Master Thesis

April 11, 2022

Diffusion of Innovation among Smallholder Farmer Households in Tanzania: An Agent-based Modelling Approach

Marc Zwimpfer

of Zürich, Switzerland (16-713-885)

supervised by

Prof. Dr. Lorenz Hilty

Dr. Matthias Huss

University of Zurich[UZH]

Master Thesis

Diffusion of Innovation among Smallholder Farmer Households in Tanzania: An Agent-based Modelling Approach

Marc Zwimpfer

University of Zurich[UZH]

Master Thesis

Authors: Marc Zwimpfer

Project period: 11.10.2021 - 11.04.2022

Informatics and Sustainability Research Group
Department of Informatics, University of Zurich

Acknowledgements

First of all, I would like to thank Dr. Matthias Huss for his support throughout my master's thesis. I especially thank him for all meetings during my thesis in which he provided me with valuable background information on the field study in Tanzania.

I would like to thank Prof. Dr. Lorenz Hilty for his valuable feedback during the course of this thesis and for allowing me to write my thesis with the Informatics and Sustainability Research Group.

Lastly, I thank my family and friends who supported, motivated and helped me throughout my time at the University of Zurich.

Abstract

Poverty and hunger still prevails in many parts of the world. Sub-Saharan countries are particularly affected by severe food insecurity due to heavy reliance on agriculture and lack of sufficient farming equipment. Thus, finding a cost-effective method to introduce improved equipment to a high percentage of farmers in reasonable time is crucial. For this reason, we develop a modular agent-based model which simulates innovation diffusion among smallholder farmer households in Tanzania. It is based on proven innovation research as well as recent findings in an ongoing field study of the University of Zurich in Tanzania. Furthermore, we define different intervention strategies to accelerate the diffusion rate of an innovation among farmers. We outline how our model can be utilized to evaluate and compare such strategies in various ways. By applying algorithms from machine-learning on our diffusion simulation results, we show how key factors behind the performance of a strategy can be determined and demonstrate possibilities to predict the success rate of a strategy. Our findings present researchers an inexpensive alternative to assess intervention strategies effectiveness before launching them in the field.

Zusammenfassung

In vielen Teilen der Welt herrschen immer noch Armut und Hunger. Die Länder südlich der Sahara sind besonders stark von Ernährungsunsicherheit betroffen, da sie fest von Agrikultur abhängig sind und nicht über ausreichende landwirtschaftliche Ausrüstung verfügen. Daher ist es von entscheidender Bedeutung, eine kosteneffiziente Methode zu finden, um einen hohen Prozentsatz der Kleinbauern in angemessener Zeit mit verbesserter Ausrüstung auszustatten. Aus diesem Grund entwickeln wir ein modulares agentenbasiertes Modell, dass die Innovationsverbreitung unter Kleinbauernhaushalten in Tansania simuliert. Es basiert auf bewährten Erkenntnissen der Innovationsforschung sowie auf aktuellen Ergebnissen einer laufenden Feldstudie der Universität Zürich in Tansania. Darüber hinaus definieren wir verschiedene Interventionsstrategien, um die Verbreitungsgeschwindigkeit einer Innovation unter Kleinbauern zu beschleunigen. Wir erläutern, wie unser Modell genutzt werden kann, um solche Strategien auf verschiedene Weise zu bewerten und zu vergleichen. Durch die Anwendung von Algorithmen aus dem Bereich des maschinellen Lernens auf unsere Diffusionssimulationsergebnisse zeigen wir, wie die Schlüsselfaktoren für die Leistung einer Strategie bestimmt werden können, und demonstrieren Möglichkeiten, die Erfolgsrate einer Strategie vorherzusagen. Unsere Ergebnisse bieten Forschern eine kostengünstige Alternative, um die Wirksamkeit von Interventionsstrategien zu bewerten, bevor sie in die Praxis umgesetzt werden.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Poverty and hunger still prevails in many parts of the world. Even though ending it and providing secure access to food is a major objective of the 2030 Agenda for Sustainable Development, the number of people suffering from hunger did not decrease since its publication despite various measures worldwide [1,2]. Therefore, continuous efforts to reach the goals of the agenda must be made.

One of the regions which is heavily affected by hunger and poverty is Sub-Saharan Africa (SSA), where almost one third of the population suffer from it. There are various environmental, political, and social reason that this region and the countries in it are situated at the bottom of worldwide health and poverty rankings. One reason for this and the consequential food insecurity, is the strong reliance on agriculture in SSA. In average, more than 60% of the population are involved in smallholder farming which is their only source of income [3,4]. However, a lack of adequate farming equipment and techniques leads to seasonal food insecurity. Especially in the time directly before the harvest, called "lean" season, this condition becomes most severe [5, 6]. There are multiple reasons which lead to seasonal food shortages. One string of research focuses on the lack of adequate post-harvest storage possibilities, which results in price fluctuations throughout seasons of a crop [5, 7, 8]. Right after the harvest, there is an oversupply of the respective crop, which causes the market price to drop and forcing the farmers to sell their crops at a lower price. Due to the lack of adequate on-farm storage, small scale farmers cannot store their crops to be self-sufficient, or only at a risk of losing it to pests. As a consequence, market prices rise for the following months, reaching their peak before the next harvest. Due to this, small scale farmer who were not able to store enough of their harvest, must buy crops back at a much higher price point or, in many cases, suffer from hunger. Preliminary studies have shown that there exist cost-effective solutions which could prevent post-harvest losses and stabilize food insecurity as a result [5,7,9]. An example of such a solution are hermetic bags which allow for substantial longer on-farm crop storage at a very low price point. However, to stabilize the whole regional market and decrease the overall food insecurity of the population, these solutions must experience widespread adoption. This poses the question of how to achieve the most effective and efficient distribution of awareness about innovative products in SSA. Innovation diffusion has been studied for decades and in numerous fields of study [10–12]. This research focuses on the processes happening when an innovation is introduced into a system on a macro- and micro-level. Diffusion of innovation can be analyzed and described in different ways but in the last years the appearance of modern tools enable large-scale simulations on a micro-level [13,14]. This allows observing innovation diffusion as a result of a multitude of decisions of independent individuals, rather than describing the diffusion on a system level. Agent-based modelling is a type of modelling technique which is based on the concept of individuals acting for themselves without outside guidance. Agent-based models (ABM) have been used extensively in simulations and predictions [15–17]. An ABM consists of autonomous agents which together from a complex system. From the interactions of these agents, an understanding of the system can be obtained which otherwise would not be possible. As these agents move freely, the outcome of the model is not predictable and may lead to system states which were not thought of beforehand. While the significance of results generated by such models has been questioned, they provide a straight-forward way to create models and obtain initial findings [14].

In the scope of this thesis, we combine agent-based modelling techniques with innovation research all in the context of SSA regions. We present an agent-based model which depicts innovation diffusion in rural areas of Tanzania among smallholder farmer households. Our model will be based on findings from Tanzania, which is a country located in SSA and heavily reliant on farming [3,18]. Furthermore, a large part of its population, living in rural areas, have been more or less frequently exposed to food insecurity [5]. This thesis is embedded in an ongoing field study of the University of Zurich based on Brander et al. [5,7] previous research which analyzes the effectiveness of improved on-farm storage in reducing food insecurity. This leads to the first research question of this thesis.

RQ1 How can innovation diffusion among smallholder farmer households be modelled in Sub-Saharan regions?

We create an ABM which incorporates the social and geographical features of Tanzanian smallholder farmers. It is based on a social network between farmers which autonomously interact with each other and propagate awareness of the innovation through the system. In addition, many processes and agent features are based on core findings from innovation diffusion research. The conceptual model design is based on observations and interviews with Tanzanian farmers participating in the current field study.

In order to reduce food insecurity, innovative products and techniques supporting sustainable farming must be adopted by a large number of smallholder farmers. However, due to monetary reasons, a widespread adoption should be achieved with an cost-effective approach. With our implemented ABM, different intervention strategies can be evaluated and compared, which leads to the second research question of this thesis.

RQ2 How can different intervention strategies be evaluated and compared?

Finding intervention strategies which are cost-effective and successful is crucial for achieving widespread adoption of farming advancements and as a result, reduce food insecurity in a sustainable manner. The major contribution of this master thesis is an ABM which, in a generic way, models the knowledge and innovation diffusion in an SSA system. Because of the modular way, in which it is built, this model can be easily adapted for other research purposes and extended with further findings from field research. Moreover, we provide an inexpensive possibility to test different intervention strategies before executing them in the real world. As stated by Rand and Rust [14,19], predictions obtained from agent-based models should be treated with caution, but are proven to reveal certain trends in the system. Hence, our model can be utilized, for example, to inexpensively narrow the potential range of strategies down. The remaining strategies than may be tested in the real world.

Overview This thesis is structured into six chapters. After this introduction, we review relevant literature related to the topic of this thesis in chapter two. The third chapter establishes the necessary theoretical background for the implementation of our ABM. The fourth chapter introduces the agent-based model developed in this thesis and the chosen design decisions. In the fifth chapter, the results obtained by running simulations on the model are analyzed and interpreted. Lastly, with the findings of the previous chapters, the research questions are answered and a critical review of the results is performed.

Chapter 2

Related Work

We present other research related to the topic of this thesis in this chapter. Firstly, literature on the topic of innovation diffusion with focus on agriculture and developing countries is presented. Afterwards, research regarding ABM related to agricultural innovations is discussed.

Innovation research, especially the topic of diffusion of innovation, is a very active field of study. Most of nowadays research is based, at least partially, on Rogers famous publication "Diffusion of Innovations" in 1962 [10]. Furthermore, extensive research is conducted on the diffusion of agricultural innovations (e.g. farming techniques or sustainable land-use) in all regions of the world.

Meijer et al. [11] propose an analytic framework to examine the adoption process of agricultural innovations. They argue that besides extrinsic factors, like farmer characteristics and the external environment, intrinsic factors also influence the adoption decision of farmers. Intrinsic factors include knowledge, perception and attitude towards an innovation of potential adopters. In essence, their framework states that extrinsic variables influence intrinsic variables of potential adopters and hence both types of variables work in tandem.

