Improving and Comparing Data Collection Methodologies for Decision Rule Calibration in Agent-Based Simulation. A Case Study of Dairy Supply Chain in Indonesia

Dhanan Sarwo Utomo

This thesis is submitted for the degree of Doctor of Philosophy

September 2019

Lancaster University Management School

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated. Many of the ideas in this thesis were the product of discussion with my supervisor Dr. Stephan Onggo and Dr Stephen Eldridge.

Excerpts of this thesis have been submitted or published in the following conference academic publications.

- Utomo, D. S., Onggo, B. S., & Eldridge, S. (2018). Applications of agent-based modelling and simulation in the agri-food supply chains. European Journal of Operational Research. 269 (3), 794-805.
- Utomo, D. S., Onggo, B. S., Eldridge, S., Daud, A. R., & Tejaningsih, S. (2018). Using Scenario-Based Questionnaire for Agent-Based Simulation Model Calibration and Validation: An Agri-food Supply Chain Example. Submitted to European Journal of Operational Research.
- Utomo, D. S., Onggo, B. S., Eldridge, S., Daud, A. R., & Tejaningsih, S. (2018). Improving the Credibility of Agent-Based Simulation using Role Playing Games: A Case Study in Dairy Supply Chain. In preparation for submission to European Journal of Operational Research.

Dhanan Sarwo Utomo

Lancaster University, UK

Abstract

This study contributes to human behaviour (decision rule) modelling in the agent-based simulation, by improving the existing data collection methodologies and comparing their benefits. Improving data collection methodologies can help in developing a more realistic agent's decision rule and increasing the validity and credibility of the final model. This study uses a dairy supply chain case because the actors in this context can have one to one correspondence with the agents in the simulation.

This study begins by presenting a literature review on the applications of agent-based simulation in the agri-food supply chain. This literature review highlights existing agent-based modelling practices in the agri-food supply chain such as the scope of the modelling, data collection, validation and sensitivity analysis techniques. This study then proposes some improvements to the existing data collection methodologies namely questionnaire survey and role-playing game. This study proposes the use of a scenario-based questionnaire to improve the benefits of a questionnaire survey for decision rules calibration. While to extend the usefulness of role-playing game this study propose the use of the design of experiment, and game scaling based on empirical probability distribution.

The improved data collection methods are then used to calibrate a base model that was developed from the previous literature. Primary data from 16 villages in Indonesia is used to elicit empirical decision rules in this calibration process. The result from simulation experiments shows that the improved data collection methods can produce models with higher operational validity. This study is concluded by evaluating the advantages and disadvantages of each data collection methodology.

Acknowledgements

Thank you to Dr. Stephan Onggo and Dr Stephen Eldridge for the excellent supervision, support and advice they have offered me throughout my PhD. I would also like to thank Mrs Gay Bentinck who has helped me to deal with all administrative and studentship matters during my PhD. And finally to my family, for allowing me to choose my own path and giving me their unconditional support

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ABS: Agent-based simulation

ASC: Agri-food Supply Chain

RPG: Role Playing Game

SBQ: Scenario-Based Questionnaire

List of Appendices

1 Introduction

1.1 Background and Motivation

Agent-based simulation (ABS) is an operational research/management science (OR/MS) technique that is gaining popularity in supporting decision making. ABS attempts to open the black box of modelling by allowing researchers to identify both the agents' behaviours and how the agents make decisions within the system. Therefore, it allows researchers to relate the human decision rules to the emergent/macroscopic patterns. ABS does not necessarily guarantee an improvement in the model's predictive capabilities, but it does provide the opportunity to generate more insights into how the system works from the perspective of the agents' behaviours. These insights can have a practical value for policymakers because they enable the evaluation of interventions aimed at modifying agent behaviours even if other system parameters are beyond the policymakers' direct control.

However, the use of ABS to support decision making is still hindered because the complex modelling process is often considered not transparent and its results are difficult to explain. Because it aims to simulate micro process (decision rules and behaviour), there are more elements in ABS which require calibration (Robinson et al., 2007), and the empirical validation process for ABS also tends to be more complicated (Heath et al., 2009). Unfortunately, behavioural modelling is still a challenge in ABS methodology owing to the difficulties in acquiring the data necessary to develop a more realistic agent's behaviour (Macal, 2016). This approach to behavioural modelling is also important in other research fields such as operations management (Bendoly et al., 2006, Bendoly et al., 2010).

Data collection is also the most common problem in any simulation project, especially when modelling systems with high complexity (Onggo et al., 2013). There are previous studies aimed at comparing the benefits of different data collection methods in a simulation study and in developing ABS.

- Eldabi et al. (2002) compared the benefits and potential biases of quantitative and qualitative data collection methods in a simulation study.
- Janssen and Ostrom (2006) summarised the benefits of different empirical methods to calibrate ABS in social sciences (i.e., survey with a close-ended questionnaire, case studies, stylized facts, role-playing games, and laboratory experiments) from various studies.
- Robinson et al. (2007) summarised experiences from many researchers in using various data collection methods (i.e. sample surveys, participant observation, field and laboratory experiments, companion modelling and GIS) to calibrate ABS in land use science. They presented one case study for each data collection method.
- Yang and Gilbert (2008) highlighted the benefits of ethnographic data in building ABS when compared to quantitative data collected using a questionnaire.

- Smajgl et al. (2011a) mapped the usefulness of various data collection methods (survey, interview, experiment, observation, RPG, and expert knowledge) in each step of model parameterisation. They presented five case studies to demonstrate different steps of model parameterisation.

- An (2012) reviewed applications of ABS to model human decision in coupled human-natural systems, including in term of data collection methods.

The previous literature has shown that each data collection method has its advantages and disadvantages. However, ABS research which aims to improve a data collection method and test its benefits in increasing a model's validity is considerably rare, especially in the field of agri-food supply chains that become the context of this study.

Additionally, when comparing the benefits of various data collection methods, previous literature mainly draws lesson learned from different case studies. This practice can be biased because the complexity of the case under study may influence the benefits of a data collection method in calibrating an agent's decision rule. Hence, this research seeks to improve and compare different data collection methods in the same case study and the same target population. Since this study uses the same case study, the calibrated models should produce the same outputs. This strategy enables us to examine and compare the model validity resulting from different calibration approaches. It also allows us to compare respondents' experience (from the respondents' perspective) during their participation in the two data collection processes.

1.2 Research Objectives

Firstly, this study aims to identify the limitations of the existing empirical data collection methods in calibrating agents' decision rules in ABS. Two data collection methods were selected namely questionnaire survey which is a quantitative deductive approach, and role-playing game (RPG) that is more inductive qualitative. Further reasons why these two data collection methods were selected are explained in Chapter 2.

Secondly, this study proposes potential improvements to the questionnaire survey and RPG methods, in order to reduce their weaknesses and extend their benefits. Improvement on the questionnaire survey method that was done by incorporating scenarios is discussed in Chapter 3. The process to improve the RPG by incorporating the design of experiment and increasing the correspondence between RPG and the reality is discussed in Chapter 4.

Thirdly, this study compares the benefits of these data collection methods in calibrating decision rules in ABS. In this comparison the following hypotheses are discussed:

- H1: Different data collection methods can produce empirical decision rules with different properties. Properties in this study include the structure of the decision rule, whether it can incorporate the context of an agent's decision, and whether the decision rules can be related to the previous theories.

- H2: Different data collection methods can produce empirical models with different levels of operational validity. A model's operational validity in this study is measured based on the match between the model's outputs and the real world data.

- H3: Different data collection methods have different benefits for decision rule calibration in ABS. Benefits in this study include the potential biases that can be eliminated and how each data collection method may help the researchers to develop a more realistic decision rule.

Finally, through computer experiments, this study proposes several behavioural interventions for the real system.

This study uses a context of dairy supply chain in Indonesia to test the benefits of the data collection methods mentioned above. This context is appropriate because the respondents (smallholder farmers) mainly controlled their own decisions. Hence, it is likely that their responses reflect their behaviour in reality. Chapters 2 and Chapter 3 further explain the uniqueness of the context discussed in this study compared to previous ABS applications.

1.3 Research Methodology

This study began by selecting a case study appropriate to test the research hypotheses. The case study that was selected is the dairy supply chain in Indonesia. The reason for this case study selection is explained further in Chapter 2.

Based on the literature collected in Chapter 2, a base model was developed (see Figure 1.1). All of the assumptions in the base model were face validated by the experts in the case study site. The assumptions in this base model serve as hypotheses to be tested with empirical data collection. Important parameters to initiate the base and calibrated models were also identified through empirical data collection. Empirical data collection using the scenario-based questionnaire was done in 2016. 153 farmers in 16 villages in West Java were involved in this data collection. The RPG data collection was done in 2017 and involved 24 farmers. In each data collection, the stakeholder's experiences and behaviours were recorded.

A variety of analyses were carried out on the farmer behaviour data in order to extract their decision rules. Data from the scenario-based questionnaire was mainly analysed using statistical (quantitative) techniques, while both semi-quantitative and quantitative analysis were done to analyse the data from RPG. The empirical decision rules found in the two data collection exercises are used to calibrate the decision rules in the base model. Seven combinations of possible calibration are found from each data collection process.

Each calibrated model was then validated against the secondary data obtained from the farmers' cooperative. From the validation result, the usefulness of each data collection method in developing a realistic model of human behaviour can be compared. Finally, the experiences (from the researcher and participant point of view) during the data collection process are compared and discussed to highlight the advantages and disadvantages of each data collection method.

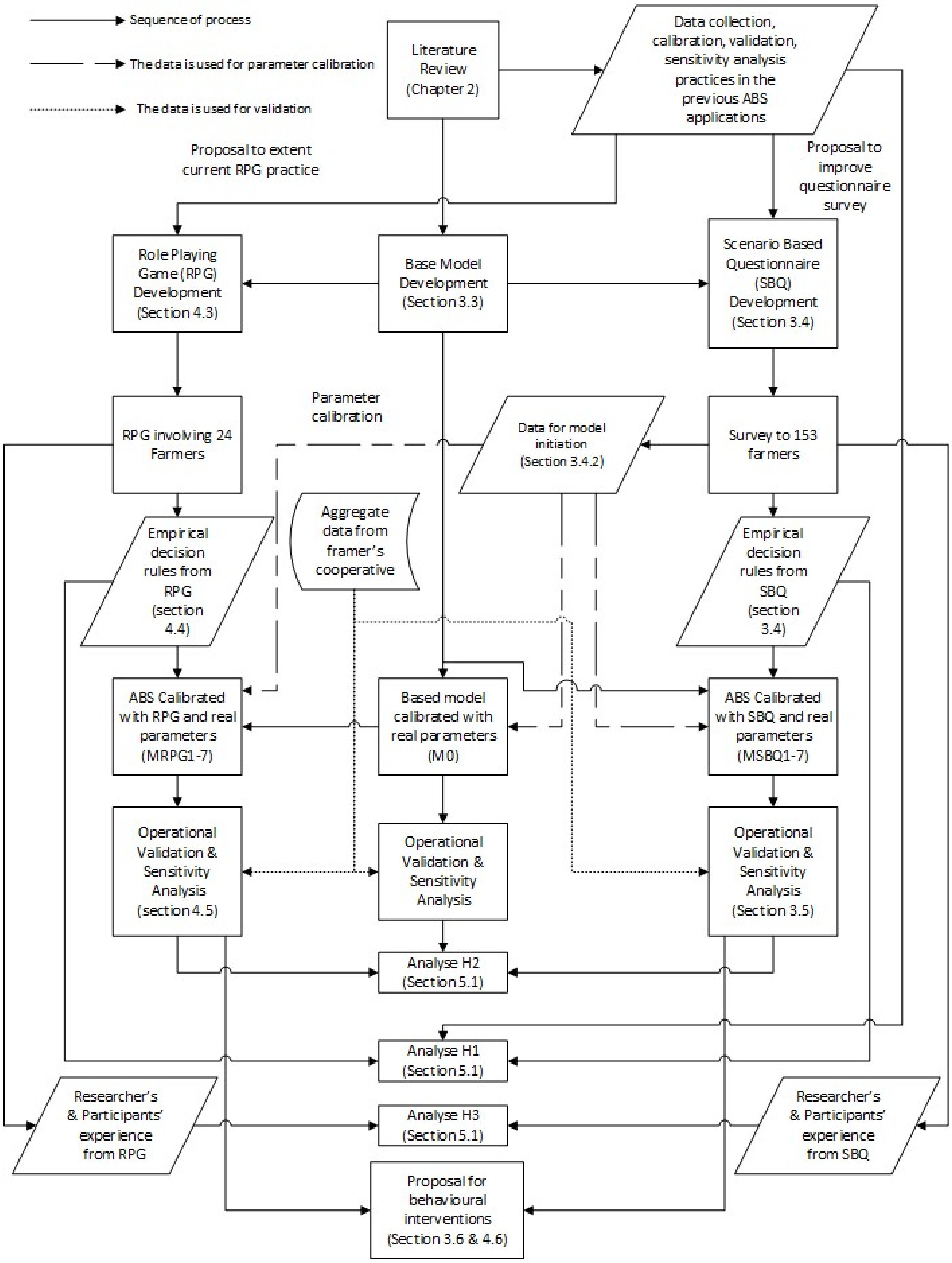

Figure 1. 1: Research methodology

1.4 Thesis Structure

Having discussed the motivation, the objectives and the research methodology, Chapter 2 of this thesis presents a literature review paper, regarding the ABS application in the agri-food supply chain. This discussion gives justification that the two data collection methods to be compared are relatively new and have rarely been used in the previous ABS applications in the agri-food supply chain. Therefore, their potential benefits, especially in developing models of the actor's behaviour, need to be investigated further. This chapter also highlights the uniqueness of the case study being used, compared to the previous ABS applications.

The specification of the base model that was developed to represent the phenomena occurring in the case study is then explained in Chapter 3. This section also discusses the process to develop a scenario-based questionnaire to identify farmers' decision rules and a questionnaire to collect empirical data for model initiation. Finally, findings from the data collection, calibration results, and lessons learned from the use of scenario-based questionnaire are discussed.

Chapter 4 focuses more on the RPG data collection and calibration. It starts by describing the way an RPG is commonly developed. This discussion shows that an RPG is mainly used to facilitate discussion and support the learning process for real-world actors, but not for developing a realistic representation of human decision rules. In order to use an RPG for this purpose, several modifications to the common practice are proposed. This section then discusses the findings, the calibration results and the lessons learned from the RPG data collection.

Chapter 5 summarises the lessons learned from the two data collection process. It starts by analysing the three research hypotheses and then discusses the potential contributions of this study.

Finally, Chapter 6 presents the conclusion of this study and highlights potential future research.

2 Applications of Agent-Based Modelling and Simulation in the Agri-Food Supply Chains

This chapter is taken from an invited review paper authored by Dhanan Sarwo Utomo, Dr Bhakti Stephan Onggo, and Dr Stephen Eldridge, from Lancaster University and Trinity Business School. This paper has been published in the European Journal of Operational Research, volume 269, page 794-805, 2018. Adjustments are made, and commentaries are added to the original manuscript to improve the coherence with other parts of this thesis.

Abstract

This paper provides a review of ABS applications in the agri-food supply chain. It begins by analysing the characteristics of the models and modelling reported in the literature. It illustrates that existing modelling research features extensive use of: single echelon supply chains; cases from high and middle income countries; unprocessed food products, empirical (as opposed to hypothetical) data; decision-making related to production planning and investment; and the use of black box validation. The second part of this paper uses bibliographic mapping to analyse areas in ASC research which are yet to be addressed using ABS. The findings from bibliographic mapping show that areas such as collaboration and competition, buyer-seller relationships, and service are under-researched. In addition, key actors in ASC such as food processors, supermarkets and retailers have not been included in the ABS models reported. Furthermore, important supply chain management theories, such as Transaction Cost Economics and Resource-Based View, are not used in the existing models.

Keywords: Literature review, Agent-based modelling, Agri-food supply chain, bibliographic mapping

2.1 Introduction

Agri-food supply chains (ASC) comprise a network of heterogeneous actors working together in different processes and activities to deliver products and services to the market and satisfy customers' demands. Actors in ASC include various organisations from producers, distributors, processors and consumers (Ahumada and Villalobos, 2009, Higgins et al., 2010, Pla et al., 2014, Borodin et al., 2016). The actors in ASC do not usually form linearly integrated businesses (Kutcher and Norton, 1982, Higgins et al., 2010). They have a high degree of autonomy with objectives that may conflict with those of the other actors. Consequently, this limited perspective makes it difficult for them to envisage how their individual decisions may affect the performance of the whole supply chain (Higgins et al., 2010). Furthermore, the dynamics in ASC are often influenced by social factors (e.g. lifestyles, personal values, safety concerns) (Busby and Onggo, 2013, Busby et al., 2016, Chebolu-Subramanian and Gaukler, 2015), economic factors (e.g. price) and the environment (e.g. climate variability) (Borodin et al., 2016). Actors in ASC have to adapt to these external factors in order to survive. In the light of these characteristics, it is not surprising that some authors (e.g., Ahumada and Villalobos (2009)) argue that ASC are complex and hard to manage.

The complexities of ASC have attracted the interest of Operational Research and Management Science (OR/MS) researchers since the late 1940s (Borodin et al., 2016) and they have been the subject of a number of reviews. Ahumada and Villalobos (2009) reviewed the application of mathematical models in agricultural production and distribution planning while Janssen and van Ittersum (2007) reviewed the use of optimisation models (known as bio-economic farm model in the agriculture literature) to assess farm innovations and responses to policies. More recently, Soto-Silva et al. (2016) reviewed the applications of OR/MS methods in fresh fruit supply chain and Borodin et al. (2016) reviewed the methods to handling uncertainty in ASC. The OR/MS techniques in these reviews include Agent-Based Simulation (ABS) and the benefits of using ABS in ASC have been highlighted by a number of authors (e.g., Higgins et al. (2007), Nolan et al. (2009), Higgins et al. (2010), Krejci and Beamon (2012) and Pla et al. (2014)).

In common with other OR/MS techniques, ABS is being continually developed and enhanced. Our paper provides a review of the ABS methods used in ASC in order to identify topics in ASC that merit further research using ABS. Our review is complementary to earlier reviews of the application of ABS in related agriculture fields. These include the environment (Kelly et al., 2013), climate adaptation (Berger and Troost, 2014) and land use (Robinson et al., 2007, Matthews et al., 2007). Furthermore, we demonstrate how bibliographic mapping can supplement a conventional literature review to identify research opportunities for the application OR/MS techniques.

Initially, the literature search methodology is described in Section 2.2. We then present an overview of the application of ABS in ASC based on our literature review and discuss the models and modelling approaches reported. In particular, the modelling objectives, application context, models (inputs, outputs, actors, rules and interactions), output analysis, experimentation, validation and model representation are discussed (in Section 2.3). Subsequently, we present a bibliographic mapping analysis and discuss the ASC topics that are yet to be addressed by ABS researchers (in Section 2.4). Finally, the conclusions of this literature review are presented in section 2.5.

2.2 The Literature Search Methodology

The literature search employed the following databases: ABI/INFORM, Academic Search Complete, Business Source Complete, Science Direct and Web of Science. The search is restricted to articles published in international peer-reviewed journals that were written in English and published before February 2016. The keywords used in literature search and the results returned from the search are presented in Table 2.1. The keyword search was applied to the content of the articles (i.e., not limited to title and abstract only).

The approach taken for the literature review is illustrated in Figure 2.1 using a PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) four-phase flow-diagram (Moher et al., 2009, Vrabel, 2015). It is a format to report literature collection and analysis process that is starting to become a standard in medicine and healthcare. It begins with Dataset D from Table 2.1. Dataset D contains articles relating to ASC. Duplicate articles along with editorials, news, announcements, proceedings and dissertations were removed to create dataset D1 comprising 16,538 articles. Dataset D1 would be used for the bibliographic mapping analysis and did not require further filtering.

Table 2. 1 Keywords used in database searching and the number of returned articles

Code	Keywords	ABI / INFORM	Academic Search Complete	Business Source Complete	Science Direct	Web of Science	Total
A	("agent based" OR "multi agent") AND (“simulation” OR (“modelling” OR “modelling”))	6,360	5,029	2,014	27,148	11,885	41,736
B	"supply chain" OR "supply chains"	44,606	9,774	23,230	43,972	23,100	144,682
C	“agriculture” OR “agricultural” OR “food” OR “agri-food” OR “livestock” OR “fisheries”	276,808	1,139,158	116,263	1,877,194	716,254	4,125,677
D	B AND C	13,608	2,000	1,849	16,035	2,444	35,936

From dataset D1, articles on ABS applications in ASC were filtered (*i.e.*, using keyword “A and D”). The number of articles retained was 251. These articles were then screened individually to ensure relevance using the following the following criteria. Firstly, the article must be accessible to the wider academic community. Secondly, the article must feature a complete ABS model rather than simply an unimplemented conceptual ABS model. Thirdly, we excluded literature review papers. Fourthly, we excluded articles that focus only on nonhuman actors and articles in which the keywords only appear in the reference section. Finally, the article must address research questions related to supply chain topics (e.g., processes and production systems, inventory management, demand management and improving the performance in the supply chain (Oliveira et al., 2016) and include one or more ASC actors (e.g., producers, harvesting & transport, food processor & storage, packaging & handling, distributors, retailers, consumers, and waste management). Similar to Cunningham (2001), Da Silva and de Souza Filho (2007), Webber and Labaste (2009) and Higgins et al. (2010), we include articles that discuss livestock, crops, fisheries, and food products in our agri-food supply chain review. Using these screening criteria, 15 articles were retained. Next, backwards and forward citation analysis of these articles was conducted using Google Scholar and Web of Science. After applying the same screening criteria, the number of articles increased to 58. These comprise dataset D2 that was used for our review.

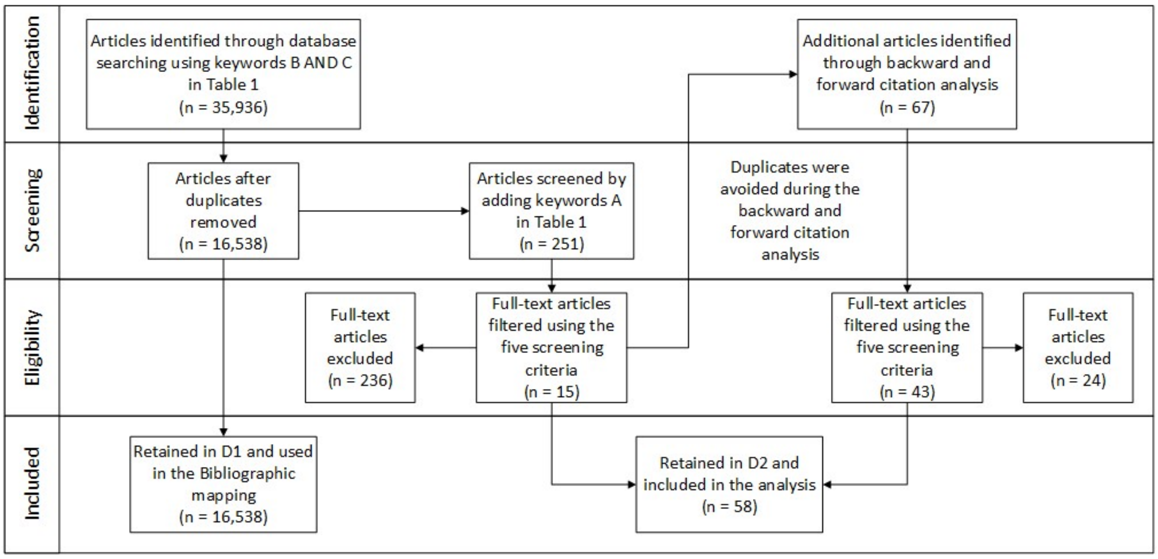

Figure 2. 1: PRISMA flow diagram of publication data collection process

2.3 Agent-based simulation applications in agri-food supply chain

This section provides a summary of research into the application of ABS in ASC based on dataset D2. As shown in Figure 2.2, the number of articles reporting on the application of ABS in ASC has been increasing, especially during the last four years (2013-2016). These articles are published in a variety of journals in the fields of environmental science, agriculture, computer science and operational research (see Table 2.2).

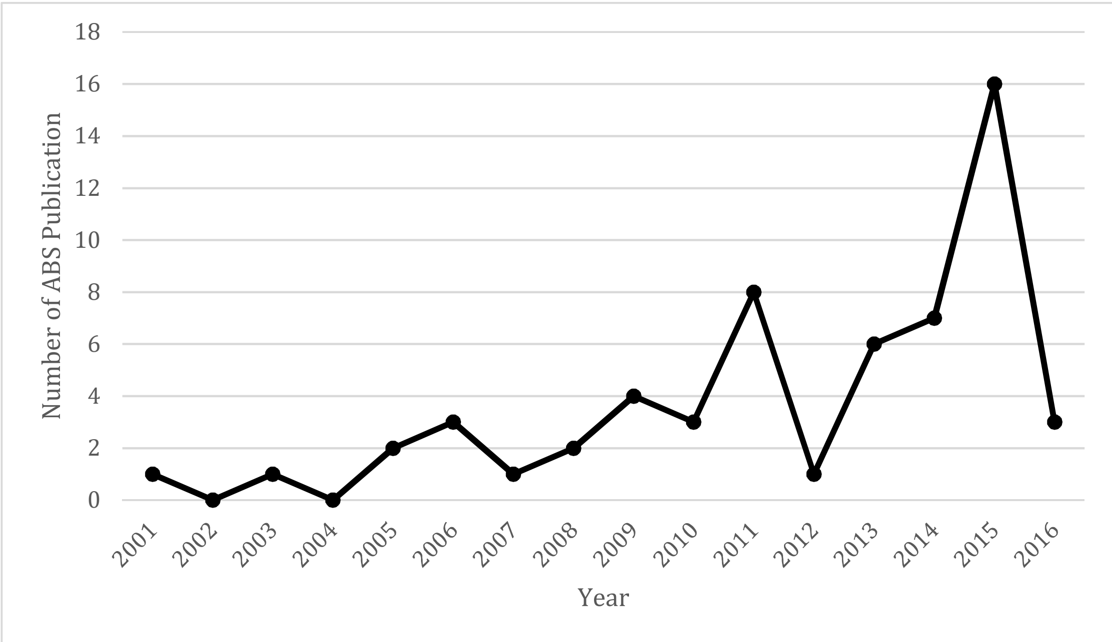

Figure 2. 2: Publication of ABS application per year (2016 contains two-month worth of data)

The most active discussion on this topic takes place in environmental science journals (e.g., Ecological Economics, Ecological Modelling, Ecology and Society, Environmental Modelling & Software and Environmental Science & Policy) and agriculture journals (e.g., Agricultural Economics, Agricultural Systems, American Journal of Agricultural Economics, and Canadian Journal of Agricultural Economics). Although ASC is an important and interesting application domain for ABS, the number of publications in OR/MS specific journals is relatively low. One of the possible explanations is that many authors have focused on ASC specific research questions and ABS is used as a modelling tool to answer those questions. Hence, it makes more sense to discuss the findings in domain-specific journals such as environmental science and agriculture.

Table 2. 2 Distribution of journals that publish ABS applications in ASC

Journal Type	Number of papers (percentage)
Environmental science	23 (39.65%)
Agriculture	20 (34.48%)
Computer science	7 (12.06%)
Operational research	1 (1.72%)
Other	7 (12.06%)

2.3.1 Research context of the previous ABS applications in ASC

This section explores the context of the research reported in the literature. The studies were divided into real and hypothetical cases and Table 2.3 shows that most studies use real cases. The real cases were subdivided into categories based on the geographical location and the economic development level of the country as described by the World Bank (2016) (i.e., high income, middle income and low income). Most of the studies took place in Europe (35.7%) and Asia (25%) and in high income (57.7%) and middle income (36.5%) countries. This geographical and economic development categorization is important because empirical evidence suggests that ASC actors from different geographical regions or different economic development levels may behave differently. For example, differences in contract farming participation between people living in Ghana, India, Madagascar, Mozambique and Nicaragua are noted (Barrett et al., 2012). Furthermore, Meijer et al. (2006) observe the differences in preference when choosing transaction governance mechanisms (i.e., market, hierarchy and network) which could be explained by different cultural backgrounds. Similarly, Godfray et al. (2010) explain different practices in developed and developing countries associated with food waste production. This suggests that a need to extend the reach of ABS research in ASC, particularly in low income countries.

Table 2. 3 Summary of the research context

Classification	Category	Number of papers
Types of case study	Real case	52
	Hypothetical case	6
Geographical Context	Africa	6
	Asia	14
	Australia	3
	Europe	20
	North America	7
	South America	6
Economic Development	High-income countries	30
	Middle-Income Countries	19
	Low-Income Countries	3
Number of food and agricultural products modelled	Single products	17
	Multiple products	35
Type of food and agricultural products	Unprocessed product	50
	Processed product	2

Alternative classifications of the real cases were also adopted relating to the number of different food and agricultural products that were studied. This classification is important because the number of products affects the modelling techniques used. This will be discussed further in section 2.3.4. As shown in Table 2.3, the majority of the existing studies incorporate various food and agricultural products. Table 2.3 also shows that the studies predominantly feature fresh or unprocessed food products.

2.3.2 Objectives of the previous ABS applications in ASC

This section discusses the objectives of research in ASC that uses ABS as the main modelling method. The research objectives can be categorized into theory development (i.e., to explain ASC phenomena or to test a theory in the context of ASC), methodology development (i.e., to improve existing methods or to propose new methods for ABS in the context of ASC) or policy development (i.e., to predict or to analyse the impact of a management/policy decision). As shown in Table 2.4, all three types of objective feature in dataset D2 and it should be noted that a paper could contribute in more than one category. Interestingly, between 2001 and 2005, most ABS applications focused on theoretical contribution while in more recent years, between 2011 and early 2016, ABS applications for policy development (38.3%) and methodology development (36.2%) studies become increasingly popular. This trend occurs because the benefits of ABS in supporting decision making have started to be acknowledged in the 2010s.

Table 2. 4 Classification based on the type of research objectives

Research Objectives	2001-2005	2006-2010	2011-early 2016	Total
Theory development	4	6	12	22
Methodology development	0	1	17	18
Policy development	1	5	18	24

One typical research objective in the theory development category is to build a model to explain the behaviour of agents and its consequence (e.g., Becu et al. (2003), Bharwani et al. (2005), Tykhonov et al. (2008), Ross and Westgren (2009), Zhang and Brorsen (2010), Graubner et al. (2011), Udumyan et al. (2014), Krejci et al. (2016), Malawska and Topping (2016)). Another popular research objective is to build a model that explains the impact of decisions by individual agents on the dynamics of a supply chain's structure. These include: the emergence of diversity; the dynamics of cooperation networks; the formation of clusters; and change in market power (e.g., Castella et al. (2005a), Ng (2008), Følgesvold and Prenkert (2009), Bert et al. (2011), Boero (2011), Ross (2011), Boyer and Brorsen (2013), Bakker et al. (2015), Krejci and Beamon (2015), Albino et al. (2016)). There is also a significant number of studies related to how innovation is adopted and spread among ASC actors (e.g., Berger (2001), Kaufmann et al. (2009), Olabisi et al. (2015)).

The methodological papers propose either new or improved methods in ABS using ASC case studies with a view to possible extensions beyond ASC. Methodological developments are proposed in a range of topics including: agents' decision-making rules (e.g., Schreinemachers and Berger (2011) and Morgan et al. (2015)); simulation parameterization (e.g., Berger and Schreinemachers (2006), Schreinemachers et al. (2009), Nainggolan et al. (2012), Troost and Berger (2015), Zimmermann et al. (2015)); sensitivity analysis (e.g., Schouten et al. (2014), Brändle et al. (2015)); model validation (e.g., Smajgl et al. (2011b), Bert et al. (2014), Kaye-Blake et al. (2014), Ge et al. (2015a)); and hybrid modelling approaches (e.g., Happe et al. (2011), Aurbacher et al. (2013), Marohn et al. (2013), Reidsma et al. (2015)).

In the policy development category, most of the studies focus on finance or the use of new technology and innovation. Financial policy is the most popular including aspects such as: credit (e.g., Berger et al. (2006), Schreinemachers et al. (2007), Wossen and Berger (2015), Schreinemachers et al. (2009)); payment schemes (e.g., Happe et al. (2006), Uthes et al. (2011), Schouten et al. (2013), Brändle et al. (2015)); incentives and subsidies (e.g., Smajgl et al. (2011b), Quang et al. (2014), Zheng et al. (2015)); pricing (e.g., Morgan and Daigneault (2015)); and compensation schemes (e.g., Troost and Berger (2015)). Policies related to technological and innovation policies include the use of: fertilizers (Berger et al., 2006, Schreinemachers et al., 2007); improved seed

(Schreinemachers et al., 2007); tree crop innovations (Schreinemachers et al., 2010); organic agriculture (Gagliardi et al., 2014); and technology standards (Zheng et al., 2015). There are a number of smaller groups of research related to policy making such as how supply chain actors should cooperate to adapt to climate change (Wang et al., 2013) or minimise the risk of bioterrorist attack (Chaturvedi et al., 2014). Other studies explore: suitable inspection policies to improve product quality (Ge et al., 2015b, Ge et al., 2015a) or food safety (McPhee-Knowles, 2015); appropriate harvesting management plans (Worrapimphong et al., 2010); and policies for manure handling (Zheng et al., 2013) or animal welfare (Osinga et al., 2015).

Research in the policy category can be further divided into categories based on the scope of the policies. Julka et al. (2002) propose that the scope of supply chain policies can be classified into: intra-enterprise policies that cover departments within an organisation and their interface with other organisations; inter-enterprise policies that cover an organisation and its supply chain; and cluster policies that cover all industries in a sector including their suppliers and customers and government. Table 2.5 shows that the majority of the studies have focused on the cluster category (please note that a paper may analyse more than one policy scope).

Table 2. 5 Classification based on policy scope

Scope of policy	Number of Papers
Intra-enterprise	5
Inter-enterprise	5
Cluster	16

2.3.3 The use of data in the previous ABS applications in ASC

This section discusses the input data (for ABS development) and output data (collected during the simulation experiment) used by the models featuring in the studies. Table 2.6 shows that most of the input data are empirical. Related to Table 2.3, there are four studies that use a real case study but use hypothetical data. This is because these studies incorporate hard to measure variables, even though they aim to replicate real world supply chains. The most popular empirical data sources are from secondary sources such as Farm Accountancy Data Network (FADN) and agricultural census data (e.g., Happe et al. (2006), Happe et al. (2011), Zimmermann et al. (2015)). These types of publicly available data are usually presented in aggregate owing to confidentiality concerns. Aggregated data poses a challenge to ABS modelling when generating a representative population though approaches such as that proposed by Troost and Berger (2015) can mitigate this limitation. Hypothetical data are used when researchers build a stylised model to test theories (e.g. Krejci and Beamon (2015)) or when empirical data are difficult to collect (e.g. food contamination McPhee-Knowles (2015)).

Table 2. 6 Input and output data in the previous ABS applications

Classification	**Category**	**Number of papers**
Data type	Hypothetical data	10
	Empirical data	48
Input data source	Secondary data	38
	Primary: Survey	18
	Primary: Interview	11
	Primary: Participatory Modelling	6
Output data	Production measures	24
	Financial measures	22
	Environmental measures	12
	Trust & relationship among agents	5
	Quality & safety	4

ABS can be used to produce and analyse various output data relevant to ASC as shown in Table 2.6. Output data related to production, finance and environment described in the papers can be easily measured and this reflects the prevalence of their use in ABS studies. However, output data related to trust, relationship, quality and safety are more difficult to measure objectively and, similarly, these measures are considered to be more difficult to model (Tykhonov et al., 2008). Nevertheless, ABS is being used to model both types of measures.

The most popular output data related to aspects of production such as yield and produced quantity (e.g., Bharwani et al. (2005), Happe et al. (2006), Berger and Schreinemachers (2006), Zheng et al. (2013), Zimmermann et al. (2015)) and finance such as income and wealth (e.g., Berger (2001), Becu et al. (2003), Bharwani et al. (2005), Berger et al. (2006), Schreinemachers et al. (2007), Marohn et al. (2013)). Examples of models that produce environmental metrics include Schreinemachers et al. (2009), Uthes et al. (2011) and Quang et al. (2014). Examples of models that use difficult-to-measure output data include: trust and honesty, measured using a probability (Tykhonov et al., 2008); the stability of symbiotic relationship, measured by relationship duration (Albino et al., 2016); cooperation, measured by how many times agents decide to work together with others (Krejci and Beamon, 2015, Boero, 2011); and inspection quality, measured by the probability of product misclassification (Ge et al., 2015b, Ge et al., 2015a).

2.3.4 Agents, their decision-making rules and their interactions

This section discusses the key model design features (i.e., agents and their rules for decision-making and interactions) used in the previous studies. Actors in ASC include producers, post-harvest processors, retailers, consumers and others (e.g., Higgins et al. (2010), Pla et al. (2014), Borodin et al. (2016)). Table 2.7 shows that the producer (i.e., farmer) is included in most ABS models because it is considered the most important actor, especially in agriculture and environmental science journals. This can also explain why the existing reported studies focus on unprocessed agricultural products (see Table 2.3).

Table 2.7 shows that the scope of most ABS models is one echelon. This is likely to reflect a need to keep the models simple and, in agriculture studies, the focus of the analysis is often on the producer. However, our sample shows that the number of ABS models that incorporate multiple echelons is increasing (2001-2005: one paper, 2006-2010: three papers, 2011 onward: nine papers). This is a welcome trend as modelling multi-echelons should provide more insights for supply chain research (e.g., van der Vorst et al. (2000)).

Table 2. 7 Summary of model details

Classification	Category	Number of papers
Agents in the model	Producer	55
	Post-harvest	5
	Processor	5
	Retailer	6
	Consumer	5
	Other	3
Number of ASC echelons	1	44
	2	9
	3	2
	4	2
	5	1
Decision-making rule	Rule-based	46
	Equation-based	22
Type of interactions	Narrowcast	35
	Broadcast	20
Type of agent decisions	Production planning	34
	Investment	20
	Technology choice and adoption	11
	Cooperation	10
	Product tracing or quality	5
	Selling	5
	Product delivery	4
	Other	9

How decision-making rules are represented in an ABS model can be divided into two categories: equation-based and rule-based (both categories may be used in one model).

The equation-based representations use mathematical equations such as: linear programming (e.g., Becu et al. (2003)); mixed integer programming (e.g., Berger (2001), Happe et al. (2006) and Schreinemachers and Berger (2011)); and regression modelling (e.g., Bakker et al. (2015)). The rule-based representation uses declarative languages such as "if then else" rules (e.g., Morgan and Daigneault (2015)), threshold models (e.g., Quang et al. (2014)) and imitation models (e.g., Osinga et al. (2015)).

The interactions among agents in an ABS model can be divided into narrowcast (i.e., an agent only interacts with its neighbours) or broadcast (*i.e.*, an agent interacts with all or most of other agents) (Onggo et al., 2014). An example of narrowcast interaction is described in Zheng et al. (2015) in which agents gain knowledge regarding innovations from their neighbours. An example of broadcast interaction is presented in Quang et al. (2014) in which agents monitor the adoption rate in the population and compare it to their willingness to take risk when deciding to adopt an innovation.

In terms of types of decision, we proposed the following categories:

- Production planning: determining the type and quantity of commodities to be produced, land allocation and resource allocation. All these decisions are usually modelled together (e.g., Berger (2001), Happe et al. (2006) and Krejci and Beamon (2015)).

- Investment: deciding to buy or sell land, adding or selling machinery (e.g., Schreinemachers et al. (2009), Schouten et al. (2014)).

- Technology choice and adoption: deciding when and how to share knowledge with other agents and adopt a new innovation (e.g., Olabisi et al. (2015), Berger et al. (2006)).

- Cooperation: deciding when and how to cooperate with other agents (e.g., Krejci and Beamon (2015)) and selecting partners with whom to cooperate (e.g., Gagliardi et al. (2014)).

- Product tracing or quality: deciding how to control product quality and how to trace a product's source (e.g., Ge et al. (2015b), Ge et al. (2015a)).

- Selling: deciding where the agent will sell its products (e.g., Krejci et al. (2016)).

- Product delivery: deciding, for example, whether the agent will send products according to the specifications agreed with the buyer or not (e.g., Tykhonov et al. (2008)).

- Others: decisions including managing irrigation in Becu et al. (2003) and pricing in Graubner et al. (2011).

Table 2.7 shows the dominance of production planning decisions and reflects the focus on just the producer as the only supply chain echelon (31 of 34 studies). Even so, understanding how farmers determine their production strategy and its consequences is important and interesting for researchers.

2.3.5 Validation and Sensitivity Analysis

Macal (2016) observes that what ABS gains in its ability to model complexity is offset by losses in its analytical tractability which includes issues relating to experiment design, output analysis and validation using empirical data. Our literature review illustrates that this observation is particularly true for ABS applications in ASC. This section discusses how ABS applications in ASC deal with validation and experimentation issues. In this discussion, validation techniques are classified into theoretical (i.e., comparing model behaviours with theory), and empirical validation (i.e., comparing model behaviours with observation or expert judgement).

It is also possible to classify the validation techniques of simulation models, including ABS models, into black box and white box validation (Kleijnen, 1995a, Montanola-Sales et al., 2011). Black box validation evaluates whether the model outputs either reflect the empirical observations for the same set of inputs (e.g., Malawska and Topping (2016), Berger (2001)) or are consistent with the result from a mathematical model (e.g., Onggo and Karatas (2016), Ge et al. (2015a)). White box validation evaluates whether the decision rules of agents represent the decision rules of actors in the real world and whether the structure of the model (such as the network between agents) represents reality. This includes techniques such as examining the validity of the model structure, i.e. static logic and the dynamic logic of the model components and behaviours (Pidd, 2004, Montanola-Sales et al., 2011) and interactive modelling sessions (Berger and Troost, 2014, Arnold et al., 2015).

Table 2.8 shows that the validation process used by 27 studies (i.e. 47%) is unclear or unspecified. However, it is encouraging that the proportion of papers in this category may be declining (57% between 2001 and 2006, 42% between 2007 and 2012, and 47% from 2013 onward). This finding is consistent with the earlier observations by Heath et al. (2009) that the number of ABS papers without validation continues to decrease every year. In our sample, 67% of studies using hypothetical cases and 40% of studies using real cases are not validated. In addition, many theory development studies are not accompanied by a validation process (52%). However, fewer policy development and methodology development studies are without validation details (37% and 22%, respectively).

