investment relaxes the corresponding capacity constraint in the agent decision problem,

$$\cdots - a_b^{iB} x_b^{iB} \leq b_b^B \forall b \tag{B.99}$$

and b_b^B is increased by $a_b^{iB} x_b^{iB}$ after the decision has been taken and before entering the production decision stage for the current year.

Agent investments in those assets that are split into fixed and size-dependent parts are usually subject to the following condition:

$$x_b^{i\beta B} - Mx_b^{i\alpha B} \leq 0 \forall b \in M, Ue \tag{B.100}$$

In the investment decision, the production problem has been formulated for an average year in the near future. The objective function represents the annualized total farm gross margins of the next years, and consequently the objective function coefficients of the investment activities are the annualized investment cost, calculated as

$$c_b^{iB} = -\left(\eta_b \frac{a_b^B}{\lambda_b} + (1-\eta_b)a_b^B \iota_f \frac{(1+\iota_f)_b^\lambda}{(1+\iota_f)_b^\lambda - 1}\right) \tag{B.101}$$

with η being the share of the investment paid from equity, a^B the investment cost, λ the lifetime of the asset, and ι_f the interest rate on borrowed capital.

Investments are restricted by the liquid means available to the agent,

$$\sum_b \left(\eta_b a_b^B x_b^{iB}\right) - x^{tC} \leq 0 \tag{B.102}$$

further the continuous cash demand by equity fixed in the asset is considered in the general liquidity restriction (eq. B.45)

$$\sum_l a_{l,c}^{LC} x_l^L - x^{tC} - x^{bC} - \sum_b \left(a_b^{BC} x_b^{iB}\right) \leq 0 \tag{B.103}$$

with the corresponding coefficient (a^{BC}) being calculated as

$$a_b^{BC} = \eta_b a_b^B \left(\frac{(1+\iota_e)_b^\lambda}{(1+\iota_e)_b^\lambda - 1} - \frac{1}{\lambda_b \iota_e}\right) \tag{B.104}$$

with ι_e being the discount factor applied to equity, which is assumed to be equal to the interest rate on short-term deposits ($\iota_e = c^{dC}$)

B.1.12 Post harvest decisions

After harvest, the production decision problem would usually be solved again, with all land use activities fixed at the areas determined in the pre-season production decision, all expected yields replaced by the actual yields obtained and all expected prices replaced by the actual prices realized in the markets. This gives the agent the opportunity to adapt the production plan to the production results: Buying less or more feedstock on the market, or increasing or reducing animal or biogas production.

In the case of all the simulations run in this thesis, perfect foresight of prices and yields was assumed and this step could be omitted as the agent harvest results correspond to their planned values.

B.2 Demography: The farm household and farm succession

All farms are modeled as family farms, as this remains the predominant form of farming enterprise in the study area.

B.2.1 Composition of the farm household

Each agent household in the model consists at least of one male or female household head, the farm manager. Further, it may comprise the farm manager's spouse, their children, a retired household head and his/her spouse (usually the household head's parents), and in some cases also siblings of the household head.

New members enter the agent household either by birth or by marriage. All female household members between 15 and 49 have a positive probability of giving birth. A newborn household member is randomly assigned a gender and a career path as young farmer or young non-farmer. Nonfarmer household members have no interest in farming as their profession. They may work on the agent farm between the age of 14 and 19, but leave the household with 20 years. Young farmer household members, on the other hand, are eligible to succeed the current household head once they surpassed the age of 22. Whether they are employed on the farm or work somewhere else is part of the agent production decision. If they are over 23 and employed on the farm, they have to be paid and their remuneration is accounted as labor cost. A young farmer household member that did not become household head automatically retires and becomes ineligible for employment on the farm at the age of 65.

All unmarried household members (except young nonfarmers, children and seniors above 70 years) have a positive probability of marrying. The status of new household members marrying into the agent household is determined by his/her spouse: The spouse of a young farmer household member will be a young farmer household member, the spouse of a retiree is a retiree and the spouse of the household head is a specific type of its own.

Like marriage and giving birth, also the death of agent household members is determined randomly based on their current probability of dying, which depends on their gender and current age. Fertility, mortality and marriage probabilities have been calculated using statistical information from destatis [2012*a*,*b*,*c*]. The probability for a male newborn agent household member to be interested in farming (*potsuc_prob_male*) is assumed to range between 0.5 and 1, while the probability for a female newborn agent household member to become a young farmer is only 0.1, unless the household has a female household head, in which case it is 0.5. (This gender bias follows the patterns observed e.g. by Mann 2007)

The minimum household consumption in the model is 26,000 Euro for the household head and 8,000 Euro for each retiree (former household head or spouse of household head). If the agent income is higher than the minimum consumption, a certain percentage determined by parameter *sconextra* of the additional income is consumed in addition. When agent income falls below minimum consumption, it is consumed entirely. A certain share of minimum consumption determined by the parameter *sconred* is consumed even if income does not suffice to cover it. Cash is then either taken from the agent farm cash reserves or if no cash is left the agent is bankrupt and leaves the model.

In the model, household heads and retirees until the age of 70 as well as young farmer household members over the age of 19 count as full workers. Household members between 14 and 18 years of age have a labor provision of 30% of a full worker. Spouses of household heads provide 60%, retirees between 70 and 75 years 50% and retirees between 76 and 80 years 20% of the labor capacity of a full-time worker.

B.2.2 Household head succession

Mann [2007] groups the factors influencing the decision to take over a farm business into identity-related and environmental factors. Following this concept, the identity-related factors are represented in the model by the distinction between young farmer and young nonfarmer household members, which is modeled as a purely statistical relationship.

The environmental factors are mainly related to the economic situation of the farm and determine whether a potential successor who is generally interested in farming finds it worthwhile to take over the farm once the decision has to be taken. In the model, farm succession, i.e. passing the responsibility of the household head to another household member, can be triggered by either death or retirement of the current household head. Succession requires the availability of a potential succeeding household member and the fulfillment of certain economic preconditions for the successor to accept the succession. These preconditions, the eligibility of household members for succession and the consequences of an unsuccessful succession depend on the event which triggered the attempt for succession.