Another interesting aspect of innovation diffusion and adoption in rural African areas is determined by Iiyama et al. [12]. Following interviews with smallholder farmers - potential adopters - they noted that, even though preferred by farmers, some products do not have high adoption rates. This is caused by the very limited resources of farmers in such regions and hence, removing uncertainty about the innovation from farmers should be a key objective of successful launch of new products.

Various research shows that decision of farmers to adopt a agriculture related innovation is not solely based on economic considerations but on other psychological factors too [20–22]. It is shown that the fundamental opinion of the farmer regarding the context of the innovation, for example organic farming, has significant impact on the likelihood of adoption. In addition, social pressure to adopt the innovation, from other farmers as well as the public, should not be underrated concerning the adoption rate. In summary, their research reveals that, even though being the predominant decision factor, other non-economic aspects should be considered as well.

Using ABM in the context of diffusion of innovation has been successfully achieved in various field of studies. In the last years, different frameworks for ABMs simplified their utilisation [23–25]. With more widespread adoption and larger models, design and implementation guidelines become more important. Sun et al. [19] state the difference between model "complexity" and "complicatedness". A system's complexity describes how difficult it is to analyse the behavior of an ABM system. A model becomes more complicated with an increasing number of agent types, parameter and other entities. Hence, this describes the level of difficulty of the structure of the model. They conclude that with more empirically grounded models, the "complicatedness" of model tends to increase and thus the generalizability of the results decreases if not careful. Hence models, in order to make predictions or used as decision making support tools, should be kept as simple as possible.

Rand and Rust [14] propose guidelines which should be used to rigorously create and use agent-based models in the context of marketing. They note that without adequate rigour of the model, it is not possible to compare results from different ABMs in order to obtain signification findings. Thus, they present, and illustrate with an example, a set of standards which should be used when implementing an ABM with a focus on verification and validation of the model.

Other research focuses on social interactions and agriculture and how to incorporate these topics in ABM. The application of ABM in innovation research was reviewed multiple times [26, 27]. It shows that there is a major interest in modelling innovation diffusion based on micro-level decision, as done in ABM, compared to more traditional mathematical or statistical approaches.

Manson et al. [15] present a novel approach to model the adoption of rotational grazing techniques in farming with help of an ABM. Their model is based on strong and weak links between different agents and the concept of creating a social network of all involved parties. Furthermore, they classify what type of information can be obtained from different links and how this influences the final adoption decision. In a qualitative approach by interviewing farmer, they further show how important a social network as well as communication is in innovation diffusion and in the agriculture sector [28]. Similarly, Diemer et al. [29] determined that credibility of the informant is crucial and that agents tend to pay more attention to information about the innovation gathered from reliable and known sources. In addition, they find that agents enter an "information need" phase, in which they are actively searching information about an innovation and are more prone to accept information from other sources which matches with Rogers research [10].

Deffuant et al. [16] proposed an agent-based model to investigate innovation diffusion of organic farming. The context of their model is the empirical data of organic farming techniques in the French department of Allier. A central part of their model is the social interactions between farmers and how opinions change accordingly. Furthermore, the agent network used in the model is inspired by the concept of "small-world" networks [30]. In such networks, peers with close spatial proximity are significantly more likely to know each other than ones which are far apart from each other. Their selection and calibration of parameters followed thorough experimenting in an attempt to correlate with real-world data. However, compared to the gathered reference data, the adoption level predictions with their model were in average too high. They reasoned that characteristics and overall initial attitude towards the innovation in discussion heavily influences the actual rate of adoption. Later on, Deffuant et al. [31] present an evolution of their model [16] which is applicable to a more general process of innovation diffusion. In their model, agents and their attitude towards an innovation follow a fixed state transition scheme. The states of the agents are based on their individual interest and their changing level of uncertainty over time. At its base, the model reflects the ideas of social opinion and individual advantage in the decision making process as already introduced by earlier threshold models [10, 32]. Threshold models introduce the concept that the decision outcome of an individual in a network is based on the percentage of other peers in the network which already made their decision. Hence, every peer in the network posses an individual threshold when it considers accepting a change already performed by others. Under these assumption, their model indicates that innovation with low social value and high individual benefit perform worse than innovations with high social value and low individual benefit. In addition, they model communications between peers which have a negative impact on the attitude towards the innovation. Many other models tend to focus only on positive word-of-mouth (WoM) between peers. The high importance of WoM is undisputed in innovation diffusion research and is recognized as a key driver of fast diffusion [33–35]. Hence, animate individuals to mention the innovation to others should be at the core of all marketing strategies.

In a similar, more recent study, Ambrosius et al. [17] discuss the diffusion of organic farming in the case of Dutch pig farmers. The foundation of the model is a socio-spatial network of pig farmers with different sized farms and farming styles. Moreover, there exists a measure of similarity between farmers in terms of farming style, farm size and whether to sell to organic or conventional food markets. In case two farmers in an interaction are similar, their attitude towards the innovation converges and vice-versa. By modelling a realistic meat market, they show that the most important factor to adopt organic farming is the actual demand of organic meat and therefore increasing the profit margin. This decision process of the farmers in the model is outlined in detail in a previous work of Ambrosius et al. [22] and shows that, although being mostly influenced by the consumer demand, the diffusion also can be guided by directly addressing the heterogeneous groups of farmer with different farming styles.

Chapter 3

Theoretical Background

Designing a model concerning the innovation diffusion in smallholder agriculture requires detailed background information about various related topics. This chapter provides an information base and literature review on which the later design decisions of the model are based. Firstly, relevant information on Tanzania, as an example for an SSA country, is presented which influences the structure of the ABM. This follows a summary of the most relevant findings of innovation research. Eventually, the base concepts of ABM as well as the most important frameworks to implement an ABM are presented.

3.1 Tanzania

The United Republic of Tanzania (Tanzania) is a sovereign Sub-Saharan country which is situated in East-Africa and borders the Indian-Ocean [36]. Tanzania has a population of about sixty million people, which makes it the largest East-African country by population. Furthermore, as many states in this region, the largest part of the population is young (below 25 years of age) caused by the high fertility rates. The official languages are Swahili and English, however, many regional dialects and languages exists, which may differ radically between different regions of the country. Most of the residents follow a Christian religion followed by a Muslim belief [18].

Administration

Tanzania is governed by an elected president and a cabinet. An exception is Zanzibar, which is an autonomous island state in the Indian ocean, which has its own parliament. The United Republic of Tanzania, which includes mainland Tanzania as well as Zanzibar, is divided into 31 regions, each with its own administration as shown in figure 3.1. Furthermore, regions are divided into a total of 169 districts where each district contains its own district council. Large urban areas and cities are considered as districts themselves. Therefore, there are different types of districts, namely Cities, Municipalities, Towns and "plain" districts. These districts vary a lot in terms of population count and area.

Districts are split further into wards. The term "ward" does mean something different in urban and rural districts. While in urban districts a ward is normally a set of streets, in rural districts a ward is composed of villages. Villages can than be further divided into Hamlets, which is the smallest settlement group in Tanzania. Administratively, a village is the smallest considered structure. There exist some exceptions to this structure, in which the administrative division does not follow the above explained logic (e.g. the capital city Dar es Salaam) [4, 37]. The Tanzania

Figure 3.1: Map of Tanzania with marked districts Kondoa and Kilosa, in which the ongoing field study of the University of Zurich is located. Source: `https://www.openstreetmap.org/`

National Bureau of Statistics provides geographical data of administrative areas up to a ward level [4].

Economy

Compared to other countries in this region, Tanzania achieved an over-average economic growth in the last years of about 6% per year [18]. Only the recent pandemic stopped the growth rapidly, but it is projected to return to its original value in the foreseeable future. The economy of the Tanzania and its GDP is heavily reliant on the Agricultural Sector which makes up more than a quarter of the annual GDP. The primary agricultural product is maize followed by variety of other food crops. Coffee and cotton are the two most produced cash crops with are exported in large quantities. Most agriculture activity is performed by smallholder farmers [38] which are organized in farm groups. Furthermore, smallholder farmer household use a large portion of their harvest as self-sufficiency. Besides agriculture, mining for rare metals and gemstones for export make up a big proportion of the primary work sector. The secondary work sector is made of mostly of processing goods from agricultural resources.

Additionally, tourism is a major source of income for some regions of Tanzania, especially Zanzibar. Other regions which profit from tourism mostly are located in northern regions because the Kilimanjaro mountain, which is the highest mountain on the African Continent and very frequently visited by Tourists. Additionally, a number of National Parks are located in this region which act as tourist attractions too.

Poverty

Although the country has a steady decline of people living in poverty, still around a quarter of the population live below the poverty line [37]. However, the over-average economic growth did not lead to an equal decrease of people living in poverty. All in all, around 26% of the population are considered poor (people who are not able to reach their basic consumption need) of which 8%

are considered as extremely poor (people who do not take at least 2'200 kcal in per day, which is the daily minimum nutritional need of an adult).

These values differ heavily between rural and urban areas as well as different regions. While the percentage of poor people in urban areas lies just below 16%, in rural areas more than 31% of the populations is considered as poor. Urban and rural differences can be observed in various factors, for example education level, easy access to water or electricity and level of available sanitation. In all these categories, urban areas perform considerably better than rural areas. Differences can also be observed between regions in Tanzania, where the poverty ranges from above 50% to as low as 8%. The poverty of a region correlates in many cases to the level of urbanization of it, which causes different level of market access. Lower market access has several disadvantages, for example higher transportation costs and less option to sell output to the market. This can result in heavily fluctuating market prices for goods and rural farmers may be forced to sell their goods at prices below the actual market price. Low market access in rural regions is mostly caused by the poor infrastructure in Tanzania. Again, there are major differences between road and public infrastructure in rural and urban regions [37]. While in urban areas different kinds of transportation exist, in rural areas the road infrastructure is in a very poor conditions and public transport is unreliable [39]. Therefore, the level of travel by inhabitants of these regions is rather low [40]. For farmers, this lack of reliable transportation means is very problematic as they are not able to sell their products in reasonable time or acceptable prices. As a result, either they have to sell their harvest below market-price due to the dependency on transporter or the harvest goes bad if no acceptable seller is found in time. Hence, the lack of transportation possibility and overall bad road infrastructure lead to inefficient farming in those regions.