Table 2. 8 Paper classification based on the validation technique

Validation information	Classification	Category	Number of papers
Absent or unclear	N/A	N/A	27
Present	Based on data	Theoretical	7
		Empirical	25
	Process	Black box	30
		White box	2

Empirical validation is the most popular means to validate ABS models in ASC. This method includes: visual comparison between the trends produced by simulation and the actual trends (e.g., Brändle et al. (2015)); statistical comparison (e.g., Malawska and Topping (2016)); and fitting a regression line between the simulated and actual data (e.g., Berger (2001), Schreinemachers et al. (2007), Schreinemachers et al. (2010) and Marohn et al. (2013)). Alternatively, theoretical validation is adopted frequently when validating difficult-to-measure qualitative behaviours (e.g., Tykhonov et al. (2008)). Reasons for this include the lack of widely available historical qualitative data or this data may not be in a form that can be readily used for simulation. Furthermore, standard statistical techniques may not be suitable for validating qualitative behaviour. Only two papers in our dataset employs white-box validation (i.e., Bert et al. (2014) and Arnold et al. (2015)) which is a concern, especially considering that black-box and white-box validations are complementary activities (Montanola-Sales et al., 2011).

To interpret the result of a simulation study, it is important to describe the statistical features of its outputs (Hamill, 2010). Furthermore, the shape of a simulation output distribution is usually a priori unknown and an appropriate number of replications is needed to produce meaningful statistics (Lee et al., 2015). Table 2.9 shows that 23 of 58 papers do not explicitly mention the number of replications used in the experiment or report the confidence interval of their simulation. The table also shows that a subjective method using researcher judgement is by far the most common with a range of 10 to

350 replications. The only exception is the study by Osinga et al. (2015) which uses an objective method based on the coefficient of variation (Lorscheid et al., 2012).

Sensitivity analysis can be used to understand the risk from making a decision based on a model because many input parameters are estimated and, for ABS models, the parameters are often estimated subjectively. Sensitivity analysis can also be used as a form of model validation by checking that the model reacts correctly to changes in input parameters (Sargent, 2013). In the context of ABS, sensitivity analysis can also be used to gain insight into the patterns and emergent properties of the model (ten Broeke et al., 2016). Sensitivity analysis techniques can be broadly categorised into one factor at a time (OFAT) and global sensitivity analysis (GSA). OFAT sensitivity analysis requires us to select a set of parameter values (baseline) and then vary one parameter at a time while keeping all other parameters fixed. Hence, OFAT does not take into account the possible interaction effects between parameters. On the other hand, GSA includes the interaction effects by sampling a model's outputs over a wide range of parameter values and then fitting a regression function or calculating sensitivity indices for these outputs (Sobol′, 2001, ten Broeke et al., 2016). Table 2.9 shows that most of the studies use OFAT. It should be noted that the proportion of papers that do not apply sensitivity analysis is relatively high (28%). However, our sample suggests that the proportion of studies reported without sensitivity analysis is decreasing (75% between 2001 and 2005, 23% between 2006 and 2010, 24% from 2011 onward). Those without sensitivity analysis are mostly theoretical papers (63%). Although ABS models used for theory development are not directly used for policy decision making, sensitivity analysis is still important to either ensure that the proposed theories are robust or find the parameter boundaries for which the proposed theories are valid.

Table 2. 9 Summary of output analysis

Output analysis	Category	Number of papers
Number of replications	Determined with subjective method	28
	Determined with objective method	1
	N/A but confidence intervals were given	6
	N/A, confidence intervals not given	23
Sensitivity analysis	OFAT	38
	GSA	4
	N/A	16

2.3.6 Model representation methods

A good model representation is important for the communication between stakeholders which affects the credibility of a model (Onggo, 2013). It is also important to ensure that the model can be duplicated and developed further by other researchers (Collins et al., 2015). Table 2.10 shows that most papers do not use any structured model representation techniques (i.e. they describe the model in unstructured text). Simple flowcharts, Overview Design and Details protocol (ODD) (Grimm et al., 2010) and Unified Modelling Language (UML) are the most popular in those papers that use a structured representation technique. In our sample, flowchart representation has been used for a long time and ODD has started to gain popularity since 2011. ODD representation is particularly popular for papers published in environmental science journals.

Table 2. 10 Classification based on model representation techniques

Scope of policy	Number of Paper (percentage)
Flowchart	13 (22%)
ODD	11 (19%)
UML	1 (2%)
N/A	33 (57%)

2.4 Discussion

In this section, we discuss the main findings from our review with findings from similar reviews and identify research areas in the ASC that have not taken the advantage of

ABS even though ABS has been shown to be useful in those areas in other application domains.

2.4.1 The state of research in ASC that uses ABS

There have been a number of reviews on the application of ABS in related application domains such as land use (Robinson et al., 2007), environmental science (Kelly et al., 2013) and forest product supply chains (Vahid et al., 2016). Table 2.11 summarises and compares these related reviews. We include Oliveira et al. (2016) because it provides the latest review on supply chain simulation (which includes 34 ABS papers). Consistent with our observations, Oliveira et al. (2016) and Vahid et al. (2016) both note the increase in the numbers of papers that use ABS which demonstrates that that ABS has been accepted as one of the analytical tools in these domains.

Table 2.11 shows that both hypothetical and real case studies are used in the literature. The number of cases reported from low income countries is low. The objectives of the ABS models reported in these earlier reviews are for theory and policy development. This outcome contrasts with our observation that there have been a significant number of papers that seek to improve ABS modelling methodology. In terms of data, most papers that use ABS for policy development in these earlier reviews use empirical data and those for theory development (e.g., most papers in Oliveira et al. (2016)) use hypothetical data. This is consistent with the finding from our review. It should be noted that in their review, Robinson et al. (2007) indicate that there are other empirical data collection techniques that could be used for ABS. For example, discrete choice experiment has been used in the forestry domain (e.g., Holm et al. (2016)). Another method is the use of a social experiment which can help researchers understand how humans behave and has a strong grounding in economic theories (e.g., Barreteau et al. (2001)). Qualitative data collection methods such as the monographic case study have also been used in natural resource management domain (Castella et al., 2005b). Qualitative data collection methods are very important since they may provide a deeper understanding of how actors make decision for a given context (Robinson et al., 2007). Our findings also confirm those of previous literature reviews that illustrate that most models use easy-to-measure output data.

Table 2. 11 Summary of discussion from other literature review

Category	Robinson et al. (2007)	Kelly et al. (2013)	Vahid et al. (2016)	Oliveira et al. (2016)
Review Purpose	To compare strengths and weaknesses of five data collection methods for ABS	To compare five approaches to model complex trade-offs in land use	To identify challenges in forest products supply chain research	To identify developments and advancements in supply chain simulations
Review Domain	Land use science	Environmental science	Forest products supply chain	Supply chain
Review Scope	Five ABS papers (2001 – 2006)	Various methods including 11 ABS papers published between 2006 and 2012	Various methods including five ABS papers published between 2007 and 2012	Simulation papers including 34 ABS papers published between 1992 and 2014
Research Context	Five real cases from one low-income, three middle-income and one high-income countries	Four hypothetical models; one low-income, two middle-income, and four high-income countries	Real cases from high income countries	From all papers (including ABS), 57% use hypothetical case and 43% use real case
ABS Research Objective	Mainly for policy development	Theory & policy development	Not discussed	Mainly theory development
Data collection method	Qualitative & quantitative data collection methods	Not discussed	Not discussed	42% use empirical data (the method was not discussed)
Output data analysis	Not discussed	Mainly easy-to-measure	Easy-to-measure	Not discussed
Model design	Not discussed	Not discussed	Not discussed	Not discussed
Validation & Sensitivity Analysis	Not discussed	Not discussed	Not discussed	Not discussed
Model Representation	Not discussed	Not discussed	Mainly ODD	Not discussed

Our review differs from the previous literature reviews by considering the model designs used in ABS studies (e.g., number of echelons, type of agents, agent's decision rules, and types of interaction). We find that most ABS applications in ASC focus on one echelon (i.e., the producer) and the simulation of production planning and investment decisions. The agents in the models mostly incorporate rule-based decision-making and narrowcast interactions. Furthermore, our review is exceptional in that it considers how the experimentation and model validation have been conducted and demonstrates that there has been an increase in the number of studies that carry out validation and sensitivity analysis. Finally, in terms of model representation, the most commonly used methods for model representation are the flow chart and ODD but our findings also show that the majority of articles do not use any method for model representation. Overall, when compared with these earlier reviews, our review provides a more detailed analysis of the characteristics of ABS model design (number of echelons, type of agent, agent's decision, model representation and interaction), validation and sensitivity analysis. In other words, our review is done from the perspective of an ABS modeller.

2.4.2 The gap between ASC and ABS research topic

This section highlights those topics within ASC in which ABS has not yet been used, even though ABS may have been used to address similar topics in other supply chain domains. To achieve this, VOSviewer software (van Eck and Waltman, 2009) was used to create a co-occurrence network of the terms obtained from the titles, abstract and keywords in dataset D1. Two terms are said to co-occur if they both occur on the same line. Terms with similar meaning were grouped together using the VOSviewer thesaurus (e.g., "agent-based" and "ABM"). VOSviewer places the terms in the network in such a way that the distance between two terms indicates the number of cooccurrences of those terms. Based on this network, VOSviewer identifies a number of clusters. Figure 2.3 shows the co-occurrence network of the terms used in ASC literature and the clusters identified (each colour represents one cluster and we have also added the circles to make the cluster more visible). There are six clusters in Figure 2.3. There are many intersections between these clusters. This indicates that many articles discuss multiple aspects of ASC. This also indicates that many articles were published in multi-disciplinary journals.

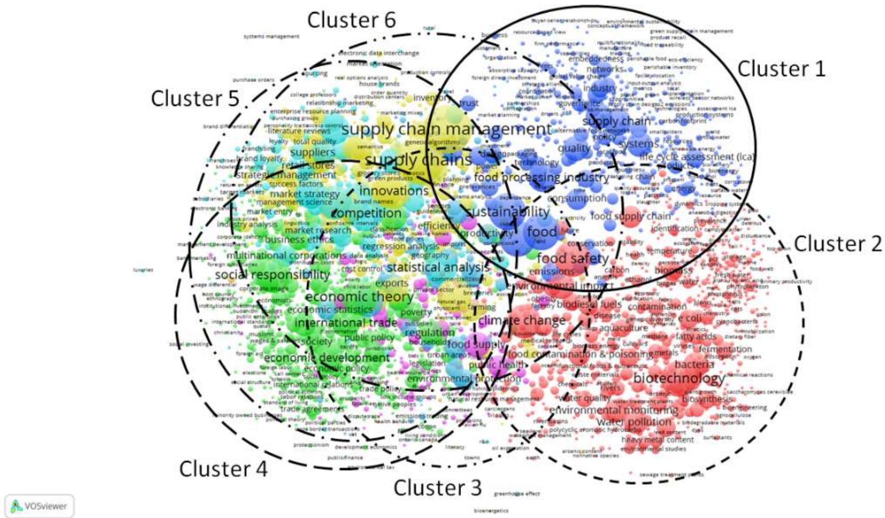

Figure 2. 3: Bibliographic mapping of ASC research

To understand what these six research clusters represent, the most popular keywords in each cluster are identified and listed in Table 2.12. Based on these keywords, these clusters represent ASC-related research from: logistics, supply chain and management science (cluster 1); natural sciences e.g. biotechnology, microbiology and environmental science (cluster 2); humanitarian aid and public health (cluster 3); political economics (cluster 4); marketing (cluster 5); and general management (cluster 6).

Table 2. 12 Top 10 popular keywords in each cluster of ASC research

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
supply chain management	biotechnology	food supply	economic theory	competition	sustainability
supply chains	food safety	public health	globalization	marketing	Food
innovations	climate change	certification	cooperation	services	agriculture
logistics	environmental impact	risk management	international trade	SME	food processor
sustainable development	bacteria	pharmaceutical industry	social responsibility	competitive advantage	management
decision making	environmental monitoring	disasters	regulation	consumer behaviour	value chain
operations research	food supply chain	humanitarian aid	impact analysis	CSR	collaboration
buyer-seller	e-coli	obesity	business ethics	retailing	supermarkets
information systems	microbiology	emergency preparedness	economic development	market strategy	transaction cost economics
management science	contamination & poisoning	nutrition research	politics	statistical analysis	resource-based view

Figure 2.4 shows the co-occurrence network of the terms used in the use of ABS in the ASC literature (i.e., dataset D2) and five research clusters were identified (each colour represents one cluster and we have also added the circles to make the cluster more visible). Cluster 1 represents papers focusing on understanding the agricultural system, including land use and crop production. Cluster 2 consists of papers aiming at modelling climate change adaptation, proposing mitigation policies and assessing their impact. Cluster 3 is the group of methodological papers including sensitivity analysis and parameter uncertainty handling. Research in cluster 4 focuses on modelling the complexity in food supply chains including interaction with the environment, social network and heterogeneity. Cluster 5 includes those studies modelling the diffusion of innovation.

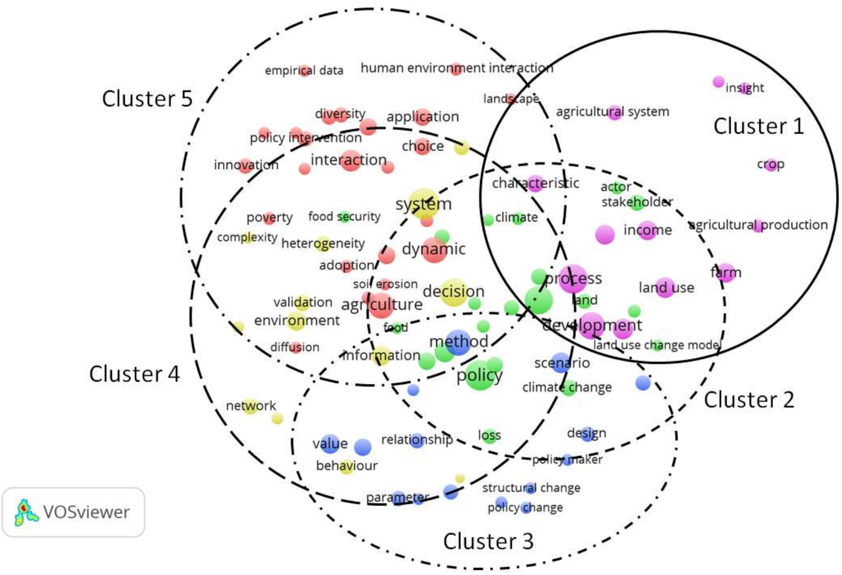

Figure 2. 4: Bibliographic mapping of ABS in ASC

Table 2. 13 Top 10 popular keywords in each cluster of ABS research

Cluster 1 (18 papers)	Cluster 2 (20 papers)	Cluster 3 (15 papers)	Cluster 4 (22 papers)	Cluster 5 (15 papers)
process	policy	method	system	dynamic
development	farmer decision making	scenario	decision	agriculture
land use	market	value	information	interaction
structure	Price	uncertainty	environment	choice
Farm	Role	relationship	social network	resource
Income	stakeholder	design	heterogeneity	farm household
Production	assessment	management	complexity	application
Characteristic	climate change	parameter	comparison	innovation
agricultural system	adaptation	sensitivity analysis	behaviour	policy intervention
Crop	individual	policy change	food supply chain	diffusion

In this section, clusters and popular keywords in ASC research and in ABS research have been presented. By comparing the keywords presented in Table 2.12 and Table 2.13, we can identify areas in ASC research (dataset D1) that are under-represented or missing in ASC research that uses ABS (dataset D2). These gaps may arise because ABS is not the right tool to research into these areas. However, further consideration of these areas and the benefit from the application of ABS is warranted. Based upon this review, the following research opportunities are proposed:

- *Cooperation, competition and collaboration*: These keywords appear in clusters 4, 5 and 6 in Table 2.12 but are missing from Table 2.13. ABS has been used to study cooperation, competition and collaboration in other domains. For example, in innovative product supply chains, Arvitrida et al. (2015) used ABS to explain the effect of competition and collaboration on supply chain performance (supply chain's survival and profit). He et al. (2013) provide an example of an ABS application in a retail supply chain. Specifically, they study the optimal strategy to respond to competition in the retail industry and find that everyday-low-price strategy is the best. From data set D2, we can only find two papers studies the collaboration between ASC actors (Krejci and Beamon (2015) and Boero (2011)). One main advantage of ABS is its ability to model the interactions between actors in a social network. Hence, ABS should play more important role in the research into cooperation, competition and collaboration in ASC. For example, we could use ABS to study the effect of the collaboration strategy between farmers and supermarkets on supply chain survivals.

- *Buyer-seller relationship*: The keyword buyer-seller appears in cluster 1 in Table 2.12 but is missing from Table 2.13. This keyword is a result of thesaurus grouping, and includes sub-keywords such as supplier, supplier relationship and buyer-seller in ASC research which are important concepts (e.g., Emanuela (2012)). ABS has already been used to study buyer-seller relationships in ASC (i.e., Tykhonov et al. (2008)). This is the only example from dataset D2. ABS was used to simulate how different levels of trust and honesty affect the interactions between buyers and sellers in ASC. We know that ABS has been used to study buyer-seller relationship in the supply chain literature. For example, ABS has been used to study the general partner selection problem in a supply chain (e.g., Schieritz and Grobler (2003)). In another context, Franke et al. (2005) use ABS to demonstrate how buyer's trust and seller's reputation can lead to more stable supply chains and in some cases, monopoly to arise. Hence, we believe that ABS has potential for research into different aspects of buyer-seller relationships in ASC. For example, we can use ABS to study the impact of strength of buyer-seller relationship (e.g., level of loyalty) on the average market price and production volume of an agri-food product.

- *Service*: This keyword is a result of thesaurus grouping, and includes sub-keywords such as customer service, food service, service industries and service architecture. The keyword service appears in cluster 5 of Table 2.12 but is missing from Table 2.13. ABS is one of the methods that is suitable for research into services in supply chains (Lusch, 2011). For example Rouzafzoon and Helo (2016) use ABS to study the service distribution and location problem in a healthcare supply chain. We believe that ABS is also relevant to researching services in ASC. For example, we can use ABS to study the effect of locations of service providers (e.g. post-harvesting, handling, bottling and packaging) and the level of service provision on ASC performance.

In addition to the research areas highlighted above, we also found that food processors, supermarkets and retailers have not yet been widely considered in ASC research that uses ABS. This is illustrated by keywords related to food processor, supermarket and retail being used frequently in ASC research (Table 2.12) and indicates the importance of these actors in ASC. However, these keywords are missing from Table 2.13. This may be due to the commercial confidentiality of the data related to these ASC actors. Publicly available retail data sources include the IRI Marketing Data Set (Bronnenberg et al., 2008) and initiatives such as the Consumer Data Research Centre (CDRC, 2017) may be useful in providing to access individual data. As for the aggregated data, methods such as that proposed by Troost and Berger (2015) are needed to calibrate an ABS model. Another technique that can be used is to ask these agents only to disclose reasonable assumptions regarding confidential information and parameters. The researchers then can perform sensitivity analysis on these parameters and check whether the corresponding variances are still within acceptable boundaries (see, for example, Sonderegger-Wakolbinger and Stummer (2015) in luxury goods context).

We also observed that ASC research frequently uses Transaction Cost Economics and/or the Resource-Based View to provide its theoretical foundation. These keywords are also missing from Table 2.13 which indicates that they have not yet been used in the ABS models for ASC despite there being a number of examples that illustrate how these theories can be incorporated, in ABS in general (e.g., Klos and Nooteboom (2001), Bylund (2015)).

2.5 Conclusions

We have presented a literature review of research in Agri-food supply chain (ASC) that uses agent-based simulation (ABS) as the main modelling tool. Our findings demonstrate that the number of papers addressing ASC policies has increased which suggests that researchers have started to apply ABS to real world decision-making related to ASC. Similarly, there has been an increase in the number of papers addressing the methodological aspect of ABS for ASC research, which indicates that ABS has gained acceptance as a modelling tool in this application domain. The increase in the number of papers with model validation is another positive development.

ABS research in this area has been dominated by the following characteristics: single echelon supply chains; cases in high and middle income countries; unprocessed food products; use of empirical data (especially from secondary sources, surveys and interviews); decisions related to production planning and investment; and the use of black box validation (especially in combination with empirical data). In comparison to earlier reviews of the use of ABS in other related domains, this review encompasses more papers and, more importantly, it provides a comprehensive review of ABS model design and ABS modelling approaches in ASC research.

We have also demonstrated how the bibliographic mapping technique can be used to highlight potential research areas within ASC that have not yet taken advantage of ABS despite ABS being shown to be valuable in similar research areas in other application domains. The identified research areas are: *cooperation and competition*; *buyer-seller relationships*; and *service in supply* chains. We have highlighted that some important actors in ASC, such as food processors and supermarkets, are rarely modelled using ABS. Furthermore, general theoretical frameworks such as Transaction Cost Economics and the Resource-Based View could potentially be incorporated into the design of these models.

2.6 Reflection from the Excluded Literature

We acknowledge that there are several studies associated with ABS applications on agri-food supply chains which are not discussed in this chapter. These studies were excluded from this literature review solely because they are not indexed in the scientific databases used in this study. Including these articles in this literature review would provide more ABS applications using real case studies in high-income countries, for example Etienne (2003), Millington et al. (2008), and Bommel et al. (2014). This supports our findings in Table 2.3. In terms of geographical context, these studies mainly discussed case studies in Europe (for example Etienne (2003) and Millington et al. (2008)), Africa (for example D'Aquino et al. (2003)) and South America (for example Bommel et al. (2014)). Hence, the proportion of studies using Asian context is lower than what is reported in Table 2.3. In terms of types of agricultural product modelled, most of the excluded papers also focussed on unprocessed products (for example D'Aquino et al. (2003), Etienne (2003), Cockburn et al. (2013) and Bommel et al. (2014)).

In terms of research objectives, most of the excluded papers focussed on developing theory (for example Cockburn et al. (2013)) or developing policies (for example D'Aquino et al. (2003), Etienne (2003) and Bommel et al. (2014)). If these papers were considered in this literature review, then the proportion of ABS methodological research is lower than what is reported in Table 2.4. This finding supports the claim in this study that more methodological researches are needed to develop the ABS methodology further.

In terms of the types of data, most of the excluded papers utilised empirical data. Popular techniques to collect empirical data were survey (for example Courdier et al. (2002)) and participatory modelling (for example D'Aquino et al. (2003), Etienne (2003) and Bommel et al. (2014)). Consistent with the finding described in Table 2.6, most of the excluded papers focussed on analysing production measures. Nevertheless, there were papers that focus on analysing the degree of specialisation among agents (Cockburn et al., 2013) and quality measures (Courdier et al., 2002).

In terms of the type of agent’s decisions, most of the excluded papers focussed on modelling production planning and investment decision of one supply chain tier, namely the producer. These decisions mainly modelled using rule-based decision rule, except a study by Millington et al. (2008) which employed equation-based decision rule. These findings are consistent with Table 2.7

In terms of model validation, only half of the excluded papers considered model validation. Theoretical validation is the most common technique used in the excluded papers. In terms of sensitivity analysis, the excluded papers had limited discussion on details such as how many times the simulation was replicated and how the experimentation was conducted. In conclusion, incorporating the excluded papers had no significant impact on findings of this chapter.

3 Using Scenario-Based Questionnaire for Agent-Based Simulation Model Calibration and Validation: An Agri-food Supply Chain Example

This chapter is adapted from a paper authored by Dhanan Sarwo Utomo, Dr Bhakti Stephan Onggo, Dr Stephen Eldridge, Dr Andre Rivianda Daud and Safitri Tejaningsih. The first three authors contributed in the study design and major part of manuscript preparation. The last two authors enriched the original manuscript using their field experiences. The primary data collection and analysis were done by the first and the last two authors. Finally, the simulation model and experiments presented in this chapter were developed and analysed by the first author. The original manuscript has been submitted to the European Journal of Operational Research, in January 2018. The result of the first review process has been released and it is very likely that this paper will be accepted, subject to revisions. Adjustments are made and commentaries are added to the original manuscript to improve the coherence with other parts of this thesis.

Abstract

A scenario-based questionnaire is a survey method that aims to identify the respondents' decision rules using their responses to a series of scenarios. It is rarely used in ABS with most researchers preferring a survey with closed questions as the data collection method. This is particularly true for ABS studies in ASC. In our paper, we explore how to design and deploy the scenario-based questionnaire in ABS using the case of a dairy supply chain. Our findings suggest that the decision rules extracted using a scenario-based questionnaire can improve ABS validity. Furthermore, we demonstrate that the decision rules extracted using this approach highlight opportunities for behavioural interventions to improve system performance.

Keywords: OR in Agriculture, calibration, validation, agent-based simulation, agri-food, dairy

3.1 Introduction

Questionnaire surveys are a popular data collection method and their value in scientific research is well established. Researchers typically use the data collected to develop black box models such as statistical models to predict the output of a system based on a variety of controllable and uncontrollable inputs. ABS attempts to open this black box by allowing researchers to identify both the agents' behaviours and how the agents make decisions within the system. ABS does not necessarily guarantee an improvement in the model's predictive capabilities but it does provide the opportunity to generate more insights into how the system works from the perspective of the agents' behaviours. These insights can have a practical value for policy makers because they enable the evaluation of interventions aimed at modifying agent behaviours even if other system parameters are beyond the policymakers' direct control. Furthermore, this approach to behavioural modelling is also important in other research fields such as operations management (Bendoly et al., 2006, Bendoly et al., 2010). However, behavioural modelling is still a challenge in ABS methodology owing to the difficulties in acquiring the data necessary to develop better representations of agent behaviour (Macal, 2016).

Researchers commonly use a questionnaire survey method with closed questions as a means of data collection to develop their ABS. This approach enables the measurement of emergent behaviours (*i.e.*, system outputs), the value of controllable and uncontrollable parameters, the agents' attributes, and the agents' decision parameters (*i.e.*, information used by the agents when making their decisions). However, the use of this type of questionnaire is problematic in respect of the validation and calibration of the agents' behaviours and their decision rules (*i.e.*, how they process information when making a decision) which are also necessary within an ABS. This validation and calibration of agents' decision rules is especially important if researchers would need truer representations of human behaviour in the ABS (Macal, 2016).

Our study aims to affirm the benefits of a questionnaire survey in an ABS but, unlike the majority of existing studies that use simple closed questions, we focus on the use of a scenario-based questionnaire design to validate and calibrate agent's decision rules. By using the case of a dairy milk supply chain in Indonesia, we propose a sequence for the design of the questionnaire and then use the data acquired to validate and calibrate our agents' decision rules. We perform a series of experiments using these calibrated models in order to identify both the decision rule(s) that are most influential on the model output and the decision rules that are valuable in improving the validity of the

ABS. We then explore the potential impact of policy interventions that could influence behaviours in the real system. Using these findings, we seek to highlight the benefits of a scenario-based questionnaire for ABS calibration.

To achieve these objectives, we begin, in section 3.2, by discussing how the scenario-based questionnaire might be useful to calibrate decision rules in ABS. In section 3.3, we discuss the base model in this study and its validation process. In section 3.4, we describe the sequence followed in order to develop a scenario-based questionnaire from the base model, the survey process, the data analysis and the derivation of the empirical decision rules. We then describe, in section 3.5, the simulation experiment process adopted to test the effects of these empirical decision rules on the model outputs. In section 3.6, we discuss the insights obtained from the scenario-based questionnaire survey and simulation experiments. Example of the scenarios and statistical analysis tables can be found in the appendices.

3.2 Literature review

In this section, we discuss how the questionnaire survey method has been used in ABS modelling in areas relevant to the case study (*i.e.*, ASC) prior to reviewing the use of scenario-based questionnaires to identify human decision rules in a variety of management studies.

3.2.1 The use of the questionnaire survey method in ABS studies in ASC

A dairy supply chain is but one of a variety of ASCs. As mentioned in chapter 2, ASCs are complex and hard to manage because they comprise a network of heterogeneous actors such as producers, distributors, processors and consumers (Ahumada and Villalobos, 2009, Higgins et al., 2010, Pla et al., 2014, Borodin et al., 2016). The dairy supply chain, in particular, is an economically important part of agriculture that is influenced by several factors including internationalisation (Glover et al., 2014) and consumer perceptions towards food safety (Ge et al., 2015a). This degree of complexity leads to ABS being considered suitable as a research methodology for studying and supporting decision making in supply chains in general and especially in ASC (see, for example, Utomo et al. (2018), Taticchi et al. (2015)).

Macal (2016) suggests behavioural modelling (i.e., developing more realistic models of human behaviours) as one of the challenges facing the ABS methodology. One means that can be taken to overcome this challenge is to validate and calibrate the agents' behaviour using empirical data (Bankes, 2002, Macal, 2016) though discussion in Chapter 2 shows that only 46% of ABS studies in the ASC area use primary data collection. The remaining studies relying on hypothetical or secondary data, which suggests that ABS modelling in ASC is facing a similar challenge.

The questionnaire survey, usually as part of a case study design, is one of the important research methods used for ABS studies in social science (Janssen and Ostrom, 2006). The survey usually employs a series of closed questions that aim to measure parameters quantitatively. The survey responses are usually used to determine coefficients and constraints in an equation-based ABS based on microeconomic theory (Robinson et al., 2007). For example, Happe et al. (2011) use a farm survey to identify available resources and their potential use in a linear optimisation matrix that describes plant and livestock production activities. The survey responses can also be used to generate statistical descriptions of agents' attributes in a population (Robinson et al., 2007). For example, Morgan et al. (2015) use a survey to estimate the key characteristics such as the demographics, income, risk tolerance and current farm practices of human actors. Researchers can also use survey data to construct an agent typology. For example, Valbuena et al. (2008) use data concerning demographics, perceptions and farm structures from their survey to classify clusters of agents. Similarly, Krejci et al. (2016) develop agent classifications based on their respondents' personae.

However, the use of questionnaire data to calibrate decision rules (i.e., modifying predefined rules or derive new rules) in ABS models related to ASC is rare.

3.2.2 The use of scenario-based questionnaires in eliciting human decision rules

Observing respondents' responses to a written scenario is one way of identifying real actors' decision rules. Researchers have used this data collection method in a variety of business and management studies. For example, in operations management, it has been used to explore the factors influencing the decision to outsource the manufacture of a component (Mantel et al., 2006). Urda and Loch (2013) use scenarios to explain how emotions and social preferences influence decision making. Choo et al. (2015) investigate how knowledge accumulation and manufacturing improvements are influenced by the style of executive problem solving adopted by an organisation. More recently, Azadegan et al. (2017) use scenarios to identify the drivers for managers in developing countries to increase their environmental investments. Similarly, Su et al. (2017) use a scenario-based questionnaire to investigate the effects of individual negotiation styles on the opportunism and compliance behaviours of buyers and suppliers.

Nevertheless, researchers can use standard closed question surveys to elicit human decision rules. For example, a farmer can be asked to rate (e.g., from "strongly disagree" to "strongly agree") the factors that influence him to sell his cow. However, bias due to memory loss makes this kind of survey question unreliable except for very salient events (Janssen and Ostrom, 2006). A scenario requires the respondent to solve a current and representative decision problem rather than recall a previous event. Hence, in contrast to retrospective self-reports, scenarios may reduce the biases from memory loss (Grewal et al., 2008, Su et al., 2017).

The scenarios used by researchers are usually based upon a theory (e.g., the theory of planned behaviour (Jafarkarimi et al., 2016)) or a number of null hypotheses to be tested. Responses to the scenario are then used to support or reject these null hypotheses. Unlike in other quantitative methods, researchers using the ABS methodology will hypothesize mechanisms and decision rules (Axelrod, 1997) rather than factors or parameter values. However, we believe that these hypothesized decision rules can be equivalent to null hypotheses when constructing scenario-based questionnaires. Hence, in this study, we developed the scenarios in our survey using the hypothesized decision rules created for the base model that we developed earlier from the findings of a literature review. We used a narrative for each scenario that was adapted from the real world farmers' experience. A previous study highlights that scenarios that are designed using real world situations allow the researcher to make generalizations or draw conclusions about an individual's or a group's behaviours in reality (Cowlrick et al., 2011). We then used the data collected from the survey to calibrate (accept or reject and adjust) the decision rules in the base model.

3.3 ABS base model of dairy supply chain in West Java

In this section, we present the ABS base model that we subsequently calibrated using the scenario-based questionnaire.

The typical dairy supply chain in Indonesia is composed of many tiers comprising farmers (producers), cooperatives (collector and handler), milk processing industries (processors), retailers and consumers. In common with earlier studies, the number of farmers is large while the number of processors is very small (Glock, 2012). Most farmers are smallholders with low production levels. Owing to population pressures, the land they own is relatively small and, usually, is only sufficient to build a pen for their cattle. For reasons of security, the pens are usually located next to the farmers' houses in the middle of residential areas. The forage grows along the road and river banks. It is difficult for the farmers to herd their cattle through the residential area so the farmers need to gather the forage from outside of their village and transported back using carts or motorcycles. In this sense, forage is a common resource for all these farmers and, when the forage availability is low, the competition between farmers to obtain forage becomes more intense.

In this supply chain, the milk produced by the farmers is collected and transported to the milk processors by farmers' cooperatives. The role of a farmers' cooperative is important because it is cheaper for the milk processing industries to buy milk in large quantities and, also, because it is highly perishable, the milk must be transported efficiently and refrigerated at all times (Glover et al., 2014, Manish and Sanjay, 2013) which is prohibitively expensive for the smallholder farmers. However, the cooperative's decisions are not fully controlled by the farmers. The cooperative also has external investors, shareholders and employs professional managers and workers. Hence, the cooperative operates like an independent company with smallholder farmers acting as suppliers who have little influence on the cooperative's decisions.

In our research, we modelled a dyadic interaction between smallholder farmers and the cooperative in West Java using ABS. The dairy supply chain in the case study area is one of the biggest in Indonesia and we considered it representative of other dairy supply chains in the country. Furthermore, we believe the case of the dairy supply chain to be suitable to demonstrate the benefits of a scenario-based questionnaire because the smallholder farmers (i.e., the respondents in our study) usually control their own decisions. Hence, the respondent's answer will correspond directly with the agent's decision rules in the simulation. This is in contrast to supply chains featuring large organisations in which the decisions are more likely to be made by a management team or via group agreement.

In order to develop the base model, we followed the suggestions of Gilbert (2004) and collated the relevant body of knowledge from previous studies. During this literature review, two sets of models relevant to the dairy supply chain were found. The first set of models assumes that farmers have a land endowment. They maximize their income by allocating their land to produce multiple crops. If they decide to produce milk then they allocate some of their land to grow the forage. Examples of these models are provided by Happe et al. (2009), Happe et al. (2011), Marohn et al. (2013) and Quang et al. (2014). The second set of models comprise grazing models in which the farmers herd their livestock to a common source of forage (*i.e.*, the rangeland). Examples of these models are provided by Boone et al. (2011), Rasch et al. (2016), Martin et al. (2016) and Rasch et al. (2017). In our case study area, the farmers also mainly rely on their surrounding environment as a common source of forage hence the second set of models was considered to be more suitable as the foundation for our base model. The main difference is that the farmers in our case need to transport forage for their cattle while the cattle do not move at all. This introduced more production constraints into our modelling such as labour, working hours and transport capacity.

In accordance with Macal and North (2010), the agents, their attributes, relationships and behaviours were then defined based on this body of knowledge. The conceptual model was implemented using the NetLogo programming platform. After the simulation implementation, verification and sensitivity analysis were carried out to eliminate errors in the base model. The base model was face-validated by presenting and discussing it with an expert panel in Indonesia. The expert panel comprised university researchers and policymakers from the Animal Husbandry Department, and an experienced farmer. The first aim of the face-validation is to justify that the boundary of the system being modelled is sufficient to replicate the real system behaviour (namely, the dynamics of milk production, cow and cattle population, and the number of farmer households). Secondly, this process aims to ensure that the base model has some correspondences with the reality and that its behaviour can be accepted rationally (Schmid, 2005) by the expert. In accordance with Sonderegger-Wakolbinger and Stummer (2015), the experts were encouraged to suggest revisions to the model boundary, assumptions, the agent's behaviour and the parameter values used in the base model. These suggestions were used to adjust and to improve the base model described below.

3.3.1 Description of the base model

The base model aimed to replicate the milk production, cow population and number of farmer household trends in the case study area. These model outputs were selected because of their importance to policymakers as indicated by the annual reporting of dairy industry statistics. To produce these outputs, the model uses several inputs such as the initial number of farmer households, number of family labour, cattle ownership, and cow productivity. The actual values of these parameters were identified using the survey described later.

There are three types of agent in the base model, namely: a number of separate farmer households; a cooperative and forage patches. The farmer household's role is to produce milk and supply the cooperative. The cooperative sets the milk price based on the milk quality and then sells the milk to the milk processing industry. The farmers interact with the patches whose main function is to provide forage for their cows. The conditions in the case study area are representative for the typical supply chain in Indonesia though the configuration of the agents in the system may vary. We did not set the farmer's location in the simulation based on the actual farm location. By running the simulation with many replications we aim to observe the system's behaviour under a variety of agent configuration. Hence, we expect that the average value from many simulation replications is representative for all dairy supply chains in the country. The simulation operates on daily time step, although some processes occur on a monthly and annual schedule.

3.3.2 The patch agent

One patch represents one kilometre square area and, in total, there are 306 patches in the base model. In the simulation, there are three types of patch (*i.e.*, used patch, unused patch and forage patch). Used patches represent the land area that has been occupied by building, houses, roads, etc. Unused patches represent empty land areas that are not suitable to grow forage but can be used to build new cattle pens. Forage patches represent land areas that are currently overgrown with forage.

Every day the forage patches produce forage. The amount of forage production $\frac{dF}{dt}$ on these patches ($Kg/km^2 day$) is defined as a function of the amount forage grow and forage taken, as described in equation 3.1.

$$\boldsymbol{\frac{dF}{dt} = Min\big((F_{max} - F_t - Fc_t), (F_t - Fc_t) * (1 + G)\big)} \tag{3.1}$$

F_{max} represents the maximum amount of forage (Kg) per kilometre square area. There are various forage grass species in the case study area. The details of the actual composition are not available. However, Bahar (2014) estimates the total weight of forage (consisted of various type of grass) that can grow in one kilometre square area in Indonesia is between 270 and 734 ($tonnes/km^2$). Hence, in each run, the maximum amount of forage that can grow on a patch is randomized within this range. We used uniform distribution in this randomisation process. This distribution was selected because information regarding the mean, mode or standard deviation of the forage density is not available. Similarly, we could not obtain any information regarding where or on what soil type the forage grows (spatial variation). F_t is the initial forage level at day t and Fc_t is the amount of forage taken by the farmers on day t. G represents the forage growth rate, which average value is 1.1% (per day) (Bahar, 2014) and it is taken as a constant.

There are many factors that influence the dynamics of forage availability such as precipitation (Gross et al., 2006) , land capability, soil type, gradient and seasonality. However the experts have agreed that considering these factors may make the model too complex to analyse. In addition, the data provided by Bahar (2014) was taken from regions with different precipitation, land capability, soil type and gradient. Therefore, these factors are neglected in current model version.

3.3.3 The farmer household agent

A farmer household agent consists of several family members who work together to rear cattle. Each farmer household has several attributes. Some of the farmer's attributes are modelled as variables (e.g., money, number of cattle, pen area and type of transportation mode). Other farmer attributes are modelled as lists (e.g., family members' age, cattle gender, cattle age, the percentage of fodder fulfilment, services per conception and maximum milk production). Each element in the services per conception and maximum milk production list represent the fertility and the maximum milk that can be produced by each cow respectively. The elements in these lists only have a non-zero value for the cows.

In line with the previous studies, we assumed that the farmers accumulate their assets over time (Gross et al., 2006, Boone et al., 2011). Farmers' assets consist of money and cattle. Farmers' income comes from milk and cattle selling and they use their money to pay monthly living expenses. According to our experts, rearing dairy cattle is very time consuming so very few farmers have sources of income other than producing and selling milk. Therefore, in this model, we assumed that farmers do not produce crops or have other jobs away from the farm. In common with earlier studies, we assumed that farmers increase their assets by using strategies to collect forage, sell milk, sell cattle, buy cattle and expand pen area (Gross et al., 2006, Boone et al., 2011).

Every day farmers collect forage to feed their cattle. They scan forage patches around their house. The maximum distance they can travel is limited by the number of working hours and the speed of the transportation mode at their disposal. Each farmer household typically has 8 hours per day to collect forage during the period between the cooperative's milk collections (*i.e.*, 7 am and 3 pm). In the case study area, the farmers collect forage on foot or by motorcycle or truck. In common with Martin et al. (2016), we assumed that farmers prioritize the location with the highest forage level when choosing the location to collect forage. If there is more than one location with the highest forage level then farmers prioritize forage collection from the closest location to their house.

Having decided on the location to collect forage, the agents move to the designated patch. Their travel time is taken away from their remaining working hours. The amount of forage they can collect from the given patch is constrained by the patch's forage level, the amount of family labour, their remaining working hours and their transport capacity. Actual measurements regarding these variables were not available so we asked our expert to suggest reasonable approximations based on their experience. The expert suggested that each family labourer could harvest 40 kg of forage per hour. Furthermore, the expert suggested that they could carry 40 kg of forage per person per trip if they transport the forage on their back or using a cart, 60 kg of forage per trip if they use motorcycle and 600 kg of forage per trip if they use a truck.

Farmer agents use the forage to feed their cattle. The cattle require 40 kg of fodder each per day that comprises forage and additional fodder. The expert suggested that, to stay healthy, the cattle require 30 kg of forage each per day. For the cows, the forage fulfilment also affects the quantity of the milk they produce. However, the expert suggested that the farmers usually substitute forage with additional fodder whenever they cannot collect enough forage for their cattle. This is consistent with a previous study that assumed the level of additional fodder used is affected by the forage availability (Gross et al., 2006).