In the case of the death of the current household head, potential candidates for succession in our model are the young farmer household members with at least 23 years of age and the spouse of the deceased household head, in case this member is not older than 65 years. Succession succeeds if at least a share (*suc_mincons*) of the minimum household consumption is achieved. If succession fails in the case of death, the agent farm will be shut down.

Two different cases of retirement of the current household head are distinguished in the model: Between the age of 55 and 70, household heads may make a voluntary attempt to retire. Household heads between 55 and 64 attempt to retire in a given year with a probability of 10%, household heads above 65 will attempt to retire every year. Only young farmer household members with at least 23 years of age are eligible for succession. Succession will succeed only if they have been employed on the agent farm and the agent income covers at least *suc_mincons* times the minimum household consumption of the farm household after succession. If succession fails, the current household head will remain farm manager and try to retire later. Household heads above 70 are obliged to retire, and if they do not find a successor or the later does not accept, the agent farm is shutdown. Only young farmer household members of at least 23 years of age are eligible for succession and these will succeed if the farm income covers at least *suc_mincons* times the minimum consumption of the farm household after succession.

This tiered system of retirement was implemented in the model to make succession independent of a casual bad year, and let agents choose a suitable situation for succession.

If several potential succeeding household members are available, the one with the highest priority becomes the new household head. The priority ranking is as follows: the oldest male young farmer household member between 23 and 45 years has highest priority, followed by the youngest male young farmer between 46 and 65 years, the oldest female young farmer between 23 and 45 years, and the youngest female young farmer between 46 and 65 years. In the case of death of the current household head, the spouse of the deceased household head follows with lowest priority.

B.2.3 Influence on investment and production decisions

Apart from determining labor provision, household consumption and a potential closing of the agent farm business due to the death or retirement of the farm manager, the household composition also affects the agent production and investment decisions in two other ways:

First, employment of a young farmer household member – though considered labor cost in the financial accounting of the agent farm –is not considered a cost by the agent farm manager during planning as soon as the minimum consumption of the household is expected to be covered by the agent farm income. This model implementation reflects the empirical observation that farm managers actually tend to enhance their business in order to be able to employ their potential successors, potentially even reducing their own income.

This condition is implemented in the model by distinguishing between ordinary household labor (b^{H1}) and young farmer labor (b^{H2}) . Employing young farmer labor x^{H2} is associated with a cost (c^{H2}), while employing other household labor x^{H1} is not. x^{H2} is an integer activity corresponding to 25% of a full workload, making sure, that the amount of hours worked by the young farmer household member is meaningful. If the total farm gross margin surpasses the sum of minimum consumption, depreciation, rental payments and other fix costs, the cost for employing young farmer household members can be (partially) offset or even overcompensated depending on the value of the parameter ζ_{H2ut}. (The either-or condition is implemented using the two auxiliary integer activities x^{yH2ut}, x^{nH2ut} .)

$$\begin{array}{lllllll}
\sum cx & & -0.25c^{H2}x^{H2} & +\zeta_{\mathrm{H2ut}}c^{H2}x^{H2ut} & & & \rightarrow & \mathrm{obj} \\
& x^{H1} & & & & & \leq & b^{H1} \\
& & 0.25x^{H2} & & & & \leq & b^{H2} \\
& & -0.25x^{H2} & +x^{H2ut} & & & \leq & 0 \\
& & & x^{H2ut} & -Mx^{yH2ut} & & \leq & 0 \\
& & & & x^{yH2ut} & +x^{nH2ut} & \leq & 1 \\
\sum cx & & -0.25c^{H2}x^{H2} & & & +Mx^{nH2ut} & \geq & b^{Cmc} + b^{Cf}
\end{array} \tag{B.105}$$

Second, the age of the household head and the availability of a potential succeeding household member both influence the investment horizon of the farm. To avoid that farm managers close to retirement without successor make investments that pay out only over a long time, the expected remaining farm life is used in the agent investment calculus instead of the expected lifetime of an asset, whenever the later is greater than the former. The expected remaining farm life is the remaining time until the current household head turns 65. In an alternative implementation, in case a potential successor is present, the time until the potential successor will turn 65 is used instead.

B.3 Land markets

The lowest spatial unit in MPMAS is a map cell, which in the case of the Central Swabian Jura model corresponds to one hectare of land. Each map cell is associated to an owner and a user, which may be identical, but can also differ if some form of land market is used in the simulation.

The cells used by an agent form the land available for farming and appear as capacity on the right hand sides of the corresponding soil constraints of the MIP problem. For every cell that the user does not own itself, a rent must be paid to the owner.

If land markets are activated, agents may decide to rent in land or to offer part of their land on the market. To determine potential suppliers and renters, MPMAS determines the average shadow price for each type of soil. The average shadow price for a soil type is estimated by solving the production decision problem of every agent with an additional unit of land of the respective type. The resulting total gross margin is compared with the original total gross margin achievable on the land currently in use by the agent and this additional total gross margin is averaged over all agents.

Agents will offer those cells of a given soil type for which the shadow price times a *markup* parameter minus the assumed transport cost from farmstead to cell is lower than the average regional shadow prices. A special 'other land owner'-agent represents the owners of all map cells that do not belong to any of the farm agents in the model. These cells are always on offer.

Agents that decided not to offer cells of a given soil type will try to rent them in in case their individual shadow price multiplied with a *markdown* parameter is greater than the average shadow price. At the same time a maximum distance for cells considered for rental is

Table B.3: *Rental prices for rental contracts existing at simulation start depending on start year.*

Start year	Arable land	Grassland
1999	199	119
2003	207	112
2007	210	116
>=2010	221	117

calculated based on the difference between individual and average shadow price. An overall maximum can be specified by the modeler.

MPMAS then goes through all offered cells and tries to find a potential renter by identifying the bidder with the highest bid among those bidders whose farmstead is closer to the cell than their maximum renting-in distance. The bid is calculated from the individual shadow price of the potential tenant by subtracting transportation cost. If this bid is higher than the minimum rent expected by the owner a rental contract is made and the rental payment is set to the average between bid and expected minimum rent. Both, owner and tenant reassess their shadow prices and decide whether to offer or bid for further cells.