Even though many programs to reduce poverty exist, the decline of poor people in Tanzania slowed down in the last years. The poverty percentages decrease overall, however, in absolute numbers more people live in poverty than before [37]. This is caused by an high population growth - especially in rural areas. Due to the lack of old-age insurances in these areas, it is common that children are considered as such an insurance and a bigger family is planned. This results in a viscous circle, as these children than grow up and live in poverty.

Ongoing Field Study to Reduce Food Insecurity

This thesis is embedded in an ongoing field study of the University of Zurich in Tanzania analysing the adoption of novel on-farm storage methods [5, 7]. The field study is situated in two districts of Tanzania shown in figure 3.1. In the previous phase of the project, the advantages of improved on-farm storage were analysed in a field study with over a thousand participating smallholder farmers. They provided households with hermetic storage bags which is an inexpensive on-farm storage technology. Hermetic storage bags allow harvested grains to be stored in a, from atmospheric oxygen sealed, way which lead to the suffocation of vermin contained in the harvested crops. Hence, the post-harvest loss is minimized in contrast to the conventionally used storage techniques.

Without adequate storage option, farmers are forced to either sell their harvest shortly after harvest or store crops but risk damage by rodents or other types of pests. This leads to an oversupply of crops right after the harvest and a consequential market price drop. Consequentially, crop prices rise steadily until the next harvest. Farmers, which could not store enough crop for self-sufficiency, are forced to buy overpriced grains which leads to food insecurity.

In a randomized control trial, the research team found evidence that the seasonal food insecurity was significantly reduced in household using hermetic storage bags. With SMS-based surveys, they found that especially during lean season, the proportion of household suffering food insecurity could be reduced by 38% on average. However, even though the advantages of hermetic storage bags are undisputed, adoption rates remained low which must change in order that food insecurity can be reduced on a large scale.

3.2 Innovation Research

Diffusion of innovations (DoI) has been researched for more than a century. Everett Rogers is accounted for as the popularizing this theory in 1962 [10]. DoI describes the adoption process of a social system of an new product over time. At its core, Rogers defines DoI to be universal processes of social shift rather than be caused by peculiarities of the innovation itself. Hence, it is possible to abstract the diffusion process from the innovation itself and define more generic approaches.

There are certain key features Rogers and other researches have defined which are of value to this thesis and explained in more detail in the following sections.

3.2.1 Key Components

Rogers defined four main components of DoI which cause the diffusion to happen in a particular manner [10].

Innovation At the core of the process lies the innovation itself. An innovation is an object which is perceived as new by potential adopters. It is important to note that the innovation itself must not be entirely new, just perceived as new by intended consumers. The key characteristics of an innovation, which influence its diffusion, are its relative advantage for the adopter, its complexity, its compatibility with already existing objects in the same system as well as its testability by potential adopters. The perceived sum of these characteristics decide whether an individual adapts or rejects an innovation. The less uncertainty in each category is present, the likelier an adoption may be. Hence, reducing uncertainty should be a priority when releasing an innovative product.

Communication Channels According to Rogers, the final adoption decision is not only based on your own preference but largely influenced by interactions with other individuals in your social system. Potential adopters share information about the innovation via various channel in order to reach mutual understanding of it. Different types of communication exists which can either be one-way channels, like mass media, or two-way channels, like known as word-to-mouth communications.

Time Prior to Rogers, most behavioral research neglected the importance of time on the adoption process. Adoption does not happen at an instant but is continuous over time. Most notably, time is prevalent in the innovation to decision process which is the interval between first knowledge of the innovation to the opinion-formation of the potential adopter.

Social System DoI does not happen in a vacuum but is embedded in a social system. A social system is a set of interrelated units which try to accomplish a common goal [10]. As diffusion of an innovation is located in a social system, it is influenced by the social structure of said system. For an innovation to be successful and prevail, it must reach a certain level of adoption in a social system which is called the "critical mass". Failing to reach this leads to a non-sustainable adoption and the innovation is lost over time and not adopted by the majority. Additionally, the type of social system, its "culture", influences to chance of widespread adoption heavily.

3.2.2 Innovation-Decision Process

The innovation-decision process presented by Rogers consists of five different phases which follow each other mostly chronologically. This process outlines through which stages an individual undergoes when deciding over the adaption of an innovation. In figure 3.2, the different stages of the decision process are shown and summarized. This process happens at the micro level of potential adopters.

In the *knowledge (or awareness)* stage, an individual first hears about an innovation through different communication channels. Rogers differentiates between three types of knowledge of an individual. Awareness-knowledge is the simple knowledge about the existence of an innovation and lies at the foundation of the further process. How-to-knowledge describes how to use the innovation correctly and could influence the adaption decision in a later stages. Lastly, principles-knowledge may not be important for the decision process but could cause the individual to discontinue using the innovation.

During the *persuasion* phase, an individual forms his own opinion of the innovation. The positive or negative attitude towards the innovation is based on the knowledge gained in the first stage as well as the subject feeling of the individual regarding the innovation. In this step, is information obtained by "friends" or trusted peers is relevant as on average this type of information is considered more trustworthy.

At this point, the individual makes a decision whether to adopt or reject an innovation. Both decisions must not be final as they can be changed later on, based on newly obtained knowledge. Rogers defines two types of rejection: active and passive rejection. While in passive rejection, the individual simply does not consider adopting an innovation at all, in active rejection, the innovation was tried or evaluated but could not convince the individual. An active rejection may also happen in a later stage when already using the innovation but discontinued after a while.

In the *implementation* stage, an individual, who decided to adopt the innovation, actually starts using the innovation as it is intended. During this stage, the effectiveness and usefulness of the innovation is still actively evaluated by the individual. If the innovation cannot convince the adopter, it may be still actively rejected and decided to discontinue to use it.

Eventually, in the *confirmation* stage, the adoption decision already happened. However, the adopter now searches for information to support his decision. On average individual in this stage are more receptive to information which justifies their decision than contradicting one. The attitude of the social network of an individual concerning an innovation may also influence the final continued usage decision of an potential adopter.

3.2.3 Diffusion Process on a Macro-Level

The previous section introduced the innovation decision process which outlines the adoption process on a micro-level. Additionally, it is possible to analyze a typical innovation diffusion on a macro-level. Two properties of this process are interesting to point out: the rate of adoption and the different types of adopters.

Rate of Adoption

The rate of adoption indicates the speed of the innovation diffusion. Hence, it resembles the percentage of the population which adopts an innovation over time. Figure 3.3 shows the typical S-shaped curve for the adoption rate.

First, only few individuals adopt the innovation. However, by increasing communication via various channels, the rate of adoption starts to grow exponentially after some time. The percent-

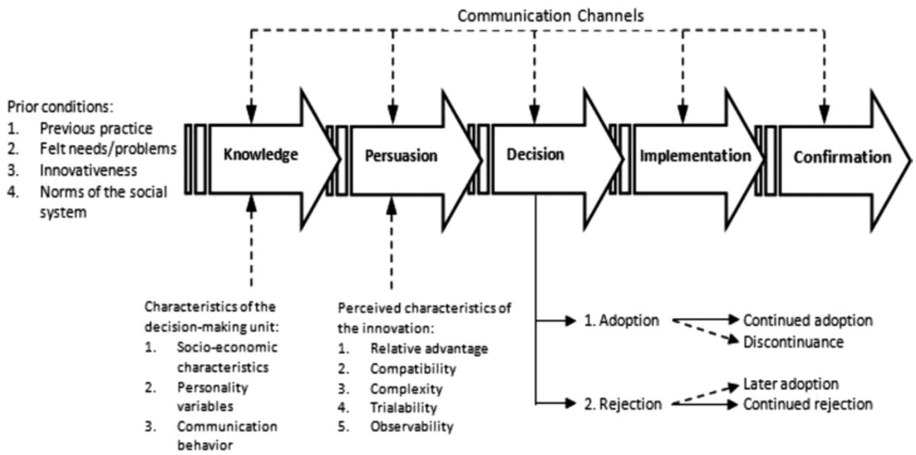

Figure 3.2: Visualized Innovation-Decision Process According to Rogers.

age of adopters in the population where the curve becomes exponential is referred to as "critical mass". After this point, the innovation diffuses on its own and does no longer need active marketing efforts. Reaching a higher percentage of adopters, the curve begins to flatten until the slowest adopters in the population use the innovation. Hence, reaching the "critical mass" rapidly is a major task in the process of introducing an innovation. Especially, in the beginning, a well defined marketing strategy is needed to reach this threshold.

One possible approach is to convince opinion leaders in the society to spread favourable information about the innovation in the system. Opinion leaders are individuals which have more influence on other participants than average. Furthermore, they tend to be more communicative over different channels and reach more potential adopters in shorter time.

Adopter Types

In the last section, the concept of opinion leaders was introduced. However, the individuals can be more generically grouped into five groups. The main difference between the groups is their respective time of adoption. Opinion leaders are mostly included in the first two adopter types. Figure 3.3 shows the frequency of each adopter type. Note that the distribution follows a normal distribution and correlates with the market share over time curve. The critical mass is reached at the conjunction between early adopters and the early majority.

Innovators Innovators are the first 2.5% of the population which adopt an innovation. They have a high willingness to experience new ideas and overall low risk aversion. In contrast to other types of adopters, they are willing to take the risk that the innovation is unsuccessful and make a loss. On average, innovators are over average communicative about the innovation and try to convince other groups to adopt the innovation. Mostly, innovators have a thorough (technical) understanding of the innovation. Innovators have a variety of contacts outside of the social system in which an innovation is introduced.

Early Adopters Early adopters pose the second type of adopters. Still, they are opinion leaders like innovators. Compared to innovators, they are more integrated in the targeted social system of the innovation. Moreover, they have a good understanding of the innovation and pose as role models for later types of adopters.