Farmers' cows which have been pregnant can produce milk. The first pregnancy usually occurs after the cow's age reaches two years. The quantity of milk produced is determined by several factors (*i.e.*, age, genetics and forage) as described in equation 3.2.

$$\boldsymbol{Qm_i = \begin{cases} MaxProd_i * ProdEff(NumPreg_i) * \overline{Forage_i} & , PregPeriod < 7\ month \\ 0 & , PregPeriod > 7\ month \end{cases}} \quad (3.2)$$

Qm_i denotes the quantity of milk produced by cow i in a day. The $PregPeriod$ variable represents how long the given cow has been pregnant. The farmers usually stop milking a cow which has been pregnant for 7 months and restart the milking process after it gives birth. Hence the milk production during this period is zero. $MaxProd_i$ denotes the maximum milk production and reflects the genetic attributes of the given cow. $NumPreg_i$ indicates how many times the given cow has ever been pregnant and it represents the age factor. A cow's milk production is not constant throughout its lifetime. The expert suggested that a cow achieves its maximum milk production after the second pregnancy and then decreases linearly after the subsequent pregnancies. $ProdEff$ represents the percentage of $MaxProd_i$ which is currently produced by the given cow. We also assumed that the milk production is proportional to $\overline{Forage_i}$, which represents the average forage fulfilment (between 0 and 1) of cow i. The average forage fulfilment of 1 means that the given cow always obtains sufficient forage throughout its lifetime.

In addition to the quantity, the base model also takes milk quality into account. This variable determines the milk price per litre received from the cooperative. The expert suggested that the average proportion of forage in the total fodder determines the milk quality. Accordingly, the highest milk quality is achieved when the average proportion of forage is 75%. Hence, whenever the farmer agents substitute forage with additional fodder, the milk quality decreases and this leads to them receiving a lower milk price. We assumed the relationship between forage proportion and milk quality is a linear function in which the milk quality value is 100% when the forage proportion is 75% or higher.

Decisions regarding how many cattle should be retained are the most important decision made by the farmer agents because it would affect the amount of forage required, cattle weight, mortality and the amount of additional fodder used (Gross et al., 2006). Three separate processes determine how the farmers buy or sell their cattle in the base model. In the first process, the decision to sell or buy cattle is triggered by the forage availability (Gross et al., 2006, Lie and Rich, 2016, Lie et al., 2017). In the second process, this decision is triggered by the cattle's age (Rasch et al., 2016, Rasch et al., 2017). Finally, in the last process, it is triggered by farmers' financial condition (Boone et al., 2011).

In common with earlier studies (Gross et al., 2006, Lie and Rich, 2016, Lie et al., 2017), for the first process, we assumed that the forage availability is a trigger for the farmers to sell or buy cattle. When the forage is less available (e.g., during a drought), they sell some of their cattle and, conversely, buy new cattle (cows in particular) when the forage becomes more available. We assumed that the farmers sell or buy their cattle to an external agent outside the system and not to other farmers in common with earlier studies (Boone et al., 2011, Lie and Rich, 2016, Rasch et al., 2016, Lie et al., 2017, Rasch et al., 2017).

In making this decision, we assumed that the farmers could make a short-term forecast of forage availability. They calculate the average forage they obtain each day. When the average forage collected is not sufficient to feed all of their cattle they will start to sell their cattle. According to the experts, the farmers will prioritize the sale of the bulls first because they do not generate routine income. They will start to consider selling their cows only when they do not have any more bulls. When selling the cows, farmer agents compare the potential income they can get by feeding less forage but retaining all of their cows (equation 3.3) with the potential income they can get by feeding sufficient forage but selling some of their cows (equation 3.4).

$$\boldsymbol{income^{retain}} = \boldsymbol{n_{cow}}\left(\left(\overline{\boldsymbol{Qm}} * \overline{\boldsymbol{MP}} * \frac{\textbf{rounddown}\left(\frac{\overline{\boldsymbol{Fc}}}{\boldsymbol{30}}\right)}{\boldsymbol{n_{cow}}}\right) - \left(\left(\boldsymbol{10} + \boldsymbol{30} - \frac{\overline{\boldsymbol{Fc}}}{\boldsymbol{n_{cow}}}\right) * \boldsymbol{AfP}\right)\right) \tag{3.3}$$

$$income^{sell} = \text{rounddown}\left(\frac{\overline{Fc}}{30}\right) * \left((\overline{Qm} * \overline{MP}) - (10 * AfP)\right) \tag{3.4}$$

In equations 3.3 and 3.4, n_{cow} denotes the number of cows currently owned by a farmer. $\overline{Fc}$ represents the average forage obtained by the farmer and $\overline{Fc}/_{30}$ represents the maximum number of cows the farmer can retain for the given forage availability. $\overline{Qm}$,

$\overline{MP}$ and AfP represent the average milk production per cow, the average milk price per litre and the additional fodder price respectively. In equation 3.3, the farmer has more cows to produce milk but suffers a production penalty owing to the lack of forage and must pay more for additional fodder. In equation 3.4, the farmer has fewer cows but each cow can produce more milk and the agent does not need to buy additional fodder. If $income^{sell} > income^{retain}$ then the farmer will decide to sell the cows and vice versa.

Figure 3.1 illustrates how the farmer's decision is influenced by the average of forage obtained. This figures assumes that the farmer currently own 10 cows, the $\overline{Qm}$ is 30 litre per cow, the $\overline{MP}$ is 4,275 IDR/kg, and the AfP is 2,400 IDR/kg.

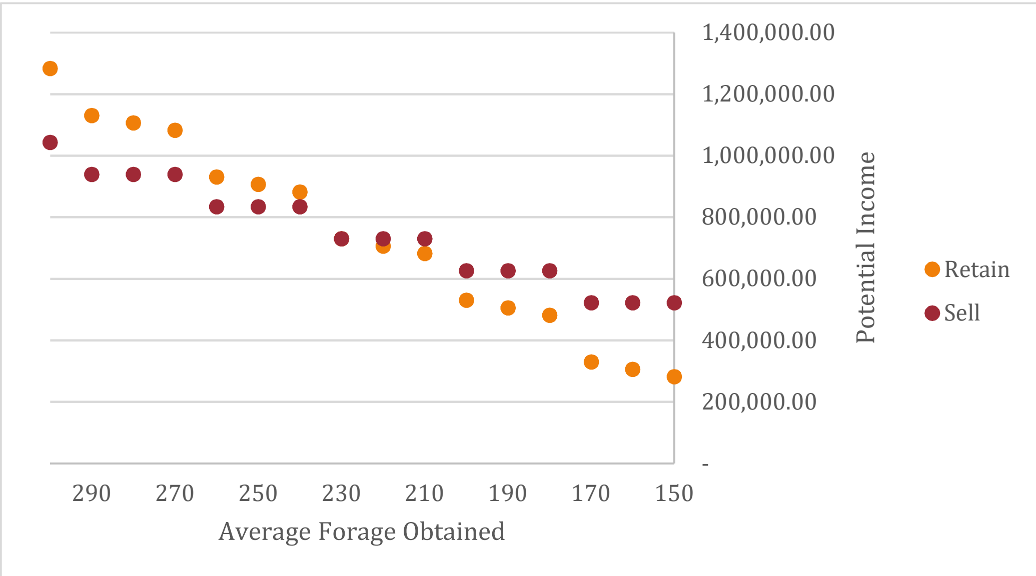

Figure 3. 1 Income from retaining and selling the cows

When selling cattle, we assumed that the farmers will prioritise the sale of the oldest cattle first as described in earlier studies (e.g., Boone et al. (2011)). For the bulls, an older bull usually has more live weight (Quang et al., 2014) and is more valuable. Older cows have more live weight but older cows are usually considered less productive.

On the other hand, if they can collect more forage than is needed then the farmers start to consider buying more cows. The number of new cows a farmer is willing to buy is proportional to the additional cows that can be fed using the excess forage. For example, if the forage excess is only sufficient to feed one more cow then the farmer agent will consider buying one additional cow. The constraints in the buying decisions are the pen capacity and the farmer's money. If the farmer agent owns sufficient pen capacity to contain all of its current cows and the new cows then it just needs to have sufficient money to buy the cows. However, if the farmer agent does not have sufficient pen capacity then it must have sufficient money to buy the cows and to increase the pen capacity. The farmer agent's ability to increase pen capacity is also limited by the land availability on the patch where it is living. The fertility and productivity of newly bought cows are assumed to be random.

In the second process, the selling decision is based on the cattle's age. In general, the bulls are sold at two years old. According to experts, farmers believe that the bulls have reached their optimum live weight at this age. Meanwhile, the cows are culled when they reach the age of 10 years. At that age, it is believed that the milk productivity of the given cow has become too low. The ages at which the bulls are sold and the cows are culled are subject to the calibration process.

In the third process, the cattle selling decision is triggered by the farmer's financial condition. Each month, the farmer agent forecasts the amount of money it will have at the end of the month by taking into account the income it earned in the previous month and the living expenses it must pay. The living expense value is calculated by multiplying the number of the farmer agent's family members and the standard cost of living in the area. If the forecasted amount of money is less than the living expense value then the farmer agent starts to consider selling its cattle. As in the first process, the farmers were assumed to sell the bulls first. In this process, we also assumed that the farmers select the cattle to sell based on the age. This selling process is repeated until the farmer agent's money deficit is covered.

In all of those three processes, we identified the maximum amount of money that can be earned by farmers by selling their cattle using the survey data. The actual price received by the farmer agent was assumed to be proportional to the age of cattle being sold. On the other hand, the price that must be paid by the farmers to buy a new cow was assumed constant and the value was determined using the survey data.

Cow reproduction is the next process executed by the farmer household agent. Prior studies incorporate a fixed time schedule (e.g., annually) or growth rate (e.g., increase the population by 10% every year) for cow reproduction (e.g., Gross et al. (2006), Rasch et al. (2016), Martin et al. (2016), Rasch et al. (2017)). In our study, we considered a heterogeneous cow fertility factor. In the cow reproduction process, the farmers artificially inseminate those cows who are two years old or older and not pregnant at the beginning of each simulation month. The successfulness of the artificial insemination process depends on the cow's fertility, which is represented by the services per conception variable. If the artificial insemination fails then this process would be repeated in the subsequent month.

If the artificial insemination process is successful then the pregnancy process lasts for nine months. The cow then gives birth to either a male or a female calf, each with 50% probability. If the cow gives birth to a female calf then the newborn calf inherits the milk productivity and fertility of its mother.

The next procedure related to farmer households retirement and succession. There are two main factors affecting retirement and succession of farmer household agent (*i.e.*, age and financial condition). At the end of each simulation year, all farmer household members who are older than the productive age are removed from the farmer household family member list and the amount of family labour decreases. A farmer household agent can also acquire a new family member with a probability of 1.2% (the average population growth in Indonesia). A farmer household agent is deleted from the simulation if it does not have any family members left or if it runs out of money and cattle.

Probabilistically, a new farmer household can be generated in the simulation. We defined its attributes based on the input parameters, as in the initiation process of farmer household agents. However, as we mentioned earlier, owing to population growth, the farmland that was once located in the rural area is currently surrounded by residential areas. The non-farmers tolerate the existence of a farmer household who continues dairy farming because they are native to the area while the non-farmers are mainly newcomers. The cooperative's database also showed that all of its members were farmer families from generation to generation. However, conflict with non-farmers could spark easily if a newcomer tries to start dairy farming. This conflict is usually triggered by pollution caused by manure production and potential water contamination. When a farmer household decides to stop dairy farming, their land will usually be sold and converted into residential area settlement or another business. Our simulation aimed to replicate the reality in the case study area so the probability of a new farmer agent entering the system is set to be equal to zero. However, a sensitivity analysis can be run on this probability value if the survey data indicated that the emergence of new farmers was quite possible.

3.3.4 The cooperative agent

The cooperative agent collects and grades milk from all farmer household agents. It is assumed that the cooperative determines the milk-buying price as a linear function of milk quality, ranging from Rp 3350 to Rp 5200 per litre (Rp is Indonesian currency). Based on the discussion with the experts, the cooperative sells the milk to the milk processing industry at a fixed price. The actual buying price from the milk processing industry is unknown but the experts estimated that it is approximately 5500 (Rp per litre). The experts agreed that the cooperative's daily operational costs can be assumed to be fixed regardless of the total volume of milk they handle. Hence, it is more profitable if they can operate at full capacity.

3.4 Survey Instrument Design, Survey Process and Survey Findings

Section 3.4.1 discusses the process to develop the survey instrument in this study, the survey process and findings. The survey instrument included closed questions and a scenario-based questionnaire. The findings that support the assumptions used in the base model are discussed in section 3.4.2, while section 3.4.3, section 3.4.4 and section 3.4.5 discuss new decision rules found from the survey result. In addition, we also discuss the respondents' perception of the scenario-based survey in section 3.4.6.

3.4.1 Survey instrument design and survey process

The purpose of our survey was the collection of data that we could use to calibrate the input parameters and decision rules of the farmer agents in the base model. The cooperative was excluded from the survey because its decisions are made by many decision makers collectively. Figure 3.2 describes the process we adopted to design our survey instrument. This figure shows that questions related to input parameters calibration are grouped in Part 1 of the survey instrument, while scenarios to calibrate the agents' decision rules are grouped in Part 2.

To develop the questions in Part 1, we began by listing the parameters used in the base model. These parameters included demographic (*e.g.*, age and education), socio-economic (e.g., income and off-farm jobs) and technical factors (e.g., cattle ownership and cow productivity). The questions to identify the value of each parameter were developed by taking examples from the previous studies and surveys such as the agricultural census (Statistics Indonesia, 2017).

To develop scenarios in Part 2 of our survey instrument, we listed all the types of decisions that can be encountered by the farmer agent. The purpose of the scenarios was to calibrate the decision rules so it was important that this list included not only actions that can be taken by the farmer agent in the base model, but also other actions that may be performed in the real world. We also provided an option where the respondents could explain actions that were not represented by other options. We have included an example for Scenario 2 in Appendix 2.

Next, we listed the decision parameters for each decision. This represented the information considered by the agent to select its action. We then determined the range of decision parameter values that would be used to make variations of a scenario. It is important that the decision parameter range included all values that can occur in the simulation and the real world (i.e., collectively exhaustive). This was done to avoid bias owing to extrapolation. Extrapolation bias may happen when a decision parameter value that occurs in the simulation goes beyond the range of data obtained from the respondent. If this happens then the agent's decision rules in the simulation are not representative of the real world actor anymore. For example, in this study, we set the cattle mortality range in our scenario between 0% - 100%. Information from experts' observations can be useful to establish these ranges especially when the decision parameter distribution is a priori unknown. For example, the experts observed that there are farmers who start to sell their cattle when experiencing forage shortage for a week. However, they also observed that some farmers will retain their cattle for two months even though they are facing a forage shortage. Based on this information, we set the range of farmer's forecast horizon in our scenario between one week and two months. If the decision parameter proves to be significant and it is not possible to specify a collectively exhaustive range for it, then a special error message would be created to warn when its value in the simulation violates the data boundary. The corresponding run would then be excluded from further analysis because it may contain bias.

We combined these actions and decision parameters with a story to develop each scenario. This scenario guides the respondents to choose their actions by considering the given decision parameters. The story in a scenario is based on real farmers' experiences that are observed and retold by the experts. We then use the minimum, maximum and mean value of each decision parameter range to vary one scenario into several sub-scenarios using permutation. Presenting several scenario variants is important to identify the sensitivity of a real actor's actions toward the changes in decision parameter value.

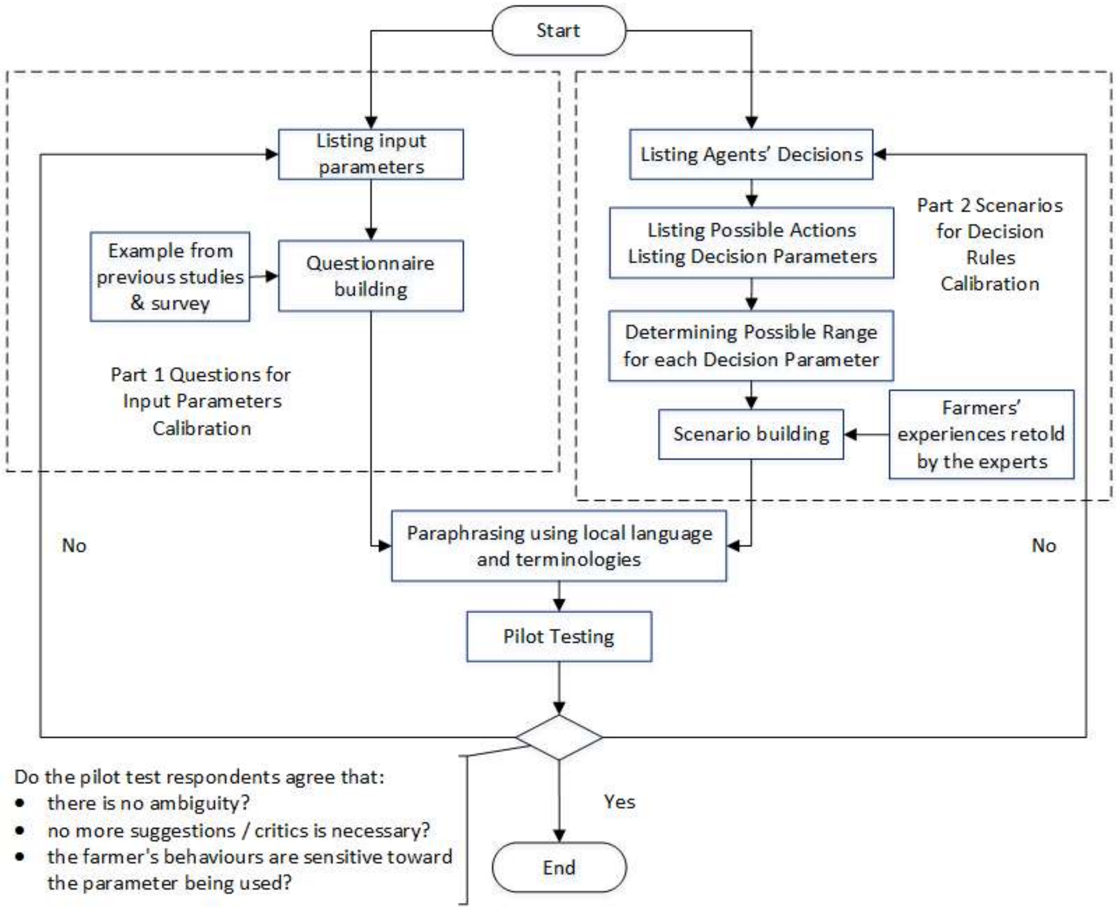

Figure 3. 2 Process used in developing the scenario-based questionnaire

The questions in Part 1 and Part 2 were then translated using local language and terminologies. We also used traditional measurement units in all of the survey instrument questions to ensure that the respondents could understand all questions easily.

Pilot tests were then conducted with lecturers and graduates from the Animal Husbandry Department and a farmer. These respondents were chosen because they had experience in interacting with the farmers in the case study area. There were several objectives of this pilot test. Firstly, it aimed to minimize errors and ambiguity by asking the respondents to propose revisions to the questions or scenarios that were ambiguous or difficult to be understood. Secondly, the pilot testing aimed to ensure that respondents' behaviours were sensitive to the scenarios presented while keeping the questionnaire as short as possible. This was important because the permutation technique we used initially resulted in a massive number of scenarios in the initial survey instrument design. We asked the respondents to suggest new parameter values if their behaviour was not sensitive toward the decision parameter values presented. Respondents could also propose new action options and decision parameters for a scenario. The proposed action would be considered in the revised survey instrument if it was mutually exclusive to the existing options and it was plausible for a real farmer. If two adjacent sub-scenarios were considered too similar and had no effect on the decision then the respondents could propose the elimination of one of the scenarios. When changing the decision parameter values as well as eliminating a sub-scenario, it was necessary to keep the combination of decision parameter value across all sub-scenarios collectively exhausted. By using respondent's suggestions, we improved the survey instrument design over the course of three iterations.

The full scale survey was carried out from 1st to 31st of August 2016. The respondents comprised 153 farmer households located in 19 villages in the West Java Area. The scenario-based questionnaire is generally more complex than a standard closed question questionnaire so each respondent was accompanied when completing it. The survey was conducted from house to house in the evening after the respondents had finished all of their daily activities to ensure there were minimal distractions for the respondents. The questionnaire used hard copy format. On average, each respondent required two hours to complete all the survey questions. Randomly, we interviewed several respondents after they completed all the questions in the survey instrument. In these interviews, we asked about their perceptions regarding the survey instrument that we used. If they have participated in similar surveys (e.g. agricultural census), we also asked them to compare the usefulness of our survey instrument.

After the survey was completed, the respondent's responses were converted into an electronic format using Microsoft Excel and SPSS. Codes are used to record responses to closed answer questions in part 1 and part 2. If the respondent gave open answers, the respondent’s answers were then transcribed as sentences. This transcript was then read in turns by the research team. The research team then agreed whether the respondent’s answer could be classified into one of the codes used in the closed answer option. Our analysis shows that all respondents' open answers, actually, can be classified into one of the answers provided.

Similarly for open answer responses, the respondent's response to interview questions was transcribed into sentences. In turns, the transcript was read by the research team. The research team then classified the respondents’ response, for example, as to whether the respondent can easily understand the survey instrument and whether the respondent prefers the scenario-based questionnaire to the standard questionnaire.

3.4.2 Empirical data for input parameter calibration and assumption validation

This section discusses how we used the survey data to validate assumptions in the base model and to calibrate input parameter values. The data discussed in this section were extracted from the first part of the survey that consists of demographic, socio-economic and technical parameter questions. Examples of questions used are presented in Appendix 1.

The first assumption in the base model was that dairy farming was the sole income for farmer households in the case study area so they did not aim to maximize their income by combining various on-farm and off-farm jobs. 85.6% of our sample stated that dairy farming is their sole income. Using a 95% confidence interval, we estimated that 80% - 91.2% of farmers in the area also focused on rearing dairy cattle. For those who have income other than dairy farming, this income contributes between 12% and 19.3% of their total income. This first assumption distinguishes our study from earlier studies (e.g., Boone et al. (2011), Rasch et al. (2016), Rasch et al. (2017)) in which the farmers are assumed to have multiple sources of income.

The second assumption in the base model was the reliance on family labour in dairy farming. Our data showed that, at a 95% level of confidence, only 14.9% to 28.1% of farmer households have ever hired labour other than family members. In this subset, the involvement of outside labour was also very low, with an average between 6.2 to 16.5 person-hours per month (rounded up to 1 to 2 person-days in a month) at a 95% confidence level. Consequently, the data supported the assumption that non-family member involvement could be ignored.

Our third assumption was that farmers rely on a forage-commons to feed their cattle. 100% of the responses showed that the farmers obtain forage by collecting it from the areas surrounding their village. There were some farmers who could produce additional forage but the proportion of the forage they produced compared with the total forage they needed was very small (at a 95% level of confidence, the average is between 5.5% and 13.3%) so this assumption was also supported by the data.

The survey data showed that, at a 95% level of confidence, between 98.1% and 100% of farmers routinely used additional fodder. In line with the assumption in the base model, 100% of respondents stated that they obtained this additional fodder from a supplier. Our data also showed that, on average, the farmers gave 0.4 kg more additional fodder for every 1 kg decrease in forage availability, which was also in line with the assumption in the base model.

Regarding the decision to quit from dairy farming, 86.3% of the respondents agreed that the absence of a successor was the most important contributing factor. The second contributing factor was financial difficulties (bankruptcy) which were agreed by 71.3% of the respondents. Only 8.4% of the respondents perceived a possibility of the rise of new farmers as their competitors. These data supported the assumptions in the base model regarding the farmer household retirement and succession.

The survey was also used to identify the value of the input parameters for the model. We used STATFIT software to identify the most appropriate theoretical distribution for the survey data in order to improve the correspondence between the characteristics of the agent population in the simulations and the farmer household population in reality. Table 3.1 presents the empirical data for the main model parameters.

Table 3. 1 Empirical parameters for model initiation

	Descriptive statistics					
Variable Name	Min	Max	Mode	Mean	Std. Dev	**Distribution**
Agent attributes						
Farmer Age (years)	22	74	38	46.17	10.98	Triangular
Family Labour (person)	0	4	1	0.92		Binomial
Number of Cow (heads)	0	18	3	4.10		Poisson
Number of Bull (heads)	0	5	0	0.81		Poisson
Peak Milk Prod (litre/day)	10	35	20	20.81	19.35	Normal
Service per conception (times)	1	8	2	2.38		Poisson
Constants						
Cow Selling Price (millions IDR/head)				13.1		
Bull Selling Price (millions IDR/head)				16.4		
Heifer Buying Price (millions IDR/head)				9.6		
Minimum Milk Price (IDR/litre)				3,350		
Maximum Milk Price (IDR/litre)				5,200		
Additional Fodder Price (IDR/Kg)				2,400		

In Table 3.1, the farmer age indicates the age of the head of the farmer household. When initiating our simulation, we first generated the age of the head of the farmer household. Subsequently, the age of the remaining family members was generated between the age of the head of the household and the minimum value of Farmer Age. Family Labour indicates the number of family members involved in dairy farming. The structure of the cattle population owned by the farmer household is represented by the Number of Cow and Number of Bull variables. Please note that the minimum value of Number of Cow parameter is zero. This happens because there was a farmer who was replacing all of his cows (he had culled all of his cows and was searching new cows to buy). Peak Milk Prod indicates the highest amount of milk that was ever produced by a cow. Service per Conception represents how many times a cow should be given artificial insemination until it gets pregnant.

Cow Selling Price and Bull Selling Price indicate the highest selling price ever received by the farmer. In the simulation, the farmers receive these prices when they sell the cattle at optimum age, which was assumed equivalent to the optimum weight. If farmers sell a cow or bull before it reaches the optimum age, then the selling price received decreases proportionally with the animal's age. Inflation was not considered in this model and the cattle selling prices are considered as constants. Heifer Buying Price represents the average price that should be paid by the farmers when buying a new heifer. Maximum and Minimum Milk Price cover the price range received by the farmers when selling milk. Milk processing industries buy milk at a fixed price so the range of milk buying price given by the cooperative was also considered constant. Additional Fodder Price is the cost to the farmer for one kilogram of additional fodder. This was considered as a constant because the cooperative provides subsidies to maintain the additional fodder price.

3.4.3 Empirical buying decision making rule

The base model assumed that the farmers' willingness to buy new cows was influenced by the excess forage they obtain. The initial scenario was therefore designed to identify farmers' behaviours toward different levels of forage excess. However, respondents during pilot testing suggested the milk price as a factor that possibly can affect farmers' willingness to increase the number of their cows. Hence, we combined these two variables to represent the farmers' buying rule. This combination produces four sub-scenarios (see scenario 1 in Appendix 2). In each sub-scenario, we asked the respondent to state how many cows they were willing to buy assuming that they had sufficient money.

We used multiple regression analysis to extract the empirical buying decision making rule (The complete statistical analysis is presented in Appendix 3). Owing to the high skewness in the excess forage data, we transformed it using a square root function to obtain a better fit. In the current model version, all control variables (e.g., socioeconomic and demographic information) were ignored. The ANOVA result shows that a linear function is appropriate to explain the relationship between the dependent variable (number of new cows) and the independent variables (number of excess forage obtained and milk price). In the best fit model, the only significant predictor is the square root of additional forage obtained, with a coefficient of 0.095. This variable explains 14.7% of the variation in the number of new cows a farmer wants to buy. The final regression model is shown in equation 3.5.

$$\boldsymbol{Add_{Cow} = -1.603 + 0.095 Add\_Forage^{\frac{1}{2}}} \tag{3.5}$$

In the base model, we assumed that farmers would be willing to buy one new cow every time the excess forage they obtain is sufficient to feed one additional cow. Therefore, the regression coefficient indicates that the real farmers are more risk averse than the theoretical farmers.

3.4.4 Empirical selling decision making rule

In the base model, we assumed that the farmers would sell their cows when they are experiencing forage deficit. Hence our initial scenarios were designed to identify farmers' response to the length and magnitude of forage deficit. However, the respondents in the pilot testing suggested the possibility that farmers would not react to these scenarios. According to them, the farmers often substituted forage with vegetable waste or even banana trees. The farmers would start to consider their actions after they observed the impact on their cattle's health. The farmers obtained this information from a veterinarian who visits them weekly. Based on this suggestion, we modified the scenario design to identify the farmer's response to the information provided by the veterinarians concerning the cattle's health condition.

In these scenarios, we asked the farmers to decide whether to sell their cattle or not when they are facing various drought intervals (i.e., to represent the period during which they usually experienced forage shortage), the level of forage deficit and the cattle's health condition (e.g., Scenario 2 in Appendix 2). The options in these scenarios are nominal so we used multinomial logistic regression to extract the empirical selling decision rule. We present the complete statistical analysis result in Appendix 4. The model fitting information shows that a logistic regression model is appropriate to represent the relationship between the independent and dependent variables (with significance value < 5%). The Nagelkerke pseudo R^2 shows that the model can predict 75% of respondents' response.

The parameter estimates show that the length of drought and the level of forage deficit do not affect the farmers' decision to sell their cows (with significance value of 0.922 and 0.873 respectively). In the open answer field, some of the farmers also explained the possibilities of using forage substitutes. This supported our pilot test respondent's observation.

In contrast to the forage deficit and drought period scenarios, farmers' responses were more sensitive toward the cattle's health condition scenarios (with significance value < 5%). The regression coefficient of 11.442 shows that as the likelihood for the cow to become sick and die increases the more the farmers choose to sell the cow. Equation 3.6 shows the final regression model.

$$\ln \frac{P_{sell}}{1 - P_{sell}} = 11.442 P_{die} - 6.342 \tag{3.6}$$

Our study's approach differs from the assumptions in earlier similar models (e.g., Gross et al. (2006), Lie and Rich (2016), Lie et al. (2017)) in which the farmers sell cattle when they are experiencing forage deficit. The fact that length of drought and the level of forage deficit are not significant is very beneficial for a later experiment. The range of these two variables' values is a priori unknown and their value in the simulation may go beyond the range of the empirical data but, because these two variables are not significant, we did not need to select outputs that may contain bias due to extrapolation.

3.4.5 Empirical sorting decision making rule

The base model assumed that, when selling cows, the farmers prioritize the sale of the oldest cows first. However, in the pilot testing our respondents also proposed other characteristics that might be considered by the farmers, namely: cow fertility and whether it is pregnant or not. Hence, we also incorporated these characteristics into our scenario-based questionnaire (e.g., Scenario 3 in Appendix 2). In each scenario, we asked the farmers to compare two cows with different characteristics. We then asked them to choose which cow they prefer to sell. From this pairwise comparison, the surveyor then helped the respondents to order their preference from 1 (the most preferred) to 8 (the least preferred).

We used regression analysis to describe the farmers' preference based on the cow characteristics. The cow characteristics become dummy variables in the regression model (i.e., a code of zero is used for the old cows, cows with low fertility and pregnant cows). The regression model shows that all three characteristics are significant predictors of farmers' priority and the model R^2 is 97.8%. Equation 3.7 shows that age becomes the first criteria in farmer selection process followed by pregnancy and fertility factors, respectively. The farmers place higher priority on selling a cow that is older, not pregnant and with low fertility.

$$\boldsymbol{Priority = 2.76 + 4 * Young\ Cow + 1.21 * High\ Fertility - 1.75 * Not\ Pregnant} \tag{3.7}$$

The first criteria in farmers' empirical decision making rule is similar to the decision making rule in the base model. Nevertheless using the scenario-based questionnaire we are able to obtain more detail information regarding how the farmers select the cow to be sold.

3.4.6 Respondents' perception toward the scenario-based questionnaire

Most of our respondents had taken part in previous studies that also used questionnaires as a data collection instrument. An example of these studies is the agricultural census conducted by the Indonesian Statistical Bureau annually (Statistics Indonesia, 2017). To reveal their perception toward this questionnaire design compared to the design in the previous studies, we conducted a short interview with some of them after they completed the questionnaire.

More than 80% of the interviewees felt that they could understand the scenarios presented. This is because the scenarios were written in their daily language and terminology. According to them, this questionnaire was different to the questionnaires in the previous studies. In the previous surveys, it was difficult for them to imagine how the data would be used and how the research outcome would be beneficial for them (partly because the surveys often used technical terms and concepts which they did not always understand). In contrast, some of the interviewees could guess how the data from the scenario-based questionnaire could be used to select interventions that might help them. For example, one of the interviewees said that "If the government or cooperative know that we decide to sell our cows because it is very difficult to collect sufficient forage, then they could help us to import forage from other regions".

The interviewees also found that the scenarios had occurred or were very likely to occur in the real world. Those who ever faced similar situations claimed that their responses to our questionnaire were similar to their actual actions back then. Those who had never faced similar situations claimed that it was very likely that they would take similar actions to their responses in the questionnaire. They also considered this design to more beneficial for them because it stimulated them to think about their action if they were to face a similar real scenario in the future. These interviews provide an additional form of face-validation that gave us more confidence that the decision rules revealed by the respondents reflected what they actually do.

3.5 Experiment Results and Sensitivity Analysis

The experiment in this section aims to test whether introducing the empirical decision rules can improve the external validity and lead to significant changes in the base model's behaviours. We also analyse the effects of different decision rules on the system's performance and propose the decision rules that might be preferable for the real-world actors.

In this experiment, the decision rules from the empirical survey data were combined into seven model variants. We refer to these models as empirical models. In the experiment, all models (base model (M_0) and empirical models (M_{SBQ}1-7), M_{SBQ} stands for model calibrated with scenario based questionnaire) use the same input parameters and the initial population of farmers is set based on the real data in January 2010. Table 3.2 presents the combination of decision rules used in each model variant. We control the random number to ensure fair comparison so that the difference in the model outputs is solely caused by the different decision rules used by agents. Each model was run for five simulation years (from January 2010 to December 2014) and replicated 25 times.

Table 3. 2 Decision rules in models calibrated with scenario-based questionnaire.

	M_0	M_{SBQ1}	M_{SBQ2}	M_{SBQ3}	M_{SBQ4}	M_{SBQ5}	M_{SBQ6}	M_{SBQ7}
Buying	Base	SBQ	Base	Base	SBQ	SBQ	Base	SBQ
Selling	Base	Base	SBQ	Base	Base	SBQ	SBQ	SBQ
Sorting	Base	Base	Base	SBQ	SBQ	Base	SBQ	SBQ

In Table 3.2 base label indicates that the decision rule from the base model is used in the calibrated model. Conversely, SBQ label means that the empirical decision rule from scenario-based questionnaire data collection is being used.

3.5.1 Impacts of empirical decision rules on ABS operational validity

In this section, we investigate whether adding the empirical decision rules obtained from the scenario-based questionnaire data improves base model validity. Two techniques are used for the validation process (i.e., mean error estimation and regression analysis). The mean error estimation aims to measure the magnitude of model output deviations from the real data and the regression analysis results show how well the model outputs represent the trends in the real data. We used the real cattle population, cow population, milk production and the number of farmer households data obtained from the farmer cooperative (KPBS, 2016) to validate the ABS models. These variables are considered to be important by both the government and cooperative when recording their statistics.

Table 3. 3 Cattle population, cow population, average daily milk production and the number of farmer households in Pangalengan West Java 2010-2016 (KPBS, 2016)

Year	Cattle population (head)	Cow population (head)	Average daily Production (litre)	Farmer Household
January – 2010	21,322	21,083	159,333	5072
January – 2011	21,438	20,960	136,694	4204
January – 2012	22,366	22,073	138,904	3439
January – 2013	16,173	16,080	97,476	3053
January – 2014	13,415	13,399	84,207	2888
January – 2015	12,563	12,555	76,372	2852

To estimate the mean error, first, the difference between model outputs at the end of each simulation year and the real data, from January 2010 until December 2011 (i.e., $Error_i = Data_i - Simulation_i$ where i = 2011 … 2012) was measured. This time interval is chosen because a drastic decline occurred in cow population, cattle population and average daily milk production in 2012 (which appears in the data for January 2013). This decline occurred owing to an external factor that was not considered in the model (i.e., the policy to stop beef imports). This policy created an incentive for the farmers to sell their productive cows as meat.

We then computed the mean error (ME) from 2011 to 2012 (i.e. $ME = \sum_{2011}^{2012} Error_i/2$). Table 3.4 shows the average ($\overline{ME}$) and standard deviation (S_{ME}) of outputs from 25 replications. A t-test was then carried out to infer whether, in the long run, the model's average ME is zero. The two-tailed significance (sig. column) of the t-test at 95% confidence level is also presented in Table 3.4.

Table 3. 4 The descriptive statistics and t-result of ABS models' error from the real data

Model Name	Cattle Population		Cow Population		Daily Milk Production		Farmer Households	
	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.
M_0	(-2272.3, 4395.4)	0.02	(-1443.1, 4025.3)	0.09	(20600.2, 13421.1)	0.00	(62.6, 130.6)	0.02
Empirical models with one empirical decision rule								
M_{SBQ1}	(-1494.8, 5075.0)	0.15	(-876.3, 4703.7)	0.36	(16811.3, 15151.1)	0.00	(48.3, 125.1)	0.07
M_{SBQ2}	(-1458.1, 5095.7)	0.17	(-867.4, 4573.9)	0.35	(6267.3, 18333.9)	0.10	(422.4, 313.5)	0.00
M_{SBQ3}	(-1755.2, 4643.9)	0.07	(-975.3, 4244.8)	0.26	(20693.8, 13108.0)	0.00	(61.0, 122.3)	0.02
Empirical models with two empirical decision rules								
M_{SBQ4}	(-1506.9, 5088.8)	0.15	(-876.0, 4718.4)	0.36	(16943.8, 15206.1)	0.00	(45.1, 124.1)	0.08
M_{SBQ5}	(-1523.6, 5144.1)	0.15	(-965.3, 4593.2)	0.30	(5912.9, 18434.2)	0.12	(408.8, 315.7)	0.00
M_{SBQ6}	(-1472.9, 5104.2)	0.16	(-874.6, 4586.1)	0.35	(6359.1, 18349.6)	0.10	(422.3, 311.9)	0.00
Empirical model with three empirical decision rules								
M_{SBQ7}	(-1504.0, 5116.5)	0.15	(-904.4, 4588.1)	0.33	(6118.6, 18383.3)	0.11	(411.1, 310.0)	0.00

In Table 3.4, a lower $|\overline{ME}|$ value indicates that on average the model output is closer to the real data. While, a significance value higher than 5% indicates that we fail to reject the null hypothesis that the simulation output reflects the real world data (i.e., a valid model). Table 3.4 shows that the base model is only valid for prediction of the cow population. However, Table 3.4 also shows that the model's operational validity can be improved by using the empirical decision rules. Buying, selling and their combinations are the decision rules that can improve the model's operational validity on most output variables while the empirical sorting decision rule can only increase the model's validity in predicting cattle and cow population. Table 3.4 also shows that the significance of the base model and model that use empirical sorting decision rules are not very different.

Table 3.5 summarizes the regression analysis results between simulation outputs and real data. In this regression analysis, the mean of simulation outputs from 25 replications (for example the mean of simulated cow population in 2012, $\overline{Cow}^{2012} = \sum_{i=1}^{25} Cow_i^{2012} / 25$, with i represents the replication) was used as the independent variable and real data was used as the dependent variable. This regression analysis focused more on the match between the trends produced by the simulation and the trend in real data rather than the accuracy of the predicted value. Consequently, the external factor mentioned earlier is not very influential and all data from 2011-2015 can be incorporated. The significance column (Sig) in Table 3.5 shows the significance of the ANOVA test and confirms the validity of the regression analysis. A lower significance value indicates a smaller probability that the relationship between the average simulation outputs and the real data occurs by chance. The positive regression coefficient value, presented in column B, indicates that the simulation outputs and the real data have a similar trend (i.e., they move in the same direction). The R^2 values show the proportion of variation in the real data that can be explained by the simulation outputs variation. A higher R^2 value indicates that a particular model has a better fit to the real data.

Table 3. 5 Summary of regression analysis between the simulation outputs and the real data

Model Name	Cattle Population			Cow Population			Daily Milk Production			Farmer Households		
	Sig.	B	R^2	Sig.	B	R^2	Sig.	B	R^2	Sig.	B	R^2
M_0	0.00	2.77	0.98	0.00	3.27	0.96	0.03	1.44	0.83	0.00	0.60	0.97
Empirical models with one empirical decision rule												
M_{SBQ1}	0.00	3.00	0.99	0.00	3.75	0.98	0.02	1.66	0.86	0.00	0.62	0.97
M_{SBQ2}	0.00	4.11	0.95	0.00	3.89	0.96	0.00	5.31	0.97	0.00	0.67	0.99
M_{SBQ3}	0.00	2.70	0.99	0.00	3.21	0.97	0.03	1.43	0.83	0.00	0.60	0.97
Empirical models with two empirical decision rules												
M_{SBQ4}	0.00	3.04	0.99	0.00	3.82	0.98	0.02	1.66	0.86	0.00	0.62	0.97
M_{SBQ5}	0.00	4.16	0.95	0.00	3.98	0.96	0.00	5.41	0.97	0.00	0.67	0.99
M_{SBQ6}	0.00	4.09	0.98	0.00	3.93	0.97	0.00	5.29	0.98	0.00	0.67	0.99
Empirical model with three empirical decision rules												
M_{SBQ7}	0.00	4.10	0.95	0.00	3.95	0.97	0.00	5.40	0.98	0.00	0.66	0.99

Table 3.5 shows that for all output variables all models have significantly linear relationships with the real data. All models are also able to imitate the trends in the real data. However, the models that use the empirical decision rules often have a better fit to the real data. Specifically, the empirical buying decision can increase the R^2 value for most output variables.

3.5.2 Impacts of empirical decision rules on model outputs

In this section, we evaluate whether adding empirical decision rules significantly affects the behaviours of the base model. We use the same approach as the previous experiment, except that we measure the difference between empirical model and the base model outputs (i.e., $Error_i = Output_i^{base} - Output_i^{empirical}$ where i = 2011 … 2015). For each empirical model, we measure its impact by using the mean of difference for the five-year period (i.e. $ME = \sum_{2011}^{2015} Error_i/5$). A high ME value indicates that the output of a particular empirical model is different from the base model. Table 3.6 shows the average ($\overline{ME}$) and standard deviation (S_{ME}) of outputs from 25 replications.

Table 3.6 also shows the significance value from the paired t-tests between each empirical model and the base model (at 95% confidence interval).