If no contract could be established, the cell is taken out of the auction process. Auctions continue until either no offered cells or no bids remain.

The original land market implementation of MPMAS was designed for a Chilean application, where only short-term rental contracts were relevant: Rental contracts had a duration of one year, and, since no long-term investment decisions could be taken on this basis, the rental decisions were taken based on the production decision problem for the current period after the investment decision was already taken.

For the Central Swabian Jura, the MPMAS land market was adapted to be able to simulate also long-term rental contracts. In this enhanced version, the shadow price is assessed using the investment decision problem and rental contracts have a standard duration that can be defined by the modeler. Further, the model can be initiated with existing rental contracts.

For most of the simulations in this thesis, only rental contracts existing at simulation start are considered and assumed to have unlimited duration. The owner of all rented cells is the 'other land owner' agent. Agents return cells to their owner in case they are not able to pay the land rent or sell owned land to avoid bankruptcy. The land of agents that give up farming and exit the simulations is also included into the portfolio of the 'other land owner' agent. Rental payments for arable and grassland have been set according to state-wide averages reported in public statistics [Statististisches Landesamt Baden-Württemberg, 2013] and shown in table B.3. Additional transport costs were not considered.

In the final recursive dynamic simulations with activated land market, agents were allowed to expand farm area by renting land from the 'other land owner' agent, but could not offer land themselves.

B.4 Yield simulation with Expert-N

Within the joint research project of Hohenheim University and Helmholtz Center Munich, fellow researchers employed the plant and soil modeling package Expert-N [Stenger et al., 1999] to simulate crop yields. Expert-N is a modularized framework that provides a selection of different models for plant and soil processes: DAISY [Müller et al., 2003] was used to simulate nitrogen and carbon dynamics in the soil. Wheat, maize and barley yields were simulated using CERES [Ritchie et al., 1998; Ritchie, 1991], and rapeseed using GECROS [Yin and van Laar, 2005].

Expert-N models were calibrated using year-round field observations from three sites and two years in the study area. Spatial information of soil properties in the study area was obtained using the LUBW [2007] digital soil map. The nine relevant soil mapping units were linked to reference soil profiles from the experimental sites and the profile database of the Institute of Soil Science and Land Evaluation of Hohenheim University. Pedo-transfer functions were used to estimate soil properties. A meteorological time-series for the years 1951-2010 from the weather station in Stötten was used as weather input [Calberto Sanchez, 2015].

While a run-time coupling between MPMAS and Expert-N has been implemented for the research project [Troost et al., 2012, 2010], in the present study the two models are used as a model cascade (fig. B.1).

The models use a shared schedule of crop management including timing and amount of tillage, sowing, fertilization, and harvesting, which is stored in a shared MySQL database. Results from Expert-N simulations, i.e. the predicted crop yields for each potential combination of weather conditions, soil type and crop management, are fed back into the database and then used in MPMAS.

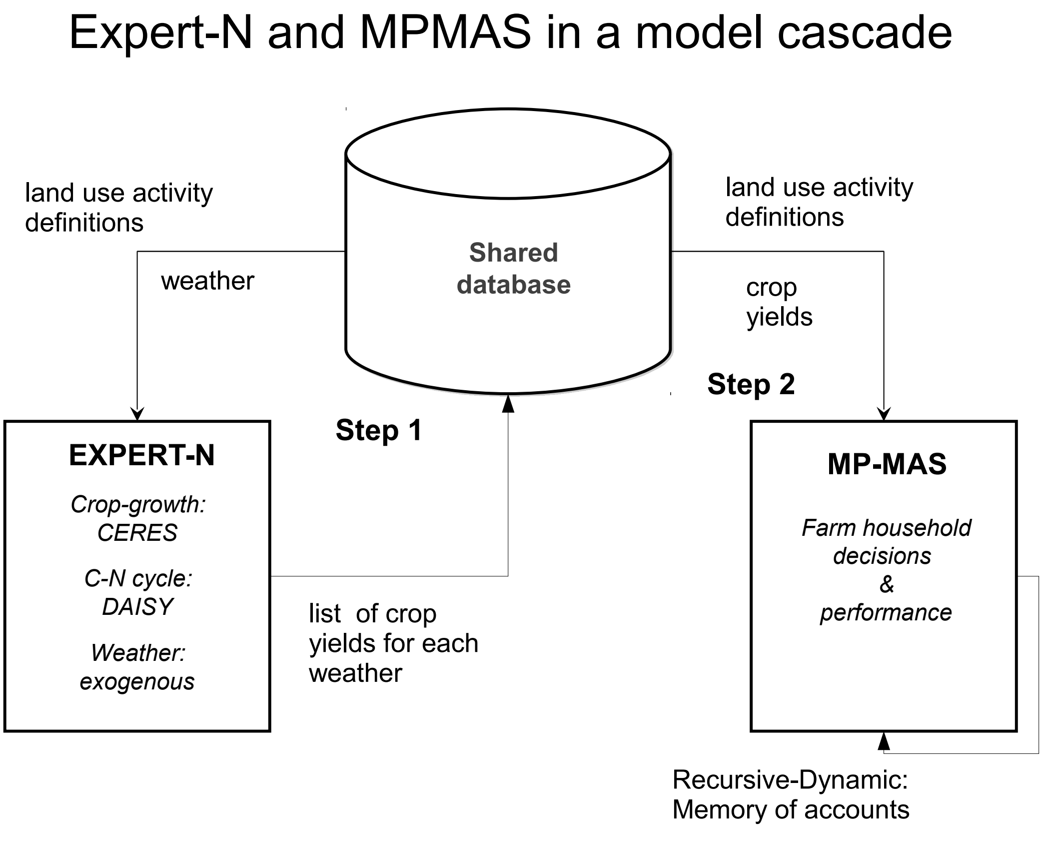

Figure B.1: *Integration of MPMAS and Expert-N in a model cascade*

Appendix C

Data

Table C.1: *Suitable days for field work by work season in different KTBL zones for medium soil resistance (Source: KTBL [2010], own classification of seasons)*