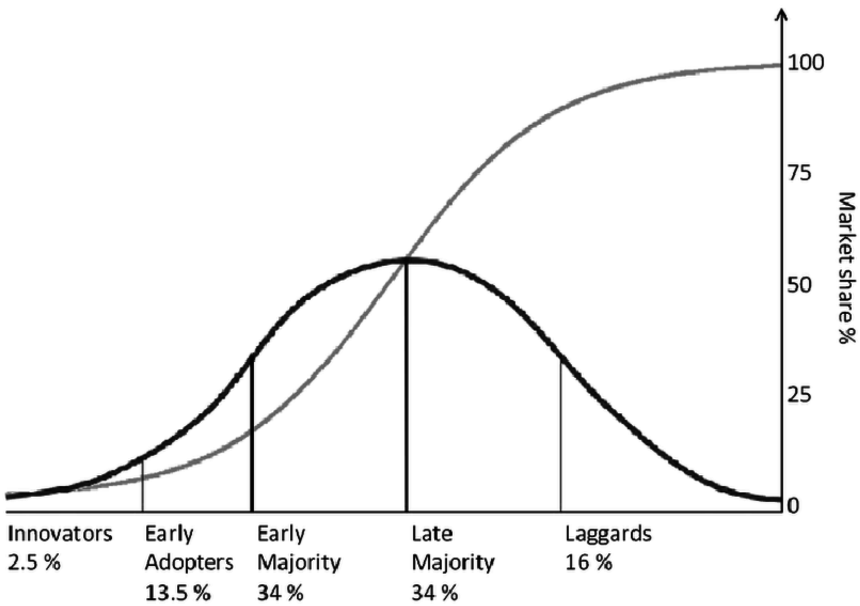

Figure 3.3: Distribution of Adopter Types and s-Shaped Adoption Curve According to Rogers.

Early Majority When the critical mass for an innovation diffusion is reached, the early majority begins deciding and adopting an innovation. They no longer are opinion leaders but are well integrated in the system and further facilitate diffusion. Furthermore, their adoption decision is based on lower risk tolerance than previous types and based on their personal advantage by it.

Late Majority The late majority is more sceptic of the innovation. They are also well integrated in the system but they have fewer resources compared to previous groups. Therefore, the only adopt innovations with low risk and low uncertainty. Often, their decision is based on social or economic pressure to accept an innovation not to feel left out.

Laggards Lastly, laggards may adopt an innovation. However, at this point the adoption rate already decreased significantly. They are often conservative and isolated in the social system. Additionally, their resources are even more limited than the late majority's and therefore they are not willing to take any risks when deciding whether to adopt an innovation.

3.3 Agent-based Modelling

With agent-based models, complex systems are modelled as a result of interactions between autonomous individuals [41]. In complex systems, many heterogeneous elements and their actions cause macro-level dynamics of the whole system. Even though the ideas behind ABM were found many years ago, only computational power from the last years made large-scale ABM simulations possible [42].

While traditional modelling techniques, for example equation-based modelling, try to grasp all system dynamics in a top-down approach by determining rules which are valid for the whole system, in ABM only the actions at a micro level are defined. Furthermore, equation-based modelling expects change to be continuous and the individuals to be homogeneous. However, considering models based on individual entities, their actions mostly are discrete and the individuals are heterogeneous.

In ABM, the modeler has to define decisions and interactions at an agent-level in contrast to other modelling approaches where macro-level phenomena must be directly modelled. In many cases, for example in early research, it is straightforward to deduct agent-level properties and behaviour from the real-life data. However, deriving aggregate behaviour and actions of a complex system as a whole is not as simple in most cases. Many trends or rules may not be determinable at the beginning, which may lead to a model distortion when not taken account for.

Another concept in which ABM differs from other modelling methods, is that it does not have to be deterministic but decisions can be based on randomness. In the real world, decisions of individuals may be mostly deterministic, however, they are influenced by many contextual factors which cannot all be incorporated in a feasible manner into the model. Especially, in early stages of a model design, not all factors and parameters of a system, which lead to a final decision of an agent, are known. This set of parameters can be taken account for in a best-effort approach by making decision be based on randomness or probability to some extend.

3.3.1 Core Entities in Agent-Bases Models

Agents At the base of an ABM lies the agent entity. An agent is an autonomous individual in the simulation with a specific behaviour [41]. It is possible to define different types of agents, which take different roles in the simulation, like as an example buyers and sellers in a market simulation. The agent's behaviour and its properties are determined by the modeler. However, this behaviour does not have to be deterministic but include different levels of randomness. Circling back to the market simulation example, the modeler may define that buyers contact sellers regularly. However, the agent "decides" on his own, which buyer to contact and at which time. This may lead to system changes, which are unbeknownst by the modeler at design time.

Environment The second core component of ABM is the environment in which the simulation happens [41]. Environments can take any form in a simulation, but most common are spatial or network environments. Spatial environments define a simulation space which implements some sort of coordinate system and introduce the concept of physical location in a model. On the other hand, in network environments agents do not have a physical location, but are represented as nodes in a network and their respective location is determined by the set of connections to other agents. Environments may have their own properties and affect the agent's interactions. If implemented, environment may change over time caused by agents interacting with them.

Interactions An ABM consisting of only agents and environments without interactions between those entities would not provide any value. There are five general types of interactions which are found in ABM [41]. Agent-self and agent-agent interactions, are actions of an agent with itself or another agent respectively. Environment-self and environment-environment interactions are the counterparts of the previous interaction types for environments. Even though, not as intuitive as agent interaction, environment ones can happen in different models, for example diffusion models, where the neighbouring environments influence each other. Lastly, agent-environment interactions are very frequent and cover all action between those two entities. In a sophisticated ABM, there are various interactions of all types found which complement each other.

3.3.2 Verification and Validation

For a model to be a useful abstraction of a real-world system, it is important that its outputs are accurate and representative. This can be achieved in two steps by verifying and validating the model [14, 41]. It is essential that a model passes both criteria as otherwise it is not accurate.

Verification Verifying a model is the process of checking it for its correctness. To test this criteria, only little knowledge of the underlying system must be known as implementation errors should be found. There exists different ways to check for correctness, but most common is testing. After implementation, unit-tests are added which each test for a single part of the model to work correctly. Later, user testing should ensure that the single parts work together as well. Furthermore, the mapping of the conceptual model to the model implementation can be checked and verified that the correct logic is executed. In summary, the verification process checks whether the conceptual model is correctly implemented but not whether the conceptual model is accurate itself.

Validation Although a model may be verified, it is not guaranteed that it actually corresponds to the real-life system. Validation can be split into four parts [14]. Microvalidation is the check if behaviour of individual agents and environments reflect their real-life analogs. In Macrovaldation, it is tested whether the macro patterns in the model are congruent with the ones in the real system. In ABM which need input data, an empirical input validation should be performed to ensure that the input data corresponds to the real world and is not biased. Eventually, an empirical output validation is needed to check the plausibility of the output of a model in relation to the real world. Especially, if comparable data sets exist, the output can be checked with these for its validity. In summary, the validation process checks whether the conceptual model is accurate abstraction of the real world system and results from it can be considered as plausible.

3.3.3 Agent-based Modelling Frameworks

There are multiple ways to implement an agent-based model. While a proprietary solution certainly provides most flexibility, using a proven ABM framework allows for standardization and profiting from the knowledge of various developers. Furthermore, frameworks have an optimized performance tailor made for agent-based modelling. Considering that ABM's main focus lies on agent acting independently, it is apparent why performance is important for large-scale simulations with a model.

This section introduces a selection of the most frequently used ABM frameworks when implementing a model. They all provide similar functionality for a straightforward development of ABM models. However, depending on the type of model and simulation context, certain frameworks offer an advantage in some areas. Hence, after the initial design of an ABM, the selection of a suitable framework is the foundation of a successful implementation.

Repast

Repast is a framework for ABM originally developed at the University of Chicago in 2001 [24]. Over time Repast evolved from a single modelling framework into a suite containing various tools for ABM. At this point in time, the Repast Suite [1] consits of *Repast Symphony*, *Repast for High Performance Computing* and the newest addition *Repast for Python*, which is still in its beta phase. The source code of all Repast components is publicly available and actively maintained.

[1] https://repast.github.io/

Repast Symphony is the further developed version of the original modelling framework. It is based on *Java* [2] and at its core provides class library for implementing and visualizing ABM simulations. Its flexibility, ease-of-use as well as its extendability led to a wide-spread adaption. *Repast for High Performance Computing* is a C++ version of the framework which is optimized for ABM simulation on large clusters of computers. The increased performance of this framework compared to *Repast Symphony* is achieved by parallelisation. Lastly,the *Repast for Python* framework is a port of the original *Java* based library to *Python* [3]. Due to the widespread adaption of *Python*, especially in research, this should lower the initial hurdle to use Repast.

NetLogo

NetLogo is programming language and modeling environment for agent-based simulation which was first released in 1999 [13,23]. As Repast, NetLogo is a open-source software and based on *Java*. As of now, NetLogo is the most prevalent ABM software and is actively developed since 1999 by different contributors. NetLogo provides a wide range of tools and functions which allow for an efficient model implementation.

The agent model of NetLogo is quiet straightforward. An agent is referred to as a "turtle" and poses the main entity of the model. "Turtles" are able to actively move and interact with other entities. "Patches" make up the environment of the model world. Each patch has a precise location but in contract to "turtles" cannot change it. A "link" is the third type of modelable entity, which are used to connect different "turtles". "Turtles" and "links" can be of different types, each with their own set of attributes. Lastly, the "observer" acts as the controller of the simulation and tunnels the instructions of the modeler to all entities.

The instructions for a simulation are defined by NetLogo primitives which either can be "commands" or "reporters". The difference between these two types of procedures, is solely that "reporter"-procedures return a value after its execution, while "commands" do not. Hence, in the remainder of this thesis, both terms will be referred to as "procedures". The modeler defines all instructions for the simulation in such procedures. In other terms, they form the business logic of the simulation. The NetLogo environment introduces a sense of time into the model with "ticks". Each "tick" poses as an abstract time interval and every tick NetLogo starts the predefined procedures. Moreover, if the core functionality does not suffice, a set of extensions is available for more complex instructions.

Additionally, NetLogo provides out-of-the-box interface UI elements which are directly accessible from the source code of the model. This reliefs the modeler from implementing an UI. Figure 3.4 shows a basic NetLogo interface. Furthermore, a variety of tools can be used to individualize the model and match.

AnyLogic

AnyLogic is a business-grade modelling solution for agent-based-modelling as well as discrete-event and system dynamics simulations [43]. Originally developed at the St Petersburg Technical University, the software is licensed and distributed by *The AnyLogic Company* since 2000.