Table 3. 6 The descriptive statistics and t-result of empirical ABS deviation from the base model

Model Name	Cattle Population		Cow Population		Daily Milk Production		Farmer Households	
	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.
Empirical models with one empirical decision rule								
M_{SBQ1}	(-574.9, 355.9)	0.00	(-814.4, 400.7)	0.00	(-8252.5, 2121.3)	0.00	(-44.2, 13.6)	0.00
M_{SBQ2}	(-1011.8, 581.7)	0.00	(-726.4, 341.6)	0.00	(-36557.5, 11822.1)	0.00	(319.9, 230.4)	0.00
M_{SBQ3}	(85.8, 190.3)	0.03	(69.6, 169.9)	0.05	(300.0, 959.6)	0.13	(3.7, 20.2)	0.37
Empirical models with two empirical decision rules								
M_{SBQ4}	(-592.6, 370.7)	0.00	(-824.3, 410.4)	0.00	(-8166.9, 1900.5)	0.00	(-45.3, 21.1)	0.00
M_{SBQ5}	(-1140.1, 639.2)	0.00	(-844.1, 385.1)	0.00	(-37192.5, 12147.4)	0.00	(304.4, 230.6)	0.00
M_{SBQ6}	(-1012.6, 603.2)	0.00	(-728.1, 370.0)	0.00	(-36360.7, 11981.0)	0.00	(319.6, 228.4)	0.00
Empirical model with three empirical decision rules								
M_{SBQ7}	(-1099.6, 583.6)	0.00	(-808.6, 344.6)	0.00	(-36931.7, 11832.4)	0.00	(305.6, 226.2)	0.00

Except for the sorting decision, all significance values indicate that the outputs of the empirical model are different from the outputs of the base model. When we compare the empirical models with the empirical decision rule individually, we can observe that among the three empirical decision rules, the selling decision rule has the highest impact on most model outputs as confirmed by having the highest$|\overline{ME}|$. On the other hand, the sorting decision rule has the lowest impact on all model outputs. This also explains why the differences between M_0's and M_{SBQ3}'s significance in Table 3.4 are low. The empirical models with two and three empirical decision rules show that there is an interaction effect between all empirical decision rules. For example, for most outputs, adding the empirical sorting decision tends to lower the $|\overline{ME}|$ values (e.g., $|\overline{ME}|^{MSBQ7} < |\overline{ME}|^{MSBQ5}$ on cattle population, cow population and daily milk production outputs).

In validating the simulation and measuring the impact of empirical decision rules in this study, we relied very much on statistical significance tests principles. Analysing simulation outputs using statistical tests can be biased, because the p-value will show a significant result if the modeller use a very large sample size (number of replications), even though the effect size of an intervention is very small (White et al., 2014). However, the experts also argue that hypothesis testing is still useful when one desires to compare the output of a simulation's output to observed data (White et al., 2014) or a baseline value. This is what we did in this study. Also, we limited the number of replications in our experiment. We stopped adding new replications once the change of the average simulation outputs across different replications is not significant in practice anymore. For example, Figure 3.3 presents the average of cow population in 2015 across different number of replications. This figure shows that after 20 replications, the average of cow population stabilises and adding more replications does not significantly change the system's performance (± 200 heads in difference) from a practical perspective.

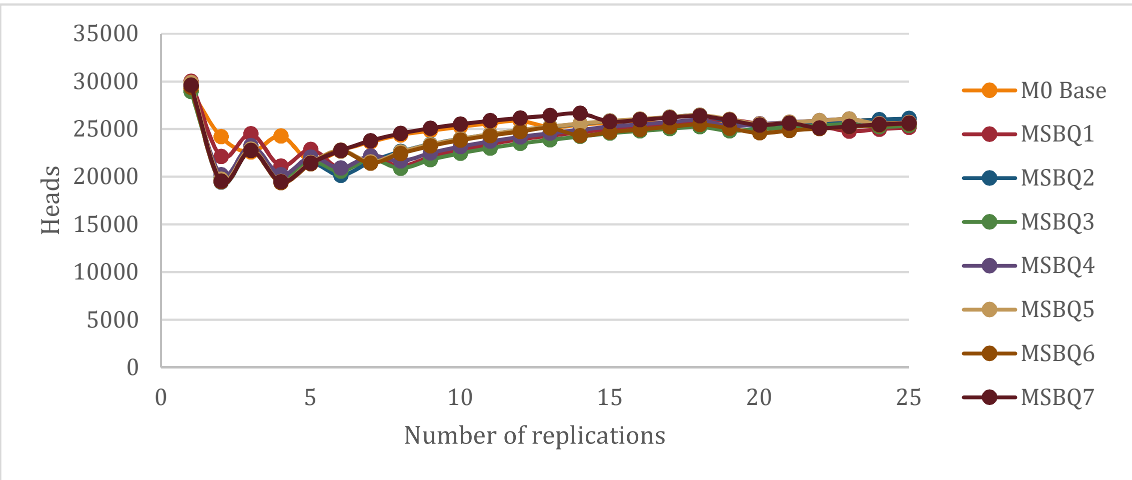

Figure 3. 3 the average of cow population in 2015 across different number of replications

3.5.3 Effects of various decision making rules on the system performance

Following on from the identification of the empirical decision making rules that might improve ABS operational validity (i.e., representativeness), we now analyse whether retaining these rules is more beneficial for the real-world actors. To answer this question, we compare the performance of models with different decision making rules calibration by plotting the mean of the simulation outputs.

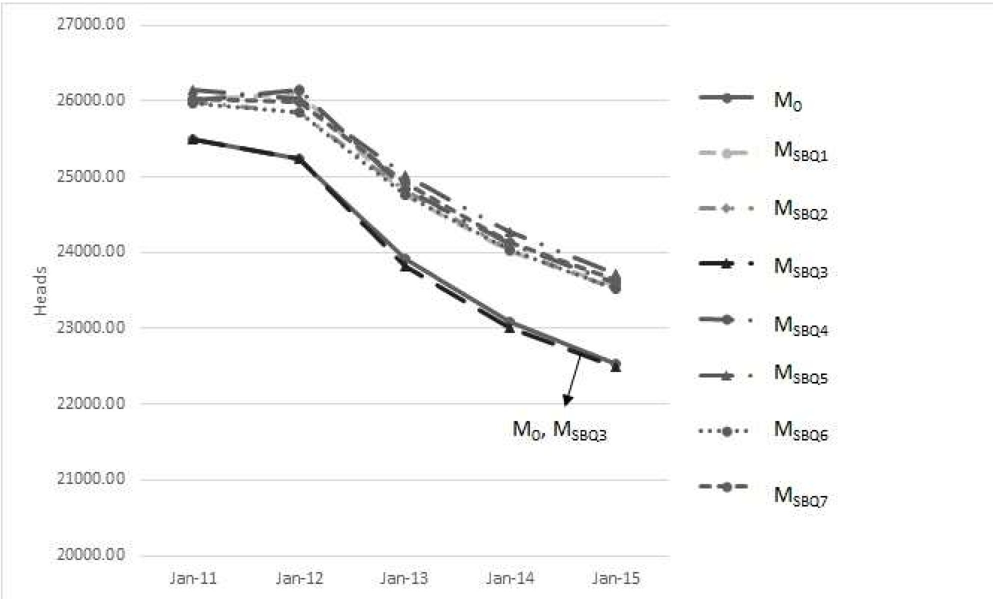

Figure 3. 4 Cow population of models calibrated with scenario-based questionnaire

Our analysis shows that the cow population level (Figure 3.4) from most models (except for the base model and model with the sorting decision) are almost the same. However, Figure 3.5 shows that the mortality rate of models with the empirical selling decision is much higher when compared with the models without it. This happens because, in the empirical selling decision rule, the farmers take risks and keep their cows even though they have insufficient forage. This causes agents in these models to spend more money to replace their cows throughout the simulation period. In addition, because the forage sufficiency also determines milk quality and buying price, the agents receive less income when they are experiencing forage deficit. These two factors lead to a higher number of retiring farmers in models with the empirical selling decision rule (Figure 3.6) and lower average farmers' assets (Figure 3.7).

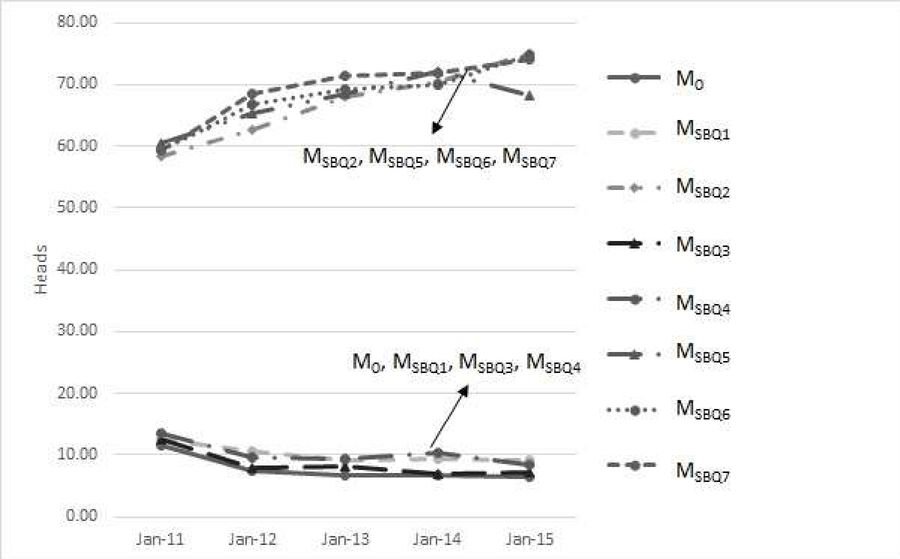

Figure 3. 5 Cattle mortality rate of models calibrated with scenario-based questionnaire

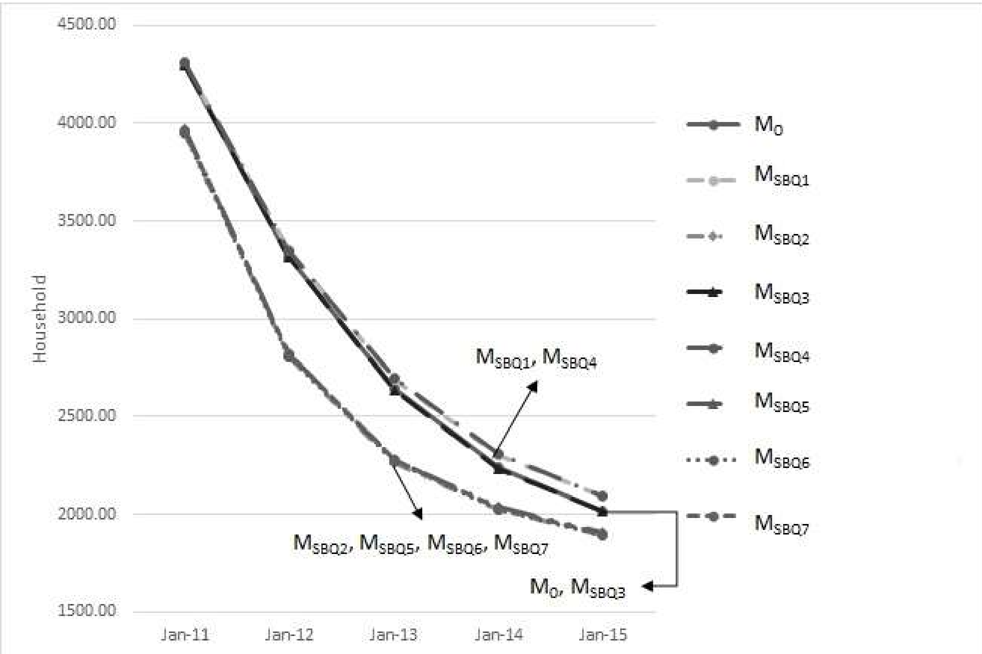

Figure 3. 6 The number of farmer households of models calibrated with scenario-based questionnaire

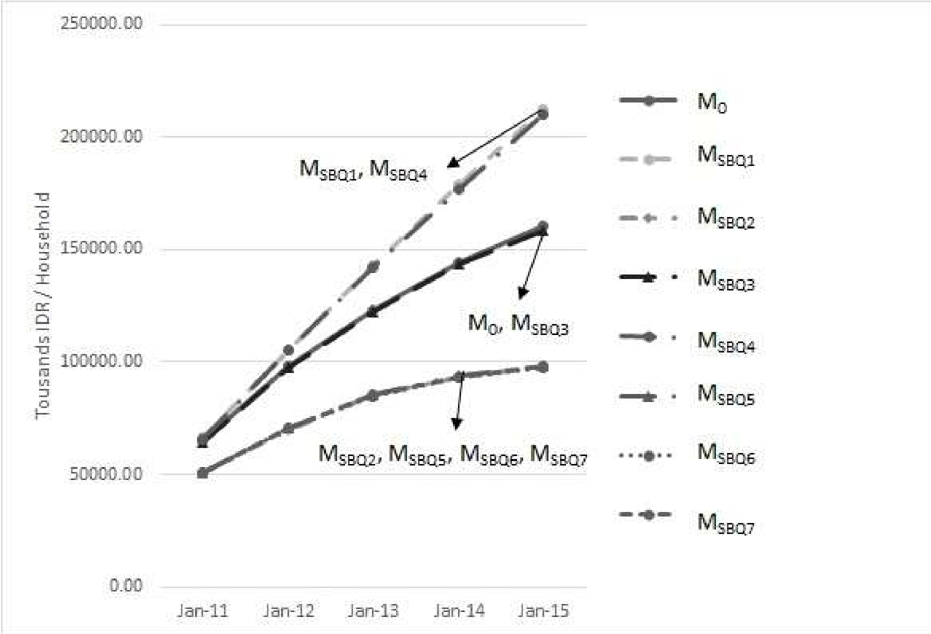

Figure 3. 7 Average farmer household's asset of models calibrated with scenario-based questionnaire

Counterintuitively, models that use the empirical selling decision rule always produce higher daily milk production (Figure 3.8). This happens because the high number of retiring farmers means the competition to gather forage becomes less intense. This leads to a higher percentage of forage fulfilment and milk productivity per cow. The less intense competition also leads to higher cow ownership per farmer household in these models. The empirical data from our survey as well as the experts' observations together confirm that it is virtually impossible for the new entrants to enter the system in our case. However, in other case studies, this condition may easily attract new entrants. This could mean that the competition intensity increases and decreases over time.

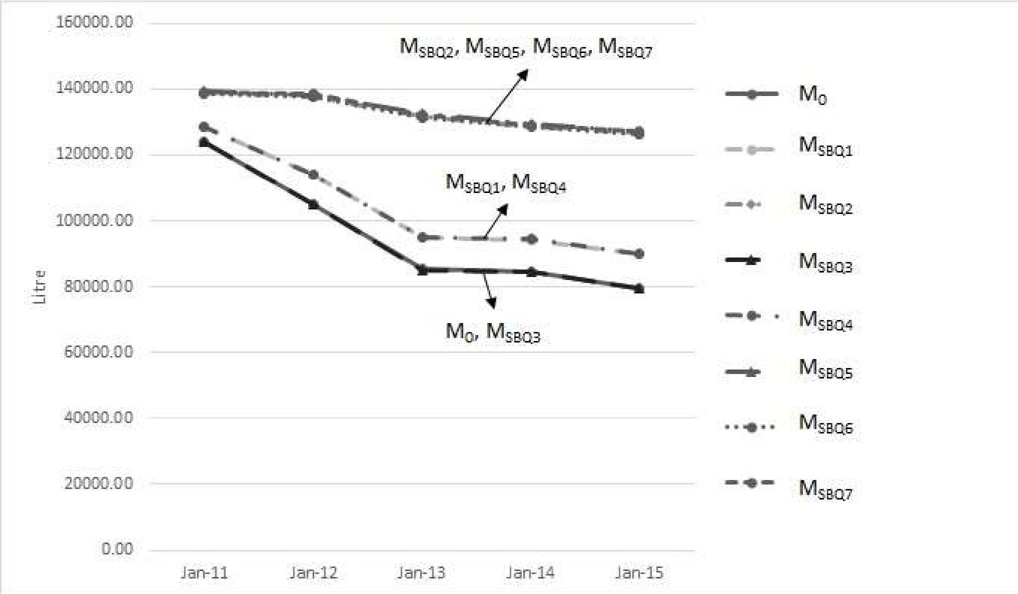

Figure 3. 8 Average daily milk production of models calibrated with scenario-based questionnaire

The dynamics in our model differ from previous modelling results (e.g., Lie and Rich (2016), Lie et al. (2017)) in which the increase in milk production is solely affected by the total cow population. In our model, we introduce an additional factor that affects the milk production (i.e., the cow’s productivity). The cow's productivity is in turn influenced by the forage sufficiency which itself is affected by the competition to obtain the forage.

To test the effects of competition among the farmers we compared several single runs from the base models that have different initial total forage values. A higher amount of forage represents less intense competition. In Figure 3.9 we plot the total cow population in each year. In this figure, the total initial forage in low competition scenario is 316 million tonnes, 315 million tonnes in medium competition scenario and 138 million tonnes in high competition scenario. Figure 3.9 shows that the cow population increases at the beginning of the simulation in the low competition scenario but is stationary in the medium competition scenario and, tends to fall in the high competition scenario.

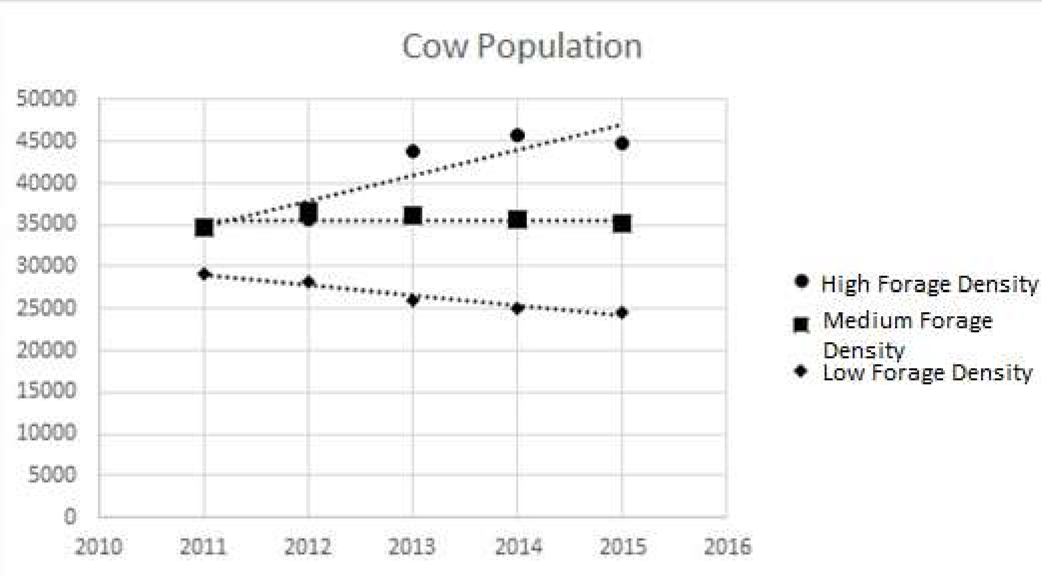

Figure 3. 9 Cow population under different forage competition scenarios

This result is important for policymakers and can be used to propose behavioural interventions. If policymakers wish to prioritise the milk production quantity, retaining the empirical selling decision rules would be more beneficial. On the other hand, if they wish to place more emphasis on the farmers' welfare then interventions to change the farmers' behaviour to the decision rule used in the base model would be more appropriate.

3.6 Discussion

3.6.1 Insights from scenario-based questionnaire data collection process

This section discusses the insights obtained from our experience in designing and using scenario-based questionnaires as a data collection method to calibrate decision rules in ABS. Based on our experience, the scenario-based questionnaire has the following benefits:

- *A scenario-based questionnaire enables the clarification of the context of the agents' decisions*. According to Yang and Gilbert (2008), surveys that are usually used to collect quantitative data place less emphasis on context (i.e., whether and how a decision rule is activated by considering an agent's current state and environment). Furthermore, An (2012) observes that the statistical methods commonly used to analyse survey data are often problematic in providing insight into an agent's motive, incentive and preferences when making a decision. Our experience shows that, with careful pilot testing, it is possible to identify the context of a decision rule using a scenario-based questionnaire. For example, we identified that cattle health conditions rather than forage shortage trigger the farmers' selling decision rule. These health condition scenarios were proposed by our pilot testing respondents. By applying statistical analysis to the survey data, we were also able to identify farmers' preferences when selecting the cow to be sold.

- *The concepts incorporated in a scenario-based questionnaire are meaningful for the respondent*. Yang and Gilbert (2008) suggest that one of the differences between qualitative data and quantitative data relates to how meaningful the concepts used are for the real world actors. Concepts used in qualitative data are usually more meaningful for real world actor (emic). On the other hand, concepts used in quantitative data are usually more meaningful for the researchers (etic). Robinson et al. (2007) suggest this is one of the disadvantages of the survey as a data collection methodology in ABS because the respondents' understanding might be different from the researchers and may bias their responses. Interviews with respondents indicate that the data from a scenario-based questionnaire, which is designed with sufficient pilot testing, is more emic than the standard closed end questionnaire. The potency of a scenario-based questionnaire to minimise potential bias in respondents' response could be beneficial to increase ABS validity and credibility. This, in turn, can increase the policymakers' willingness to use the modelling results.

- *A scenario-based questionnaire can identify how actors react to new scenarios*. Generally, the data obtained by a survey is considered as a snapshot in time. Consequently, Robinson et al. (2007) suggest that the survey method is good for capturing the existing condition but not very suitable to represent temporal variation. Longitudinal surveys are effective in capturing temporal variation but this option can be expensive and is not always feasible within the constraints of a research project. The interviewees reported that the scenarios used in our study could help them to think about the actions they would take in situations they had not yet experienced. This suggests that even though the scenario-based questionnaire survey remains as a snapshot in time, we can still obtain indications of how the real actors will choose their actions in possible future situations.

- *A variety of established statistical techniques can be used to analyse the data obtained from a scenario-based questionnaire to create decision rules*. For example, in our study, the decision to sell or retain a cow is binary and we used multinomial logistic regression to extract the decision rule and, alternatively, we could use techniques such as curve fitting. It is also possible to incorporate the effect of agent heterogeneity in the decision rule as suggested by Robinson et al. (2007). For example, this can be achieved by clustering agents' attributes (e.g., based on demography and socioeconomic parameters as by Valbuena et al. (2008)) or by using these attributes as dummy and control variables in a regression model. Using the behavioural sensitivity analysis, as we describe in section 5, we could choose the decision rule that has a better fit to the real data.

Nevertheless, the scenario-based questionnaire in this study also inherits the weaknesses of the survey method. For example, our survey assumes that the head of the farmer household is the sole decision maker in the family (Robinson et al., 2007). In reality, each family member may contribute opinions and thoughts when the head of the household make a decision. Also, we rely on statistical techniques to analyse the data and these techniques rely upon many structural and technical assumptions (Robinson et al., 2007). Similarly, extrapolation based upon statistical analyses of survey data needs care. Relationships derived from the analyses of survey data can be good at estimating values within the data range (i.e., interpolate). However, when the simulation is running there is potential for the variable values to exceed the range of empirical data. In this case, the decision rules derived from the survey are used for extrapolation. When this happens, the decision rule in the ABS is not representative of the actual agents even if, on aggregate, our simulation result is valid when compared to the real data. We attempted to minimise this potential bias by defining collectively exhaustive parameter ranges to be used during the scenario design process. There were several parameters whose range was a priori unknown (e.g., how long the farmers experience forage shortage before they eventually decide to sell their cows) but, fortunately, these parameters did not significantly affect the farmers' decisions. If these parameters were significant then we could have avoided the potential bias by excluding simulation runs in which these parameters' values exceeded the data range.

3.6.2 Insights from experiments with a variety of empirical decision rules

In section 3.5.1, we illustrate that the use of different empirical decision rules may produce models with different levels of validity. A model with higher validity is indicated by a higher significance value in the t-tests of the models' $\overline{ME}$ and also higher R^2 value. A higher t-test significance value means the mean of the model outputs are not different from the real data. A higher R^2 indicates that the model can better represent the trends in real data. The experiment results enabled us to select the models that can better represent the system (i.e., more valid). In our case, we considered a model to be better if it had significance above 5% on most output parameters and higher R^2 than the base model. When applying these criteria the models with a higher operational validity than the base model are M_{SBQ1}, M_{SBQ4}, M_{SBQ6} and M_{SBQ7}.

In section 3.5.2 we demonstrate that the greater the differences between the empirical decision rules and the base model's decision rules then the greater the changes in $\overline{ME}$ significance value. This result allowed us screen out those decision rules that did not significantly affect the final modelling results in order to keep the model simple and easier to understand (parsimony). For example, adding the empirical sorting decision rule did not produce significantly different results to the base model because this decision rule is executed only after the selling decision is made. Therefore, in order to keep the model simple, the empirical sorting decision rule (M_{SBQ3}) was set aside even though it had higher operational validity than the base model.

Our experiment results could also be used to identify the decision rules that influence the model outputs the most. In our study, the most influential decision rule was the empirical selling decision because its outputs were significantly different from the base model's output for all output variables. Furthermore, compared to the other empirical decision rules, it also has the highest $\overline{ME}$ value on most output parameters. Consequently, policymakers should pay more attention to this decision rule when designing interventions to the system.

In section 3.5.3, we describe how the model behaviours are sensitive to the empirical decision rules being used. Incorporating the empirical selling decision rule significantly changed the value of all output variables while the cattle mortality rate variable was the only model output that was not sensitive to the use of the empirical buying decision rule. All models produced similar trends. For example, the cow population and the number of farmer household variables showed declining trends towards the steady state, the cattle mortality rate tended to be stationary, and the average of farmer household assets was increasing but may have a saturation point. However, we also noted that in models with the empirical selling decision rule the cattle mortality rate was always higher. This was detrimental to the farmers and led to higher rate of retirement in these models. However, owing to the smaller number of farmers in these models, each farmer household was more able to satisfy its cattle forage requirement. Consequently, the total cow population in these models was almost similar to the other models. In addition, the milk productivity per cow in models with the empirical selling decision was higher.

We propose two overall conclusions following our discussion of our modelling results. Firstly, the data from the scenario-based questionnaire are useful not only in the validation of the decision rules in the base model but also in the calibration process which leads to models with higher operational validity. Secondly, testing the sensitivity of the resultant model to different empirical decision rules can successfully identify the most and the least influential empirical decision rules. This leads to simpler models that focus on those decision rules that are important from the policymaker/interventionist perspective.

3.6.3 Insights for policymaking

Throughout this study we worked closely with the farmers' cooperative. The cooperative provided secondary data to validate the simulation outputs. They also allow us to access their member database who were then invited as participants in this study. Therefore, in this section we analyse how our simulation can be useful for the farmers' cooperative to design new policy interventions.

Based upon our results in Section 3.5.3, we observe that if the real farmers continue using the empirical selling decision rule then the outcome is a dairy supply chain comprising a smaller number of farmers, with higher cow ownership and productivity, but lower asset value owing to high cattle mortality.

For the farmers, it is reasonable to assume that they would prefer an outcome in which they own higher asset values. Higher cow ownership does not necessarily mean that these smallholder farmers will transform into big farmers. The involvement of non-family member labour is very low, owing to their availability and the farmers' capability to hire them, so cow ownership is still constrained by family labour and forage availability. However, this outcome is also problematic for the farmers' cooperative. A smaller number of farmers but with higher cow ownership and higher productivity may mean it is more efficient for the processing industries to buy milk directly from the farmer and obviate the need for the cooperatives.

Data from FAOSTAT (FAO, 2016) and the agricultural census (Statistics Indonesia, 2017) show that there is no significant correlation between milk consumption and milk prices in Indonesia. Higher milk production can lead to a lower milk price but cheaper milk does not necessarily mean that customers consume more milk. Hence, for government policymakers, an increase in milk production is only valuable if it reduces the import quantity and contributes to the import/export trading balance. Excessive production that leads to the decrease in price may lower the farmers' income. This is a very important consideration for the government as the agricultural sector absorbs 31.7% - 37.9% percent of the workforce in Indonesia. The modelling outcome also signals a potential increase in rural unemployment and its associated problems for the government. Retiring farmers contribute to the unemployment numbers but the loss of cooperatives is also important in that they employ many people in the supply chain (e.g., milk graders, truck drivers and fodder factory workers). Furthermore, environmental degradation becomes a risk in that retiring farmers would usually sell their land that may soon be converted into residential, industrial or recreational areas that, among many concerns, will reduce the water catchment area.

In summary, even though the empirical selling decision rule incorporates reality the performance is suboptimal. On the other hand, the selling decision rule in the base model neglects some of the reality but its consequences could be more desirable. Hence, encouraging the farmers to forecast the forage availability and make selling decision based on it is more beneficial for the real world actors. To enable this shift then the farmers should be empowered, for example, by training them to record the daily forage they obtained and to forecast forage availability based on this record. The government and cooperative could help by providing equipment (e.g., weighing scales) which are not owned by all farmers in the case study area. Other possible aids could include an online climate forecast application with a forage availability projection module. The farmer's cooperative can work together with university researchers and The Agency for Meteorology and Climatology can help to provide such tools.

Another intervention would be to make forage more available by preventing land conversion from forage into residential uses or by helping the farmers to utilize abandoned land to grow forage by easing permits or supporting the farmers to lease these lands. These policies are beyond farmers' cooperative control. However, the farmers' cooperative has substantial power to lobby other policymakers such as the spatial planning department.

Further developments might include the farmers' adoption of forage preservation technology in such a way that the availability of forage becomes more stable. This technology is available but it is underutilized owing to the constraints the farmers have in terms of labour, time, and suitable storage facilities.

We plan to present these potential interventions to the farmers' cooperative. We hope that the farmers' cooperative can suggest which intervention is feasible and can be used in reality. Also, many of these interventions (i.e., forage preservation technology, preventing land conversion, and utilization of abandoned land) are beyond the current model's boundary. By presenting the modelling results, we hope that the farmers' cooperative can suggest which intervention that should be included in the future modelling.

3.6.4 Generalizability of the applicability of scenario-based questionnaire of the results

This chapter demonstrates the value of a scenario-based questionnaire in the calibration and validation of agent decision rules in an ABS model of a dairy supply chain. We have also shown that the calibrated decision rules contribute to improving the operational validity of the final model. We believe that a scenario-based questionnaire approach may also be useful in other agri-food/agricultural supply chains. For example, traditional fisheries and the natural rubber industry exhibit features and structures that are very similar to our case study. Furthermore, our case study can be interpreted from an operations management perspective in which input resources (i.e., forage and fodder) are transformed into a final product (i.e., milk) using transforming resources (e.g., labour, cows). The farmer's decision rule when collecting forage or to substitute forage with additional fodder are analogous to an operation's decisions as to how much and where the input resource can be obtained and at what cost. Similarly, the farmer's decision to buy or sell is similar to an operation's decisions regarding how much investment in transforming resource/capacity it needs both now and into the future to satisfy customer demand for the product. Clearly, the complexities of, say, a large-scale manufacturing operation are very different to those of an Indonesian farmer household. For example, the decision-making processes may feature negotiation and consensus reaching within a cross-functional management team. This would require additional behavioural modelling similar to the studies mentioned earlier in section 3.2 (i.e., Mantel et al. (2006), Urda and Loch (2013), Choo et al. (2015), Azadegan et al. (2017), Su et al. (2017)). Nevertheless, there does appear to be an opportunity to explore applications of scenario-based questionnaires to elicit operations decisions within ABS models in a broader range of industry sectors.

3.7 Conclusions

In this paper, we have presented steps to design a scenario-based questionnaire from a base model that was developed from the previous literature. We test the usefulness of scenario-based questionnaires in a case of supply chain dairy, and record respondents' perception of the survey. We show that the data obtained through the survey is useful for calibrating the input parameters and the decision rules in the base model. After running experiments with various combinations of empirical decision rules, we show that some empirical decisions rule can improve the model's validity. We also analyse the model's robustness regarding the different empirical decision rules developed and identify possible behavioural interventions in the real system. Overall, we have demonstrated that the use of a scenario-based questionnaire can enhance the value of ABS models in ASC for both researchers and users of such models.

4 Improving the Credibility of Agent-Based Simulation using Role Playing Games: A Case Study in Dairy Supply Chain

This chapter is adapted from a manuscript prepared for submission to the European Journal of Operational Research. The manuscript is authored by Dhanan Sarwo Utomo, Dr Bhakti Stephan Onggo, Dr Stephen Eldridge, Dr Andre Rivianda Daud and Safitri Tejaningsih. The first three authors contributed in the study design and major part of manuscript preparation. The last two authors enriched the original manuscript using their field experiences. The primary data collection and analysis were done by the first and the last two authors. Finally, the simulation model and experiments presented in this chapter were developed and analysed by the first author. Adjustments are made and commentaries are added to the original manuscript to improve the coherence with other parts of this thesis.

Abstract

Role playing games (RPG) have been widely used by researchers in agent-based simulation (ABS) as a means of observing player behaviour. These players may be subsequently represented as agents in an ABS model or they may be policy makers who will use the ABS model as a decision support tool. Prior research demonstrating the benefits of RPG to facilitate learning and communication among the stakeholders is plentiful. Our study extends this research by considering how the benefits of RPG in ABS model calibration can be quantified and, particularly, whether RPG can be used to develop more credible agent decision rules. Using the dairy supply chain in Indonesia as a case study, our findings suggest that the decision rules extracted using RPG can improve ABS validity. Furthermore, we demonstrate that these decision rules both highlight opportunities for behavioural interventions to improve system performance and enable a more inclusive policy planning approach that takes into account the perspective of the small-holder farmers.

Keywords: OR in Agriculture, role-play game, calibration, validation, agent-based simulation, agri-food, dairy

4.1 Introduction

Agent-based simulation (ABS) is an Operational Research (OR) method that has gained popularity as a decision support tool by owing to its ability to relate human behaviours in a system to the emerging patterns of the behaviour of the system as a whole. A key challenge for ABS researchers is to incorporate more realistic and representative models of human behaviour (Macal, 2016). This challenge is not restricted solely to ABS but also in OR generally and has led to the increasing importance of behavioural OR research (Hämäläinen et al., 2013). Modelling human behaviour is also important in other research fields such as operations management (Bendoly et al., 2006, Bendoly et al., 2010) yet the study of stakeholder behaviour in OR is currently underexplored (Hämäläinen et al., 2013).

Recent research has shown that engaging stakeholders in the modelling lifecycle is very beneficial for a simulation project (Voinov and Bousquet, 2010, Robinson et al., 2014, Scott et al., 2016, Voinov et al., 2016) and de Gooyert et al. (2017) note an increasing trend of OR studies that involve stakeholders. They suggest that such engagement enables researchers to obtain better insights regarding how the stakeholders view and structure their problems, and how they define and make trade-off decisions when faced with a variety of options. These insights are particularly valuable for ABS model developers during the conceptual modelling phase. Furthermore, this engagement with stakeholders makes policy development more inclusive. Inclusivity is important in policy making because a local community may have knowledge or wisdom that is unknown, and not considered, by policymakers (d'Aquino and Bah, 2014). Therefore, engaging a greater range of stakeholders is considered to be important for future OR practice (Higgins et al., 2010).

Role Playing Games (RPG) is one of the techniques that has been used in ABS to facilitate engagement with real world stakeholders (Janssen and Ostrom, 2006, Robinson et al., 2007, Smajgl et al., 2011a, Voinov et al., 2016). Most of the current RPG studies focus on demonstrating the benefits of RPG in facilitating learning and communication among the stakeholders. However, owing to the richness of interaction during an RPG session, it is possible that stakeholders/players devise decision rules that are not representative of those in the real system (i.e., artificial decision rules). Many of the ABS models that have been calibrated using RPG aim to explore all possible stakeholder strategies and to evaluate the consequences of these strategies so it does not matter if some of the decision rules obtained through the RPG are artificial. However, if an ABS model is to be both realistic (i.e., explains what is really happening in the real system) and predictive (i.e., able to estimate the real system behaviour) then these artificial decision rules must be screened out.

In this paper, we propose an extension to current RPG data collection practice in order to enable the calibration of a more realistic and predictive ABS model. Our extension employs operational validity measures proposed by Sargent (2013) and uses a dairy supply chain in Indonesia as a case study to illustrate how this can be conducted and evaluated. Agri-food supply chains are important application areas of OR (Ahumada and Villalobos, 2009, Borodin et al., 2016) and empirical calibration is common in ABS studies of them. However, as mentioned in Chapter 2, studies that engage stakeholders in model development and calibration process are rare. Supply chains that feature large organisations can be problematic in that decisions might be taken via group consensus and an individual player in an RPG may not represent the group decision. Consequently, the Indonesian case is useful in that the dairy supply chain comprises smallholder farmers who, usually, are the sole decision makers and it could be expected that their observed behaviours during the RPG sessions better reflect their decisions in the real world.

The remainder of our paper is organised as follows. We begin, in Section 4.2, with an overview of the use of RPG in ABS research and propose a possible extension. We follow this with an analysis to develop the hypothesised decision rules that we can use in our base model which will be used later for comparison purposes with the models calibrated using RPG. In Section 4.3, we explain how the data collection process was carried out using this extended RPG. The results from our RPG data collection (i.e., extracted decision rules) are presented in Section 4.4. In section 4.5, we quantify the benefit of RPG by reporting the findings of a set of simulation experiments that were conducted to compare the operational validity of the base model and RPG-calibrated models. These finding and their implications for behavioural intervention are discussed in Section 4.6 and we present our conclusions in Section 4.7. All appendices are available in the supplementary material.

4.2 Literature Review

4.2.1 The process to design and use an RPG for ABS calibration

RPG aims to collect information about the stakeholder's perceptions, their decision rules and behaviour in a particular context by observing how the players make decisions (individually or collectively) under various scenarios or policy interventions (Robinson et al., 2007). In representing the context being studied, RPG is considered to be more descriptive and easier to understand in comparison with laboratory experiments Janssen and Ostrom (2006). Therefore, its benefits in aiding the stakeholders to express their feelings and understanding and to overcome communication problems owing to lack of trust have been demonstrated in earlier studies (Castella et al., 2005b, Robinson et al., 2007).

There are variations in the way RPG has been developed. In agri-food supply chains, RPG was developed mainly using case studies in a specific context (e.g., Worrapimphong et al. (2010), d'Aquino and Bah (2014), Salvini et al. (2016)). In these examples, researchers use RPG to develop ABS models but others have designed their RPG based on an existing ABS model (e.g., Castella et al. (2005b), Joffre et al. (2015), Amadou et al. (2018)). In these case, the objective is to validate the decision rules in ABS (i.e., a decision rule is considered to be true if it is replicated by the player in the game). Some researchers have developed an RPG based on a theoretical perspective or by enriching a laboratory experiment with elements and information from the real world (e.g., Meijer et al. (2006), Tykhonov et al. (2008)).

The value of parameters used in an RPG (e.g., the crop yield or livestock reproduction patterns) is usually determined based on the agreement or consultation with the stakeholders (e.g., Castella et al. (2005b), d'Aquino and Bah (2014), Joffre et al. (2015), Salvini et al. (2016)). Some researchers have also utilised empirical data from field research to parameterise the RPG (e.g., Worrapimphong et al. (2010)).

To achieve the aims mentioned above, the selection of players is important in RPG data collection. The players in an RPG session are usually selected from the relevant stakeholders in the case study site. Non-probabilistic sampling methods (e.g., voluntary sampling (d'Aquino and Bah, 2014)) are usually employed to select the RPG players. If the objective of an RPG is to test a theory then stakeholders from various sectors and organizations can be invited to play the game. This enable the researchers to compare the stakeholders' behaviours and generalise the theory being tested. For example, the Trust and Tracing game (Meijer et al., 2006) which used the food supply chain context was played by university students, government officials, farmers and even primary school pupils.

Owing to the richness of interactions between players during the game, the quality of information obtained during the RPG is often considered to rely heavily on researcher skill (Robinson et al., 2007). Data regarding the players' decision rules are commonly collected through observation (e.g., d'Aquino and Bah (2014)), discussion during debriefing (e.g., Castella et al. (2005b)) and post-game interviews (e.g., Papazian et al. (2017)). The data obtained are usually analysed qualitatively and a decision rule is considered to be plausible if it is agreed by the players as a common practice, or if the rule is repeated by several players independently (e.g. Salvini et al. (2016)). This analysis usually produces a rule-based decision model (Robinson et al., 2007).

Until recently, the use of quantitative analysis to extract decision rules from RPG was rare and some authors even consider it to be difficult (Salvini et al., 2016). Furthermore, until recently, the quantitative techniques used to elicit decision rules from RPG experiment are limited to the descriptive statistics method. For example, Joffre et al. (2015) studied the preferred shrimp production system under various incentives. They counted the frequency of production system (i.e., intensive, extensive and integrated system) selected by the players under various incentives. These frequencies were the used to calculate the probability of agents to select a certain production system in the simulation.

Relying on discussions during debriefing and post-game interviews to elicit stakeholder's decision rules is not a bias-free process. One of the biases that can arise is the confirmation bias. This may happen when researchers have developed their own understanding and viewpoints about the system and its future trends (Voinov and Bousquet, 2010). This bias may lead to the observer-expectancy effect in which the researcher subconsciously influences the RPG players during the debriefing. This process is also prone to biases that come from the group process. For example, the less prominent group members tend to align their opinion to the group leader or to the majority (Kunsch et al., 2009).

4.2.2 The use and the validation of RPG calibrated ABS models

In general, a simulation model aims to understand a problem entity. A problem entity can be something realistic (e.g., a real system or phenomenon, an ongoing policy) or something that is not happening currently (e.g., a proposed system, idea, or a planned policy) (Sargent, 2013). Gilbert (2004) classifies realistic models as ABS models that aim to incorporate realistic mechanisms only while artificial models are those that also aim to incorporate unreal mechanisms. Using this classification, most ABS models that were calibrated using RPG are artificial models. The RPG was used to elicit realistic and artificial decision rules from the stakeholders as well as plausible new ideas and policies to influence the system (e.g., Castella et al. (2005b), d'Aquino and Bah (2014)). In other words, the researchers can encourage the players to perform actions they would not dare to do in the real world. The calibrated ABS models are mainly used to explore the consequences of these decision rules and ideas on system performance. For example, Campo et al. (2009) use ABS to project different amounts of resources that can be obtained by villagers under a variety of forest management strategies. These strategies were invented by the players through an RPG workshop. By knowing the potential impacts of each strategy, the stakeholder can discuss and negotiate the strategy to be adopted in the future. In this case, the objective of the ABS model is not to produce estimates of real system performance and the precise matching between simulation outputs and real-world data is less important.

If the problem entity being simulated is not realistic then, in many cases, the data to validate an artificial model are not available. The absence of real data inhibits the use of operational validation but the validity of the final simulation model can still be assessed by using conceptual model validation, theoretical validation and computerised model verification (Sargent, 2013). This process is usually carried out by presenting the final model to the stakeholders (e.g., d'Aquino and Bah (2014), Joffre et al. (2015), Papazian et al. (2017), Amadou et al. (2018)). In these examples, the final ABS model is considered as valid (conceptually and theoretically) when its specification and outputs can be rationally accepted by the researchers and stakeholders. This validation process is not free from bias because, in many cases, the RPG players also are also involved in judging the validity of the final model (Robinson et al., 2007).