			Weather sensitivity level [1]				
Field work season	Probability[2]	Zone	1	2	4	5	6
SPR	80%	4		1	3	9	27
Spring		5		0	3	11	31
early March		7		1	4	18	40
– early May	60%	4		3	5	19	40
		5		2	5	21	44
		7		3	6	28	53
ESU	80%	4		14	24	21	37
Early summer		5		12	23	28	43
late May		7		17	32	38	50
– early July	60%	4		21	31	33	46
		5		21	31	41	51
		7		27	37	49	56
HWB	80%	4	3.11	5	7	6	10
Harvest winter barley		5	1.89	4	8	7	11
late July		7	3	6	10	11	13
	60%	4	7.22	8	10	9	12
		5	6.22	7	10	11	14
		7	8.33	9	11	13	15
HWR	80%	4	1.89	5	8	6	9
Harvest winter rapeseed		5	1.89	5	8	8	11
early August		7	3.33	7	10	11	13
	60%	4	6	8	9	9	12
		5	6.33	7	10	11	13
		7	8.56	9	11	13	14

[1] 1 - cereal harvest; 2- hay harvest (soil dried); 4 - harvest of grass silage; 5 - medium sensitive activities such as harvest of silage maize, mineral fertilization, and sowing; 6 - less sensitive activities such as organic fertilization and incorporation of crop residues into the soil

[2] For sensitivity level 1, KTBL lists only the probability level for 80% for different levels of grain humidity, we use the values for 14% and for 16%, respectively.

Suitable days for field work by work season (cont.)

Season	Probability	Zone	Weather sensitivity level 1	2	4	5	6
HWW	80%	4	1.33	4	7	6	10
Harvest winter wheat		5	1.67	4	7	7	11
late August		7	2.56	5	9	10	13
	60%	4	4.89	7	9	10	13
		5	5.56	6	9	11	14
		7	7.56	8	11	13	15
SP1	80%	4		3	6	6	11
Early September		5		2	6	7	11
		7		3	7	10	13
	60%	4		6	9	9	13
		5		4	8	11	13
		7		6	9	13	14
HSM	80%	4		0	0	7	11
Harvest silage maize		5		0	0	8	12
late September		7		0	0	11	13
	60%	4		0	0	11	13
		5		0	0	11	13
		7		0	0	13	14
AUT	80%	4		0	0	10	27
Autumn		5		0	0	7	26
early October		7		0	0	17	35
– early November	60%	4		0	0	22	37
		5		0	0	19	36
		7		0	0	31	42
WIN	80%	4		0	0	0	2
Winter		5		0	0	0	2
late November		7		0	0	0	7
– late February	60%	4		0	0	0	5
		5		0	0	0	6
		7		0	0	3	9

Table C.2: *Overview of soil mapping units and their link to reference profiles (classification and link to soil profiles by Joachim Ingwersen; area shares own calculation based on LUBW 2007).*

Soil classification			Share of					
German	WRB	Area [ha]	total	arable	forest	grassland	mixed	Profile ID [a]
Rendzina, Braunerde-Terra fusca, Terra fusca-Parabraunerde	Rendzic Leptosols, Chromic Cambisols, Chromic Luvisols	77,799	60.0%	48.0%	60.4%	68.9%	67.5%	0
Pararendzina	Calcaric Regosols	5,509	4.2%	5.0%	2.8%	4.7%	5.0%	1
Braunerde-Pelosol, Pseudogley-Pelosol, Pararendzina	Vertic Cambisols, Vertisols	1,007	0.8%	1.6%	0.5%	0.0%	1.1%	2
Parabraunerde, Terra fusca, Terra fusca-Braunerde	(Chromic) Luvisols, Chromic Cambisols	5,557	4.3%	12.6%	2.6%	1.6%	0.9%	3
Parabraunerde, Pararendzina-Braunerde	Luvisols, Cambisols	506	0.4%	0.7%	0.1%	0.3%	0.2%	4
Parabraunerde	Luvisols	553	0.4%	1.4%	0.2%	0.0%	0.1%	5
Braunerde-Terra fusca	Cambisols, Chromic Cambisols	8,665	6.7%	12.8%	4.1%	3.9%	7.0%	6
Brauner Auenboden, Auengley	Fluvisols	2,826	2.2%	1.2%	0.6%	5.1%	2.3%	7
Kolluvium	Anthrosol	10,024	7.7%	14.1%	1.2%	9.7%	10.2%	8
Rendzina, Braunerde-Rendzina	Rendzic Leptosols, Mollic Leptosols	16,802	13.0%	2.5%	27.5%	5.5%	5.6%	0
Niedermoor	Histosols	97	0.1%	0.1%	0.0%	0.4%	0.0%	-
Ortslage	Build-up areas	400	0.3%	0.2%	0.1%	0.0%	0.0%	-
Total		129,746	100.0%	100.0%	100.0%	100.0%	100.0%	

[a]See Table C.3 for linked soil profiles.

Table C.3: *Reference soil profiles used in Expert-N simulations (as communicated by Joachim Ingwersen and Germán Calberto).*

Profile ID	Source	Horizon	Lower depth [cm]	Texture	Bulk density [g/cm^3]	Total C_{org} %	C/N
0	PAK-EC6	Ap1	12	Tu2	1.04	3.314	9.8
		Ap2	21	Tu2	1.29	2.540	10.3
1	DE-7817-3	Ap	30	Tu2	1.14	2.140	10.7
		2lCv	60	Lts	1.65	0.290	9.7
		mCv	82	Lt2	1.36	0.230	11.5
2	DE-7619-4	Ah	16	Tu2	1.22	2.950	10.2
		rApBv	35	Tu3	1.23	0.920	8.4
		BvP	65	Tu2	1.21	0.630	7.0
		lCvP	120	Tu2	1.21	0.340	4.9
3	DE-6623-12	Ah	6	Tu4	0.98	3.300	15.0
		AhAl	26	Tu4	1.31	1.210	15.1
		Bt	41	Tu3	1.41	0.580	9.7
		2T1	58	Tu2	1.42	0.630	10.5
4	DE-7926-204	Ah	6	Ls2	1.38	3.480	10.2
		Al	30	Ls2	1.41	1.970	12.3
		2Bt	50	Lt2	1.37	0.460	5.8
5	DE-IBS-265	Ah	5	Sl3	1.19	4.500	16.1
		Al	30	Sl4	1.45	1.200	13.3
		Btv	45	Ls4	1.56	0.500	10.0
		Bt	78	St3	1.39	0.100	10.0
		Cv	88	St2	1.39	0.010	10.0
		C	100	Sl3	1.40	0.010	10.0
6	PAK-EC4	Ap1	21	Tu3	1.31	2.630	9.5
		Ap2	29	Tu3	1.34	1.293	9.8
		Tv	41	Tu2	1.32	0.972	9.1
7	DE-7518-1	rAp	30	Lu	1.27	2.490	8.6
		M1	82	Lu	1.40	1.040	8.0
		M2	140	Lu	1.37	0.010	-
		M3	167	Lu	1.47	0.010	-
8	PAK-EC5	Ap	20	Tu4	1.37	2.171	9.4
		eM1	60	Tu3	1.40	1.063	9.3
		eM2	90	Tu3	1.51	0.380	6.3