AnyLogic offers an immense feature library which exceeds both NetLogo and Repast. Additionally, as it is not only applicable for ABM but two other types of simulations, it allows for more complex models. These different modelling approaches are combinable to form large-scale model in which sub-parts may follow different modelling paradigms. Furthermore, industry-specific extension libraries are available, like for modelling rail or road traffic.

[2] https://java.com
[3] https://www.python.org/

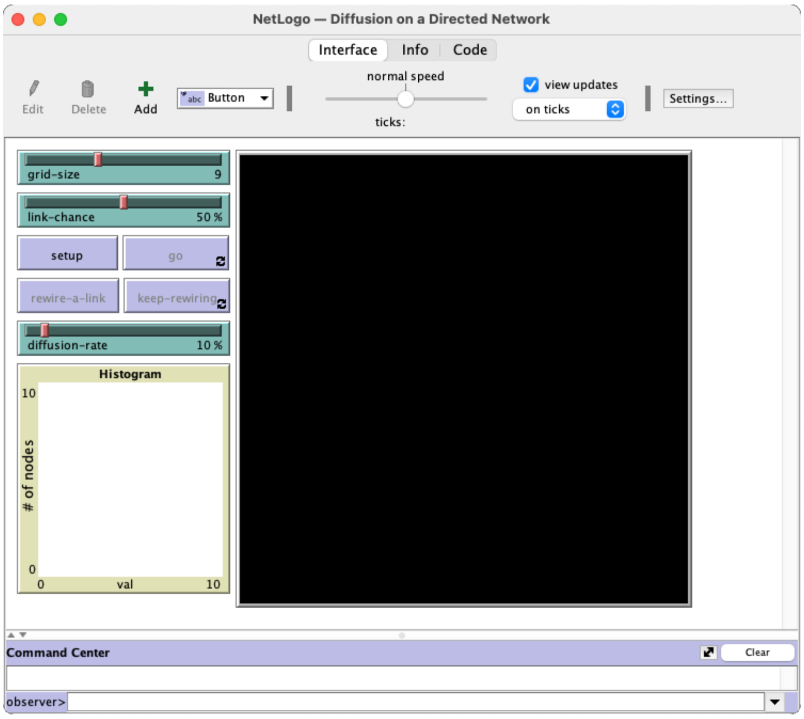

Figure 3.4: Example of a User Interface of the NetLogo Modelling Application.

Even though the AnyLogic framework provides many means to design and implement sophisticated models, it is not open-source and not free-of-use. While there is a gratis educational version, the functionality is limited. Therefore, the usage rate in research is smaller compared to NetLogo or Repast.

SARL

SARL is a general-purpose agent-oriented programming language [25]. In contrast to Repast or NetLogo, SARL is not a fully-fletched framework but a programming language which empowers efficient agent-based model implementations. SARL presents fundamental abstraction for dealing, among other things, with concurrency, distribution, interaction and decentralization. SARL incorporates modern software engineering concepts, which are not found in NetLogo or Repast but are essential for more complex applications.

SARL incorporates the agent-based programming paradigms by having the concepts of "Agent", "Space", "Capacity" and "Skill". Each agent is an autonomous entity which has a set of skills to implement its capacity. The capacity is the specification of the agents skills, which is public to its context. A "space" is an abstraction for defining interaction environment for agents, in which inter-agent interaction as well as interactions of agents with their environment can happen. These concepts should allow for a straight-forward agent-based model implementation based on modern software design approaches.

While SARL provides the means to implement an agent-based model, it does not offer an execution framework like Repast or NetLogo. However, SARL can be executed on Janus, which is multi-agent platform, which implements the infrastructural needs of running simulations with SARL [25].

Chapter 4

Modelling Innovation Diffusion

In this chapter, we introduce an agent-based model for simulating knowledge diffusion in rural Tanzania. In the first part, the conceptual model is established from findings in the literature review and in a current field study. Afterwards, the implementation of an agent-based model, deduced from the requirements of the conceptual model, is described.

4.1 Conceptual Model Design

In this section, we establish the conceptual design for a model of knowledge diffusion in rural Tanzania. The design decisions are based on findings from the ongoing field study of a research team of the University of Zurich in Tanzania presented in section 3.1. On-site observations and interviews revealed important characteristic of the social system in this region which should be depicted in our model. Furthermore, they are complemented with theory of innovation research summarized in section 3.2. A model is necessarily a simplification of its real world counterpart and hence, assumptions have to be made to achieve a simplification. All assumptions concerning a certain aspect of the model are stated in the corresponding subsections.

4.1.1 Goal of the Model

Eventually, the model will be used to compare the effectiveness of different roll-out strategies of new products in rural Tanzania. As these new products are considered as novelties in the environment of interest, they can be regarded as innovations [10]. Therefore, the goal is to design a model which depicts the diffusion of innovations in rural Tanzania in order to simulate different intervention strategies and measure their success.

The proposed conceptual model will depict the geographical and socio-economic characteristics in rural Tanzania. The core of the model is the abstraction of the interaction between different modelled parties. This is central because diffusion of innovation happens mainly through communication of individuals in their social system [33–35]. Furthermore, the demographic and geographic properties of Tanzania with relevance for the diffusion of innovation are established and applied to the model design. All design decisions, the reasoning behind them, as well as their implementation, will be explained in the following subsections.

4.1.2 Model Components

As introduced in section 3.3, agent-based models are composed of agents, an environment and interactions. Figure 4.1 shows our final conceptual model. The separate parts of the model are

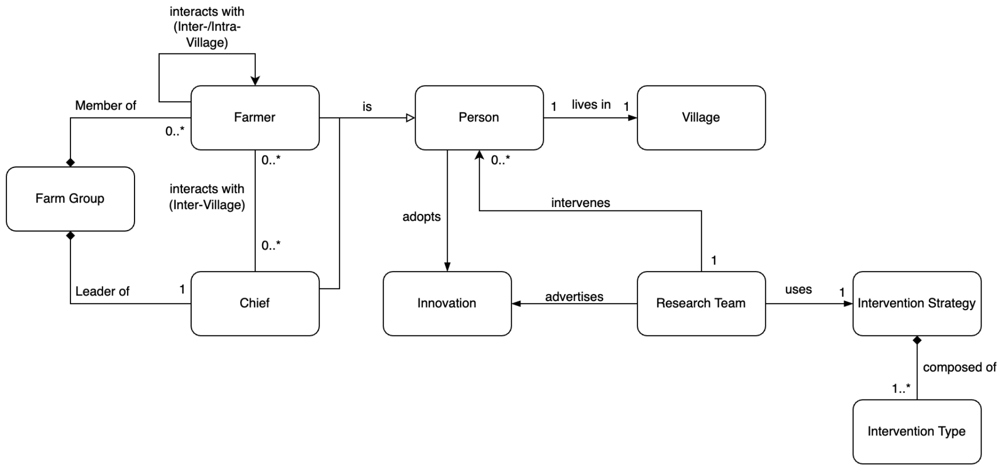

Figure 4.1: Conceptual model of the simplified social system of smallholder farmer households in Tanzania together with an intervening research team introducing the innovation. It contains all entities and their relations necessary for our agent-based model.

explained in detail in the following subsections.

4.1.3 Agents

Person A person is an abstract entity in the model which poses as the super-entity for all natural person-like entities. Farmers and chiefs are two sub-classes of a person which are explained in more detail later. A person is a central agent in the model because they form an opinion about the innovation over time. Eventually, a person makes the decision whether to accept or reject an innovation. In the classic diffusion of innovation process by Rogers, a person is referred to as an individual adopter [10].

Each person has a specific location in the model which impacts its relation to other peers in the system. In this case, the location of an individual is defined by the village it is living in. Furthermore, depending on the type of agent, it has different types of relations with other peers. As each agent is unique, they have some characteristics making them distinguishable from one another.

Farmer A farmer is a sub-entity of the person entity. Farmers can be regarded as the central type of agent in the model as the final goal of the model is to evaluate the effect of different intervention strategies on them. The target of each strategy is to convince the farmer agent to adopt the innovation and to do it as quickly as possible.

In this model, a farmer can be considered equal to a farming household as no differentiation between different members of a household is made. Since there cannot be partial adoption of an innovation in a single household of storage techniques, there would be no significant benefit from modelling different members per household. Furthermore, introducing different members each with individual behaviour, would lead the model to a complexity which would make it more difficult to make generalizable observations. Lastly, it is common that Tanzanian household are headed by a single persons which makes the final decision [4].

Nevertheless, a farmer represent a typical but simplified farmer in rural Tanzania. In this model, all farmer agents produce the same crop and hence potentially consider adopting an innovation impacting cultivation of this crop. Important to note is that the type of crop or product which is produced by the farmers does not matter as long as the innovation provides an improvement in the production of the specific crop. Hence, each farmers poses as a potential adopter of the innovation and goes through the innovation-decision process from section 3.2.2. In addition, every farmer interacts on a regular basis with other farmers and chiefs. There are three types of interactions with other agents which are explained in detail in subsection 4.1.8.

From on-site observations, the individual integration into the social system differs significantly between farmers. Whereas most farmers tend to have a similar amount of friends, single farmers in each village are very well connected to many people across other villages. Hence, this minority of farmers can be considered as opinion-leaders in the diffusion of innovation process.

Chief The chief-agent is the second sub-entity of the person-entity and quite similar to the farmer-entity. It has the same attributes and behaviour as standard farmers. In contrast to standard farmers, it poses as the head of a farm group. Hence, a chief is also a potential adopter of the innovation.

In terms of Rogers diffusion of innovation, a chief agent acts as an opinion leader. Firstly, due to its position as farming group leader, it has more interactions with other individuals in average. Furthermore, due to their position as farming group leaders, they tend to have more influence on other individuals.

4.1.4 Farm Groups

A farm group is a group of farmers which regularly meet in order to talk about agriculture related topics. All members of a group are involved in the same type of farming, for example wheat or live-stock. During the regular group meetings, they exchange on the one hand news and information about farming but on the other hand socialize in general. Each farm group has a chief which is elected by the community in most cases.