4.2.3 The purposes of the RPG extension

From this literature review we conclude that most of the existing studies use RPG to elicit as many ideas and decision rules as possible in the particular RPG setting. This information is used to calibrate an artificial ABS model, which is in turn used to project the consequences of these ideas and possible decision rules. By observing the ABS outputs, the stakeholders can discuss potential policies to be adopted in the future. For this purpose, both artificial and realistic behaviours are important. Hence, it is not necessary to filter the realistic behaviour from the artificial behaviours.

However, such a filtering process becomes important when a researcher is tasked to develop a realistic model and aim to produce estimates of the real system behaviours. This filtering process is also important for policymaking purposes because, without being able to justify the representativeness of the simulation in describing the real condition, there is no guarantee that correct intervention are identified.

Our aim is to demonstrate that RPG is also beneficial in developing a realistic model in which the agent's decision rules are more likely resemble to what the stakeholders really do. However, several steps must be added to the popular RPG data collection methodology in order to do this. Firstly, we seek to improve the correspondence between the RPG and reality in order to increase the likelihood for the stakeholders to show their real decision rules (Rungtusanatham et al., 2011, Cowlrick et al., 2011). Second, we use experiment design and incorporate quantitative observation to reduce the reliance on the debriefing and post-game interview when eliciting the stakeholder's decision rules. Finally, we evaluate the match between the simulation outputs and the real data (i.e., operational validity) to highlight the decision rules that are more likely to be realistic.

4.2.4 ABS model of dairy supply chain in West Java and the hypothesised decision rules

The base model that will be compared against RPG calibrated models in this paper is developed based on ABS models reported in the literature. The detail of the model, which is an ABS model of a dairy supply chain in West Java, has been described in Chapter 3. This ABS model aims to be empirical, and because the access of data to the other supply chain actors was unavailable, it focuses on modelling dyadic interactions between the farmers and the cooperative. The key agents in this model are the farmers, the cooperative and the physical environment. It simulates the dynamics of cow population, cattle population, milk production and the number of farmer households. Empirical data for the base model initiation was collected through a survey and secondary sources. The agents' decision rules in the base model were derived from the literature. The decision rules that will be used as the null hypotheses to be tested in this paper are:

1. *Forage collection decision*: When collecting forage, farmers are assumed to prioritize the location with the highest forage level (Martin et al., 2016) and the closest location to their house.
2. *Cow selling decision*: In accordance with Gross et al. (2006), Lie and Rich (2016), and Lie et al. (2017), the farmers are assumed to sell their cows when the forage is less available. This is done since the cattle mortality rate will increase if the cows have insufficient forage. The number of cows being sold is proportional to the level of forage deficit they are experiencing.

3. *Selling priority*: When selling the cows it is assumed that the farmers will prioritise to the sale of the oldest animal first. As mentioned in Chapter 3, in addition to having more live weight, older cows are usually considered to be less productive.
4. *Cow buying decision*: In accordance with Gross et al. (2006), Lie and Rich (2016), and Lie et al. (2017), when the farmers can collect more forage than what is actually needed and they have sufficient money, then they start to consider buying more cows. The number of new cows a farmer is willing to buy is equal to the additional cows that can be fed using the excess forage.
5. *Cow trading partner*: In common to Boone et al. (2011), Rasch et al. (2016) and Rasch et al. (2017), the base model assumes that the farmers sell or buy their cows to an external agent outside the system and not to other farmers.

4.3 RPG development and data collection process

Figure 4.1 illustrates the methodology used in our study. Steps in the dotted box have been discussed in detail in Chapter 3. Section 4.3.1 describes the process to develop an RPG from the base model. During this process we seek to improve the correspondence between the RPG and the real world. The data collection process using the RPG is described in section 4.3.2. In this data collection process, we incorporate design of experiments and quantitative observation to reduce the bias caused by the reliance on debriefing and post-game interview.

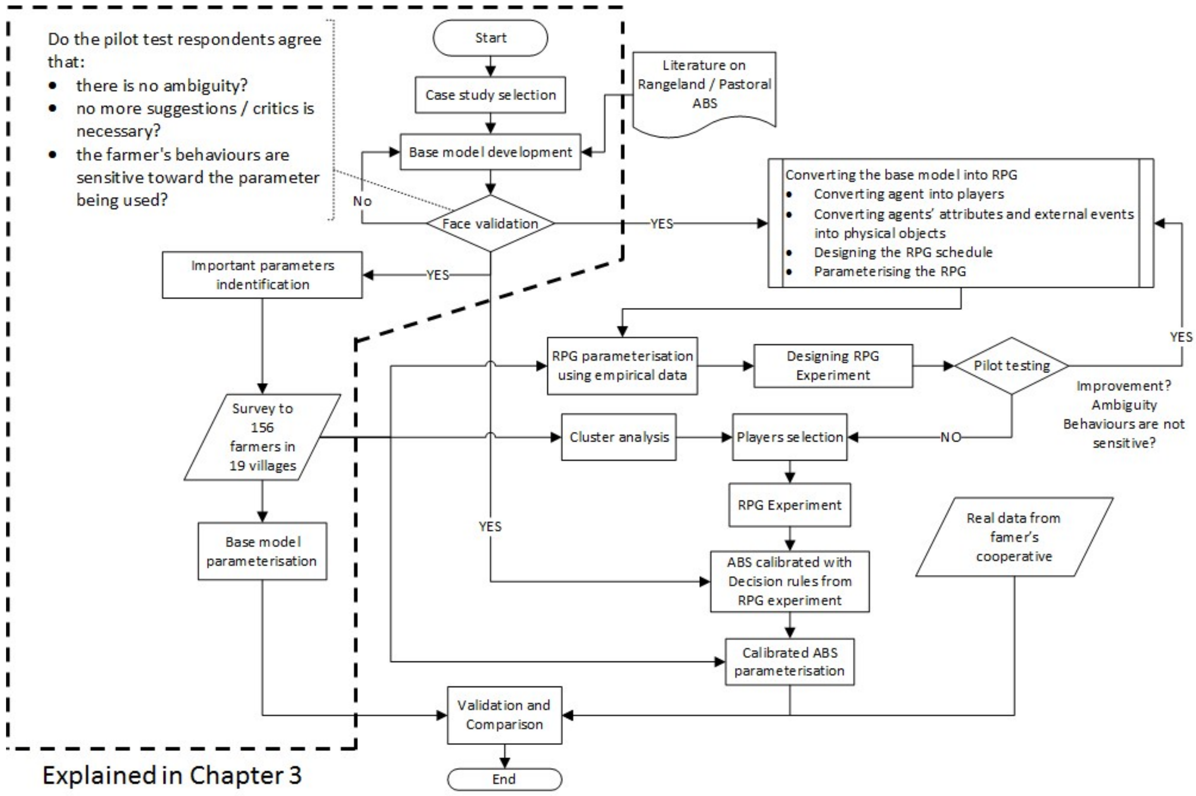

Figure 4. 1 Research methodology for extending RPG data collection method

4.3.1 RPG development process

Details describing how an RPG is designed to collect data for ABS model calibration and validation are rarely discussed in the literature. Hence, one of the contributions of this paper is to describe the RPG development and data collection process so that it can be discussed and debated to identify the best practice. It also benefits those who want to learn about how to design RPG in the context of ABS model calibration and validation. Our RPG design process consists of several steps, namely:

1. To define the objective of the RPG
2. To select appropriate game format
3. To decide RPG players whose behaviour will be observed to calibrate agents in the ABS model
4. To design RPG items that represent agents' attributes and external events.

5. To design the RPG schedule.
6. To estimate the RPG parameters.
7. To design the RPG experiments.
8. Pilot testing.

The first step in developing the RPG in this study is to clearly define its objectives. In this study the RPG aims to collect data useful to confirm or reject the decision rules used in the base model. Using this data we expect to obtain realistic decision rules so that our final model can estimate the real world data. This is different from the aim of RPG in the earlier studies which tend to explore new ideas regarding collective rules and management strategies.

In terms of the format, we design our RPG as a turn-based board game that comprises several rounds. A board game format was chosen because the farmers are more familiar with this type of game (e.g., chess and checkers) than computer games. In addition, the use of a board game also prevents disruption during the data collection process owing to, for example, an unreliable electricity supply.

Each round in our RPG represents one year. This time scaling is important to maintain the correspondence between the RPG, the simulation and the real world. This time scale was chosen because it is similar to the time scale used in our ABS outputs and in the real data for validation. In addition, the changes that occur in a shorter time scale may be too small to be substantial.

The third step is to define the RPG players whose decision rules will be observed. This step began by listing all the type of agents in the base model. As mentioned in section 4.2.4, there are three key agents in the base model, namely the farmers, the cooperative and the environment. The farmers and the cooperative are considered to be active (i.e., able to make decisions) and, hence, they were played by human players. The decisions to be calibrated belong to the farmer agents so the real farmers played these roles. The number of players was limited to four farmers to ensure the researcher/observer was not overloaded while the role of the cooperative was played by another member of the research team.

The passive agent, namely the land/forage available, was displayed on the game board. The case study site area is approximately 30,000 hectares. This area was represented by 48 cells on the game board, each representing 6.25 km^2 area in the real world (see Figure 4.2). As mentioned in Chapter 3, this resolution was chosen because the real data shows that farmers travel for at least 2.5 km per day in a round trip.

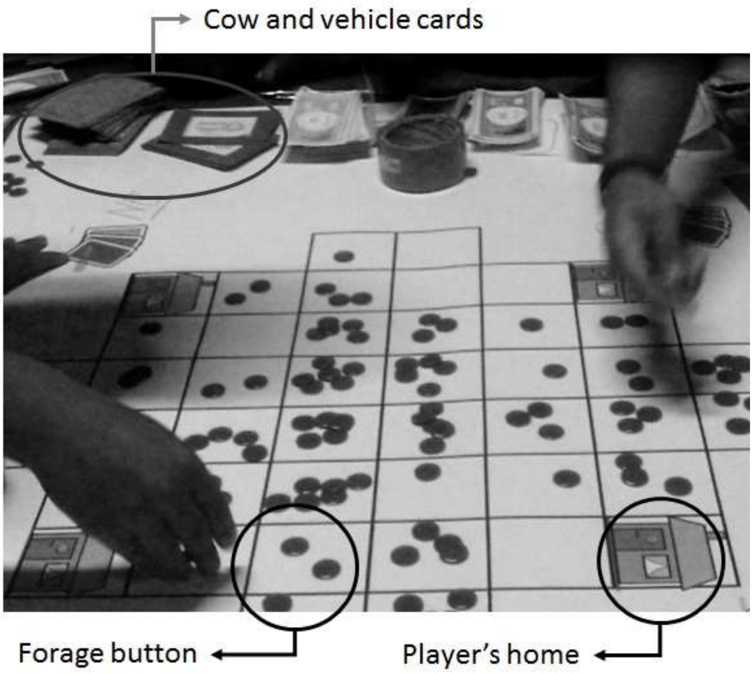

Figure 4. 2: The RPG game board

The fourth step in the RPG development is choosing items to represent the key attributes that influence players' decisions. For attributes that can change ownership in the real world (e.g., cows and forage), we designed RPG items that can be physically moved easily during the game. In our case, we used buttons to represent forage, cards to represent cows and vehicles, and a small notebook to record the players' money and milk production.

For attributes whose ownership does not normally change in the real world (e.g., the milk productivity of one cow cannot be transferred to another cow), we used items that cannot be moved during the game. For example, the milk productivity and productive lifetime of a cow are represented using information printed on the rear of the cow card. The milk productivity represents the amount of milk that can be produced by the given cow in a round while the cow's productive lifetime represents how many rounds it can produce milk for (see Figure 4.3.a). In this game, we assumed that a farmer's house cannot change ownership because it is used as a permanent base for the players. Hence, each house was represented as a home cell (see Figure 4.2).

We used cards to represent external events that may happen in the real world and may influence how decisions are made. These external events are (i) cattle reproduction; (ii) income from selling the bulls; (iii) living and other expenses that must be paid by the farmer household; and (iv) cow mortality. For example, when a player receives the card illustrated in Figure 4.3.d, then the given player can get one additional cow but must pay Rp.5000. The value of Rp.5000 represents the total of income from selling the bulls and other expenses. The sixth step explains in more detail about the process to determine the distribution of values used in these event cards.

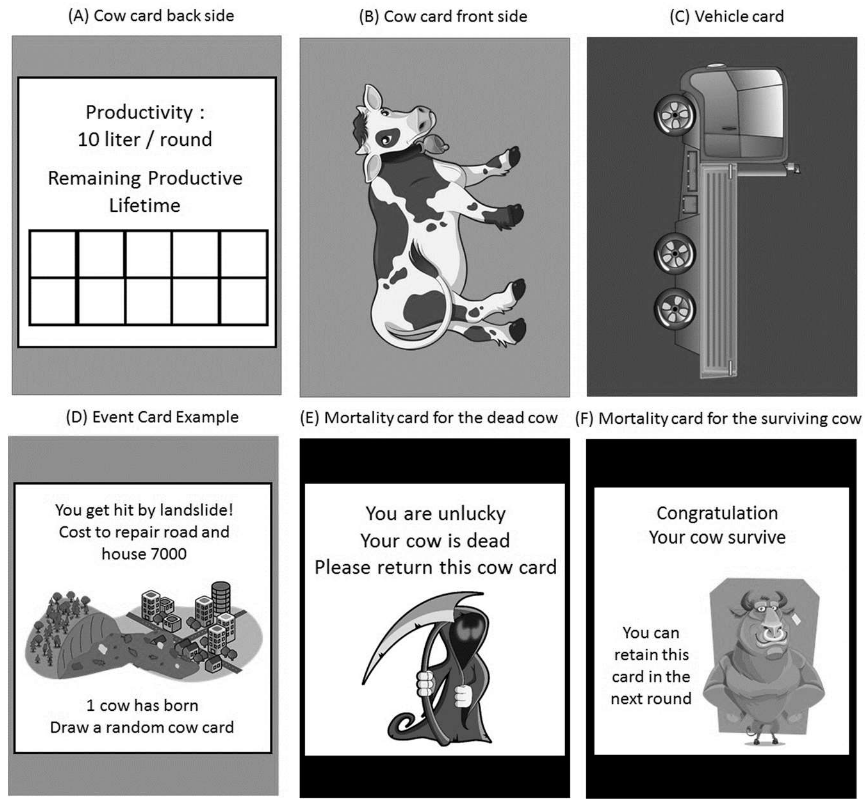

Figure 4. 3: Cow, vehicle, event and mortality cards translated to English from Bahasa Indonesia. All images used are taken from public domain

In the subsequent step we define the schedule of our RPG. It has been mentioned in Chapter 3 that the base model's schedule reflected the real farmers' daily activities and had been face validated by the domain experts in the case study site. Hence, the RPG schedule follows the base model's schedule to ensure it had a similarly strong correspondence to reality.

At the beginning of a round in the RPG, all players are given a chance to collect forage buttons from the game board by moving to a cell where the forage buttons are located (for simplicity, diagonal moves are forbidden). In reality, a farmer has 8 hours per day to collect forage. However, farmers do not always use all the time available to them on each day. The actual amount of time that is used to collect the forage is affected by the weather conditions, the farmers' physical condition and possibility of vehicle breakdown. To accommodate these uncertainties, the maximum number of moves a player can make is determined by the value obtained from rolling a dice. The dice's face value represents the actual working hours a player can use in the given round. A player does not have to use all of their moves. However, the unused moves cannot be carried forward to the next round. The move is always started from the player's home. If the player decides to take forage buttons from a certain cell then the player must return and start the next move from his home. The number of forage buttons that can be taken from a certain cell is constrained by the number of vehicles owned by the player. To remove bias, the player's turn in each round is shuffled by using permutations.

Following the collection of forage, the players must allocate the forage buttons to their cows (one forage button for each cow). If a player owns more forage buttons than cow cards then the remaining forage buttons can be carried forward to the next. Conversely, if a player own fewer forage buttons than cow cards then the player can negotiate with other players to obtain more forage buttons. The player also has an option to sell some of his cow cards to other players or to the market. If the cow card is sold to another player then the buyer must allocate a forage button to the newly bought card. Players are free to determine their negotiation strategies and price when trading forage buttons and cow cards. For example, the seller may choose to reveal their cow's productivity or not (this information is on the rear of the cow card and is only visible to the cow's owner).

Players then add up the total of their milk production and sell it to the cooperative player to earn money (only the cow cards that have been allocated a forage button can produce milk). The remaining productive lifetime on each cow card is decreased by one by ticking one of the productive lifetime boxes. At this stage, for each cow card without a forage button, the owner must draw a card from the mortality card deck. The mortality cards determine whether the cow will survive to the next round (see Figure 4.3.e and 4.3.f).

All players then draw one event card and follow the instructions written on the drawn card. For example, the card may instruct the player to pay for an unplanned expense or get a new cow card. The drawn card must be returned to the deck after the player has read it to keep the probability of each event constant during the game. This deck is also reshuffled at the beginning of each round.

All players are then given an opportunity to sell some of their cow cards (e.g., players may decide to sell the unproductive cows). They can also make investment or disinvestment decisions by selling or buying cow and/or vehicle cards. They can trade these cards with the market or with other players. As with the forage trading, a player is free decide his negotiation strategies and price. When a player decides to buy a new cow from the market, a new cow card is drawn randomly from the cow card deck.

Finally, the round ends by replacing the forage buttons that have been taken from the game board to represent forage regrowth. The cycle continues until the game is stopped by the facilitator. At the end of the game, the monetary value of all players' assets is calculated. These assets include money, cow cards and vehicle cards. The player with the highest asset value wins the RPG.

The next step that was carried out to develop the RPG in this study is to estimate the game parameters. Rather than relying on agreement and consultation with the stakeholders to parametrise an RPG (e.g., Castella et al. (2005b), d'Aquino and Bah (2014), Joffre et al. (2015), Salvini et al. (2016)), we matched the RPG's parameters with empirical data. The process to parameterise the RPG using empirical data is summarised in Table 4.1.

Table 4. 1 Summary of RPG's parameters

Parameter	Estimation	Explanation
Size of a cell	2.5 km x 2.5 km	The reason for selecting this resolution has been discussed previously
Maximum number of moves by a player	12 cells per round	By considering the uncertainty of working hours utilisation explained in section 3.1.4 farmers can travel up to 30 km per day; therefore, 12-sided dice is used.
Amount of forage in one button	10.95 tonnes	On average, a cow needs 10.95 tonnes of forage per year to stay healthy.
Number of buttons	27 to 49 buttons	The total forage production on the case study's site is estimated between 303.23 and 545.82 tonnes.
Capacity in a vehicle card	2 to 4 buttons	A farmer household is able to transport between 22.95 and 43.98 tonnes per year
Distribution of cow's productivities in cow cards (Prod)	Prod (5) = 3 cards, Prod (10) = 7 cards, Prod (15) = 51 cards, Prod (20) = 79 cards, Prod (25) = 51 cards, Prod (30) = 7 cards and Prod (35) = 3 cards	The empirical distribution of cow productivity is $\mathcal{N}(20.44; 4.33^2)$. We divide this distribution into 7 intervals and use the midpoint of each interval. The number of cow cards with a certain level of is calculated accordingly (e.g. $P(Productivity = 10) = 3.5\%$ of all cow cards)
Distribution of cow's productive lifetimes in cow cards (PL)	PL (5) = 67 cards, PL (7) = 67 cards, PL (10) = 67 cards	It is assumed to be uniformly distributed $U(4,10)$. We divide this distribution into 3 intervals and use the midpoint of each interval. The number of cow cards with a certain productivity lifetime is calculated accordingly (e.g. $P(Productivity\ lifetime = 10) = 33.33\%$ of all cow cards)
Number of events related to the birth of a new cow in event cards (birth)	Birth (0) = 149 cards, Birth (1) = 42 cards, Birth (2) = 6 cards, out of 197 event cards	The survey data shows that the birth of new cows has a Poisson distribution with mean of 0.29 per year.
Distribution of events related to cash flows (CF) in event cards	CF (-16,000) = 1 cards, CF (-13,000) = 5 cards, CF (-10, 000) = 24 cards, CF (-7,000) = 56 cards, CF (-4,000) = 64 cards, CF (-1.000) = 37 cards, CF (2,000) = 10 cards, out of 197 cards.	The typical additional income comes from selling bulls. The distribution of money received from selling bulls per year is $\mathcal{N}(977.21; 1{,}604.03^2)$. The distribution of expenses is $\mathcal{N}(7{,}370.02; 3{,}112.47^2)$ annually. By combining the two distributions altogether, the distribution of cash flow is $\mathcal{N}(-6{,}392.81; 3{,}501.48^2)$. We divide this distribution into seven intervals with mid points of -16,000; -13,000; -10,000; -7,000; -4,000; -1,000 and 2,000. The units are in Rp.10,000.
Milk selling price	Rp. 120 per litre	The average milk selling price from the survey is Rp. 4,480 per litre. A cow can produce milk for nine month in a year. Thus, in a year the selling price of one litre of milk is Rp. 1,209,600. To simplify this value we round it to the nearest hundred thousand and divide it by 10,000
Market price of a vehicle	Rp. 2,000	This price is paid when a farmer buy a vehicle from the market. If a farmer buys it from another farmer, they can negotiate the price. Based on survey data, the average vehicle price in the area is Rp. 22,600,000. To simplify this value we round it to the nearest ten millions and divide it by 10,000
Market price of a cow	Rp. 1,500	This price is paid when a farmer buy a cow from the market. If a farmer buys it from another farmer, they can negotiate the price. Based on survey data, the average cow price in the area is Rp. 15,670,000 per head. To simplify this value we round it to the nearest millions and divide it by 10,000

In our RPG, we were interested in observing the players' general strategies that they use in any given situations rather than replicating the players' real-world attributes (e.g., the number of cows they have) in the game. This approach is common in RPG (Castella et al., 2005b) so at the beginning of the game, each player receive the same two vehicle cards and four cow cards ((i) one card with productivity of 15 litre and productive lifetimes of 5 years; (ii) one card with productivity of 15 litre and productive lifetimes of 10 years; (iii) one card with productivity of 25 litre and productive lifetimes of 5 years; and (vi) one card with productivity of 25 litre and productive lifetimes of 10 years).

The next step is to design the experiments to be carried out using the RPG. The base model's decision rules being confirmed using RPG are influenced by forage availability and cattle mortality. Thus, we designed our experiments based on the two factors. We varied the values of each factor with a collectively-exhaustive range. This is to avoid bias due to extrapolation, that may happen when an intermediate parameter value (e.g., an output of a regression that is used as an input to another regression) in the simulation goes beyond the range of data obtained from the experiment. We selected three levels of forage availability and cow mortality (giving us a total of nine experiments). In our case, each RPG experiment takes on average 2.5 hours including preparation, role playing session and debriefing. Owing to time constraints, we conducted five main experiments (Experiments 1 to 5 in Table 4.2).

Table 4. 2 The RPG experiment sets

		Experiments			Additional Experiment after pilot
Main treatments		Probability of cow's mortality			
		Low (30%)	Medium (50%)	High (70%)	Very High (90%)
Forage Availability	Low (24 Forage Buttons)		Experiment 4		
	Medium (36 Forage Buttons)	Experiment 1	Experiment 2	Experiment 3	Experiment 6
	High (48 Forage Buttons)		Experiment 5		

Finally, we conducted a pilot test involving lecturers and students from the animal husbandry department of a local university and an experienced farmer. The objectives were: (i) to ensure that the RPG's guidance, rules and schedule were correct and easily understandable for the players; (ii) to ensure that the experiment sets being used had sufficient contrast so that the players may exhibit different behaviours; and (iii) to identify necessary improvement in the RPG design. All experiments 1 to 5 were tested and a pilot test respondent played in more than one experiment set.

Changes based on the feedback from the pilot include adjusting the terminologies in the RPG scripts to match the terminologies commonly used by the local farmers, replacing the notebook with toy money to reduce the mental burden in recording each transaction (and make the game more fun), changing the instruction written on the event cards to a short story (about the event) that is commonly experienced by the farmers (to make the RPG more realistic), and adding one additional experiment (experiment 6) to increase the contrast among experiment sets.

4.3.2 RPG data collection process

The first and important step in the data collection process is to select suitable players for the RPG experiments. The selection process aims to maximise the diversity of the players and to mimic the agent proportion in reality as closely as possible. We used the agglomerative hierarchical and K-means clustering techniques to group the respondents mentioned in Chapter 3 into clusters based on their demographic, socio-economic and business scale profile. Both approaches consistently showed that farmers can be grouped based on their experience and business scale into three clusters (see Figure 4.4). A farmer was considered to be highly experienced if he/she has been running dairy business for more than 20 years. If the farmer owns seven or more cows, and more than 600 m^2 of land, then the farmer's business scale was considered to be big.

The first cluster in Figure 4.4 was formed by three experienced farmers (2%) whose farming businesses are big. The second cluster was formed by experienced farmers whose farming businesses are small. There were 20 farmers in this category (13%). The majority of farmers (85%) formed the final cluster. In this cluster, the farmers are less experienced and run small-scale farms.

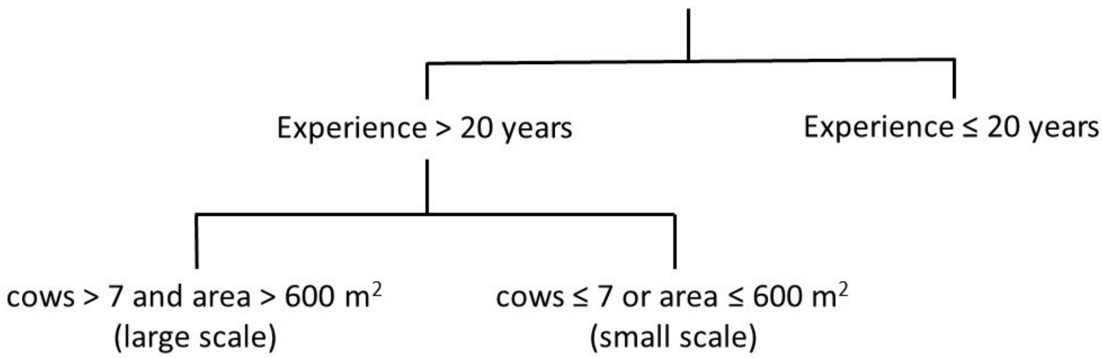

Figure 4. 4 Clustering result to select players for RPG experiments

Of the 19 villages in the case study site, we selected three villages in which there are farmers from clusters 1, 2 and 3. There are very few farmers in cluster 1 (3 people) so all of them were invited to participate in the RPG. Players representing cluster 2 and cluster 3 were randomly selected from the farmer data base in each village. Table 4.3 describes the composition of players in each RPG experiment.

Table 4. 3 Player composition in the RPG experiments

Experiment	Cluster 1	Cluster 2	Cluster 3	Location
1	1 player	1 player	2 players	Warnasari
2	0 player	1 player	3 players	
3	1 player	1 player	2 players	Sukamenak
4	1 player	1 player	2 players	Pangalengan
5	0 player	0 player	4 players	Sukamenak
6	0 player	1 player	3 players	Pangalengan
Players proportion	12.5%	20.8%	66.7%	

The next step is to carry out the RPG experiment and observation process. The RPG experiments were held in the evening to allow all the players to complete their daily tasks and evening time is usually reserved for socializing or holding meetings in the villages. Consequently, the players could concentrate fully on the game. All players formally consented to the recording of their actions via the observation table and video recorder.

In each experiment, a research team member played the role of the cooperative, as well as the market that trades cows and vehicle cards. Every two players were accompanied by one researcher who helped them in understanding the RPG rules and assisted them in organising the cards and money they own. However, the research team could not intervene the players' decision-making. Two other researchers took observer roles and recorded every decision made by the players.

The RPG began with an explanation of the purpose of the games and the equipment, rules and game schedules. Before the actual RPG sessions, the players played two trial rounds to familiarise themselves with the RPG. One experiment lasted for 10 rounds or until a one hour time limit was reached. Every player’s decision was considered as a decision situation and recorded as one data point. Hence our observation approach was similar to that used in other experimental gaming studies (e.g., Moffat and Medhurst (2009)).

To approximate the decision rule used by the players, in each round we recorded the dice value obtained by each player. Based on the dice value obtained by a player, we identify the cells that can be reached by the player and the number of forage buttons available on those cells. We then recorded which cell was selected by the player, how many forage buttons that can be taken by the players (taking into account the player's transport capacity and the number of forage buttons available in the selected cell) and the number of forage buttons that were actually collected by the given player. We then calculated whether the player has sufficient forage buttons or not. When the player does not obtain sufficient forage, we recorded whether the player was trying to buy or borrow forage from other players. If forage transactions occur, we recorded the amount of forage being transacted, the characteristics of players who are willing to sell or lend some of his/her forage button, and the agreed price. We then recorded how the players allocate their forage buttons to their cows, namely the characteristics of the cows who were prioritised to get the forage button. When a player experienced forage deficit, we recorded whether or not the particular player wanted to sell his/her cows and to whom he/she sells the animal, and the price agreed. Finally, we recorded the event cards obtained by the players and the changes in the players' money, cows and vehicles. This record is used to approximate the decision rules used by the players. This approximation was later confirmed to the players during the debriefing.

The RPG experiment was concluded with a debriefing session that began by discussing how well the players understood the RPG process. They were then asked to assess the similarity between the RPG and the real world and indicate whether they had ever encountered similar situations in their real experience. The observed decisions were then confirmed to the players. They were asked to describe the decision rules they used. By having anchors based on the observation, we reduced the reliance on the players' narration and potential biases.

We then asked whether those decision rules had been actually used in real world. If not, we asked about applying those rules in reality and the conditions that would drive them to use those rules. Finally, we elicited their perceptions toward the RPG and possible improvements for the RPG implementation. Bearing in mind that they have participated in similar studies earlier (e.g., the agricultural census and our own earlier survey), we also asked them to compare the usefulness of RPG to other data collection techniques.

4.4 Findings from the RPG data collection

We cross-examined the decision rules recorded by the researchers during the RPG and debriefing with the video recording to ensure that the validity of decision rules (i.e., ensure that what the players say during the debriefing, the researchers' observations and the video recording were consistent). After the debriefing, we excluded the data points collected from two players because they told us that they could not follow the RPG process. In total, 151 data points were retained for further analysis. In the following sections, the findings related to the hypothesised decision rules and the players' perceptions and experiences during the RPG are explained.

4.4.1 Forage collection decision

Players make the forage collection decision rule when they decide on how to collect the forage. After a player throws the dice, we recorded the position of cells with the most forage button and the closest that can be reached by that particular player. If the player follows the decision rule mentioned in the literature, the player will select one of these cells. If the player chooses another cell, we recorded the position of the selected cell, the number of steps used by the player to reach that cell, the number of forage buttons that can be taken by the player (depend on the player's transport capacity) and the number of forage buttons that is actually collected by the player. Based on this record, we can approximate the player's decision rule, which was then confirmed in the debriefing.

We have found that players choose the decision rule reported by the literature (see Section 4.2) in 45% of the instances (note that this does not mean 45% of players use this rule). However, the RPG also revealed two other variants:

i. From all cells that can be reached, the players choose a cell in which the number of forage buttons is greater than or equal to their transport capacity (not necessary the cell with the highest number of forage button) and the nearest to their home (52% of the instances).
ii. For all the cells that can be reached, the players choose the nearest cell regardless of the number of forage buttons on it (3% of the instances).

All three variants show the preference for a location close to home. This is a rational decision because farmers want to minimize their travelling time. This was confirmed during the debriefing. However, the debriefing revealed another local cultural reason. A location that is far from home is likely to be too close to other farmers' homes and, in the Indonesian context, collecting forage in this location is considered to be disrespectful and can trigger conflict because forage is a common resource. Hence, there is a social factor comes to play when choosing a location closer to home. This illustrates the value of RPG in uncovering the motives behind a decision. For example, a policy intervention that assumes farmers make the decision purely based on cost minimisation may result in unintended consequences such as inter-farmer conflict.

4.4.2 Cow selling decision

The base model used a literature-based assumption that farmers will sell their cows when forage is less available. From our discussion during the pilot, we believed that cow mortality was the principle reason rather than lack of forage (i.e., a lack of forage is not good for the cow's health and productivity). Consequently, we used the forage level scenario, the forage buttons deficit and the mortality rate scenario collected during the RPG (no significant correlation among the three independent variables) to estimate the probability that a player will sell their cow cards.

We developed a logistic regression model and the final model (equation 4.1) is significant at the 5% significance level. The Nagelkerke pseudo R^2 shows that the model can predict 53.7% of respondents' responses. The parameter estimates show that the forage level scenario and the forage button deficit do not affect the players' decisions to sell their cows (with significance value of 0.591 and 0.253 respectively). However, the players' responses were more sensitive toward the cattle mortality rate (significance value < 5%). As the likelihood for a cow to become sick or die increases, the probability that a player will sell his/her cow also increases.

$$\ln \frac{P_{sell}}{1 - P_{sell}} = 7.953 Mortality - 1.109 \quad (4.1)$$

This result confirms our belief. The players during the debriefing explained that farmers tend to keep their cows even when experiencing forage deficit provided that local veterinarian considers their cows to be healthy during his weekly visit. In this RPG, the veterinarian's assessment is represented by the cow mortality scenario being used. This shows that their main concern is the health of their cows. Hence, our RPG has uncovered a different cow selling decision rule from the base model.

4.4.3 Selling priority decision

The base model assumes that farmers choose to sell the cow card with the smallest remaining productive lifetime. The RPG data support this assumption. For 83% of the instances, the players choose to sell the cow with the lowest remaining productive lifetime. The remaining observed behaviour during the RPG can be explained by the players' strategy to win the game by selling their cows towards the end of the game. Hence, it does not reflect their real-world practice.

4.4.4 Cow buying and upgrading decision

Through the RPG experiment, we found two reasons why players buy new cow cards namely, to increase the number of cow card they own (83% of the instances) or to upgrade the quality of their cows (17% of the instances).

The base model assumes that farmers buy new cows when they have enough money and a forage surplus. The RPG result shows that players use this decision rule in 86% of the instances. The debriefing confirmed this. However, the RPG also uncovered two other variants:

i. Occasionally (10% of the instances), even when experiencing a forage button deficit, some players buy cows to replace those who die owing to the lack of forage as long as they have sufficient money to buy new cow cards.
ii. On the rare occasions when they have extra forage but not enough money, they still buy cows by borrowing some money from another player (4 % of the instances).

If the buying decision is made to upgrade their current cows (i.e., replace those with low milk productivity or are closer to the end of their productivity lifetime), the players always follow the base model's decision rule. During the debriefing, the players told us that this decision is made in order to increase their profitability and exists in the real-world. Even though there is uncertainty regarding the quality of cattle sold by the cattle traders, this strategy is considered to be effective in increasing milk productivity because it does not require the farmers to collect additional forage or to increase their pen capacity.

We included the upgrading decision rule in the model calibration and examined whether it can improve the model validity. In this new decision rule, with probability of 17%, an agent who has sufficient forage and money will sell its cow whose productivity is lower than the average milk productivity of the cow population and buy a new one from cattle trader.

4.4.5 Cow trading partner

In line with the discussion in Chapter 3 the model in this study assumes that the farmers mainly trade their cattle with a cattle trader. This is confirmed by the RPG. All cow card transactions are made with the cattle trader. The debriefing uncovered the underlying reason behind this decision rule.

Firstly, finding a farmer who is willing to sell a cow with high productivity is unlikely. Therefore, when a farmer is trying to sell a cow to another farmer, the prospective buyer tends to doubt the quality of the cow. Furthermore, it is easy to temporarily increase a cow's milk productivity by stopping milking the cow for several days prior to the transaction. To avoid this, the buyer needs to spend several days observing directly the actual productivity of the cow which may be prohibitive for a small-holder farmer.

Secondly, when selling a cow to another farmer, a seller tends to ask for a higher price than when selling to a cattle trader. A cattle trader does not intend to make profit from the cows' milk, they usually sell the cows as meat. The price offered by the cattle traders follows the meat market price fluctuation and is also influenced by the profit margin desired by the cattle traders. The milk productivity of the cow which is being traded, does not become an incentive for cattle traders to offer a higher price. However, when trading with another farmer, the seller knows that the buyer will make a profit from the milk produced by the cow. Therefore, if the seller can prove that the cow's milk productivity is high then the seller can ask a higher price than the price offered by cattle traders.

4.4.6 Players' experiences

Most players found this RPG to be very interesting and compelling, so much so that they expressed a desire to retain the RPG equipment when the experiments were completed. According to the players, the RPG is representative of their real life experience (the similarity between the game and the reality is between 70%-80%). This suggests a relatively high level of validity for the RPG as a data collection instrument.

The players felt that the RPG gave them more freedom to express their feelings and behaviours. For example, when participating in a survey, they felt that sometimes interpreting questions and options in a questionnaire can quite cumbersome, especially if the questionnaire uses different terminology to that which they are accustomed to their daily life. Furthermore, during interviews, they found it difficult to explain their perceptions in a way that could be properly understood by the interviewer. According to them, playing the RPG was fun and less boring and less intimidating than responding to surveys or interviews. They admitted that boredom, particularly, often led them to provide the quickest or easiest response rather than describe what they actually feel or do.

4.5 ABS Calibration and Simulation Experiment

In this section, we aim to illustrate how the benefits of RPG can be quantified. To achieve this, we used the decision rules found using our RPG to calibrate the base model. The base model itself was developed using NetLogo (Wilensky, 1999). The RPG has uncovered several decision rules that are different from the rules used in the base model, namely: the forage collection decision (two variants); the selling decision (one variant); and the upgrading decision (one variant). Among the two variants of the forage collection decision, we choose to implement the first variant (see Section 4.4.1) because it is the more frequently occurring. The model was designed so that we can set options for the three decision rules, (i.e., whether we use the rule from the base model or that from the RPG). Hence, there are eight possible combinations of model calibration (see Table 4.4). In Table 4.4 the base label indicates that the model uses decision rule from the base model. Conversely RPG label indicates that the particular decision rule is calibrated using RPG data.

Table 4. 4 Decision rules in models calibrated with RPG

	M_0	M_{RPG1}	M_{RPG2}	M_{RPG3}	M_{RPG4}	M_{RPG5}	M_{RPG6}	M_{RPG7}
Forage collection	Base	RPG	Base	Base	RPG	RPG	Base	RPG
Selling	Base	Base	RPG	Base	RPG	Base	RPG	RPG
Upgrading	Base	Base	Base	RPG	Base	RPG	RPG	RPG

We carried out simulation experiments to understand the impact of these decision rules on the operational validity of the model outputs. Based on the generative sufficiency principle (Epstein, 2006), a decision rule can be considered as plausible if it can grow macro patterns that are in agreement with the reality. This agreement can be measured by evaluating the fitness between the simulation outputs and the real data (Thorngate and Edmonds, 2013). We also conducted experiments to identify decision rules that might be preferable for the real-world actors.

The initial population of farmers in the experiment was set based on the real data in January 2010. The random number seed was controlled to ensure fair comparison (i.e., the difference in the model outputs is solely caused by the different decision rules used by agents). Each model was run for five simulation years (2010 - 2014) and replicated 30 times.

4.5.1 Impacts on model outputs relative to the base model

The first step in quantifying the benefit is by evaluating whether the RPG's decision rules produce model outputs that are significantly different from the outputs of the base model. For this purpose the difference between calibrated model and the base model outputs (i.e., $Error_i = Output_i^{calibrated} - Output_i^{base}$ where i = 2011 … 2015) was measured. The impact of each calibrated model was measured by using the mean of difference for the five-year period (i.e., $ME = \sum_{2011}^{2015} Error_i/5$). A high $|\overline{ME}|$ value indicates that the output of a calibrated model (M_{RPG1}, M_{RPG2}, … or M_{RPG7}) is different from the base model (M_0). Table 4.5 shows the average ($\overline{ME}$) and standard deviation (S_{ME}) of outputs from 30 replications. It also shows the significance value from the paired t-tests between each calibrated model and the base model (at the 95% confidence interval).

Table 4. 5 the descriptive statistics and t-result of empirical ABS deviation from the base model (* indicates significant at 5%)

Model Name	Cattle Population		Cow Population		Daily Milk Production		Farmer Households	
	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.
Empirical models with one RPG decision rule								
M_{RPG1}	(9456.9, 10687)	0.0*	(8611.3, 10048.7)	0.0*	(31360.9, 29518.1)	0.0*	(569.3, 249.9)	0.0*
M_{RPG2}	(131.3, 492.5)	0.2	(3.3, 348)	0.9	(27732.5, 11154.5)	0.0*	(-495.4, 346.1)	0.0*
M_{RPG3}	(93.7, 253.1)	0.1	(79.2, 227.8)	0.1	(102.3, 937.7)	0.6	(2.9, 14.6)	0.3
Empirical models with two RPG decision rules								
M_{RPG4}	(8815, 9353.3)	0.0*	(6962.7, 6987.2)	0.0*	(51652.6, 32512.3)	0.0*	(-3.4, 792)	0.9
M_{RPG5}	(9448.6, 10684.2)	0.0*	(8606.5, 10039)	0.0*	(31193.4, 29327)	0.0*	(565.4, 248.5)	0.0*
M_{RPG6}	(188.3, 496.3)	0.0*	(49, 396.6)	0.5	(28992.3, 9486.1)	0.0*	(-490.3, 337.7)	0.0*
Empirical model with three RPG decision rules								
M_{RPG7}	(8825, 9354.2)	0.00*	(6973.6, 6982.5)	0.0*	(53116.8, 31089.1)	0.0*	(4.8, 783.1)	0.9

Table 4.5 shows that all outputs produced by the model with forage collection decision rule (M_{RPG1}) were significantly different from the outputs from the base model and it had the highest $|\overline{ME}|$. The model with the RPG's upgrading decision rule (M_{RPG3}) was the only model that produced similar outputs to the base model while the remaining model produced at least one output that was significantly different from the output of the base model. Assuming that the base model has a good operational validity for all outputs, this analysis illustrates that M_{RPG2} was likely to improve the overall validity of the model because we have the evidence of the micro-behaviour validity from the RPG. This was to be confirmed in the next step.