Appendix D

Agent Population

D.1 Full-time farm classification (FADN)

Table D.1: *Full-time farm classification (FADN)*

Year	Class	Limits
1999	full-time	SBE $\geq$ 15,000
	part-time	5,000 $\leq$ SBE $\leq$ 15,000
	not-represented	SBE $<$ 5,000
2003, 2007	full-time	$\geq$ 16 ESU and $\geq$ 1 labour unit
	part-time	$\geq$ 8 ESU, but $\leq$ 16 ESU or $\leq$ 1 labour unit
	not-represented	$<$ 8 ESU
	SBE	*'Standardbetriebseinkommen', standard farm income*

D.2 Allocation of machinery

Table D.2: *Rules used to distribute machinery among the agents*

Machinery	Conditions
Tractors	
157, 102, 67 kW	arable > 160 ha, or manure $\geq$ 3200m^3
120, 83, 45 kW	arable > 70 ha or grassland $\geq$ 180 ha
102, 67, 45 kW	arable > 50 ha, or manure $\geq$ 1200m^3
83, 45 kW	arable > 20 ha, or manure $\geq$ 200m^3
45 kW	grassland $\geq$ 15 ha, or dairy cows $\geq$ 10, or medium-aged cattle $\geq$ 20
Tillage and seeding implements	
seeder 2 m, plough 0.7 m	arable 20-50 ha
seeder 3 m, plough 1.05 m	arable 50-160 ha
seeder 4 m, plough 1.75 m	arable > 160 ha
Spraying and fertilizing equipment	
15 m	arable 20-110 ha
24 m	arable > 110 ha

Table D.2: *Rules used to distribute machinery among the agents (cont.)*

Machinery	Conditions
Maize seeder	
	if biogas, or dairy cows > 30, or medium-aged cattle > 50, and …
3 m	… arable 20-160 ha
6 m	… arable > 160 ha
Manure trailer	
7 m^3	manure 200-1200 m^3
12 m^3	manure 1200-3200 m^3
20 m^3	manure $\geq$ 3200m^3
Manure drag hose	
12 m	manure 2000-3200 m^3
24 m	manure $\geq$ 3200m^3
Machinery combination for grass harvest (mowing, stirring, swathing)	
5 m	grassland $\geq$ 180 ha
3.2 m	grassland $\geq$ 90 ha, or dairy cows $\geq$ 104, or medium-aged cattle $\geq$ 150
2.4 m	grassland $\geq$ 20 ha, or dairy cows $\geq$ 10, or medium-aged cattle $\geq$ 20
Round baler	
1.2 m	grassland $\geq$ 20 ha
Self-loading trailer	
20 m^3	grassland $\geq$ 20 ha
Loader	
102 kW	if arable > 30 and (dairy cows > 60, or medium aged cattle > 120, or biogas)
Grassland cultivation	
roller 3 m, grass harrow 4 m	grassland 25-50 ha
roller 6 m, grass harrow 9 m	grassland $\geq$ 50 ha
Combine harvester	
125 kW, 4.5 m	arable 100-180 ha
175 kW, 6 m	arable $\geq$ 180 ha

D.3 Imputing the maximum of the marginal inverse cdf given population total

$$
\begin{aligned}
Total_v &= \sum_{i=1}^{N} icdf_v\left(\frac{i}{N}\right) \\
z &= \text{round}\,(.99N) \\
Total_v &= \sum_{i=1}^{z} icdf_v\left(\frac{i}{N}\right) + \sum_{j=z+1}^{N} icdf_v\left(\frac{j}{N}\right) \\
p_z &= \frac{(z-0.5)}{N} \\
\Delta &= Total_v - \sum_{i=1}^{z} icdf_v\left(\frac{i}{N}\right) \\
\Delta &= \sum_{j=z+1}^{N} icdf_v\left(\frac{j}{N}\right)
\end{aligned}
$$

Assuming linear interpolation we can set

$$
\begin{aligned}
icdf_v(p_j) &= icdf(p_z) + (j-z)*m \quad \forall\, j \in \{z+1, ..N\} \\
\sum_{j=z+1}^{N} icdf_v\left(\frac{j}{N}\right) &= \sum_{j=z+1}^{N} icdf(p_z) + (j-z)*m \\
&= (N-z)*icdf(p_z) + m*\frac{(N-z)*(N-z+1)}{2} \\
\Leftrightarrow m &= \frac{2\,(\Delta - (N-z)*icdf(p_z))}{(N-z)*(N-z+1)}
\end{aligned}
$$

and derive

$$icdf_v(1) = icdf(p_z) + (N-z)*m$$

Appendix E

Results

E.1 Turing test

Table E.1: *Turing test: Ranking of production plans by participants' choice and goodness-of-fit. (The underlined alternative is the true, observed alternative.*