In practice, in many cases farmers are members of different farm groups based on their type of farming. Most farmers do not just perform one type of farming but multiple [4]. Furthermore, it is possible that farmers from multiple villages belong to the same farm group. However, to reduce the complexity of the model, it is assumed that all farmers of one farm group live in the same village and that all farmers of a village are part of the according farm group.

4.1.5 Decision Process

Each agents undergoes a decision process whether to adopt or reject the innovation over time. The process of an agent in our model is based on a simplified version of the decision process proposed by Rogers [10]. First, we neglect the possibility that a farmer stops using the innovation after he adopted it previously. The potential innovations of interest in this thesis are new storage technologies, which are proven to bring substantial benefits [5,7,9]. Furthermore, as an adoption does not bring further costs, a discontinuation of usage would be illogical for the farmer.

Hence, our decision process is composed of three stages. First, an agent is unaware of the innovation, then it enters an consideration phase and finally, it may enter an adoption phase. Whether or not an agent adopts the innovation is based on his personal attitude towards the innovation as well as on interactions with fellow agents.

4.1.6 Villages

As introduced in section 3.1, villages are considered the second smallest form of administrative unit; only hamlets are smaller. Each farmer and chief in the model is living in exactly one village. Moreover, farmers living in the same village are considered to be member of the same farm group. For this model, we assume that all inhabitants of a single village know each other well enough to potentially mention the topic of the innovation during an interaction.

Villages are very heterogeneous in terms of size and number of inhabitants. Due to the lack of statistical data on villages, preliminary observations during the field study indicate that on average villages have between thirty and hundred inhabitants with some outliers exceeding these limits. Considering the average reported household sizes by the national census - around 5 members per household - on average villages contain between 6 and 20 households [4]. Furthermore, it is stated that approximately 90% of households are engaged in agriculture in included regions of the field study. Thus, we consider all household in a village to be engaged in the same type of agriculture for simplicity reasons.

4.1.7 Research Team

The research team is the agent in the social system which wants to establish the new innovation. By communicating with other agents in the system, the research team introduces knowledge into the modelled social system from outside. The research team achieves the diffusion of their innovation by executing intervention strategies targeting farmers of the model. Each strategy may be composed of different intervention types which themselves are configurable.

Intervention Strategy

An intervention strategy is a predefined process defining how to address potential adopters in the social system. A strategy is composed of various instructions. Some instructions specify the timing of communications, defining intervals or specific time points at which an intervention should happen. Other instructions consider the different types of potential adopters and state which set of adopters should be targeted. In regards to be coherent with the running field study, we define two types of adopter targeting.

Direct Advertisement In this approach, potential adopters, namely farmers and chiefs, are directly addressed by the research team. The type of interaction can be manifold, for example visiting the addressee in person, contacting them via telephone or send information via an SMS. This approach requires the research team to know potential adopters in person. However, not all adopters in the social system must be known just a subset of them. Intuitively, a larger set of known adopters leads to a broader spread of the knowledge of the innovation. Furthermore, the type of direct advertisement affects the uncertainty reduction of the adopter. While a generic SMS influences the adopter relatively less, a in-person communication potentially provide more value to the adopter in reducing his uncertainty towards an innovation as question may be answered bilaterally. An SMS based survey was used in previous field studies and in the currently ongoing field study of the University of Zurich in Tanzania with success [5,7].

Training of Trainers Training of trainers (ToT) is an intervention type frequently used to introduce knowledge to a society in developing countries in a sustainable manner [44,45]. In ToT, experienced trainers educate selected individuals from the population to a level that they can train others themselves and thus no more external involvement is needed. Especially in areas in which there is a shortage of professionals, this approach has been used with success [46]. Furthermore, it is shown that ToT has the potential to accelerate innovation diffusion [45,46]. The selected individuals in our ToT are chiefs of farm groups which are trained how to use the innovation and teach the members of their farm group themselves. This procedure is also currently used in the ongoing field in which this thesis is embedded [5,7].

4.1.8 Interactions

The concept of interactions in this model capsules all types of communication between different agents. This can either be in-person, in written form or through any technical means. Important to note is that, like in the real world, the topic of the innovation of interest is not addressed in every interaction. Whether or not the innovation is discussed is determined by various factors and the involved participants.

In order to collect all different types of interactions and communications, the participants of the field-study in Tanzania were observed and questioned. The key findings are the following:

- Only a negligible percentage of the population in the areas of relevance have access to internet-enabled devices.
- The usage of mass media or social media is therefore not of much significance in information diffusion. Stated by the national census, with about half of the household owning a radio, radio transmission could be used as a mass media tool [4]. However, radio transmission cannot be initiated by average individuals in the system under normal circumstances.
- A bigger percentage of the population have access to simple mobile phones, which only allow making calls and send/receive SMS. This finding is supported by the results of the Tanzanian Census which stated that only around half of all households own a mobile phone [4]. However, the percentage presented by this report is likely to be below the actual value as the adoption of mobile phones increased in the last decade.

As a result, most communication inside the system happens one-to-one and not one-to-many. Therefore, only one-to-one types of interactions are included in this model except the farm group meetings, in which a chief addresses multiple farmers. Due to this, the knowledge diffusion in the model can be reproduced more easily as there is no indirect communication where not both parties are known.

Following are the selected three types of interactions which happen between agents in our model. These interactions happen besides each other and are independent.

Inter-Village Interactions

The first type of interaction are communications between farmers which live in different villages. This interaction happens between farmers who know each other like friends or relatives. In our model, both participants of an interaction are either farmers or chiefs.

In the model, both participants are considered as equals. Therefore, each agent is equally influenced by the opinion of the other one regardless whether the agent is a chief or a farmer. How often this type of interaction happens, differs from agent to agent.

Intra-Village Interactions

This type of interaction includes all communications between farmers in the same village. In other terms, the participants in this interactions can be viewed as neighbors which are both involved in the same type of farming. Similarly to inter-village interactions, in intra-village interactions both parties are considered as equals and have the same influence on the opinion about the innovation of the counterpart.

Farm Group Meetings

The third type of interaction, which happens in the modelled system, is a farm group meeting. These are regular meetings in which farm groups gather to discuss agriculture related topics as described in section 4.1.4. The discussion are lead by the chief of the respective farm group.

In contrast to inter- and intra-village interaction, in these meetings not all participants are affected in a similar manner. According to Rogers, we define that chiefs can be considered as opinion leaders which impact the opinion of other individual over-average [10]. Hence, when a chief is interacting with other farmers and addressing the topic of the innovation, his opinion influences the other ones substantially more. This can either be in a positive or negative direction.

4.2 Implemented Model

Based on the conceptual model, an agent-based model was implemented using the NetLogo framework. In this section, we present how conceptual features are implemented in detail and how a user can interact with our model.

We chose the agent-based modelling framework NetLogo for our model implementation. As outlined in section 3.3.3, the NetLogo software contains all needed components for our model out-of-the-box. In summary, the most important features needed for this model are multi-agent modelling options, straight-forward agent linking possibilities and an intuitive user interface. NetLogo matches all these criteria and due to its widespread usage, many third-party extensions and auxiliary materials exist. Moreover, since NetLogo is an open-source software and based on Java, our model is executable on all common operating systems.

NetLogo provides an out-of-the-box concept of time in a model with *ticks*. Hence, it runs in an iterative manner repeating the same set of methods at each tick. Therefore, we use this concept and define one tick to be equal to one day in real life as this is a more intuitive measure of time to the user.

Furthermore, as this model is considered a proof of concept, it is documented thoroughly in the following sections as well as the source code itself. Additionally, the implementation follows the latest coding guidelines for it to be easily extendable and modifiable. All parts of the algorithm were built in a modular manner in order that minimal changes are needed to exclude certain parts of simulation logic if they are not needed. This is also intended to make the model effortlessly more or less complex and complicated if it was necessary for certain predictions [19].

The behaviour of simulations is controlled by values of parameter as well as pseudo-random decisions of agents. Whereas the randomness is determined by a *random-seed* of NetLogo, the parameter are set by users of the model [23]. The most important parameter are introduced in the following sections, but in total there are more parameter. The complete list of all parameter and their description is located in the appendix in table A.1. There are parameter which are only used during the setup of the model, for example the number of villages, and parameter which influence the running simulation. In the latter case, the values of the parameter may be changed while the simulation is running. Furthermore, we differ between semi-fixed and freely variable parameter. A parameter is semi-fixed when it may be changed but it should not be done regularly. These values are mostly determined early on and not changed afterwards as the behaviour of the model may change to an extend which would make new results incomparable with previous ones. These parameter define the core behaviour of the model and the environment. Freely variable parameter are intended to be changed between simulations or even during the run of a simulation. Two examples for this type of parameter are the mentioning or adoption probability of farmers.

Our model includes three different decision actions which are executed by an agent: determine whether to mention the innovation during a conversation, determine how much its own attitude changes after an innovation and the final adoption decision. All these actions are implemented according to the same principle. Each action has a configurable base value which determines the outcome of the action. This value is influenced by various features of the agent, his counterpart or the environment. How much each feature influences the action can be determined with parameter. This gives the user the possibility to adjust all actions with ease to his wishes, extend them with more influencing features or even neglect all influences and only use the base values. All three actions are explained in more detail in the following sections.

4.2.1 Agent Implementation

As determined in the conceptual model, a model should contain three types of agents: farmers, chiefs and a research team. According to the NetLogo implementation guidelines, these agents are all *breeds* of turtles [23]. Hence, they can have there own set of variables and their own behaviour.

In figure 4.2, the design of farmers and chiefs is visible. Farmer agents are displayed with a matching farmer design . Chiefs are shown as distinctive red flags and thus are easily distinguishable from other agents. Lastly, the research team is not visible to the user as it would not provide any additional value.

Farmers and chiefs have a set of variables which impact the innovation diffusion in the system. These variables are:

attitude Defines the attitude of an agent towards the innovation. The higher the value, the more likely is the adoption of the innovation.

attitude-decline-rate The rate in which the attitude of an agent declines over time if no interactions mentioning the innovation are happening.

adopter-type Assigns the agent a set of characteristics which are relevant in the innovation adoption process. Based on Rogers adopter types, there exist 5 types: innovators, early adopters, early & late majority and laggards.

innovation-related-interactions-count Counts the number of interactions of an agent in which the topic of the innovation was mentioned.

adoption-state Defines in which phase of the adoption process an agent resides. The phases are: *unaware*, *in consideration* and *adopted*.