4.5.2 Impacts on operational validity

The following experiments investigated whether adding the decision rules obtained from the RPG data could produce outputs that agreed with the real data. Two techniques were used for the validation process, namely regression analysis and mean error estimation. The former shows how good the model outputs represent the trends in the real data while the later measures the magnitude of model output deviations from the real data. These approaches correspond with the simulation validation techniques explained in Kleijnen (1995b). The real data on cattle population, cow population, milk production and the number of farmer households were obtained from the farmer cooperative (KPBS, 2016). These variables are considered to be very important and both the government and the cooperative record their statistics.

Table 4. 6 Cattle population, cow population, average daily milk production and the number of farmer households in Pangalengan West Java 2010-2016 (KPBS, 2016)

Year	Cattle population (head)	Cow population (head)	Average daily Production (litre)	Farmer Household
2010	21,322	21,083	159,333	5072
2011	21,438	20,960	136,694	4204
2012	22,366	22,073	138,904	3439
2013	16,173	16,080	97,476	3053
2014	13,415	13,399	84,207	2888
2015	12,563	12,555	76,372	2852

The regression method evaluates if the trend produced by the simulation agrees with the trend in real data. Table 4.7 summarizes the regression analysis results with the real data as the dependent variable and the mean of simulation outputs (from 30 replications) as the independent variable. The significance column (Sig) in Table 4.7 shows the significance of the ANOVA test and a lower significance value indicates a smaller probability that the relationship between the average simulation outputs and the real data occurs by chance (in other words, better model validity). A good model validity should show a positive regression coefficient value (*B*) because it indicates that the simulation outputs and the real data move in the same direction. The R^2 values show the proportion of variation in the real data that can be explained by the simulation outputs variation. A high R^2 value indicates a good fit.

Table 4. 7 Summary of regression analysis between the simulation outputs and the real data (* indicates significant at 5%)

Model Name	Cattle Population			Cow Population			Daily Milk Production			Farmer Households		
	Sig.	B	R^2	Sig.	B	R^2	Sig.	B	R^2	Sig.	B	R^2
M_0	0.00*	4.413	0.99	0.01*	5.333	0.94	0.02*	1.586	0.86	0.00*	0.523	0.98
Empirical models with one RPG decision rule												
M_{RPG1}	0.66	-1.113	0.08	0.234	-2.605	0.42	0.05	2.137	0.77	0.00*	0.752	0.98
M_{RPG2}	0.01*	4.017	0.92	0.01*	3.625	0.95	0.00*	4.367	0.97	0.00*	0.6	0.99
M_{RPG3}	0.00*	4.499	0.99	0.01*	5.515	0.95	0.02*	1.589	0.87	0.00*	0.523	0.98
Empirical models with two RPG decision rules												
M_{RPG4}	0.75	0.929	0.04	0.177	4.544	0.51	0.00*	6.127	0.97	0.00*	0.726	0.99
M_{RPG5}	0.657	-1.086	0.07	0.236	-2.552	0.42	0.04*	2.147	0.78	0.00*	0.748	0.98
M_{RPG6}	0.01*	4.192	0.93	0.00*	3.716	0.96	0.00*	4.121	0.97	0.00*	0.604	0.99
Empirical model with three RPG decision rules												
M_{RPG7}	0.734	0.988	0.04	0.171	4.405	0.52	0.00*	5.699	0.96	0.00*	0.729	0.99

Table 4.7 shows that the base model (M_0) was good (all outputs are significant, have positive coefficient and have high R^2). This indicates that the decision rules reported in the literature are useful. All output variables of the models implementing the RPG's selling and upgrading decision rules had significant linear relationships with the real data. They were also able to imitate the trends in the real data. Furthermore, their R^2 value was similar to the base model and, in several cases, slightly higher. This indicates that they were at least as good as the base model in terms of operational validity. However, the models with the RPG's decision rules have higher credibility because we have evidence that supports the micro behaviour of the agents from the RPG.

On the other hand, the model with the RPG's forage collection decision (M_{RPG1}) did not fit the real data well. This decision rule also decreased the model validity when combined with the other RPG's decision rules (e.g. compare M_{RPG2} and M_{RPG4}). Hence, this decision rule is not useful in improving the model's operational validity.

The mean error estimation method measures the accuracy of the predicted value. Unlike regression analysis, this approach cannot cope with changes in trend from the observed data. Table 4.6 shows that there was a serious decline in cow population, cattle population and average daily milk production throughout 2012 (which appears in the data for 2013). This decline occurred owing to an external factor that was not considered in the model (i.e., the policy to stop beef imports which created an incentive for the farmers to sell their productive cows as meat). Therefore, the interval used for the mean error estimation was limited to the period comprising 2010 until the beginning of 2012.

To estimate the mean error, initially, the difference between model outputs at the end of each simulation year and the real data for the selected time interval was measured (i.e., $Error_i = Simulation_i - Data_i$ where i = 2010 … 2012) (note that this is different from $Error_i$ in Table 4.5 which measures the difference between different models). The mean error (ME) from 2010 to 2012 (i.e. $ME = \sum_{2010}^{2012} Error_i/3$) was then computed. To infer whether, in the long run, the model's average ME was zero, a t-test was carried out. Table 4.8 shows the average ($\overline{ME}$) and standard deviation (S_{ME}) of mean error from 30 replications along with the two-tailed significance of the t-test at 95% confidence level. A lower $|\overline{ME}|$ value in Table 4.8 indicates that, on average, the model output was closer to the real data and a significance value higher than 5% indicates better model validity since there is insufficient evidence to reject H_0: $\overline{ME} = 0$.

Table 4. 8 The descriptive statistics and t-result of ABS models' mean error (* indicates significant at 5%)

Model Name	Cattle Population		Cow Population		Daily Milk Production		Farmer Households	
	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.	$(\overline{ME}, S_{ME})$	Sig.
M_0	(1672.7, 5023.9)	0.08	(898.0, 4576.2)	0.29	(-8475.4, 17239.6)	0.01*	(328.4, 154.5)	0.00*
Empirical models with one RPG decision rule								
M_{RPG1}	(6974.5, 10408.8)	0.00*	(5533.1, 9303.0)	0.00*	(12726.4, 37094.4)	0.07	(547.5, 221.0)	0.00*
M_{RPG2}	(1715.4, 5082.5)	0.07	(1101.7, 4514.8)	0.19	(6011.0, 22554.4)	0.16	(-158.8, 457.3)	0.07
M_{RPG3}	(1716.5, 5048.4)	0.07	(933.8, 4593.6)	0.27	(-8356.4, 17398.2)	0.01*	(331.5, 148.2)	0.00*
Empirical models with two RPG decision rules								
M_{RPG4}	(6850.5, 10255.4)	0.00*	(5249.7, 8641.6)	0.00*	(23414.9, 38925.8)	0.00*	(121.8, 662.0)	0.32
M_{RPG5}	(6963.7, 10405.4)	0.00*	(5523.4, 9296.3)	0.00*	(12625.5, 36919.6)	0.07	(547.2, 221.6)	0.00*
M_{RPG6}	(1744.6, 5084.1)	0.07	(1129.8, 4518.3)	0.18	(7272.1, 20860.4)	0.07	(-160.0, 452.4)	0.06
Empirical model with three RPG decision rules								
M_{RPG7}	(6867.2, 10246.2)	0.00*	(5270.2, 8637.8)	0.00*	(24717.6, 37328.9)	0.00*	(126.0, 656.2)	0.30

Table 4.8 shows that the operational validity of base model (M_0) and the model with upgrading decision (M_{RPG3}) was good for two outputs (cattle and cow population). The model with the RPG's selling decision rule (M_{RPG2}) had the highest operational validity as indicated by having significance value higher than 5% for all output variables. This increases our confidence in the model because we can validate both the macro-level outputs and micro-level behaviours. The operational validity of the model containing the RPG's forage collection decision (M_{RPG1}) was lower than base model since it was valid only for daily milk production.

These outcomes demonstrate how we can quantify the benefit of RPG by using operational validity measures. Our analysis shows that the model with the RPG's selling decision rule (M_{RPG2}) had a better overall validity in comparison to the base model. In the following section, we will demonstrate further benefits of RPG for ABS when the purpose is to design behavioural interventions.

4.5.3 Effects of various decision-making rules on the system performance

In this section, we compare the impact of decision rules observed during the RPG when they are used by all farmers in the real-world. These results would enable policy makers to consider possible interventions that may influence the farmers' decision rules. To satisfy this objective, we compared the performance of models with different decision rules calibration by plotting the mean of the simulation outputs.

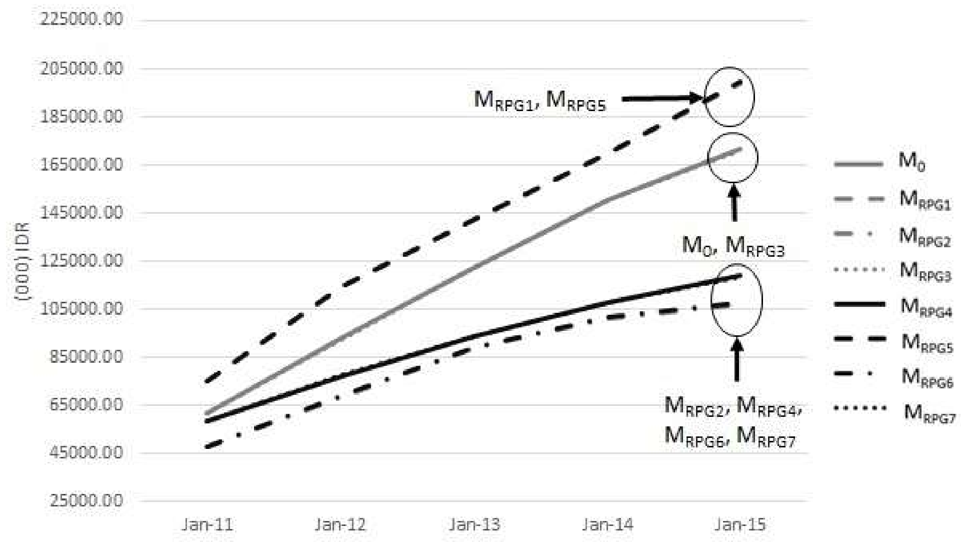

Figure 4. 5: The average farmer's asset of models calibrated with RPG

The first output to be evaluated was the average farmer's assets that represents the farmer's welfare. Figure 4.5 shows that regardless of the decision rule employed in these models, the average assets of the remaining farmers is increasing throughout the simulation. However, the increase in the average farmer's assets value has a different slope. This slope is influenced by the type of decision rule contained in the model. The slopes of models without the RPG's selling decision (M_0, M_{RPG1}, M_{RPG3}, and M_{RPG5}) are higher than the other models. Therefore these models produced higher farmer's asset value by the end of a five year simulation. This shows that the base model's selling decision rule is more beneficial than the base RPG's selling decision rule. When the base model's selling decision rule was used in combination with the RPG's forage collection decision (M_{RPG1} and M_{RPG5}), the average impact on farmer's assets is the highest. Hence, if increasing the farmers' assets is the main objective, policy makers may want to design interventions that encourage more farmers to adopt the RPG's forage collection decision and the base model's selling decision rules.

The second output that was evaluated is the cooperative's total assets. Similar to Figure 4.5, the cooperative's total asset value is also increasing regardless of the decision rule being used in the model. However, in contrast to the average farmer's welfare, Figure 4.6 shows that the cooperative earns higher total revenue when the farmer agents use either the RPG's selling or the RPG's forage collection decisions (M_{RPG1}, M_{RPG2}, M_{RPG4}, M_{RPG5}, M_{RPG6}, and M_{RPG7}). When both decision rules are used (M_{RPG4} and M_{RPG7}) the impact on total revenue for the cooperative agent is the highest.

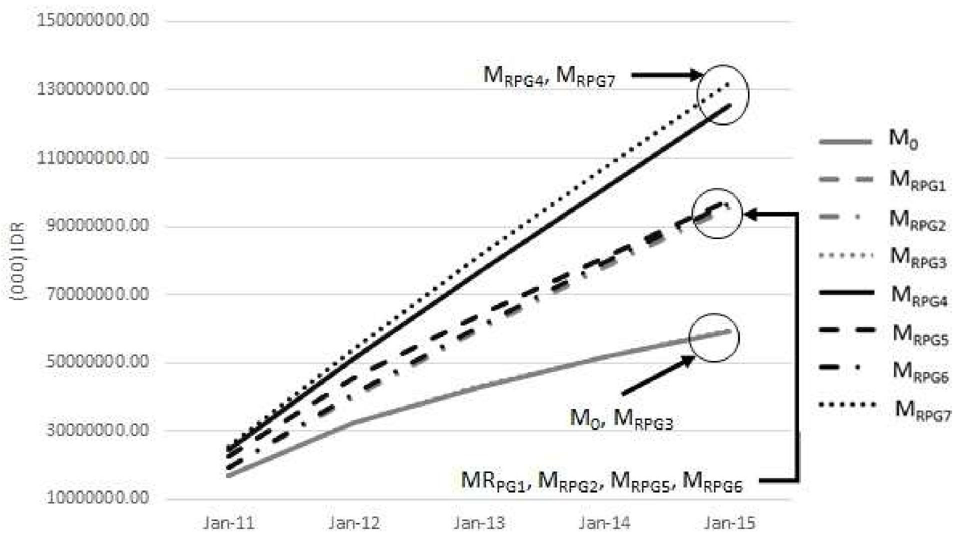

Figure 4. 6: The cooperative's total asset of models calibrated with RPG

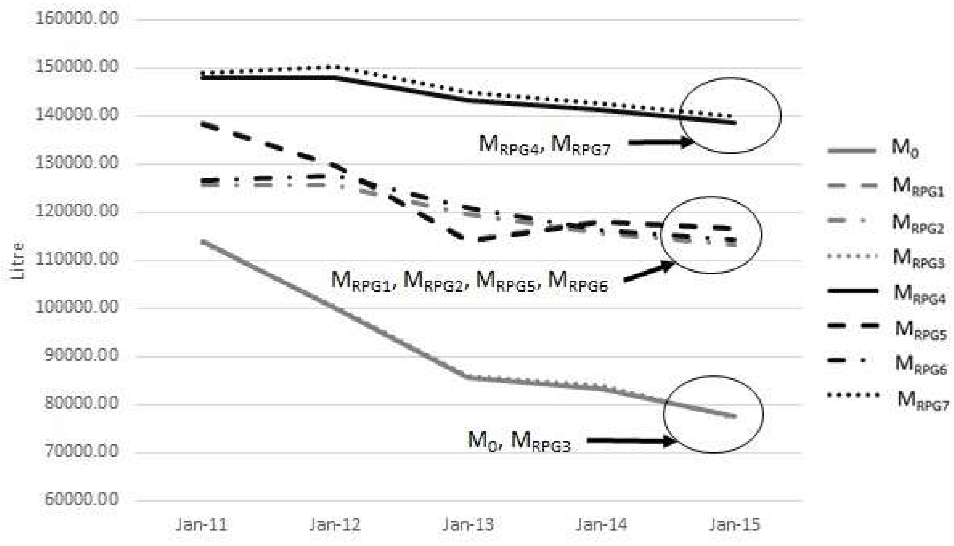

Figure 4. 7: The average daily milk production of models calibrated with RPG

We investigated further to understand why this is the case. Figure 4.7 shows that the models with the RPG's selling or forage collection decisions (M_{RPG1}, M_{RPG2}, M_{RPG4}, M_{RPG5}, M_{RPG6}, and M_{RPG7}) can produce equally high level of milk production. The optimum milk production is achieved when both decision rules are used at the same time (M_{RPG4} and M_{RPG7}). It is reasonable to assume that the level of milk production is proportional to the cow population level (i.e., the models with the RPG's selling and forage collection decisions produced similar cow population output). However, in our model, the cow population produced by the two decision rules is different. Figure 4.8 shows that models with the RPG's forage collection decision (M_{RPG1}, M_{RPG4}, M_{RPG5}, and M_{RPG7}) produced higher cow population than other models. Among the models without the RPG's forage collection decision, those using the RPG's selling decisions (M_{RPG2} and M_{RPG6}) produced the lowest cow population.

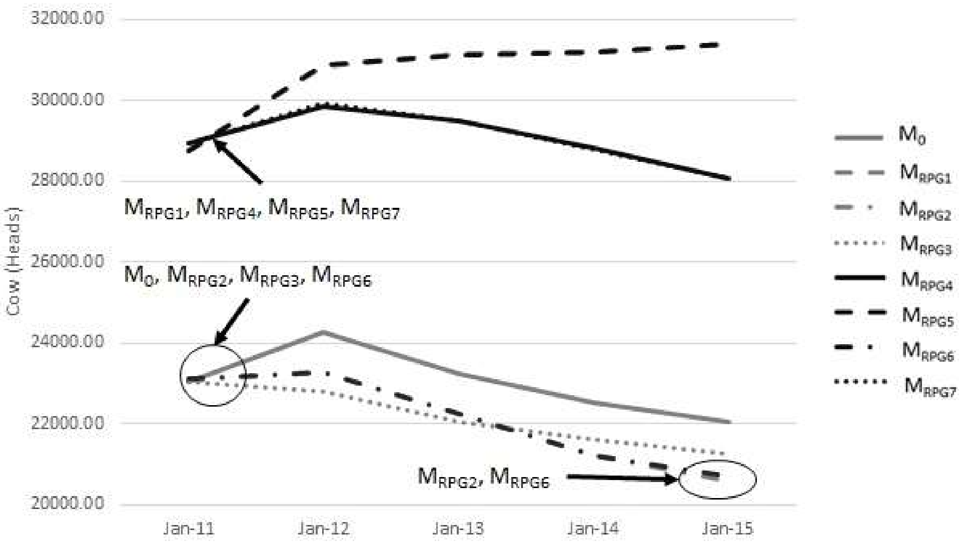

Figure 4. 8: The cow population of models calibrated with RPG

These results suggested two unusual and unexpected outcomes that needed addressing from a policy making perspective.

The first was that the models with very low cow populations (M_{RPG2} and M_{RPG6}) produced equal and sometimes even higher milk production than the other models. Our investigations led us to examine the cattle mortality rate. Figure 4.9 shows that the cow mortality in models containing the RPG's selling decision (M_{RPG2}, M_{RPG4}, M_{RPG6} and M_{RPG7}) was higher than the other models. Consequently, the farmer agents have to spend large amounts of money to replace their dead cows and this explains why, in Figure 4.5, models containing the RPG's selling decision produce lower average farmer's assets.

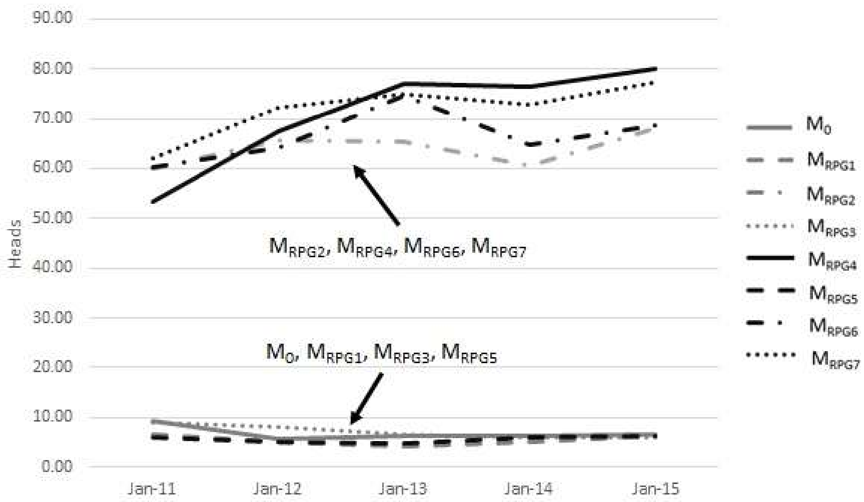

Figure 4. 9: The daily cattle mortality of models calibrated with RPG

Higher cattle mortality rate also increased the rate of retirement of the farmers. Figure 4.10 shows that the remaining farmer households were the lowest in the models with the RPG's selling decision (M_{RPG2} and M_{RPG6}).

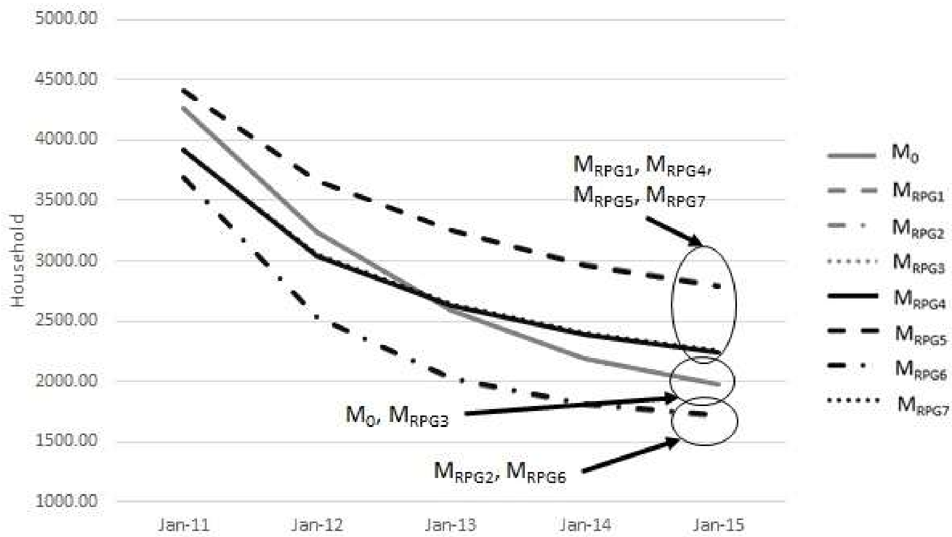

Figure 4. 10: The number of farmer household of models calibrated with RPG

The high retirement rate in models containing the RPG's selling decision but without the RPG's forage collection (M_{RPG2} and M_{RPG6}) increased the forage availability for the surviving agents. This allowed the agents to provide sufficient forage that leads to higher milk productivity per cow. Therefore even though the cow population in M_{RPG2} and M_{RPG6} is very low, the milk production in these two model are equal to M_{RPG1} and M_{RPG5} and even higher than M_0 and M_{RPG3}.

The second interesting outcome was that models containing the RPG's forage collection decision (M_{RPG1}, M_{RPG4}, M_{RPG5}, and M_{RPG7}) produced significantly higher cow population than the other models. Figure 4.11 shows the remaining forage in models containing the RPG's forage collection decision (M_{RPG1}, M_{RPG4}, M_{RPG5}, and M_{RPG7}) is lower. This means that when they are equipped with the RPG's forage collection decision, the agents can collect more forage with the same labour, transport capacity and working hours. In other words, they are more efficient. Consequently, agents in the models containing the RPG's forage collection decision were able to maintain more cows (please refer back to Figure 4.8) even when they were using the RPG's selling decision and this, in turn, led to higher cow mortality. By enabling the farmer agents to maintain more cows, the RPG's forage collection decision, even when it is combined with the RPG's selling decision, led to lower farmer retirement rate. Please refer back to Figure 4.10 which shows that the remaining farmers in M_{RPG4} and M_{RPG7} are significantly higher than M_{RPG2} and M_{RPG6}.

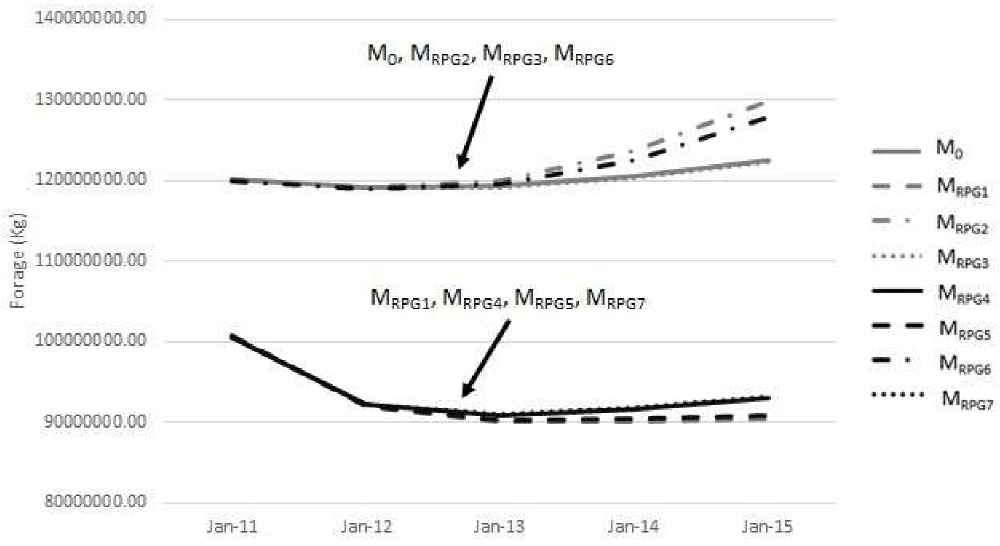

Figure 4. 11: Total amount of forage remaining on patch of models calibrated with RPG

4.6 Discussions

4.6.1 Methodological insights

In section 4.3 and 4.4 of this paper, we propose a process to develop and collect data using an RPG, with a specific purpose of calibrating decision rules in a realistic and predictive ABS model. This process started by selecting decision rules from the base model to be tested by using the RPG. We then convert the base model into an RPG. Using the assumption that the players tend to exhibit their realistic behaviours when the situations they are facing mirror the reality (Rungtusanatham et al., 2011, Cowlrick et al., 2011), the time scaling, parameterisation and player selection processes aimed to maintain the correspondence between the RPG and the real world. Finally we defined collectively exhaustive experiment sets to explore players' decision making rules and to avoid possible biases due to extrapolation. Following Moffat and Medhurst (2009), we treated every player's decision as a data point.

In section 4.5 we propose a process to filter the realistic decision rules from the artificial decision rules that can arise from simply game playing. We assume that an ABS model that contains realistic decision rules is more likely to produce outputs that are in agreement with reality (Epstein, 2006) and we evaluate the correspondence with reality by comparing the simulation outputs with the real data (Thorngate and Edmonds, 2013). In particular, we quantitatively evaluate the improvement in a model's validity caused by a particular decision rule calibration.

Our experience during this study shows that there are several benefits of the proposed process from the methodological perspective, namely:

- *Addressing problems in stakeholder engagement*: One of the challenges in simulation methodology is how to engage stakeholders who have no understanding of simulation modelling (Taylor et al., 2009). In OR more generally, approaches for structuring and understanding how stakeholders make decisions are currently underdeveloped (de Gooyert et al., 2017). Our study shows that RPG data collection can help the participants to describe more easily how they behave in reality. According to the players, unlike in interviews or surveys, in the RPG they have no difficulty in understanding the questions being asked and they do not need to try to understand the constructs being used. They also have the opportunity to relate the situations they face in the games to their reality. The players even consider the RPG to be addictive and want to own the game equipment. This provides OR researchers with a rich opportunity to obtain more insights from the stakeholders.
- *Reducing bias due to memory loss*: Memory loss bias often makes retrospective self-reports unreliable except for very salient events (Janssen and Ostrom, 2006). An RPG requires the players to solve a current and representative

decision problem in each round rather than recall a previous event to answer an interview or survey question or to explain their behaviour in a way that is understandable for the researcher. Consequently, and in common with scenario-based questionnaires, engaging stakeholders to play an RPG can reduce the biases from memory loss (Grewal et al., 2008, Su et al., 2017).

- *Gaining deeper insight into stakeholder's decision making process*. Even though some of our analysis results are consistent with the assumptions in the base model derived from the literature, we obtained more detail information regarding why the real world farmers adopt a particular decision rule. For example, we are able to explain why cattle trading among farmers rarely occurs in reality. In addition, the example of the upgrading decision rule illustrates we can obtain information which was previously unknown that can be used to calibrate the agent's decision rules in the simulation.
- *Enable one to explore a wide range of decision rules*. Maintaining the correspondence between the RPG and reality makes the relationship between the RPG and reality more vivid for the stakeholders. In particular, it improves their ability to explain how they make decisions in the real world. However, although the RPG was designed to observe realistic decision rules, it does not mean that the players are confined to their real life role. They can explore alternatives decision rules or actions that they dare not to take in reality as illustrated by the artificial decision rules found in this study. Although the RPG development process in this study seems more rigid than described in earlier studies, it still achieves the original RPG benefits claimed in earlier studies such as those of Janssen and Ostrom (2006) and Robinson et al. (2007).
- *Reducing the reliance on the researcher's skills in interpreting information obtained from the RPG*. We mention in section 4.2 that, owing to the richness of

interaction during the game, the quality of information and the capability to extract decision rules from RPG data rely heavily on the researcher's skills (Robinson et al., 2007, Salvini et al., 2016). By initially designing the experiment sets and testing the sensitivity of the players' decision through the pilot test, we were able to uncover in advance what to expect and what to record during our RPG sessions. This can help researchers to determine the questions to be explored further when the behaviours exhibited in the RPG do not match their initial expectations. Furthermore, our process allows researchers to plan their semi-quantitative and quantitative observations better.

- *Allowing researchers to develop rule-based and equation-based decision rules*. In previous studies, the RPG data are generally used to generate rule-based decision rules (Robinson et al., 2007), which is caused partially because of a reliance on qualitative analysis. This type of decision rules is useful and sufficient to capture a stakeholder's behaviours under the influence of various qualitative drivers. For example, Joffre et al. (2015) created an RPG of shrimp production system. In their RPG the players can select one of four production systems (i.e., intensive, extensive, improved extensive and integrated mangrove-shrimp) on each round. During the game, the proportion of production systems used by the players' neighbours is varied as a scenario. Players' decisions to shift from one system to another were recorded and the resulting probability was used to update the probability of agents changing from one system to another in the simulation. However, an equation-based rule can be more precise when describing a decision driven by quantitative parameters such as price, cost, land availability. By varying the value of potential drivers in each experiment set, and treating each player's decision as one observation point, our process allows the researcher to also extract equation-based decision rules.

- *Reducing reliance on debriefing and subjective validation.* The validity of an ABS model calibrated using RPG is often problematic and is commonly judged on whether the outputs can be accepted rationally by the researchers and the stakeholders. Sometimes the stakeholders in the RPG process are also involved in assessing the model validity (Robinson et al., 2007) though this approach is prone to biases that arise from a group decision making process. In addition, the use of animations as the basis for validating a simulation model with stakeholders can often produce false insights (Gogi et al., 2016). Our process can reduce these potential biases by separating the realistic and artificial decision rules and by matching the ABS outputs and the real data.

4.6.2 Insights from experiments with a variety of calibrated models

In section 4.5.1 we demonstrate that the greater the differences between the RPG's decision rules and the base model's decision rules then the greater the changes in $\overline{ME}$ significance value. This result allowed us screen out those decision rules that did not significantly affect the final modelling results and produce a parsimonious model.

In section 4.5.2, we illustrate how different decision rules may produce models with different levels of validity. A more valid model is indicated by having a higher significance value in the t-tests of the model's $\overline{ME}$ and higher R^2 value. A higher t-test significance value means the mean difference between the model outputs and the real data are not different from zero. A higher R^2 indicates that the model can better represent the trends in the real data. The experiment results enabled us to select the models that can better represent the system (i.e., more valid). In our case, we considered a model to be better if it had significance above 5% on most output parameters and higher R^2 than the base model. Our experiments enabled us to identify the decision rule with most influence on model outputs which can be valuable for policymakers when setting priorities for interventions to the system.

In section 4.5.3, we describe how the model behaviours are sensitive to the decision rules being used. All models did produce similar trends. For example, the cow population and the number of farmer household variables showed declining trends towards the steady state, the cattle mortality rate tended to be stationary, and the average of farmer household assets was increasing but may have a saturation point. However, the level of each model output was influenced by the type of decision rules being used. Models incorporating the RPG's forage collection decision produce higher levels of milk production, cow population and farmer households. This decision rule is not used by the farmers in reality but it does reveal a more efficient forage collection process and produces positive impacts on all model outputs.

We propose three overall conclusions following our discussion of our modelling results. Firstly, the data from the RPG are useful not only in the validation of the decision rules in the base model but also in the calibration process, which leads to models with higher validity. Secondly, testing the sensitivity of the resultant model to different decision rules is useful in identifying the most and the least influential decision rules. This leads to simpler models that focus on those decision rules that are important from the policymaker/interventionist perspective. Thirdly, the RPG is a useful method to explore the potential of new decision rules to improve the performance of the real system.

4.6.3 Practical insights

The ability to separate realistic and artificial decision rules is important for policymakers because the realistic decision rules describe the existing conditions that are the main target of their interventions. They are also important when we intend to develop a descriptive ABS aims at producing estimates of real world. On the other hand, the artificial decision rules present alternative scenarios that should be adopted or avoided in reality. In this section, we demonstrate the benefit of differentiating the two decision rules when proposing behavioural interventions to improve the system performance. Based upon our results in Section 4.5.3, we observe that the performance of the dairy supply chain in our case study is strongly influenced by two decision rules; namely, the RPG's selling decision rule (which is currently used by the farmers) and the RPG's forage collection decision rule (which may not be currently used by farmers). We summarise the effects of the two decision rules toward several system's parameters in Table 4.9, and then rate them from the worst to the best.

Table 4. 9 Alternative behavioural interventions

Parameters	(1) With RPG's Selling Decision Only (existing)	(2) With Both Decision Rules	(3) With RPG's Forage Collection Only
Farmer's Asset	Worst	Medium	Best
Cooperative's Asset	Medium	Best	Medium
Milk Production	Medium	Best	Medium
Cow Population	Worst	Medium	Best
Remaining farmer household	Worst	Medium	Best

Parameters that may be important for the farmers' welfare are the average value of a farmer's assets and the cow population. It is reasonable to assume that the farmers would prefer a situation in which they own higher asset values. However, other parameters may be important for the cooperative and the government. For the cooperative, low milk production could drive the milk processors to import milk rather than buying it locally. Similarly, a high rate of farmer retirement (i.e., lower remaining farmer households) may indicates that, in the long run, the cooperative is no longer needed. This also signals a potential increase in rural unemployment and its associated problems for the government. Retiring farmers contribute to the unemployment numbers but the loss of cooperatives is also important in that they employ many people in the supply chain (e.g., milk graders, truck drivers and fodder factory workers). Furthermore, environmental degradation becomes a risk in that retiring farmers would usually sell their land. This land could be converted into residential, industrial or recreational use that, among many concerns, will reduce the water catchment area.

Table 4.9 shows that the existing condition is the worst of all three possible scenarios. One immediate improvement that can be made is to encourage a change from the existing forage collection decision to the RPG's forage collection decision (scenario 2 in Table 4.9). In reality, the farmers visually assess the amount of forage they can obtain from a particular location while they are travelling. In the RPG, the players can count the number of forage buttons on each cell before they start the forage collection which, in reality, means the farmers need to be more aware of the forage availability in different locations beforehand and can plan their collection trips more efficiently. Currently, drone technology has been used to monitor deforestation and land use change in Indonesia. The same technology can be adopted to monitor the forage availability and the results communicated to farmers when, for example, the cooperative is collecting their milk. Clearly, aerial monitoring cannot give accurate estimate regarding the forage availability but it can help the farmers to make visual assessment and plan their trips accordingly.

Further improvement can be made by eliminating the RPG’s selling decision, which is suboptimal compared to the selling decision in the base model. This can be done by encouraging the farmers to forecast the forage availability and make the selling decision using this information. Farmers could be trained to record the daily forage they obtained using equipment (e.g., weighing scales) provided by the government or the cooperative (not all the farmers in the case study owned such equipment) and then forecast forage availability based on this record.

These examples illustrate that this filtering process for decision rules not only produces a more accurate model but also supports practical policy making.

4.7 Conclusions

In this paper, we present a calibration process for a realistic ABS model using the RPG method. The decision rules in the ABS, which aim at mimicking the real stakeholders' behaviours, serve as the null hypotheses to be confirmed or rejected by using RPG data. We demonstrate the usefulness of RPG to collect data regarding stakeholder's behaviours in a case of supply chain dairy and record the players' experiences of using this approach. We use a variety of approaches to extract decision rules from the RPG data and use these findings to calibrate the decision rules in the ABS model. After running experiments with a variety of calibrated models, we show that some decisions rules extracted from the RPG data can improve the model's operational validity. Hence, these decision rules are more likely to be used by real farmers. We also analyse the impacts of different decision rule calibration and identify possible behavioural interventions in the real system. Overall, we have demonstrated that the use of an RPG can enhance the value of ABS models in agri-food supply chains for both researchers and users of such models.

In this paper, we used a dairy supply chain case study to demonstrate the benefits of the proposed RPG extension. However, we believe that this methodology is also beneficial beyond dairy supply chain case study. During the pilot test, the participants have mentioned that the same game design can be directly adapted for forestry (the forage button are used to represents wood availability) and fisheries (the forage buttons are used to represent the fish reserves). By using the same methodology and modifying the game design, the RPG can be used to study, for example, a road user's decision rules. For this purpose, instead of representing the forage, the buttons can be used to represent the traffic density, and the game board can be used to represent available roads.

5 Discussion

Firstly, this chapter aims to analyse the three hypotheses mentioned in Chapter 1 by comparing the benefits of different data collection methods. Secondly, this chapter discusses the potential contributions of the three papers presented in this thesis collectively.

5.1 Comparison between data collection methods

Chapter 3 discusses the differences between the standard questionnaire and scenario-based questionnaire. Standard questionnaire in this chapter refers to quantitative methods for collecting data using mostly closed-ended questions (Robinson et al., 2007). Chapter 4 of this thesis discusses the comparison between the extended RPG and the traditional RPG. In this chapter, the traditional RPG refers to an RPG that is developed by involving the stakeholders from a particular community (Janssen and Ostrom, 2006) (not to confirm a base model) and mainly used to support learning and discussion among the stakeholders (Castella et al., 2005b).

This section compares the standard questionnaire, scenario-based questionnaire, traditional RPG and extended RPG in order to test the research hypotheses. This comparison is made by taking into account the dimensions used in the previous literature, for example Janssen and Ostrom (2006) and Robinson et al. (2007). For ease of discussion, in this chapter, respondents in the scenario-based questionnaire and players in the RPG are called participants.

5.1.1 Hypothesis 1: Different data collection methods can produce empirical decision rules with different properties

A decision rule obtained from a data collection process can have at least four properties. Firstly, a decision rule can be distinguished based on whether it contains some novelty, or merely confirms the behaviour described in previous theories (Section 5.1.1). Secondly, a decision rule can be distinguished based on whether it is grounded in theories or is purely empirical (Jager and Janssen, 2002) (Section 5.1.2). The third property is whether the empirical decision rule also explains the context in which it applies or not (Yang and Gilbert, 2008) (Section 5.1.3). Following Yang and Gilbert (2008) the context of a decision rule refers to the reason why an agent makes a particular decision and the conditions that trigger the decision to be actioned. Lastly, a decision rule can be distinguished based on how it is formulated (Section 5.1.4).

5.1.1.1 The possibility to discover new decision rules

This section focuses on discussing the relationship between the data collection method employed and the possibility to obtain new decision rules. New decision rules are those containing information (e.g., new if-then-else logic, new important variable) that have never been discussed in previous theories or literature. These rules can be realistic but unknown previously, or artificial.

Table 2.6 presents a classification of previous ABS applications in ASC based on the data collection method employed. From 18 ABS applications that use a standard questionnaire, only two papers claim that they have found new decision rules. The rest (16 papers) use a standard questionnaire to identify the value of parameters in the simulation. On the other hand, 66.7% of ABS applications that use traditional RPG (participatory modelling group in section 2.6) have claimed that the decision rules in their model have some novelty as a result of the data collection process.

This pattern may be due to the nature of the standard questionnaire that stems from the quantitative methodology. It generally focuses on theory verification and post-decision rationalisation (namely, testing whether a theory can explain a decision that has been made by stakeholders) (Eldabi et al., 2002). Therefore a standard questionnaire mainly uses retrospective self-report format and close-ended questions. Since the opportunity for the participants to express their point of view is very limited, it is hard to discover new decision rules using the standard questionnaire.

The scenario-based questionnaire presented in Chapter 3 tries to move away from the retrospective self-report format. It requires the participants to solve decision problems through a series of scenarios. Its design also gives participants more freedom to express their point of view by providing an open answer option for each scenario. The participants can use this option to describe their decision if it is not well represented by the predefined answers. However, in this study, the participants' willingness to use the open answer option is very low (only 0.4% from all responses). The low response may be caused by the additional efforts needed to describe their decision rules in detail. Therefore, even by providing open answer options, it is still difficult to discover new decision rules. Indeed, by using a scenario-based questionnaire, a new decision rule was found namely, how the farmers decide to sell a cow by considering its possibility to die. However, this information was obtained from the experts during the pilot test, not from by the participants in the data collection. In summary, the final decision rule models that can be elicited from the scenario-based questionnaire are generally confined to the concepts and variables defined from the previous theories and literature.

On the other hand, traditional RPGs that elicit participant's decision rules through the debriefing and the postgame interview are closer to the qualitative methodology (Polhill et al., 2010). Particularly in the simulation study, this methodology places emphasis on discovering how a system works (and how to intervene in it) from the participant's point of view (Eldabi et al., 2002, Polhill et al., 2010). Not only finding correlations that exist in a data set, but also using this methodology researcher hope to unfold causal mechanisms in a system, i.e., “by what intermediate steps, a certain outcome follows from a set of initial conditions” (Yang and Gilbert, 2008).

There are many examples of the successes of traditional RPG in eliciting a participant’s understanding and discovering new decision rules. For example, d'Aquino and Bah (2014) demonstrated how an RPG could be used to elicit the worldviews of drylands communities in Africa. They explained that in order to adapt to climate change these communities prefer to flexibly shift land used practice and location, rather than focusing on one land used activity. This decision rule was novel at the time.

Even though the extended RPG presented in chapter 4 incorporates experiment design and quantitative observation, this does not diminish its ability to uncover the participant's point of view. This is demonstrated, for example, in the buying decision rules elicitation. In common with the previous literature, the scenario-based questionnaire data confirms that forage availability has a significant influence on the buying decision rule. Findings from the extended RPG also support this decision rule; however, its applicability depends on the reason why the participants decide to buy new cows. When the participants aimed at increasing their cow population or replacing their dead cows, the number of new cows they buy is affected by whether they are experiencing forage surplus or not. However, when the participants buy new cows to increase the milk productivity (and at the same time they sell those cows who are less productive), it is not necessary for them to have a surplus in forage. The participants considered the second decision rule to be more effective in increasing their income since it is not necessary for them to expand their pen and to spend more time to collect additional forages.