Farm	Alternative	Rank chosen	Rank ESAE	ESAE (arable)	ESAE (arable + grassland)
1	A	1	2	0.50	0.79
	B	6	5	0.00	0.58
	<u>C</u>	5	1	1.00	1.00
	D	4	6	-0.05	0.56
	E	2	4	0.10	0.62
	F	3	3	0.20	0.77
2	A	4	6	0.01	0.45
	B	5	2	0.47	0.70
	C	1	5	0.21	0.56
	D	2	3	0.35	0.64
	<u>E</u>	6	1	1.00	1.00
	F	2	4	0.23	0.57
3	A	4	4	0.35	0.35
	B	1	2	0.70	0.70
	C	3	3	0.56	0.56
	D	2	6	0.21	0.21
	<u>E</u>	6	1	1.00	1.00
	F	6	5	0.33	0.33
4	<u>A</u>	4	1	1.00	1.00
	B	2	2	0.36	0.58
	C	3	4	-0.03	0.32
	D	5	5	-0.13	0.25
	E	1	3	0.18	0.45
	F	5	6	-0.33	0.12

E.2 Calibration & Validation

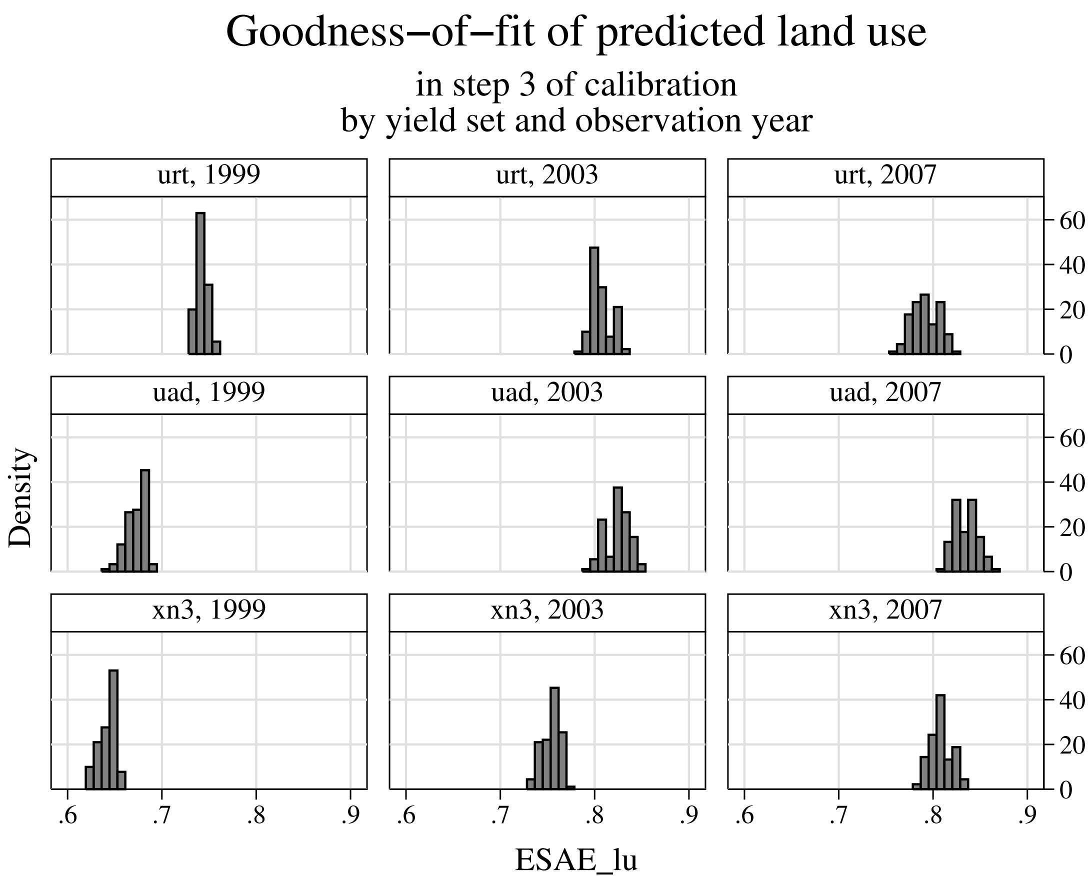

Figure E.1: *Goodness-of-fit of predicted land use (Calibration Step 3).*

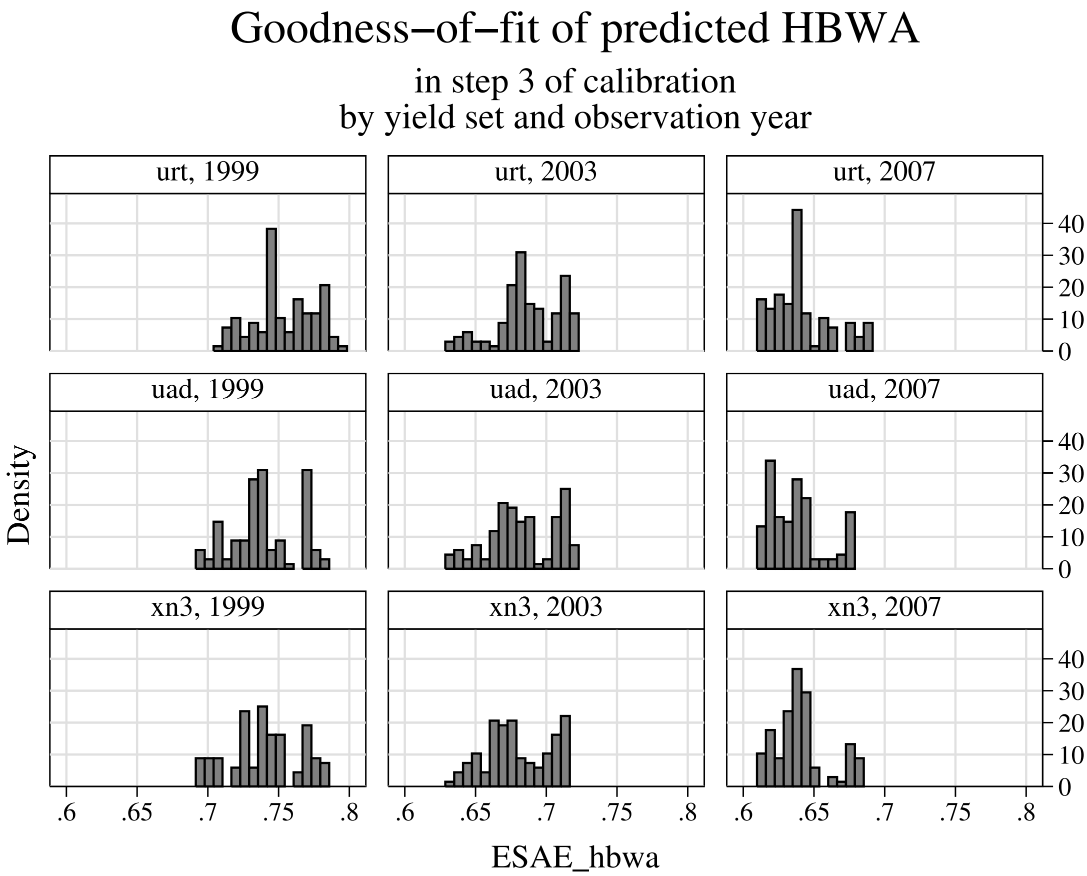

Figure E.2: *Goodness-of-fit of predicted PTOF (Calibration Step 3).*