Each agent has an attitude towards the innovation, which either can be good, bad or indifferent. This attitude is influenced by other agents during interactions in any direction. It reflects the interest of an agent to adopt the innovation and is implemented as a numerical value in the model. Furthermore, when the agent has no conversation related to the innovation with other agents, his attitude slowly decreases over time until the next interaction happens.

Before the start of a simulation, each agent is assigned an adopter type. This variable is, in its core, based on the different adopter types defined by Rogers [10]. Hence, each agent is assigned one of the five adopter types with a likelihood based on their proportion of the total population as visible in figure 3.2.3. Besides their time to adoption, Rogers identified each adopter type group by their levels of talkativeness, risk aversion and economic wealth. Therefore, this variable of an agent can be considered as the aggregated characteristic of these traits. The adopter type is used during the determination if the innovation is mentioned during a conversation, by how much the attitude of an agent is changed during an interaction as well as the adoption decision itself.

Lastly, the **innovation-related-interactions-count** variable counts the number of interactions of agent during which the innovation is mentioned. In contrast to the attitude, this value is strictly increasing over time. Besides its usage for analytic purposes during a simulation, this variable influences the mentioning probability of the innovation during an interaction.

4.2.2 Village Generation

In order to implement the concept of villages as described in section 4.1.6, we used the *patch* component of NetLogo. All villages together form the environment of the model which is a quadratic field composed of a large number of patches. Villages have a white background and their limits are marked by a black border as shown in figure 4.2.

A process creates the villages based on a set of parameter in a random manner. The parameter which affect the village are the following:

avg-nr-of-farmers-per-village Defines the average number of farmers inhabiting a village.

nr-of-villages Defines the number of villages which form the environment of the modelled system.

As the process of the village generation is random, each setup leads to newly looking environments. The process works in two steps. First, each village is given a unique ID. Then, a random patch is selected for each village and given the same ID. This identifies this patch as belonging to the village. Afterwards, while there are still patches which do not belong to a village, a random patch with an ID together with one of its neighboring patches without an ID are selected. The ID is then passed on the neighboring patch and hence the village has grown by one patch. This is repeated until all patches are allocated to a village. Afterwards, the borders of the villages are marked in order to give the user a visual impression of the village distribution.

In a second step, the number of farmers per village are determined. As villages grow in a random manner, there exist major differences in the size of the villages which reflects the heterogeneity of village sizes as stated in section 4.1.6. It is important to note that the size of a village in the model does not reflect the geographical size of the village but rather relates to the number of possible inhabiting farmers. The number of farmers therefore should correlate with the size of the village. For this reason, the average number of patches making up a modelled village is determined. Then, the ratio between the value of the parameter **avg-nr-of-farmers-per-village** and the mean number of patches per village is determined. Lastly, for each village the number of inhabitants is calculated by multiplying its number of patches with the ratio determined beforehand. Thus, villages differ greatly in their size and number of inhabitants but on average are consistent with the value specified in **avg-nr-of-farmers-per-village** and **nr-of-villages**. Eventually, the algorithm adds one chief agent at the center of each village.

4.2.3 Construction of a Social Network

There are three types of social interactions any agent can have in this model: inter-village, intra-village or farm group meeting interactions. Hence, each farmer must have a social network of friends, neighbors and a farm group over which these interaction happens. This leads to the problem of determining which agent knows which other agents. This, like the type of interactions, can be divided into three parts. The complete modelled network can be subdivided into inter-village friends networks, neighbor networks and farm group networks.

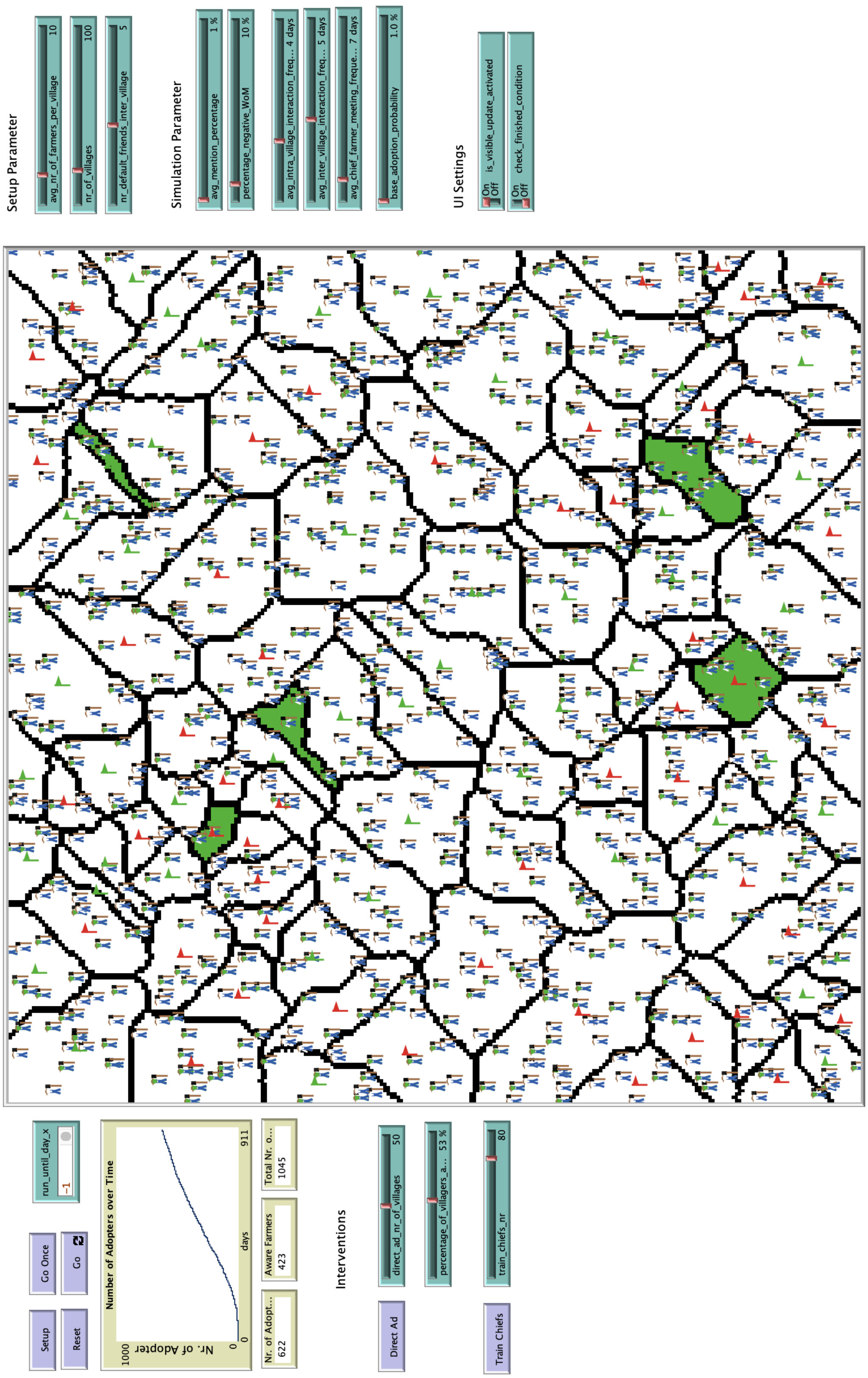

Figure 4.2: Screenshot of the final user interface of our Netlogo model displaying the map of the environment, elements to adjust parameter and a plot of the adoption rate. Visual feedback of the adoption state is given by colouring farmer & chief agents as well as villages green.

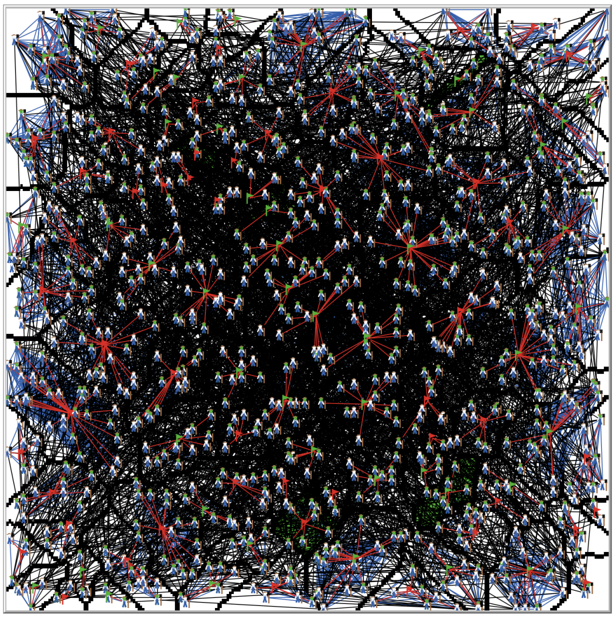

Figure 4.3: Screenshot showing the same environment as in figure 4.2 but with visible network links.

At the base of all three network lies the *link* component of the NetLogo framework. A link enables the model to connect agents either in a directed or undirected manner. In this model, all networks are undirected as it is not possible that one farmer is befriended with another one but not vice-versa. Furthermore, similarly to agents, different breeds of links may coexist. Hence, every type of network has its own type of link breed which makes them easily distinguishable. In order not to congest the user interface of the model, all links are hidden. To highlight the need for hiding the links, a model with visible links is shown in figure 4.3. In addition, each link possesses a timestamp when it was last used (an interaction happened between the two connected farmers) in order to prevent accidentally interacting over the same link multiple times in the same interval. Even though theoretically possible in the real world, the likelihood of it happening is negligible and could lead to undesirable and unrealistic behaviour in the model.

The topology of the resulting complete network composed of all three types of sub networks has heavy clustering between nodes which are from the same village and sparse links between nodes from different villages. These clusters can be considered as farmers with near geographical location in the real world. Hence, information is likely to spread reasonably faster in a cluster than between clusters.

For the generation of the three-part complete network of the model, one parameter is available to the user. This parameter can be adjusted before the model is loaded.

avg-nr-inter-village-friends Defines the average number of inter-village friends of a farmer.