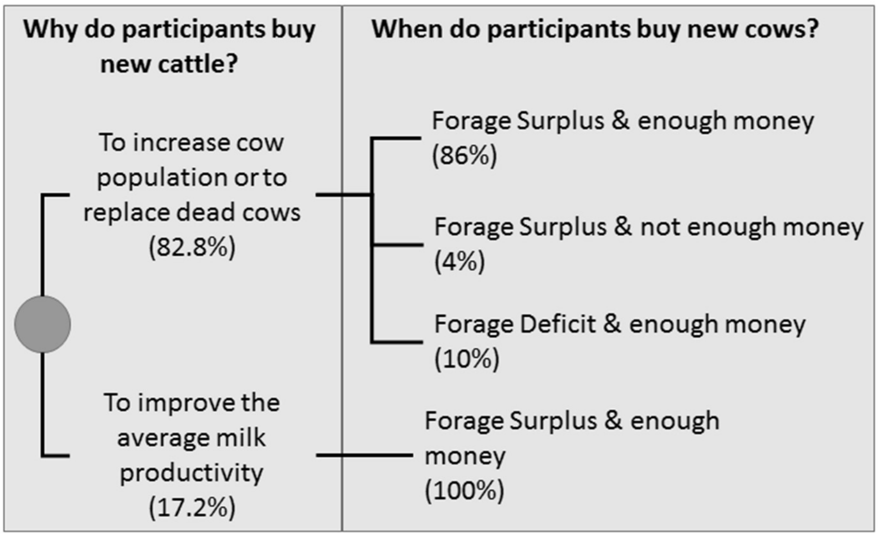

Figure 5. 1: The buying decision rule elicited from the extended RPG

Compared to equation 3.5, Figure 5.1 gives a richer description of how a participant decides to buy a cow. This decision rule has never been considered in the previous literature and demonstrates the capability of the extended RPG in discovering new decision rules. Unfortunately, the sample size is very small. From all RPG sessions there were are only 53 occasions in which the players make the buying decision. This data is not sufficient to perform statistical estimation even by using non-parametric techniques because the frequencies in some categories are less than 5.

5.1.1.2 The correspondence between empirical decision rules and the previous theories

Although an empirical decision rule that contains novel information can be an advantage, the ability to relate it to the previous theories is still important. When the decision rules in an ABS have no relationship with the previous theories at all, then the validity of the model can be considered to be low (Jager and Janssen, 2002).

Quantitative methodology (including standard questionnaires) is often considered better in maintaining the correspondence between existing theories and research than qualitative methods (including traditional RPG) (Eldabi et al., 2002). This is supported by the literature review findings in Chapter 2. The empirical decision rules in 61% of previous ABS applications that employ a standard questionnaire data collection can be related or have references to a particular theory. On the other hand, only 17% of ABS applications that employ traditional RPG contain empirical decision rules that correspond to a particular theory. This low correspondence is likely to occur when the RPG was developed based on consultation with the stakeholders, who are unfamiliar with any theories.

The scenario-based questionnaire has been used to test the applicability of theories in various fields. For example, Jafarkarimi et al. (2016) used a scenario-based questionnaire to test whether the theory of planned behaviour can explain ethical dilemmas in social networking sites. Chapter 3 presents a process to develop a scenario-based questionnaire based on a base model. If the base model has a strong correspondence to the existing theories then so do the scenarios developed from this model and the empirical decision rules produced. As an example is the empirical selling decision rules described in equation 3.6. As explained in section 3.3.3, the previous literature that becomes the basis of the base model states that the number of cows sold by the farmers is affected by the amount of forage deficit they are experiencing. Although in equation 3.6 the probability for a cow to die becomes the only variable that influences the cow selling, this decision rule still has correspondence with the previous literature, for example, Gross et al. (2006) and Lie and Rich (2016). This correspondence occurs because in the base model the independent variable in equation 3.6 is affected by the amount of forage deficit experienced by the farmers. Therefore, the scenario-based questionnaire is arguably as good as the standard questionnaire in maintaining the correspondence between the previous theories and the empirical decision rule.

In the extended RPG two elements can help to maintain the correspondence between the empirical decision rule and the previous theories, namely:

- *The base model*. In common with the scenario-based questionnaire, if the base model has high correspondence with the previous theories then so does the extended RPG, and the empirical decision rules obtained.

- *The design of experiment*. The design of experiment can become an anchor of what is being tested. For example, the relationship between cattle mortality rate, forage level and the participants' behaviour in selling their cows can be tested, although the discussion in debriefing can go beyond previous theories. Section 4.4.2 shows that the empirical selling decision rule obtained through extended RPG (equation 4.1) is similar to the results obtained through the scenario-based questionnaire (equation 3.6).

The experience obtained in this study indicates that in terms of maintaining the correspondence between previous theories and the empirical decision rule, the extended RPG is better than the traditional RPG and can be as good as the standard or scenario-based questionnaire.

5.1.1.3 Clarifying the context of an agent's decisions

This section focuses on the benefit of each data collection method in clarifying the context of an agent's decisions, namely, why and how a decision is made by considering the state of the agent's attributes and the environment (Yang and Gilbert, 2008).

Standard questionnaires commonly used to collect quantitative data are often considered to place less emphasis on the context (Yang and Gilbert, 2008). This also seems to occur in the previous ABS applications in ASC, because most models still focus on easy-to-measure parameters such as production and financial parameters (see Section 2.3.3). Statistical analysis that is commonly used to analyse data from a standard questionnaire is also considered insufficient in providing insight into an agent's motive, incentive and preferences when making a decision (An, 2012). For instance, Evans et al. (2006) mentioned that many statistical tools could be used to correlate a particular agent's attribute (e.g., age) with a specific decision. This study has not explored such relationships but this analysis can be done using the data collected through scenario-based questionnaire. Nevertheless, this kind of analysis does not necessarily answer why an agent of a certain age makes this decision.

In addition to the closed-answer options, the scenario-based questionnaire also provides opportunities for the participants to express their point of view through the open answer options. These options may help to clarify the motives underlying participant's decisions. For example in chapter 3, the participant's preference in choosing the cow to be sold can be explained using a regression model involving the cow's attributes (see equation 3.7). In addition, through the open answer option, some of the participants explained why they prefer not to sell pregnant cows. This is because they consider the newly born calf as one form of their return on investment. If the newly born calf is male, then it can be directly sold to farmers who work in cattle fattening. If the newly born calf is female then, depending on the amount of money needed, they can choose to sell the calf or sell its mother. However, as mentioned in the previous section, the chance to obtain information regarding a participant's motive using scenario-based questionnaire is low.

Some authors argue that qualitative methodologies provide better opportunities for obtaining information about the context of the agent's decisions (Yang and Gilbert, 2008). In common with other qualitative methodologies, the traditional and the extended RPGs can produce the same benefit. For example, using traditional RPG, Joffre et al. (2015) can develop a narrative of how each driver (i.e., farmer's investment capacity, willingness to shift to another production system, local policies, neighbours' production systems and the biophysical characteristics of their plot) works to determine the shrimp production system they choose. By using the extended RPG, this study can elicit the cultural reason underlying the participant's decision rule when collecting forage. The farmers collect forage from locations closest to their home not only to minimize cost and time but also because they do not want to offend other farmers. This territorial consideration can be incorporated in future modelling work.

5.1.1.4 Type of decision rule produced

It has been mentioned in Chapter 2 that in general, the decision rules in ABS can be classified into equation-based and rule-based decisions rules and that both types of decision rule have been employed in the previous ABS modelling in ASC. The literature data presented in Chapter 2 also shows that there is a relationship between the type of decision rule that can be developed and the data collection method employed. The standard questionnaire is useful to develop both types of the decision rule in the previous ABS applications (35% of paper that employ standard questionnaire use equation-based decision rule, and 65% use rule-based decision rule). On the other hand, traditional RPG seems to be more useful in developing rule-based decision rule (all papers that employ traditional RPG use rule-based decision rule).

This pattern may be caused by the techniques available to extract the decision rules in each data collection method. Many standard statistical techniques are available to analyse quantitative data. Sections 3.4 and 3.6 show that this also applies to data obtained through a scenario-based questionnaire. On the other hand, translating qualitative data into decision rules in the simulation tends to be more complex, involving more skills in knowledge elicitation and the result depends on the researcher's expertise (Agar, 2003, Polhill et al., 2010). For this reason, the rule-based decision rule becomes the most convenient way of expressing empirical decision rule obtained from traditional RPG (Robinson et al., 2007).

However, Chapter 4 shows that the extended RPG can help researchers to develop both equation-based and rule-based decision rules. This was done by incorporating experiment design and varying the value of potential drivers of participants' decisions. Each participant's decision was then treated as one observation point. This can be considered as a benefit because an equation-based rule is relatively more sensitive to the variation of potential drivers (e.g., price, land availability, cost etc).

5.1.1.5 Summary and conclusion for the first hypothesis

Table 5. 1 Summary of the first hypothesis analysis

Properties of the empirical decision rule produced	**Standard questionnaire**	**Scenario-based questionnaire**	**Extended RPG**	**Traditional RPG**
Ability to discover new decision rules	Mainly useful to confirm decision rule from theory and identify parameters value	With sufficient pilot test and open answer option it has better chance compared to standard questionnaire.	Better chance than scenario-based questionnaire. Can be as good as traditional RPG.	Big chance to discover new decision rule is.
Maintaining the correspondence with the previous theories	It is good in maintaining the correspondence between empirical decision rule and the previous theories	It is better than traditional RPG, if the base model has high correspondence with the previous theories.	It is better than traditional RPG, if the base model has high correspondence with the previous theories.	The empirical decision rule can be very contextual and has no correspondence with the previous theories
Ability capture the context of agent's decisions	Very low	By incorporating open answer option, it is better than standard questionnaire	Better than scenario-based questionnaire in clarifying the context of agent's decisions	Better than scenario-based questionnaire in clarifying the context of agent's decisions
Type of decision rule produced	Equation-based and rule-based decision rule	Equation-based and rule-based decision rule	Equation-based and rule-based decision rule	Mainly produce rule-based decision rule

Table 5.1 summarises the results of the first hypothesis analysis. This table confirms that different data collection methods can produce empirical decision rules with different properties.

5.1.2 Hypothesis 2: Different data collection methods can produce calibrated models with different levels of operational validity

Section 2.3.5 indicates that validation is still an issue that needs to be addressed further in future ABS applications, including in the context of ASC. This section focuses on analysing whether the differences in empirical decision rules produced from different data collection methods also have some influences on the level of operational validity of the calibrated model (the terminology used in Chapter 2 is empirical validity).

As mentioned in section 4.2, there are two types of ABS model, namely, realistic and artificial (Gilbert, 2004). Operational validity that measures the match between simulation output and real-world data (Sargent, 2013) is more relevant for realistic models. This is because the real world data to validate artificial models, especially those that try to incorporate new ideas and alternative decision rules, are often not available.

Section 4.2 has discussed that a traditional RPG is very useful for eliciting new decision rules based on the participant's knowledge. However, because the data obtained from traditional RPG is commonly used to develop artificial models, its benefit for increasing the model's operational validity is difficult to assess. This is supported by the findings of the literature review (Chapter 2). The literature data shows that all articles that employ traditional RPG are not accompanied by a validation process or are only validated theoretically.

Previous ABS applications in ASC show that the standard questionnaire is very useful for calibrating parameters in the model. The publication data used in chapter 2 shows that 44% of ABS applications that were calibrated using a standard questionnaire feature a model with high operational validity. Using a retrospective self-report format, it is possible to estimate the parameters value in reality, and as a consequence increase the operational validity of the model. For example, suppose a researcher is modelling the farmer's decision rule as an optimisation problem. A standard questionnaire then can be used to identify the value of the coefficients in this optimisation problem.

Experiments in Chapters 3 and 4 show that in addition to being useful for finding appropriate parameter values, the scenario-based questionnaire and extended RPG are also useful to identify appropriate modification for the decision rules in the model (decision rule calibration). For example, compared to the if-then-else rule used in the base model, the calibrated decision rules described in Chapter 3 and 4 can improve the operational validity of the model. The increase in operational validity indicates that the two decision rules have more resemblance to the farmer's behaviour in reality.

All decision rules (M_{SBQ1}, M_{SBQ2}, and M_{SBQ3}) that were elicited using the scenario-based questionnaire can improve the operational validity of most simulation outputs. However, only one empirical decision rule obtained from the extended RPG (M_{RPG2}) can increase the model's operational validity on most outputs. Perhaps this is because in the extended RPG, the participants have more opportunity to improvise and invent new strategies that they do not use in reality.

However, empirical decision rules found through the extended RPG can be considered to have higher credibility than those found through a scenario-based questionnaire. This is because the researchers also have the opportunity to directly observe the participants making a particular decision in the game. Furthermore, researchers can explore a participant's rationale in the debriefing. In contrast, the participants' responses to a scenario-based questionnaire are more than opinion or plan. They result from the participants' judgements even though the event may not have been experienced (*i.e.*, if this condition occurs then this is the action that I should take). There is no guarantee that the players will take the same action in reality.

Table 5.2 summari*s*es the analysis for the second hypothesis. This table confirms that different data collection methods can produce models with different levels of operational validity.

Table 5. 2 Summary of the second hypothesis analysis

	Standard questionnaire	Scenario-based questionnaire	Extended RPG	Traditional RPG
Types of calibration	Mainly useful to calibrate the model's parameters	Useful to calibrate parameters and decision rules in the model	Useful to calibrate parameters and decision rules in the model	Mainly useful to calibrate decision rules in the model
Improvement in the operational validity	Previous ABS applications have shown its benefit in improving the model's operational validity	All of the empirical decision rule obtained can improve the model's operational validity	Some of the empirical decision rule obtained can improve the model's operational validity	Model's operational validity is often difficult to be assessed
Credibility of the decision rule	Does not focused on identifying the participants' decision rule	Useful to identify realistic decision rules, but the credibility is lower.	Useful to identify realistic decision rules, and the credibility is higher, since all decisions have been manifested by the players during the game.	Aims at identifying participants' decision rules, but does not filter realistic decision rules from artificial decision rules.

5.1.3 Hypothesis 3: Different data collection methods have different benefits in the calibrating decision rule in ABS

Section 2.4 presents several literature review articles of ABS research in fields that are relevant to ASC. Some of these articles have discussed the benefits that can be obtained from different data collection methods, other than a scenario-based questionnaire and extended RPG. By considering the comparison criteria in the previous literature (articles presented in section 2.4 and additional articles that were found after chapter 2 was published), this section compares the benefits of the scenario-based questionnaires and the extended RPG relative to the standard questionnaire and the traditional RPG.

5.1.3.1 The meaning of the data

As is mentioned in Chapter 3, the data in a study can have different meanings for the researcher and the participants depending on the data collection methods employed (Yang and Gilbert, 2008). Because the data obtained through a standard questionnaire has a strong correspondence with the previous theories, it is usually more meaningful for the researchers. Conversely, data obtained through qualitative techniques (including traditional RPG) that try to explore the participants' point of view are certainly more meaningful for the participants.

Owing to the possibility of a misunderstanding between the researchers and the participants, some authors (see for example Robinson et al. (2007)) consider that data collected through quantitative data collection methods, such as standard questionnaires, may be biased. If it is related to decision rule elicitation, it is possible that the decision rules obtained through a standard questionnaire are not representative, or only capture a part of a participant’s actual decision rules. For example, using an economic concept one can assume that farmers are trying to maximise their profits and will, therefore, optimise the number of cattle they rear and the amount of forage they collect. Questionnaires can then be used to identify the parameters in this optimisation process. It is certain that the participants can provide information about these parameters. However, as is demonstrated in Chapter 4, the participants' decision rules in collecting forage are also driven by social and cultural concepts.

Section 3.6 shows that the data obtained from a scenario-based questionnaire can be more useful for the participants than the data from the standard questionnaire. This is because the participants can easily relate the scenarios presented to the situations they face in their daily lives. They also have opportunities to give open answers if they feel that their decision is not represented by any of the predefined options. This can minimise potential bias in participants' responses.

However, even though a scenario-based questionnaire can make concepts more meaningful for the participants, the participants still need to interpret the scenarios being used. Moreover, the participants must interpret these scenarios in a way wanted by the researcher. Section 4.4 discusses that the extended RPG and traditional RPG can give more freedom for the participants to interpret the situation they face in the game. Therefore, it is reasonable to assume that the concepts elicited through extended and traditional RPG are more meaningful for the participants than a scenario-based questionnaire.

This can be beneficial to improve the model's credibility because the participants' decision rules are made based on their perspective and understanding. When the RPG has high correspondence with the real system, it is very likely that the participants also view and understand the real system in the same way they view and understand the RPG.

5.1.3.2 The applicability of the data collection technique

This section compares the ease of different data collection techniques to be applied in a variety of cases.

The standard questionnaire is commonly used to calibrate ABS based on micro-economics concepts. In addition, as a quantitative research method, the concepts in a standard questionnaire usually have a strong relationship with theory (Eldabi et al., 2002) which can be used to explain different phenomena. Therefore it is very likely that a similar set of questionnaires can be applied for different cases.

In contrast, qualitative research methods that aim to explore the uniqueness of a case are very contextual. This also applies to traditional RPG (Robinson et al., 2007), especially if the RPG was designed by involving stakeholders who are only familiar with their local context. Therefore the qualitative data collection process, such as conventional RPG, tends to vary from one case to another.

The scenario-based questionnaire and the extended RPG proposed in this study are based on a base model. The base model may have a strong relationship with previous theories. However, in contrast to a standard questionnaire, both the scenario-based questionnaire and the extended RPG incorporate more contexts and uniqueness of the case being studied. In the scenario-based questionnaire, the contexts are embodied in the story that is used to present the scenario to the respondents. In the extended RPG the context and uniqueness are incorporated in the number of objects used in the game (e.g., the proportion of each type of card and the number of sides of the dice).

Conversely, although they are more contextual and can represent the uniqueness of the case being studied, compared to the conventional RPG, the scenario-based questionnaire and the extended RPG can be adapted more easily and applied to other cases. This is because the concepts used in the base model can be general and apply to many cases. Chapter 3 has discussed that the design of the scenario-based questionnaire in this study can be broadly applied to identify how a single decision maker makes an investment or de-investment decision. Chapter 4 has also discussed that the same RPG design can be directly applied to study how actors make decision in fishery or forestry case study.

5.1.3.3 The data collection procedure

Quantitative methods, including standard questionnaires, have an ordered and linear research procedure (from defining research hypotheses to analysing the data) (Eldabi et al., 2002). On the other hand, qualitative methodology tends to be iterative and the research questions may evolve (Polhill et al., 2010).

In general, the development of the scenario-based questionnaire presented in Chapter 3 follows the procedures in the quantitative method. However, there is an iterative part of the scenario-based questionnaire development process namely, developing stories to encapsulate the scenarios. In this process, an experienced farmer was asked to describe his daily experience in the dairy business. Probing questions were used to explore situations in which the farmers make decisions to be confirmed by using a scenario-based questionnaire. To ensure that the scenarios are making sense, the scenario draft was validated by presenting it to the participants. In this study, three iterations were required to finalise the scenarios.

Like other qualitative methods, the data collection procedure using traditional RPG also tends to be iterative. For example, Castella et al. (2005b) required six iterations to design an RPG by directly involving the stakeholders. This is important to capture the points of view from heterogeneous stakeholders. Six workshops with the stakeholders were then conducted to validate the decision rules elicited from the RPG.

The complex data collection procedure means that the reliance on the researcher's skills tends to be high in traditional RPG (Robinson et al., 2007). These skills include the ability to extract information from participants, the ability to interpret this information without bias and the ability to convert the interpretation results into simulation models. The reliance on a researcher's skills also occurs in other qualitative data collection methodologies in general. Therefore these methodologies must be carried out by fully trained researchers (Polhill et al., 2010). Some authors argue that, without sufficient training, these less structured research procedures can endanger the research such that it produces meaningless results (Eldabi et al., 2002). This makes traditional RPG seem more suitable to be used by experienced researchers.

The extended RPG introduces more order and structure to the RPG development process. Defining the hypotheses from a base model helps researchers to think in advance about the kind of information that must be extracted from the participants. The parameterisation process proposed in Chapter 4 helps the researcher to place the participants in a decision situation that want to be observed. In pure qualitative research, this framing is usually done by using probing questions. The experiment design helps researchers to plan the data tabulation process (in qualitative research coding is usually done after the data collection was finished) and to identify possible decision rule models (either equation-based or rule-based decision rules) that can be generated from the RPG. This helps the researchers to interpret information from the participants and to incorporate it in a simulation model.

The extended RPG does not completely eliminate the qualitative exploratory process common in a traditional RPG. However, in the extended RPG, this process only occurs during debriefing and it mainly aims to explore the motives behind participants' decisions during the game. The exploration was done by giving the participants open ended questions or by asking them to give examples. The cultural motive that underlies a farmer's forage collection decision (explained in Chapter 4) is one example of the qualitative exploration results. Additionally the debriefing also gives opportunity for the participants to raise additional issues. For example, in some experiments the participants expressed the importance of cow insurance schemes to reduce losses due to cow mortality. Similar to the procedure in traditional RPG, a finding in one RPG experiment can be validated by confirming it in the subsequent RPG experiments.

In conclusion, the extended RPG methodology still retains most of the benefits of traditional RPGs. However, with a more ordered and linear data collection procedure it may be easier for less experienced researchers to carry out the research.

5.1.3.4 Ability to capture how agents make decisions when facing unprecedented scenarios

This section compares the capabilities of the four data collection methods in capturing the participants' behaviour when they are confronted with new scenarios. The discussion in chapters 2 and 3 shows that the data from a standard questionnaire is, in principle, a snapshot in time and mainly used to identify the value of the real world parameters. Hence the standard questionnaire is mainly useful for describing how the participants behave under the existing condition, but not very suitable to capture how they make a decision when they are facing a new situation or scenario (Robinson et al., 2007). Using a longitudinal survey it is possible to capture how participants' decisions change dynamically owing to the changing situations they encounter. However this approach can be very expensive and is not always feasible within the constraints of a research project.

Section 3.6 reports that the use of scenarios could help the participants to think about the actions they would take in situations they had not yet experienced. Hence even though the scenario-based questionnaire survey remains as a snapshot in time, it is still possible to obtain indications of how the participants will choose their actions in possible future situations. For example, suppose owing to climate change or a certain disease, the cow mortality rate (which is currently between 0% and 50%) increases to 90%. Then, using the decision rule from the scenario-based questionnaire, we can expect that 98% of farmers will sell their cows if they cannot collect sufficient forage.

The extended and traditional RPG can also capture how the participants' decisions may change dynamically when they are facing unprecedented situations. If we use the decision rule from the extended RPG, then for the above scenario, we can expect that 99% of farmers will sell their cows when they cannot collect sufficient forage. This result is not much different from the scenario-based questionnaire result. However, the behaviour obtained through RPG can be considered more credible because in addition to the participants' answers, we also have an opportunity to observe the participants making these decisions. If the RPG has a high correspondence with the real world, then it is likely that participants will make the same decision supposing the unprecedented situation happens.

5.1.3.5 Effect of participants' boredom and fatigue

Researchers have long suspected that the participant fatigue and boredom may influence their judgment during data collection, decreasing the accuracy of the data and the resulting model (Bradley and Daly, 1994, Bijmolt and Wedel, 1995). Although some authors argue that the impact of fatigue and boredom on the quality of the final model is small (for example Hess et al. (2012)), these two factors seem to have some influence on the participants in this study.

During the scenario-based questionnaire data collection, on average, each participant spent 1.5 hours to complete all of the scenarios by themselves and, on average, one RPG session involving four participants took 2 hours (on average, each participant spent 30 minutes to think and make decisions). Therefore, the level of participants' fatigue and boredom during the scenario-based questionnaire data collection is higher than in the extended RPG. Also, the competition among participants also makes the data collection process more fun and enjoyable.

In previous studies, RPG is considered to be able to facilitate participants to show how they solve decision problems in a relaxed atmosphere (see for example Castella et al. (2005b)). The experience in this study is very much in agreement with this. As mentioned in Chapter 4, most of the participants found the extended RPG to be very interesting and addictive, so much so that they have expressed their interest in owning the RPG equipment. Those who participated in both data collection processes also considered playing an RPG is more fun and less boring than responding to a survey or interview (including the scenario-based questionnaire). They admitted that boredom often makes them give the quickest response rather than what they actually feel or do.

The impact of participant fatigue and boredom on the decision rules elicited from the two data collection methods has not been quantitatively measured in this study. If these two factors have significant impacts on the quality of the extracted decision rules, then they will also affect the model's credibility. Therefore quantitative measurement is important to identify which data collection approach can produce a more credible ABS.

5.1.3.6 Biases due to memory loss

One of the factors that affect the quality of the decision rule elicited from a data collection process is the accuracy of the information provided by the participants. The accuracy of information is influenced by the ability of participants to recall their real-world experience. Unfortunately, many cognitive biases may arise from memory loss (Schacter, 1999). These biases can reduce the accuracy of the information provided by the participants, especially if they are required to recall previous events. This is why some authors consider quantitative data collection methods (including the standard questionnaire) that rely on retrospective self-reports to be unreliable except for very salient events (Janssen and Ostrom, 2006). However, this bias may also occur in qualitative data collection approaches. In an interview, for example, participants are often asked to describe their past experiences.

The scenario-based questionnaire, traditional RPG and extended RPG require the participants to solve the current and representative decision problems rather than recall previous events. Previous studies have shown that by asking the participants to solve problems when they are being questioned can reduce bias due to memory loss (Grewal et al., 2008, Su et al., 2017).

5.1.3.7 The risk of going native

When conducting research, it is very important to describe the details and uniqueness of the case being studied. For example in ASCs, farmers in different countries may have different preference toward the production method they use, and this may be influenced by different cultures. To reveal this uniqueness, qualitative researchers often involve themselves as a part of the community being studied. By doing so, they can obtain insider view points on the research topic. However, sometimes the researcher becomes too attached to the community being studied, so that their perceptions are clouded by their personal experience and they have difficulty separating them from those of the participants (Dwyer and Buckle, 2009). Conversely, it is also possible that the participants fail to explain their individual experience fully because they assume that the researchers have a similar understanding to them (Dwyer and Buckle, 2009). This phenomena produces a bias called going native (Eldabi et al., 2002). The ability to avoid this bias depends very much on the researcher's skills and experience. Triangulation by comparing the conclusions of a researcher's observation with those of another researcher who is studying a similar phenomenon, is an effort that usually done to minimize this bias.

The scenario-based questionnaire, in common with quantitative methods, keeps the researcher detached from the community being studied. Therefore the chance of this bias occurring is low. In the extended RPG, this bias is less likely to occur than in the traditional RPG. The first reason is that the participants' influence in designing the extended RPG is smaller than in a traditional RPG. The second reason is that the extended RPG has two types of data sources namely debriefing and observation guided by the experiment design. Information from debriefing can be distorted because of the researcher's experience when interacting with the community under study. However, by using the experiment design as guidance, the observation data is objective and will not be distorted by the researcher's experience.

5.1.3.8 Summary and conclusion for the third hypothesis analysis

Table 5.3 summarises the results of the third hypothesis analysis. This table confirms that different data collection methods can have different benefit in decision rules calibration. Moreover, the scenario-based questionnaire and the extended RPG seem to be able to minimise some weaknesses in the existing data collection method. For example, they can reduce the reliance on researcher’s skills, help to give insight on how the participants make decisions under unprecedented scenarios, and reduce some of the biases that can occur in the data collection process.

Table 5. 3 Summary of the third hypothesis analysis

	Standard questionnaire	Scenario-based questionnaire	Extended RPG	Traditional RPG
The meaning of data/concepts	Data and concepts are more meaningful for the researchers	Data and concepts are meaningful for the researchers and participants	Data and concepts are more meaningful for the participants	Data and concepts are more meaningful for the participants
Applicability of the data collection technique	The same data collection instrument and procedure can be applied in a wide range of cases	The data collection instrument and procedure can many cases in which the same theoretical concept applies.	The game design and the process to collect data can be applied to other cases in which the same theoretical concept applies and with minor modifications	The game design and the process to collect data varies from one case to another
The data collection procedure	Linear and ordered	Linear and ordered	Less linear and ordered than a questionnaire, but more structured than traditional RPG	Iterative and may evolve throughout study
Reliance on the researcher's skills	low reliance	low reliance	lower than traditional RPG, can be as low as a scenario-based questionnaire	High reliance on researcher's skills
Ability to capture how agents make decisions when facing unprecedented scenarios	A snapshot in time and mainly useful for post-decision rationalisation	Can give an idea on how participants behave under unprecedented scenarios	Enable researchers to observe how participants behave under unprecedented scenarios	Enable researchers to observe how participants behave under unprecedented scenarios
Effect of participant's boredom and fatigue	Participants experience high fatigue and boredom	Participants experience high fatigue and boredom	Data collection can be carried out in a relaxed atmosphere.	Data collection can be carried out in a relaxed atmosphere.
Biases due to memory loss	High when using retrospective self-report format	Less likely to occur than the standard questionnaire	Less likely to occur than the standard questionnaire	Less likely to occur than the standard questionnaire
The risk of going native	Very unlikely to occur	Very unlikely to occur	Possible, but less likely when compared to the traditional RPG	Possible, when the researcher becomes too attached to the community being studied.

5.2 Contributions of this study

This section focuses on discussing the overall contributions of this study. These contributions can be viewed from methodological and practical perspectives.

- *In terms of the case presented*, this study has shown that most of ABS applications in ASC focus on modelling one echelon (i.e., the producer) and feature case studies from high-income countries. Therefore the model presented in this study can be considered as a contribution because it describes dyadic interactions between the farmers and the cooperative and was developed based on a case study in the middle-income country. This case study has several unique features such as the reliance on dairy farming as farmers' sole income, the reliance on family labour, and forage as a common pool resource.

- *In terms of the data collection method*, this study has discussed various attempts to improve ABS methodology. This discussion shows that efforts to improve the data collection methodology are not receiving much attention from ABS modellers in the field of ASC and, maybe, beyond ASC. Therefore, the efforts to improve the existing data collection methods that are proposed in this study can be considered as another contribution. This study has also shown that the proposed data collection methods (i.e., the scenario-based questionnaire and the extended RPG) have several advantages over the standard data collection method.

- *In terms of data collection methods comparison,* this study has discussed many studies that aim at comparing the benefits of different data collection methods in calibrating an ABS model. This discussion shows that the previous studies mainly compare experiences obtained from various case studies. This approach can be biased because the benefits and accuracy of a data collection method is affected by the complexity of the case being studied. This study uses the same case study to compare the calibration results produced by different data collection methods. Therefore, the potential bias mentioned above can be minimized. Moreover, this study can analyse the relationship between the data collection method being used and the increase in the model's operational validity. Analysing this relationship is difficult when different case studies are used.

- *In terms of sensitivity analysis*, this study has presented various sensitivity analysis techniques in the previous ABS applications namely, OFAT and GSA. These techniques mainly focus on checking whether the model reacts correctly to changes in input parameters. A sensitivity analysis process was also performed in this study. However, instead of the input parameters value, these experiments test model behaviours under various combination of empirical decision rules. By doing so, these experiments can identify the decision rules that can improve the model's operational validity and the system's performance. This approach is different from the techniques that were used in the previous studies and, consequently, the sensitivity analysis process presented in this study can also be considered a contribution.

- *In terms of policymaking,* the discussion in this study shows that the majority of previous ABS applications in the ASC focused on proposing policies at the cluster level. Policies such as credit, subsidies, pricing and compensation schemes are prevalent in the previous ABS applications. However, the experiments carried out in this study allow us to propose unique policies that target agents' behaviours. These policies include the behaviours that must be retained or discarded to achieve better system performance, and interventions that can be made to help the real farmers to change their current behaviours.

6 Conclusion and Further Research

6.1 Conclusions

This study shows that, in common to ABS modelling in other application areas, ABS modellers in the field of ASC are seeking to improve the operational validity and the credibility of their simulation model. One way to improve the operational validity and credibility of an ABS model is to calibrate the decision rules in the simulation based on data elicited from real-world actors. The role of the data collection method is critical to elicit realistic decision rules from the real world actors, and eventually to develop realistic and predictive ABS.

This study identifies questionnaire survey as the most popular empirical data collection method in the previous ABS applications. It is a deductive and quantitative approach that has many benefits, for example, it can maintain the correspondence between the ongoing research and the previous theories. However, it is often criticised due to several biases, such as bias due to memory loss. From the inductive, qualitative perspective, RPG is gaining more attention from the ABS community. This approach is considered to be able to improve trust and openness between researchers and real-world actors.

However, it is mainly used to develop artificial models whose operational validity is difficult to be assessed.

To produce ABS with higher operational validity and more credible, this study has proposed improvements and extensions for those data collection methodologies. The improved questionnaire survey is called a scenario-based questionnaire, while the improved RPG is called the extended RPG. Empirical data for calibrating the decision rules in an ABS has been collected using both improved methods from dairy farmers in Indonesia. A series of experiment were carried out to demonstrate how decision rule calibration using both improved methods can produce models with higher operational validity, compared to a model that was solely calibrated parametrically. The experiment results also allow one to propose behavioural interventions for the real system. Finally, this study has discussed the advantages and disadvantages of the improved data collection methods relative to each other and to standard questionnaire and RPG.

This research has contributed to ABS methodology by proposing improvements to the existing data collection methods, and showing how these improvements are beneficial to increase the operational validity of an ABS. This study also compares the advantages and disadvantages of different data collection methods. This may help ABS researchers to plan the data collection processes in their research. Also, unlike in the previous literature, the comparison presented in this study was produced from the same case study. When the result of various case studies are compared, the performance of a data collection methodology will be affected by the complexity of the case in which it is employed. Namely, the benefits of a particular data collection method can be seen as inferior compared to the others simply because it was employed in a more complex case study.

This study also contributes in term of the case study, by modelling dyadic interaction between farmers and cooperatives which is rare in the previous ABS applications in the ASC. Furthermore, the research was conducted in the context of a middle income country that is less commonly studied in the existing literature. It has been discussed how the farmers in this context are different, in terms of production system and cultural background, from what was described in previous studies.

Finally, the experiment results in this study can be used to propose several behavioural interventions for the real world system. These interventions include the decision rules that must be retained or changed by the real world actors, as well as the efforts that can be taken by the government to empower the dairy farmers in the case study site. Hence this study has also contributed to policy making.

6.2 Limitations and Further Research

In this study, there are at least two limitations that need to be addressed in the future research. The first limitation is related to the ability to compare the accuracy of decision rules elicited using scenario-based questionnaires and the extended RPG. The second limitation is related to the capability to assess the credibility of the calibrated model.

Accuracy refers to the degree of similarity between the empirical decision rules elicited through scenario-based questionnaire or extended RPG with the actual participants' decision rules. For instance, both scenario-based questionnaire and extended RPG can elicit the participants' selling decision rules. Calibration using selling decision rules from the scenario-based questionnaire can increase the model's operational validity on three simulation outputs while calibration using selling decision rules from the extended RPG can improve the operational validity on all simulation outputs. However, the difference in the level of operational validity is not sufficient to justify that calibration using the extended RPG is more accurate than the scenario-based questionnaire. There are many uncontrolled causes for this difference. One of them is because the participant's decision rule is time-dependent. Both data collection methods can be equally accurate, but because there is a one year interval between the two phases of data collections, the participants' decision rule may have changed. One way to overcome this limitation in the future research is to collect the data within a very short time period.

Producing a model with high credibility is another important aspect in a simulation study. Credibility refers to the users' confidence that the modelling result is correct. The initial design of this study includes quantitative assessment to measure the credibility of models calibrated with different data collection methods. This assessment was planned to be done by conducting a white-box validation process involving potential users of the calibrated model (e.g., experts from the animal husbandry department). However, owing to the time constraint, the difference in the calibrated model's credibility was only assessed subjectively based on the experience during data collection.

In addition to research that aims to overcome the limitations of this study, there are other potential future studies that can be developed. One of them is to evaluate whether or not the proposed data collection techniques are also useful in other sectors beyond agriculture. Many decision making in small medium enterprises, for example, involve a single decision maker. Hence, these data collection techniques may also be useful to elicit the real actors' decision rules in this context.

It is also interesting to evaluate the benefits of the proposed data collection techniques for decision-making processes in an organisation that involves many decision makers. In this case study, the farmers' cooperative is such an organisation. For example, policies that are being planned by the farmers' cooperative can be formalised as scenarios in the questionnaire. We can then use the scenario-based questionnaire to identify how the farmers' behaviours change owing to the planned policies. Computer simulation can then describes the impacts of these behavioural changes toward the whole system. Another alternative is to use the planned policy as one of the experiment sets in the RPG. A representative from the framers' cooperative can be involved as one of the facilitators in the RPG experiments. By doing so the cooperative's representative can observe the differences in the players' behaviour under different policy setting. The representative's experience can then enrich the discussion during decision making process in the cooperative.

7 References

AGAR, M. 2003. My kingdom for a function: Modeling misadventures of the innumerate. *Journal of Artificial Societies and Social Simulation,* 6.

AHUMADA, O. & VILLALOBOS, J. R. 2009. Application of planning models in the agri-food supply chain: A review. *European Journal of Operational Research,* 196**,** 1-20.

ALBINO, V., FRACCASCIA, L. & GIANNOCCARO, I. 2016. Exploring the role of contracts to support the emergence of self-organized industrial symbiosis networks: an agent-based simulation study. *Journal of Cleaner Production,* 112, Part 5**,** 4353-4366.

AMADOU, M. L., VILLAMOR, G. B. & KYEI-BAFFOUR, N. 2018. Simulating agricultural land-use adaptation decisions to climate change: An empirical agent-based modelling in northern Ghana. *Agricultural Systems*.

AN, L. 2012. Modeling human decisions in coupled human and natural systems: review of agent-based models. *Ecological Modelling,* 229**,** 25-36.

ARNOLD, R., TROOST, C. & BERGER, T. 2015. Quantifying the economic importance of irrigation water reuse in a Chilean watershed using an integrated agent-based model. *Water Resources Research,* 51**,** 648-668.

ARVITRIDA, N. I., ROBINSON, S. & TAKO, A. A. How do competition and collaboration affect supply chain performance? An agent based modeling approach. Winter Simulation Conference (WSC), 2015, 2015. IEEE, 218-229.

AURBACHER, J., PARKER, P. S., CALBERTO SÁNCHEZ, G. A., STEINBACH, J., REINMUTH, E., INGWERSEN, J. & DABBERT, S. 2013. Influence of climate change on short term management of field crops – A modelling approach. *Agricultural Systems,* 119**,** 44-57.

AXELROD, R. 1997. Advancing the art of simulation in the social sciences. *Simulating social phenomena.* Springer.

AZADEGAN, A., KACH, A., GOLARA, S. & MOUSAVI, N. 2017. Corporate environmental investments: A cross-national study on managerial decision making. *International Journal of Production Economics*.

BAHAR, S. 2014. Produktivitas hijauan pakan untuk produksi sapi potong di Sulawesi Selatan. *JITV,* 19.

BAKKER, M. M., ALAM, S. J., VAN DIJK, J. & ROUNSEVELL, M. D. A. 2015. Land-use change arising from rural land exchange: an agent-based simulation model. *Landscape Ecology,* 30**,** 273-286.

BANKES, S. C. 2002. Agent-based modeling: A revolution? *Proceedings of the National Academy of Sciences,* 99**,** 7199-7200.

BARRETEAU, O., BOUSQUET, F. & ATTONATY, J. M. 2001. Role-playing games for opening the black box of multi-agent systems. *Journal of Artificial Societies and Social Simulation,* 4.

BARRETT, C. B., BACHKE, M. E., BELLEMARE, M. F., MICHELSON, H. C., NARAYANAN, S. & WALKER, T. F. 2012. Smallholder Participation in Contract Farming: Comparative Evidence from Five Countries. *World Development,* 40**,** 715-730.

BECU, N., PEREZ, P., WALKER, A., BARRETEAU, O. & PAGE, C. L. 2003. Agent based simulation of a small catchment water management in northern Thailand: Description of the CATCHSCAPE model. *Ecological Modelling,* 170**,** 319-331.

BENDOLY, E., CROSON, R., GONCALVES, P. & SCHULTZ, K. 2010. Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations Management,* 19**,** 434-452.

BENDOLY, E., DONOHUE, K. & SCHULTZ, K. L. 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management,* 24**,** 737-752.

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**,** 245-260.

BERGER, T. & SCHREINEMACHERS, P. 2006. Creating agents and landscapes for multiagent systems from random samples. *Ecology and Society,* 11**,** 19.

BERGER, T., SCHREINEMACHERS, P. & WOELCKE, J. 2006. Multi-agent simulation for the targeting of development policies in less-favored areas. *Agricultural Systems,* 88**,** 28-43.

BERGER, T. & TROOST, C. 2014. Agent-based Modelling of Climate Adaptation and Mitigation Options in Agriculture. *Journal of Agricultural Economics,* 65**,** 323-348.

BERT, F. E., PODESTÁ, G. P., ROVERE, S. L., MENÉNDEZ, Á. N., NORTH, M., TATARA, E., LACIANA, C. E., WEBER, E. & TORANZO, F. R. 2011. An agent based model to simulate structural and land use changes in agricultural systems of the argentine pampas. *Ecological Modelling,* 222**,** 3486-3499.

BERT, F. E., ROVERE, S. L., MACAL, C. M., NORTH, M. J. & PODESTÁ, G. P. 2014. Lessons from a comprehensive validation of an agent based-model: The experience of the Pampas Model of Argentinean agricultural systems. *Ecological Modelling,* 273**,** 284-298.

BHARWANI, S., BITHELL, M., DOWNING, T. E., NEW, M., WASHINGTON, R. & ZIERVOGEL, G. 2005. Multi-Agent Modelling of Climate Outlooks and Food Security on a Community Garden Scheme in Limpopo, South Africa. *Philosophical Transactions: Biological Sciences,* 360**,** 2183-2194.

BIJMOLT, T. H. A. & WEDEL, M. 1995. The effects of alternative methods of collecting similarity data for multidimensional scaling. *International Journal of Research in Marketing,* 12**,** 363-371.

BOERO, R. 2011. Food quality as a public good: cooperation dynamics and economic development in a rural community. *Mind & Society,* 10**,** 203-215.

BOMMEL, P., DIEGUEZ, F., BARTABURU, D., DUARTE, E., MONTES, E., PEREIRA MACHÍN, M., CORRAL, J., PEREIRA DE LUCENA, C. J. & MORALES, H. 2014. A further step towards participatory modelling. Fostering stakeholder involvement in designing models by using executable UML. *Journal of Artificial Societies and Social Simulation,* 17.