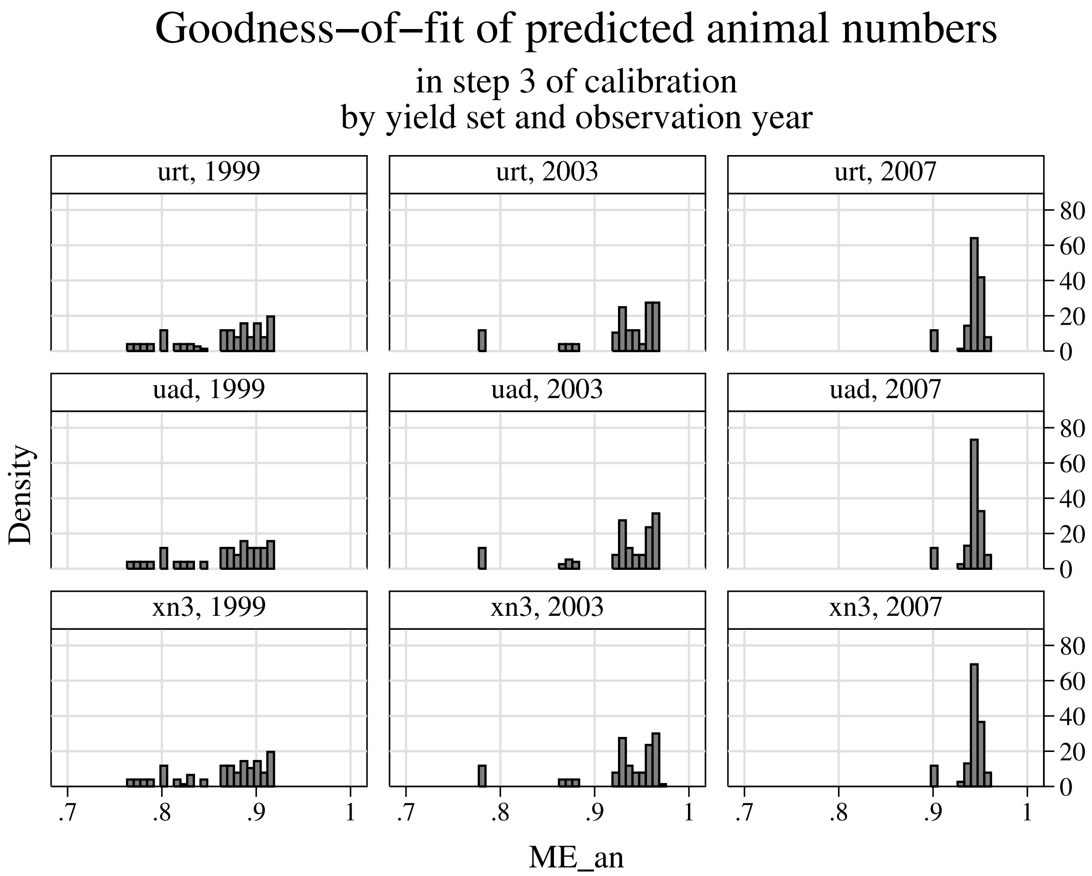

Figure E.3: *Goodness-of-fit of predicted animal numbers (Calibration Step 3).*

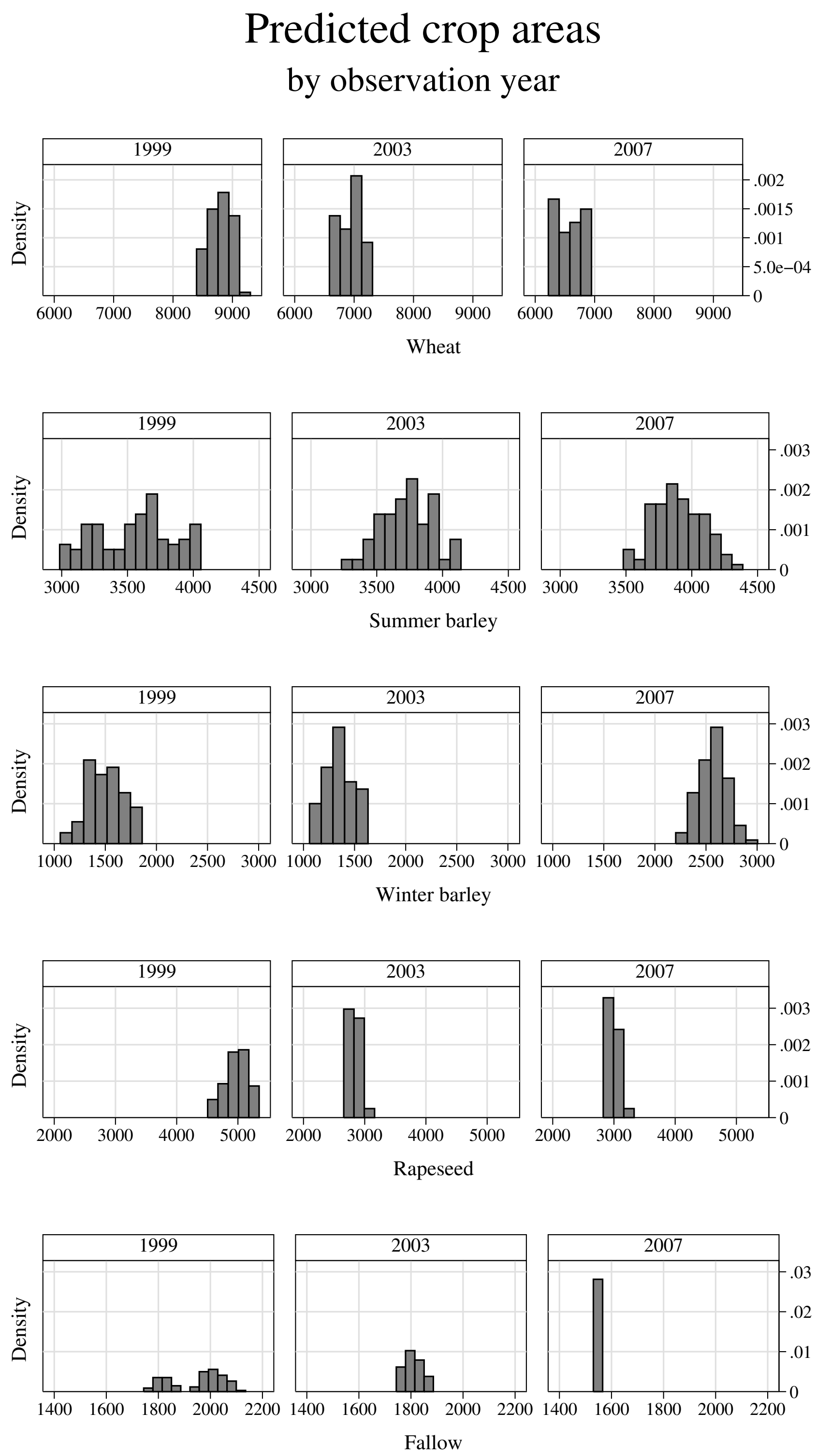

Figure E.4: *Distribution of predicted cereal, rapeseed and fallow areas over 96 validation runs.*

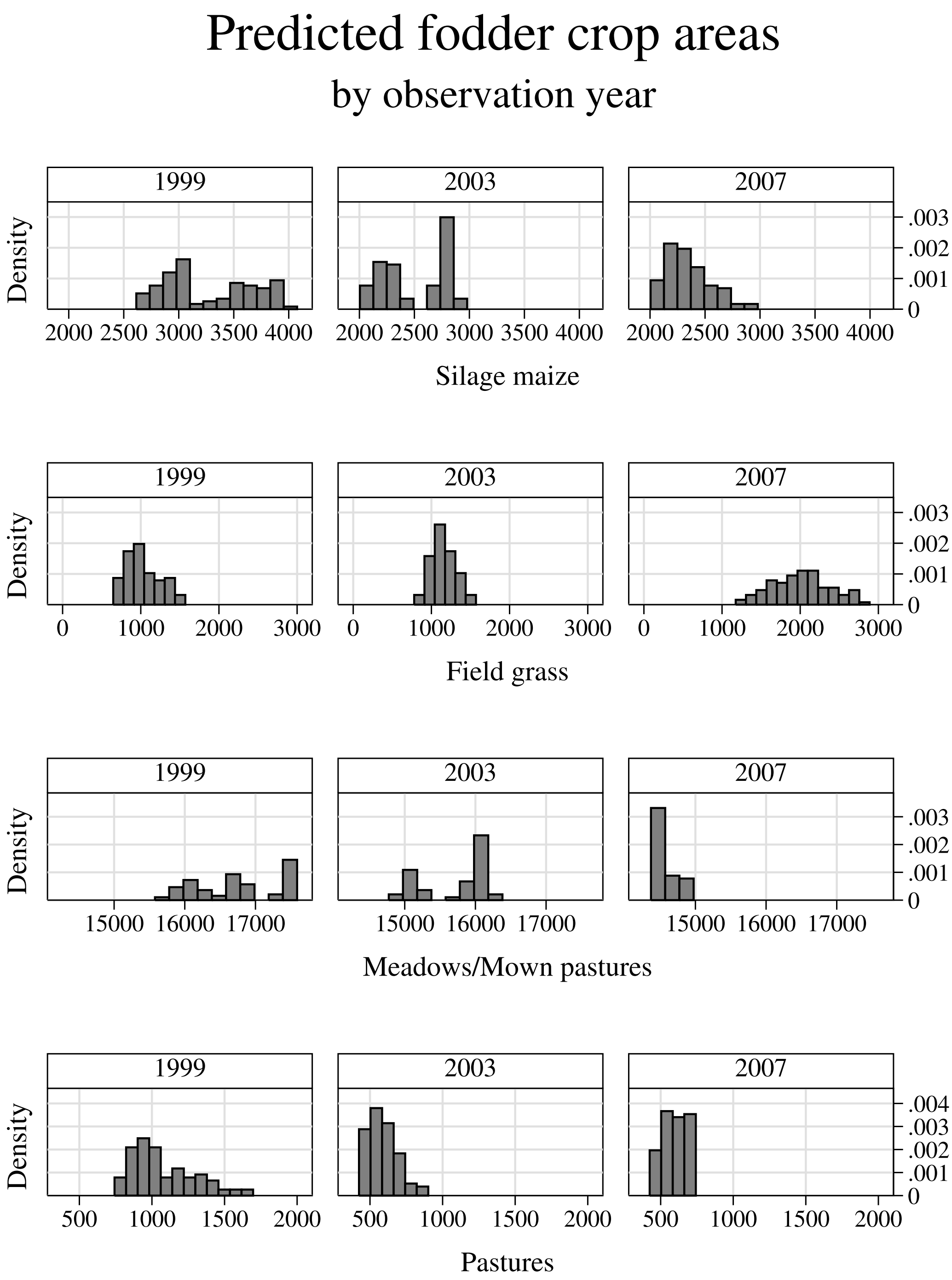

Figure E.5: *Distribution of predicted forage areas over 96 validation runs.*