BOONE, R., GALVIN, K., BURNSILVER, S., THORNTON, P., OJIMA, D. & JAWSON, J. 2011. Using coupled simulation models to link pastoral decision making and ecosystem services. *Ecology and Society,* 16.

BORODIN, V., BOURTEMBOURG, J., HNAIEN, F. & LABADIE, N. 2016. Handling uncertainty in agricultural supply chain management: A state of the art. *European Journal of Operational Research,* 254**,** 348-359.

BOYER, C. N. & BRORSEN, B. W. 2013. Changes in Beef Packers' Market Power after the Livestock Mandatory Price Reporting Act: An Agent-based Auction. *American Journal of Agricultural Economics,* 95**,** 859-876.

BRADLEY, M. & DALY, A. 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation,* 21**,** 167-184.

BRÄNDLE, M. J., LANGENDIJK, G., PETER, S., BRUNNER, H. S. & HUBER, R. 2015. Sensitivity Analysis of a Land-Use Change Model with and without Agents to Assess Land Abandonment and Long-Term Re-Forestation in a Swiss Mountain Region. *Land,* 4.

BRONNENBERG, B. J., KRUGER, M. W. & MELA, C. F. 2008. Database paper—The IRI marketing data set. *Marketing science,* 27**,** 745-748.

BUSBY, J. S., ONGGO, B. S. S. & LIU, Y. 2016. Agent-based computational modelling of social risk responses. *European Journal of Operational Research,* 251**,** 1029-1042.

BUSBY, S. J. & ONGGO, S. 2013. Managing the social amplification of risk: a simulation of interacting actors. *Journal of the Operational Research Society,* 64**,** 638-653.

BYLUND, P. L. 2015. Signifying Williamson's Contribution to the Transaction Cost Approach: An Agent-Based Simulation of Coasean Transaction Costs and Specialization. *Journal of Management Studies,* 52**,** 148-174.

CAMPO, P. C., MENDOZA, G. A., GUIZOL, P., VILLANUEVA, T. R. & 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**,** 3607-3615.

CASTELLA, J.-C., BOISSAU, S., TRUNG, T. N. & QUANG, D. D. 2005a. Agrarian transition and lowland–upland interactions in mountain areas in northern Vietnam: application of a multi-agent simulation model. *Agricultural Systems,* 86**,** 312-332.

CASTELLA, J.-C., TRUNG, T. & BOISSAU, S. 2005b. Participatory Simulation of Land-Use Changes in the Northern Mountains of Vietnam: the combined Use of an Agent-Based Model, a Role-Playing Game, and a Geographic Information System. *Ecology and Society,* 10**,** n/a.

CDRC. 2017. Available: https://www.cdrc.ac.uk/ [Accessed 7 July 2017].

CHATURVEDI, A., ARMSTRONG, B. & CHATURVEDI, R. 2014. Securing the food supply chain: understanding complex interdependence through agent-based simulation. *Health and Technology,* 4**,** 159-169.

CHEBOLU-SUBRAMANIAN, V. & GAUKLER, G. M. 2015. Product contamination in a multi-stage food supply chain. *European Journal of Operational Research,* 244**,** 164-175.

CHOO, A. S., NAG, R. & XIA, Y. 2015. The role of executive problem solving in knowledge accumulation and manufacturing improvements. *Journal of Operations Management,* 36**,** 63-74.

COCKBURN, D., CRABTREE, S. A., KOBTI, Z., KOHLER, T. A. & BOCINSKY, R. K. 2013. Simulating social and economic specialization in small-scale agricultural societies. *Journal of Artificial Societies and Social Simulation,* 16**,** 4.

COLLINS, A., PETTY, M., VERNON-BIDO, D. & SHERFEY, S. 2015. A Call to Arms: Standards for Agent-Based Modeling and Simulation. *Journal of Artificial Societies and Social Simulation,* 18**,** 12.

COURDIER, R., GUERRIN, F., ANDRIAMASINORO, F. H. & PAILLAT, J.-M. 2002. Agent-based simulation of complex systems: application to collective management of animal wastes. *Journal of Artificial Societies and Social Simulation,* 5.

COWLRICK, I., HEDNER, T., WOLF, R., OLAUSSON, M. & KLOFSTEN, M. 2011. Decision-making in the pharmaceutical industry: analysis of entrepreneurial risk and attitude using uncertain information. *R&D Management,* 41**,** 321-336.

CUNNINGHAM, D. C. 2001. The distribution and extent of agrifood chain management research in the public domain. *Supply Chain Management: An International Journal,* 6, 212-215.

D'AQUINO, P. & BAH, A. 2014. Multi-level participatory design of land use policies in African drylands: A method to embed adaptability skills of drylands societies in a policy framework. *Journal of Environmental Management,* 132, 207-219.

D'AQUINO, P., LE PAGE, C., BOUSQUET, F. & BAH, A. 2003. Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: the SelfCormas experiment in Senegal. *Journal of artificial societies and social simulation,* 6.

DA SILVA, C. A. & DE SOUZA FILHO, H. M. 2007. *Guidelines for rapid appraisals of agrifood chain performance in developing countries*, Food and Agriculture Organization of the United Nations Rome.

DE GOOYERT, V., ROUWETTE, E., VAN KRANENBURG, H. & FREEMAN, E. 2017. Reviewing the role of stakeholders in Operational Research: A stakeholder theory perspective. *European Journal of Operational Research,* 262, 402-410.

DWYER, S. C. & BUCKLE, J. L. 2009. The space between: On being an insider-outsider in qualitative research. *International journal of qualitative methods,* 8, 54-63.

ELDABI, T., IRANI, Z., PAUL, R. J. & LOVE, P. E. D. 2002. Quantitative and qualitative decision-making methods in simulation modelling. *Management Decision,* 40, 64-73.

EMANUELA, D. 2012. Outcomes of inter-organizational trust in supply chain relationships: a systematic literature review and a meta-analysis of the empirical evidence. *Supply Chain Management: An International Journal,* 17**,** 377-402.

EPSTEIN, J. M. 2006. *Generative social science: Studies in agent-based computational modeling*, Princeton University Press.

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**,** 5.

EVANS, T. P., SUN, W. & KELLEY, H. 2006. Spatially explicit experiments for the exploration of land-use decision-making dynamics. *International Journal of Geographical Information Science,* 20**,** 1013-1037.

FAO. 2016. *FAOstat* [Online]. Available: http://www.fao.org/faostat/en/#data.

FØLGESVOLD, A. & PRENKERT, F. 2009. Magic pelagic — An agent-based simulation of 20 years of emergent value accumulation in the North Atlantic herring exchange system. *Industrial Marketing Management,* 38**,** 529-540.

FRANKE, J., STOCKHEIM, T. & KÖNIG, W. 2005. The impact of reputation on supply chains. An analysis of permanent and discounted reputation. *Information Systems and E-Business Management,* 3**,** 323-341.

GAGLIARDI, D., NIGLIA, F. & BATTISTELLA, C. 2014. Evaluation and design of innovation policies in the agro-food sector: An application of multilevel self-regulating agents. *Technological Forecasting and Social Change,* 85**,** 40-57.

GE, H., GRAY, R. & NOLAN, J. 2015a. Agricultural supply chain optimization and complexity: A comparison of analytic vs simulated solutions and policies. *International Journal of Production Economics,* 159**,** 208-220.

GE, H., NOLAN, J. & GRAY, R. 2015b. Identifying Strategies to Mitigate Handling Risks in the Canadian Grain Supply Chain. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie,* 63**,** 101-128.

GILBERT, N. 2004. Agent-based social simulation: dealing with complexity. *The Complex Systems Network of Excellence,* 9**,** 1-14.

GLOCK, C. H. 2012. Coordination of a production network with a single buyer and multiple vendors. *International Journal of Production Economics,* 135**,** 771-780.

GLOVER, J. L., CHAMPION, D., DANIELS, K. J. & DAINTY, A. J. D. 2014. An Institutional Theory perspective on sustainable practices across the dairy supply chain. *International Journal of Production Economics,* 152**,** 102-111.

GODFRAY, H. C. J., BEDDINGTON, J. R., CRUTE, I. R., HADDAD, L., LAWRENCE, D., MUIR, J. F., PRETTY, J., ROBINSON, S., THOMAS, S. M. & TOULMIN, C. 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science,* 327**,** 812.

GOGI, A., TAKO, A. A. & ROBINSON, S. 2016. An experimental investigation into the role of simulation models in generating insights. *European Journal of Operational Research,* 249**,** 931-944.

GRAUBNER, M., BALMANN, A. & SEXTON, R. J. 2011. Spatial Price Discrimination in Agricultural Product Procurement Markets: A Computational Economics Approach. *American Journal of Agricultural Economics,* 93**,** 949-967.

GREWAL, D., ROGGEVEEN, A. L. & TSIROS, M. 2008. The Effect of Compensation on Repurchase Intentions in Service Recovery. *Journal of Retailing,* 84**,** 424-434.

GRIMM, V., BERGER, U., DEANGELIS, D. L., POLHILL, J. G., GISKE, J. & RAILSBACK, S. F. 2010. The ODD protocol: a review and first update. *Ecological modelling,* 221**,** 2760-2768.

GROSS, J. E., MCALLISTER, R. R. J., ABEL, N., SMITH, D. M. S. & MARU, Y. 2006. Australian rangelands as complex adaptive systems: a conceptual model and preliminary results. *Environmental Modelling & Software,* 21**,** 1264-1272.

HÄMÄLÄINEN, R. P., LUOMA, J. & SAARINEN, E. 2013. On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research,* 228**,** 623-634.

HAMILL, L. 2010. Agent-based modelling: The next 15 years. *Journal of Artificial Societies and Social Simulation,* 13**,** 7.

HAPPE, K., HUTCHINGS, N. J., DALGAARD, T. & KELLERMAN, K. 2011. Modelling the interactions between regional farming structure, nitrogen losses and environmental regulation. *Agricultural Systems,* 104**,** 281-291.

HAPPE, K., KELLERMANN, K. & BALMANN, A. 2006. Agent-based Analysis of Agricultural Policies: an Illustration of the Agricultural Policy Simulator AgriPoliS, its Adaptation and Behavior. *Ecology and Society,* 11**,** 49.

HAPPE, K., SCHNICKE, H., SAHRBACHER, C. & 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/Revue canadienne d'agroeconomie,* 57**,** 497-511.

HE, Z., WANG, S. & CHENG, T. C. E. 2013. Competition and evolution in multi-product supply chains: An agent-based retailer model. *International Journal of Production Economics,* 146**,** 325-336.

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

HESS, S., HENSHER, D. A. & DALY, A. 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice,* 46**,** 626-644.

HIGGINS, A., THORBURN, P., ARCHER, A. & JAKKU, E. 2007. Opportunities for value chain research in sugar industries. *Agricultural Systems,* 94**,** 611-621.

HIGGINS, A. J., MILLER, C. J., ARCHER, A. A., TON, T., FLETCHER, C. S. & MCALLISTER, R. R. J. 2010. Challenges of operations research practice in agricultural value chains. *Journal of the Operational Research Society,* 61**,** 964-973.

HOLM, S., LEMM, R., THEES, O. & HILTY, L. M. 2016. Enhancing Agent-Based Models with Discrete Choice Experiments. *Journal of Artificial Societies and Social Simulation,* 19.

JAFARKARIMI, H., SAADATDOOST, R., SIM, A. T. H. & HEE, J. M. 2016. Behavioral intention in social networking sites ethical dilemmas: An extended model based on Theory of Planned Behavior. *Computers in Human Behavior,* 62**,** 545-561.

JAGER, W. & JANSSEN, M. The need for and development of behaviourally realistic agents. International Workshop on Multi-Agent Systems and Agent-Based Simulation, 2002. Springer, 36-49.

JANSSEN, M. & OSTROM, E. 2006. Empirically based, agent-based models. *Ecology and Society,* 11.

JANSSEN, S. & VAN ITTERSUM, M. K. 2007. Assessing farm innovations and responses to policies: A review of bio-economic farm models. *Agricultural Systems,* 94**,** 622-636.

JOFFRE, O. M., BOSMA, R. H., LIGTENBERG, A., TRI, V. P. D., HA, T. T. P. & BREGT, A. K. 2015. Combining participatory approaches and an agent-based model for better planning shrimp aquaculture. *Agricultural Systems,* 141**,** 149-159.

JULKA, N., SRINIVASAN, R. & KARIMI, I. 2002. Agent-based supply chain management—1: framework. *Computers & Chemical Engineering,* 26**,** 1755-1769.

KAUFMANN, P., STAGL, S. & FRANKS, D. W. 2009. Simulating the diffusion of organic farming practices in two New EU Member States. *Ecological Economics,* 68**,** 2580-2593.

KAYE-BLAKE, B., SCHILLING, C. & POST, E. 2014. Validation of an Agricultural MAS for Southland, New Zealand. *Journal of Artificial Societies and Social Simulation,* 17**,** 5.

KELLY, R. A., JAKEMAN, A. J., BARRETEAU, O., BORSUK, M. E., ELSAWAH, S., HAMILTON, S. H., HENRIKSEN, H. J., KUIKKA, S., MAIER, H. R., RIZZOLI, A. E., VAN DELDEN, H. & VOINOV, A. A. 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software,* 47**,** 159-181.

KLEIJNEN, J. P. C. 1995a. Statistical validation of simulation models. *European Journal of Operational Research,* 87**,** 21-34.

KLEIJNEN, J. P. C. 1995b. Verification and validation of simulation models. *European journal of operational research,* 82**,** 145-162.

KLOS, T. B. & NOOTEBOOM, B. 2001. Agent-based computational transaction cost economics. *Journal of Economic Dynamics and Control,* 25**,** 503-526.

KPBS 2016. Data Populasi dan Penghasilan Anggota KPBS. *Unpublished Dataset.*

KREJCI, C. & BEAMON, B. 2015. Impacts of Farmer Coordination Decisions on Food Supply Chain Structure. *Journal of Artificial Societies and Social Simulation,* 18**,** 19.

KREJCI, C. C. & BEAMON, B. M. Modeling food supply chains using multi-agent simulation. Simulation Conference (WSC), Proceedings of the 2012 Winter, 9-12 Dec. 2012 2012. 1-12.

KREJCI, C. C., STONE, R. T., DORNEICH, M. C. & GILBERT, S. B. 2016. Analysis of Food Hub Commerce and Participation Using Agent-Based Modeling: Integrating Financial and Social Drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 58**,** 58-79.

KUNSCH, P. L., KAVATHATZOPOULOS, I. & RAUSCHMAYER, F. 2009. Modelling complex ethical decision problems with operations research. *Omega,* 37**,** 1100-1108.

KUTCHER, G. P. & NORTON, R. D. 1982. Operations research methods in agricultural policy analysis. *European Journal of Operational Research,* 10**,** 333-345.

LEE, J.-S., FILATOVA, T., LIGMANN-ZIELINSKA, A., HASSANI-MAHMOOEI, B., STONEDAHL, F., LORSCHEID, I., VOINOV, A., POLHILL, G. J., SUN, Z. & PARKER, D. C. 2015. The Complexities of Agent-Based Modeling Output Analysis. *Journal of Artificial Societies and Social Simulation,* 18**,** 4.

LIE, H. & RICH, K. M. 2016. Modeling Dynamic Processes in Smallholder Dairy Value Chains in Nicaragua: A System Dynamics Approach. *International Journal on Food System Dynamics,* 7**,** 328-340.

LIE, H., RICH, K. M. & BURKART, S. 2017. Participatory system dynamics modelling for dairy value chain development in Nicaragua. *Development in Practice,* 27**,** 785-800.

LORSCHEID, I., HEINE, B.-O. & MEYER, M. 2012. Opening the 'black box'of simulations: increased transparency and effective communication through the systematic design of experiments. *Computational and Mathematical Organization Theory,* 18**,** 22-62.

LUSCH, R. F. 2011. Reframing supply chain management: a service-dominant logic perspective. *Journal of Supply Chain Management,* 47**,** 14-18.

MACAL, C. & NORTH, M. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation,* 4**,** 151-162.

MACAL, C. M. 2016. Everything you need to know about agent-based modelling and simulation. *Journal of Simulation,* 10**,** 144-156.

MALAWSKA, A. & TOPPING, C. J. 2016. Evaluating the role of behavioral factors and practical constraints in the performance of an agent-based model of farmer decision making. *Agricultural Systems,* 143**,** 136-146.

MANISH, S. & SANJAY, J. 2013. Agri-fresh produce supply chain management: a state-of-the-art literature review. *International Journal of Operations & Production Management,* 33**,** 114-158.

MANTEL, S. P., TATIKONDA, M. V. & LIAO, Y. 2006. A behavioral study of supply manager decision-making: Factors influencing make versus buy evaluation. *Journal of Operations Management,* 24**,** 822-838.

MAROHN, C., SCHREINEMACHERS, P., QUANG, D. V., BERGER, T., SIRIPALANGKANONT, P., NGUYEN, T. T. & 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.

MARTIN, R., LINSTÄDTER, A., FRANK, K. & MÜLLER, B. 2016. Livelihood security in face of drought – Assessing the vulnerability of pastoral households. *Environmental Modelling & Software,* 75**,** 414-423.

MATTHEWS, R. B., GILBERT, N. G., ROACH, A., POLHILL, J. G. & GOTTS, N. M. 2007. Agent-based land-use models: a review of applications. *Landscape Ecology,* 22**,** 1447-1459.

MCPHEE-KNOWLES, S. 2015. Growing Food Safety from the Bottom Up: An Agent-Based Model of Food Safety Inspections. *Journal of Artificial Societies and Social Simulation,* 18**,** 9.

MEIJER, S., HOFSTEDE, G. J., BEERS, G. & OMTA, S. W. F. 2006. Trust and Tracing game: learning about transactions and embeddedness in a trade network. *Production Planning & Control,* 17**,** 569-583.

MILLINGTON, J., ROMERO-CALCERRADA, R., WAINWRIGHT, J. & PERRY, G. 2008. An agent-based model of Mediterranean agricultural land-use/cover change for examining wildfire risk. *Journal of Artificial Societies and Social Simulation,* 11**,** 4.

MOFFAT, J. & MEDHURST, J. 2009. Modelling of human decision-making in simulation models of conflict using experimental gaming. *European Journal of Operational Research,* 196**,** 1147-1157.

MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. & GROUP, P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine,* 6**,** e1000097.

MONTANOLA-SALES, C., ONGGO, B. S. S. & CASANOVAS-GARCIA, J. 2011. Agent-based simulation validation: A case study in demographic simulation. *Proceedings of the Third International Conference on Advances in System Simulation (SIMUL 2011)***,** 109-115.

MORGAN, F. J. & DAIGNEAULT, A. J. 2015. Estimating Impacts of Climate Change Policy on Land Use: An Agent-Based Modelling Approach. *PLoS ONE,* 10**,** e0127317.

MORGAN, J. F., BROWN, P. & DAIGNEAULT, J. A. 2015. Simulation vs. Definition: Differing Approaches to Setting Probabilities for Agent Behaviour. *Land,* 4.

NAINGGOLAN, D., TERMANSEN, M., FLESKENS, L., HUBACEK, K., REED, M. S., DE VENTE, J. & BOIX-FAYOS, C. 2012. What does the future hold for semi-arid Mediterranean agro-ecosystems? – Exploring cellular automata and agent-based trajectories of future land-use change. *Applied Geography,* 35**,** 474-490.

NG, D. W. 2008. Structural Change in a Food Supply Chain. *International Food and Agribusiness Management Review,* 11.

NOLAN, J., PARKER, D., VAN KOOTEN, G. C. & BERGER, T. 2009. An Overview of Computational Modeling in Agricultural and Resource Economics. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie,* 57**,** 417-429.

OLABISI, S. L., WANG, Q. R. & LIGMANN-ZIELINSKA, A. 2015. Why Don't More Farmers Go Organic? Using A Stakeholder-Informed Exploratory Agent-Based Model to Represent the Dynamics of Farming Practices in the Philippines. *Land,* 4.

OLIVEIRA, J. B., LIMA, R. S. & MONTEVECHI, J. A. B. 2016. Perspectives and relationships in Supply Chain Simulation: A systematic literature review. *Simulation Modelling Practice and Theory,* 62**,** 166-191.

ONGGO, B. S., BUSBY, J. & LIU, Y. 2014. Using agent-based simulation to analyse the effect of broadcast and narrowcast on public perception: A case in social risk amplification. *Proceedings of the Winter Simulation Conference 2014***,** 322-333.

ONGGO, B. S. & KARATAS, M. 2016. Test-driven simulation modelling: A case study using agent-based maritime search-operation simulation. *European Journal of Operational Research,* 254**,** 517-531.

ONGGO, B. S. S. 2013. Agent-based simulation model representation using BPMN. *Formal Languages for Computer Simulation: Transdisciplinary Models and Applications. Hershey, PA: Information Science Reference***,** 378-399.

ONGGO, S., HILL, J. & BROOKS, R. A survey on data identification and collection in simulation projects. The 27th European Simulation and Modelling Conference, 2013 Lancaster, UK. EUROSIS-ETI.

OSINGA, S. A., KRAMER, M. R. & HOFSTEDE, G. J. 2015. Sustainable animal welfare: does forcing farmers into transition help? *AI & SOCIETY,* 30**,** 403-413.

PAPAZIAN, H., BOUSQUET, F., ANTONA, M. & D'AQUINO, P. 2017. A Stakeholder-oriented Framework to Consider the Plurality of Land Policy Integration in Sahel. *Ecological Economics,* 132**,** 155-168.

PIDD, M. 2004. *Computer simulation in management science,* Hoboken, NJ, Hoboken, NJ : Wiley.

PLA, L. M., SANDARS, D. L. & HIGGINS, A. J. 2014. A perspective on operational research prospects for agriculture. *Journal of the Operational Research Society,* 65**,** 1078-1089.

POLHILL, J. G., SUTHERLAND, L.-A. & GOTTS, N. M. 2010. Using qualitative evidence to enhance an agent-based modelling system for studying land use change. *Journal of Artificial Societies and Social Simulation,* 13**,** 10.

QUANG, D. V., SCHREINEMACHERS, P. & 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.

RASCH, S., HECKELEI, T., OOMEN, R. & NAUMANN, C. 2016. Cooperation and collapse in a communal livestock production SES model – A case from South Africa. *Environmental Modelling & Software,* 75**,** 402-413.

RASCH, S., HECKELEI, T., STORM, H., OOMEN, R. & NAUMANN, C. 2017. Multi-scale resilience of a communal rangeland system in South Africa. *Ecological Economics,* 131**,** 129-138.

REIDSMA, P., BAKKER, M. M., KANELLOPOULOS, A., ALAM, S. J., PAAS, W., KROS, J. & DE VRIES, W. 2015. Sustainable agricultural development in a rural area in the Netherlands? Assessing impacts of climate and socio-economic change at farm and landscape level. *Agricultural Systems,* 141**,** 160-173.

ROBINSON, D. T., BROWN, D. G., PARKER, D. C., SCHREINEMACHERS, P., JANSSEN, M. A., HUIGEN, M., WITTMER, H., GOTTS, N., PROMBUROM, P.,

IRWIN, E., BERGER, T., GATZWEILER, F. & BARNAUD, C. 2007. Comparison of empirical methods for building agent-based models in land use science. *Journal of Land Use Science,* 2**,** 31-55.

ROBINSON, S., WORTHINGTON, C., BURGESS, N. & RADNOR, Z. J. 2014. Facilitated modelling with discrete-event simulation: Reality or myth? *European Journal of Operational Research,* 234**,** 231-240.

ROSS, R. B. 2011. Entrepreneurial behaviour in agri-food supply chains: the role of supply chain partners. *Journal on Chain and Network Science,* 11**,** 19-30.

ROSS, R. B. & WESTGREN, R. E. 2009. An Agent-Based Model of Entrepreneurial Behavior in Agri-Food Markets. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie,* 57**,** 459-480.

ROUZAFZOON, J. & HELO, P. 2016. Developing service supply chains by using agent based simulation. *Industrial Management & Data Systems,* 116**,** 255-270.

RUNGTUSANATHAM, M., WALLIN, C. & ECKERD, S. 2011. The vignette in a scenario-based role-playing experiment. *Journal of Supply Chain Management,* 47**,** 9-16.

SALVINI, G., LIGTENBERG, A., VAN PAASSEN, A., BREGT, A. K., AVITABILE, V. & HEROLD, M. 2016. REDD+ and climate smart agriculture in landscapes: A case study in Vietnam using companion modelling. *Journal of Environmental Management,* 172**,** 58-70.

SARGENT, R. G. 2013. Verification and validation of simulation models. *Journal of Simulation,* 7**,** 12-24.

SCHACTER, D. L. 1999. The seven sins of memory: Insights from psychology and cognitive neuroscience. *American psychologist,* 54**,** 182.

SCHIERITZ, N. & GROBLER, A. 2003. Emergent structures in supply chains-a study integrating agent-based and system dynamics modeling. 9-pp.

SCHMID, A. 2005. What is the Truth of Simulation? *Journal of Artificial Societies and Social Simulation,* 8.

SCHOUTEN, M., OPDAM, P., POLMAN, N. & WESTERHOF, E. 2013. Resilience-based governance in rural landscapes: Experiments with agri-environment schemes using a spatially explicit agent-based model. *Land Use Policy,* 30**,** 934-943.

SCHOUTEN, M., VERWAART, T. & HEIJMAN, W. 2014. Comparing two sensitivity analysis approaches for two scenarios with a spatially explicit rural agent-based model. *Environmental Modelling & Software,* 54**,** 196-210.

SCHREINEMACHERS, P. & BERGER, T. 2011. An agent-based simulation model of human–environment interactions in agricultural systems. *Environmental Modelling & Software,* 26**,** 845-859.

SCHREINEMACHERS, P., BERGER, T. & AUNE, J. B. 2007. Simulating soil fertility and poverty dynamics in Uganda: A bio-economic multi-agent systems approach. *Ecological Economics,* 64**,** 387-401.

SCHREINEMACHERS, P., BERGER, T., SIRIJINDA, A. & PRANEETVATAKUL, S. 2009. The Diffusion of Greenhouse Agriculture in Northern Thailand: Combining Econometrics and Agent-Based Modeling. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie,* 57**,** 513-536.

SCHREINEMACHERS, P., POTCHANASIN, C., BERGER, T. & ROYGRONG, S. 2010. Agent-based modeling forex anteassessment of tree crop innovations: litchis in northern Thailand. *Agricultural Economics,* 41**,** 519-536.

SCOTT, R. J., CAVANA, R. Y. & CAMERON, D. 2016. Recent evidence on the effectiveness of group model building. *European Journal of Operational Research,* 249**,** 908-918.

SMAJGL, A., BROWN, D. G., VALBUENA, D. & HUIGEN, M. G. A. 2011a. Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling & Software,* 26**,** 837-844.

SMAJGL, A., HOUSE, A. P. N. & BUTLER, J. R. A. 2011b. Implications of ecological data constraints for integrated policy and livelihoods modelling: An example from East Kalimantan, Indonesia. *Ecological Modelling,* 222**,** 888-896.

SOBOL′, I. M. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation,* 55**,** 271-280.

SONDEREGGER-WAKOLBINGER, L. M. & STUMMER, C. 2015. An agent-based simulation of customer multi-channel choice behavior. *Central European Journal of Operations Research,* 23**,** 459.

SOTO-SILVA, W. E., NADAL-ROIG, E., GONZÁLEZ-ARAYA, M. C. & PLA-ARAGONES, L. M. 2016. Operational research models applied to the fresh fruit supply chain. *European Journal of Operational Research,* 251**,** 345-355.

STATISTICS INDONESIA. 2017. *Agricultural and Animal Husbandry Census* [Online]. Available:

https://www.bps.go.id/subject/24/peternakan.html#subjekViewTab3 [Accessed 22 December 2017].

SU, H.-C., CHEN, Y.-S. & RO, Y. K. 2017. Perception differences between buyer and supplier: the effect of agent negotiation styles. *International Journal of Production Research,* 55**,** 6067-6083.

TATICCHI, P., GARENGO, P., NUDURUPATI, S. S., TONELLI, F. & PASQUALINO, R. 2015. A review of decision-support tools and performance measurement and sustainable supply chain management. *International Journal of Production Research,* 53**,** 6473-6494.

TAYLOR, S. J. E., ELDABI, T., RILEY, G., PAUL, R. J. & PIDD, M. 2009. Simulation Modelling Is 50! Do We Need a Reality Check? *The Journal of the Operational Research Society,* 60**,** s69-s82.

TEN BROEKE, G., VAN VOORN, G. & LIGTENBERG, A. 2016. Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation,* 19**,** 5.

THORNGATE, W. & EDMONDS, B. 2013. Measuring Simulation-Observation Fit: An Introduction to Ordinal Pattern Analysis. *Journal of Artificial Societies and Social Simulation,* 16**,** 4.

TROOST, C. & BERGER, T. 2015. Dealing with Uncertainty in Agent-Based Simulation: Farm-Level Modeling of Adaptation to Climate Change in Southwest Germany. *American Journal of Agricultural Economics,* 97**,** 833-854.

TYKHONOV, D., JONKER, C., MEIJER, S. & VERWAART, T. 2008. Agent-based simulation of the trust and tracing game for supply chains and networks. *Journal of Artificial Societies and Social Simulation,* 11**,** 1.

UDUMYAN, N., ROUCHIER, J. & AMI, D. 2014. Integration of Path-Dependency in a Simple Learning Model: The Case of Marine Resources. *Computational Economics,* 43**,** 199-231.

URDA, J. & LOCH, C. H. 2013. Social preferences and emotions as regulators of behavior in processes. *Journal of Operations Management,* 31**,** 6-23.

UTHES, S., PIORR, A., ZANDER, P., BIEŃKOWSKI, J., UNGARO, F., DALGAARD, T., STOLZE, M., MOSCHITZ, H., SCHADER, C., HAPPE, K., SAHRBACHER, A., DAMGAARD, M., TOUSSAINT, V., SATTLER, C., REINHARDT, F.-J., KJELDSEN, C., CASINI, L. & MÜLLER, K. 2011. Regional impacts of abolishing direct payments: An integrated analysis in four European regions. *Agricultural Systems,* 104**,** 110-121.

UTOMO, D. S., ONGGO, B. S. & ELDRIDGE, S. 2018. Applications of agent-based modelling and simulation in the agri-food supply chains. *European Journal of Operational Research,* 269**,** 794-805.

VAHID, S., LEHOUX, N., DE SANTA-EULALIA, L. A., D'AMOURS, S., FRAYRET, J.-M. & VENKATADRI, U. 2016. Supply chain modelling frameworks for forest products industry: a systematic literature review. *INFOR: Information Systems and Operational Research,* 54**,** 52-75.

VALBUENA, D., VERBURG, P. H. & BREGT, A. K. 2008. A method to define a typology for agent-based analysis in regional land-use research. *Agriculture, Ecosystems & Environment,* 128**,** 27-36.

VAN DER VORST, J. G. A. J., BEULENS, A. J. M. & VAN BEEK, P. 2000. Modelling and simulating multi-echelon food systems. *European Journal of Operational Research,* 122**,** 354-366.

VAN ECK, N. J. & WALTMAN, L. 2009. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics,* 84**,** 523-538.

VOINOV, A. & BOUSQUET, F. 2010. Modelling with stakeholders. *Environmental Modelling & Software,* 25**,** 1268-1281.

VOINOV, A., KOLAGANI, N., MCCALL, M. K., GLYNN, P. D., KRAGT, M. E., OSTERMANN, F. O., PIERCE, S. A. & RAMU, P. 2016. Modelling with stakeholders – Next generation. *Environmental Modelling & Software,* 77**,** 196-220.

VRABEL, M. Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Oncology nursing forum, 2015. 552-554.

WANG, J., BROWN, D. G., RIOLO, R. L., PAGE, S. E. & AGRAWAL, A. 2013. Exploratory analyses of local institutions for climate change adaptation in the Mongolian grasslands: An agent-based modeling approach. *Global Environmental Change,* 23**,** 1266-1276.

WEBBER, C. M. & LABASTE, P. 2009. *Building competitiveness in Africa's agriculture: a guide to value chain concepts and applications*, World Bank Publications.

WHITE, J. W., RASSWEILER, A., SAMHOURI, J. F., STIER, A. C. & WHITE, C. 2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos,* 123**,** 385-388.

WILENSKY, U. 1999. NetLogo: Center for Connected Learning Comp.-Based Modeling. *Evanston, IL: Northwestern Univ,* 158.

WORLD BANK. 2016. *World Bank Country and Lending Groups* [Online]. World Bank. Available: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519.

WORRAPIMPHONG, K., GAJASENI, N., LE PAGE, C. & BOUSQUET, F. 2010. A companion modeling approach applied to fishery management. *Environmental Modelling & Software,* 25**,** 1334-1344.

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

YANG, L. & GILBERT, N. 2008. Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in complex systems,* 11**,** 175-185.

ZHANG, T. & BRORSEN, B. W. 2010. The Long-Run and Short-Run Impact of Captive Supplies on the Spot Market Price: An Agent-Based Artificial Market. *American Journal of Agricultural Economics,* 92**,** 1181-1194.

ZHENG, C., LIU, Y., BLUEMLING, B., CHEN, J. & MOL, A. P. J. 2013. Modeling the environmental behavior and performance of livestock farmers in China: An ABM approach. *Agricultural Systems,* 122**,** 60-72.

ZHENG, C., LIU, Y., BLUEMLING, B., MOL, A. P. J. & CHEN, J. 2015. Environmental potentials of policy instruments to mitigate nutrient emissions in Chinese livestock production. *Science of The Total Environment,* 502**,** 149-156.

ZIMMERMANN, A., OUML, HRING, A., MACK, G., FERJANI, A. & MANN, S. 2015. Pathways to Truth: Comparing Different Upscaling Options for an Agent-Based Sector Model. *Journal of Artificial Societies and Social Simulation,* 18**,** 11.

8 Appendices

Appendix 1 Questions to validate model assumptions and identify input parameters

This section presents some of the questions that are used to validate assumptions in the base model.

1. How old are you now?
2. Is dairy cow farming the only source of income for your family? (a) Yes; (b) No
3. If you have other sources of incomes, please estimate the proportion of these other incomes from your total income?
4. Do your family members assist you in dairy farming? (a) Yes, How many people? (b) No
5. Did you ever hire outside workers (not your family member) to assist you in dairy farming? (a) Yes; (b) No
6. If yes, then how many outside workers did you usually hire?
7. If yes, then for how many days in a month did you hire this outside worker?
8. How many cows do you have? Please also mention the age of the cows that you can remember well.
9. How many bulls do you have? Please also mention the age of the bulls that you can remember well.
10. Please choose one of your cows which characteristics you can remember best. What is the highest milk production ever produced by that cow? At what age does that cow reach this level of production? How old is that cow now? How much milk does it produce currently? How many times should the artificial insemination be given to make that cow pregnant?
11. How do you acquire the forage for your cattle (you can give more than one answer)? (a) I grow the forage; (b) I collect the forage from the areas

surrounding my village; (c) I hire other people to collect the forage; (d) I buy the forage from a trader; (e) other, please explain

12. If you grow the forage, then how much land do you have (your own land and/or land that you rent)?
13. Can you produce sufficient forage to feed all of your cattle? (a) Yes; (b) No
14. If your answer is no, then how much land you supposedly have to produce sufficient forage (assume that the forage growth rate is constant)?
15. On average how much forage do you give to all of your cattle in one day?
16. Do you also feed your cattle with additional fodder? (a) Yes; (b) No.
17. If your answer is yes then how do you acquire the additional fodder for your cattle (you can give more than one response)? (a) I produce it by myself; (b) I buy it from supplier; (c) other, please explain
18. If you buy the additional fodder from the supplier, what is the price per kilogram of additional fodder?
19. On average how much additional fodder do you give to your cattle in one day?
20. If you have sold a cow in the past year, then what is the highest selling price you receive? How old was your cow when it is sold?
21. If you have sold bull in the past year, then what is the highest selling price you receive? How old was your bull when it is sold?
22. If you have bought heifer in the past year, then what is the highest buying price you pay?
23. From your experience, what is the highest milk price you received in the past year?
24. From your experience, what is the lowest milk price you received in the past year?

25. The followings are some factors that might force you to quit dairy farming. Please state how big the likelihood of these factors to force you to quit from dairy farming (5 very likely, 1 very unlikely)

a) My children do not want to continue my dairy farming business	1	2	3	4	5
b) The income from dairy farming is not sufficient for daily living	1	2	3	4	5
c) It is become more difficult to obtain sufficient forage	1	2	3	4	5
d) Competition with new farmers become more intense	1	2	3	4	5
e) The cattle quality (genetic breed) is decreasing	1	2	3	4	5

Appendix 2 Scenarios validate and calibrate agent's decision rules

Scenario 1: These scenarios are used to validate and calibrate buying decision rules

Scenario 1a: In the current condition in which you can collect forage (respondent's answer to question 15 in Appendix A) and milk price of (the mid value of respondent's answer to question 23 and 24 in Appendix A), how many more cows do you want to buy, suppose you have enough money to buy the cows and to increase your pen capacity?

Scenario 1b: Please imagine a condition in which the forage availability has increased drastically. With the same amount of labour and time, you can collect twice as much forage as the forage you can collect at this time. However, the milk price you receive stays the same. If you have enough money to buy new cows and to increase your pen capacity, then how many new cows do you want to buy?

Scenario 1c: Suppose the forage availability stays the same but the milk price is double. If you have enough money to buy new cows and to increase your pen capacity, then how many new cows do you want to buy?

Scenario 1d: Please imagine a condition in which the forage availability has increased drastically. With the same amount of labour and time, you can collect twice as much forage as the forage you can collect at this time. In addition, the milk price is also double. If you have enough money to buy new cows and to increase your pen capacity, then how many new cows do you want to buy?

Scenario 2: These scenarios are used to validate and calibrate selling decision rules

Please imagine that you only have one cow. Unfortunately, you are facing drought in the last 7 days and during this period you can only satisfy 75% of the forage needed by your cow. When the veterinarian come for his regular visit, he tells you that there is 25% chance of your cow will be sick and die tomorrow. Soon after the veterinarian leaves, you receive a call from a butcher, offering to buy your cow for 15 million. This price is acceptable considering your cow live weight. If you accept the butcher's offer while the veterinarian's prediction does not happen then you lose your potential future income. On the other hand, if you decline this offer and the veterinarian's prediction happen then you will not get anything. In this condition which action will you take?

(a) To sell your cow; (b) to retain your cow; (c) Other, please explain

Notes: The drought period variation is 7 days, 1 month, and 2 months. The forage sufficiency variation is 0%, 50% and 75%. The probability to die variation is 0%, 25%, 50% and 75%. If it is difficult for the respondent to imagine probability using percentage, then the information is rephrase using odds (e.g., in one occasion your cow will die and in 3 occasions your cow can survive).

Scenario 3: These scenarios are used to validate and calibrate cow selection decision rule

Please imagine that you have only two cows. You are currently experiencing financial difficulties and are unable to get help, hence you need to sell one of your cows. The money from selling one of these cows can meet your current needs. The first cow is young, currently, it is not pregnant but it can get pregnant easily when given artificial insemination. Your second cow is old, from your record it is hard to get pregnant when it is artificially inseminated, but currently it is pregnant. Which cow do you prefer to sell?

Notes: The age variation is old and young. The pregnancy variation is pregnant and not pregnant. The fertility variation is easy and hard. To make the comparison easier, sometimes the respondent must be asked to define the cut-off point between old and young or easy and hard to get pregnant (e.g., less than four years is considered as young).

Appendix 3 Statistical analysis of empirical buying decision rules

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Milk Price Per Liter (IDR/Liter), Sqrt Additional Forage Obtained (Kg-2)[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Additioanal Cow Wanted (Head)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.384[a]	.147	.136	3.75653

a. Predictors: (Constant), Milk Price Per Liter (IDR/Liter), Sqrt Additional Forage Obtained (Kg-2)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	361.219	2	180.609	12.799	.000[a]
	Residual	2088.503	148	14.112		
	Total	2449.722	150			

a. Predictors: (Constant), Milk Price Per Liter (IDR/Liter), Sqrt Additional Forage Obtained (Kg-2)

b. Dependent Variable: Additioanal Cow Wanted (Head)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.603	4.889		-.328	.743
	Sqrt Additional Forage Obtained (Kg-2)	.095	.019	.376	4.953	.000
	Milk Price Per Liter (IDR/Liter)	.001	.001	.084	1.108	.270

a. Dependent Variable: Additioanal Cow Wanted (Head)

Appendix 4 Statistical analysis of empirical buying decision rules

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	612.398			
Final	82.996	529.402	3	.000

Pseudo R-Square

Cox and Snell	.554
Nagelkerke	.750
McFadden	.601

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	249.166	166.171	1	.000
Prob_Die	612.353	529.357	1	.000
Drought_Length	83.005	.010	1	.922
Forage_Sufficientcy	83.021	.025	1	.873

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

Parameter Estimates

Decision[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Sell	Intercept	-6.342	.646	96.408	1	.000			
	Prob_Die	11.442	.897	162.722	1	.000	9.317E4	16061.248	540526.900
	Drought_Length	.001	.006	.010	1	.922	1.001	.988	1.013
	Forage_Sufficientcy	.106	.666	.025	1	.873	1.112	.301	4.106

a. The reference category is: Retain.

Appendix 5 Statistical analysis of empirical sorting decision rules

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	High_Fertility, Not_Pregnant, Young_Cow[a]	.	Enter

a. Tolerance = .000 limits reached.

b. Dependent Variable: Priority

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.989[a]	.978	.978	.33753

a. Predictors: (Constant), High_Fertility, Not_Pregnant, Young_Cow

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2630.125	3	876.708	7.695E3	.000[a]
	Residual	57.875	508	.114		
	Total	2688.000	511			

a. Predictors: (Constant), High_Fertility, Not_Pregnant, Young_Cow

b. Dependent Variable: Priority

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.766	.030		92.701	.000
	Young_Cow	4.000	.030	.873	134.076	.000
	Not_Pregnant	-1.750	.030	-.382	-58.658	.000
	High_Fertility	1.219	.030	.266	40.851	.000

a. Dependent Variable: Priority

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Old_Cow	.[a]	.	.	.	.000
	Pregnant	.[a]	.	.	.	.000
	Low_Fertility	.[a]	.	.	.	.000

a. Predictors in the Model: (Constant), High_Fertility, Not_Pregnant, Young_Cow

b. Dependent Variable: Priority