Inter-Village Network

The first type of social network contained in the model is the inter-village network over which interactions as specified in section 4.1.8 happen. Determined by the value of the parameter **avg-nr-inter-village-friends**, each farmer and chief is given a number of friends from other villages (friends score). As observed in the current field study, not all farmer have the same amount of

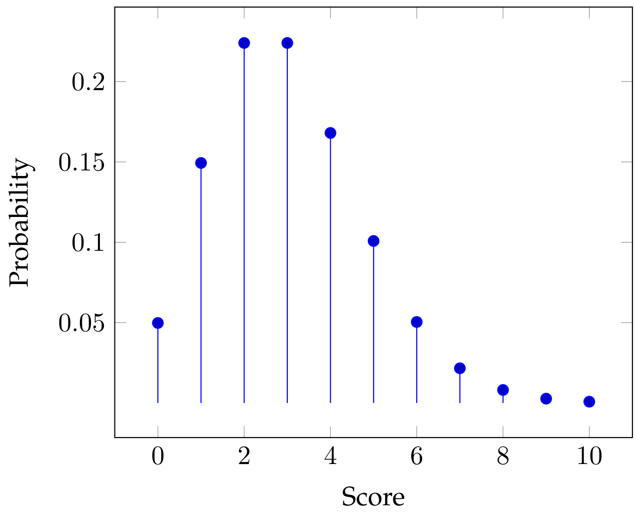

Figure 4.4: A Poisson probability distribution with a mean of 3. This distribution is utilized in various processes of our model.

friends. However, the majority of villagers share a similar number of friends which is represented by the parameter **avg-nr-inter-village-friends**. Only a small number of farmers have significantly more or less friends than the average number of friends.

Poisson Distribution Hence, we suggest using a Poisson distribution to determine the friend scores of all farmers and chiefs. A Poisson distribution is a discrete probability distribution of the number of independent events in a fixed interval [47]. Furthermore, with increasing number of modelled events, a Poisson distribution resembles the typical bell-shaped curve of a normal distribution. However, in regard to the friends score, two aspects suggest using the Poisson distribution over a normal distribution. Firstly, the friends score is a discrete variable as it is not possible to have 2.5 friends. Secondly, the friends score cannot be negative which is another feature of the Poisson distribution. NetLogo provides the command *random-poisson (μ)* which returns a random variable according to the probability of a Poisson distribution with mean μ [23].

The friends score of a single farmer is accordingly chosen by using the *random-poisson* command with Poisson distribution

$$X \sim Pois(avg\text{-}nr\text{-}inter\text{-}village\text{-}friends) \tag{4.1}$$

An example Poisson distribution with a mean of three is shown in figure 4.4. With a percentage of around 22.5%, the likelihood of choosing the number two and three are the highest and together make up almost half of all chosen values. However, with a percentage of roughly 2%, it is also possible to end up with a value of seven. This provides the demanded heterogeneity in the number of inter-village friends of each farmer.

After each farmer is given a friend score, randomly chosen farmers from different villages are linked. The process starts by selecting a random farmer which has a friend score above 0. Then, a random set of other farmers from different villages are chosen with the size of the friend score of the current farmer. Links of breed *inter-village friend* are added between the current farmer and all his friends. Lastly, the friend score of all other selected farmers is reduced by one. This procedure is repeated until no farmer remains with a friend score above 0 or in very rare cases, that no farmers from other villages can be found with a friend score higher than 0 themselves. Eventually, a network of inter-village friends over the whole environment is established.

Intra-Village Network

Another option in which knowledge diffuses in the modelled system is via an intra-village network which is formed by farmers and chiefs living in the same village. The generation of the "neighborhood" network is straightforward. For each village in the system, all farmers and chiefs are linked by *intra-village* links and thus each intra-village network has a full mesh topology. In case, it is decided at a later stage that not all villagers know each other, the process is implemented in such a way that this is easily modifiable.

Farm Group Network

The last type of network is the farm group network. In this case, this network is composed of the same farmers and chiefs as the according intra-village network for a given village. However, while an intra-village network has a full-mesh topology, the farm group network has a star topology in which the central node is the chief. Furthermore, it is important to differentiate between the intra-village network and the farm group network as the type of interactions as well as the their frequency differ.

4.2.4 Implementation of Interactions

As outlined in the previous section, for each type of interaction, there is a certain type of network. In this section, the implemented logic of interactions between agents is explained. An interaction can be split into two parts. Firstly, an algorithm decides when and between whom the next interaction will take place. Afterwards, another algorithm decides based on a set of parameter and variables whether the innovation topic is mentioned and how the participants are affected by it. For the latter one, there is no difference between different interaction types. However, the determination of the next interaction varies between the intra-/inter-village interactions and the farm group meetings.

Inter- and Intra-Village Interaction Frequency

This section explains the algorithm which handles the calculation when an agent will make his next interaction. A rudimentary approach would be to define a parameter which states in which intervals an agent interaction with another one and leave no space for decision by the agent himself. However, this does not reflect the real world as, at best, an average interaction frequency may be determined. Still this does not mean that these interactions are uniformly distributed over time and follow a strict interval. Hence, we let the agent himself decide when and with whom it wants to interact next.

For this decision process, following parameter are provided to the user:

avg-intra-village-interaction-frequency Defines the average time span which elapses between intra-village interactions started by an agent.

avg-inter-village-interaction-frequency Defines the average time span which elapses between inter-village interactions started by an agent.

As the process is equal for intra- and inter-village interactions, following definitions are valid for both. Nevertheless, both interaction types have separate intervals defined by the respective parameter.

After each interaction, an agent decides when he wants to make the next interaction in the future. In order to introduce a certain amount of randomness, the time until the next interaction is not fixed but decided by a probability distribution. As in the determination of the "friend score" of an agent, the Poisson distribution is used for this reason [47]. However, as the time until the next interaction cannot be zero, a special variant of the Poisson distributed is used: the zero-truncated Poisson distribution. In this distribution all values are greater or equal to one. Furthermore, as time is considered as a discrete variable in in this model, a normal distribution could not be applied.

Hence, after each interaction, the time interval until the next interaction is determined by choosing a random variable following a zero-truncated Poisson distribution

$$X \sim Pois_0(avg\text{-} * \text{-}village\text{-}interaction\text{-}frequency) \tag{4.2}$$

Important to note is that the calculated interval only defines when the agent is starting an interaction. Thus, an agent may be contacted by another agent at any time. This leads to the expected theoretical probability of an agent to interact with another one at any given time to be higher than

$$\frac{1}{freq} \tag{4.3}$$

which is the actually the probability that an agent starts a conversation at any given time. In this equation, *freq* is the fixed parameter value of *avg-*-village-interaction-frequency*.The actual expected probability of an agent interacting with another agent at a random point in time, is

$$\frac{1}{freq} + \sum_{i=1}^{|fr_a|} \frac{1}{freq * |fr_i|} \tag{4.4}$$

The term $|fr_a|$ is the number of friends of the agent under observation and $|fr_i|$ is the number of friends of friend i. Therefore, the probability that an agent interacts at any given time with another agent is influenced by the behaviour of other agents too in this model. Considering the real world, this reflect the probability of calling a friend or being called by a friend.

Selecting a Counterpart The previous process explains how an agent decides when he will start the next interaction. However, it is not decided who is his vis-à-vis for the next interaction. The agent makes this decision just before his next interaction. The potential counterparts are determined by the linked peers of the agent over the according networks established beforehand. It then randomly picks one of the linked agents on the correct network and starts a conversation with it. Furthermore, this procedure is implemented in such a way that the same link cannot be executed twice at the same point of time. Now, the next step is to determine whether the innovation is mentioned during the interaction.

Farm Group Meeting Frequency

The frequency of farm group meetings is devised in an similar manner as the frequency of inter- and intra-village interactions. After every meeting the chief selects the time until the next meeting based on a zero-truncated Poisson distributed variable. The only difference is that the mean of the distribution is determined by the value of the parameter **avg-farmgroup-meeting-frequency**.

avg-farmgroup-meeting-frequency Defines the average time span which elapses between farm group meetings.

farmgroup-meeting-attendance-percentage Defines the percentage of farm group members which actually attend the meeting.

As in reality, we added some fluctuation to the percentage of farmers attending the meeting. This is determined by the value of the parameter **farmgroup-meeting-attendance-percentage**. Before the meeting starts, the farmers, which attend the meeting, are chosen according to the parameter at random. Besides not attending, farmers which are excluded, could also be considered as farmers which indeed attend the meeting but do not pay enough attention to the chief due to distractions. By the nature of the topology of the underlying network, only the chief interacts with the participants and participant do not interact with each other.

Mention Probability

As in reality, having a conversation with another person does not guarantee that a certain topic is mentioned. We incorporate this in the model by determining a probability, whether the innovation is mentioned during an interaction based on characteristics of the two participating agents. The probability is based on a base probability given by a parameter. This base probability is influenced by the adopter types of the participants, the number of previous interactions mentioning the innovation and whether the innovation is already adopted by the participant. A set of parameter is available to the user with which he can adjust the mentioning probability calculation.

avg-mention-percentage Defines the base probability of the innovation being mentioned in an interaction. For example, if set to 10%, the innovation is mentioned on average in one out of ten interactions.

avg-farmgroup-mention-percentage Defines the base probability of the innovation being mentioned during a farm group meeting.

max-influence-adopter-type Defines the maximum influence of the participants adopter types on the mentioning probability in percent.

max-influence-prev-interactions Defines the maximum influence of participants previous, innovation-related interactions on the mentioning probability in percent.

opt-number-previous-interactions Defines the optimal number of previous innovation related interactions for an agent, which did not adopt yet, to mention the innovation again.

max-influence-adoption-state Defines the maximum influence of the participants adoption state on the mentioning probability in percent.

The process which determines whether the innovation is mentioned is implemented in a modular way. This allows, if the simulation context requires it, to exclude or include certain influences on the mentioning probability with minimal change. Furthermore, new influences may be added effortlessly to the model.

Firstly, the process confirms that at least one of the participants of the interaction already is aware of the innovation. Naturally otherwise, it would be impossible to mention the innovation. Afterwards, the influences of the participating agents characteristics are evaluated.

At the start of a simulation, each adopter type is given a numerical value from 1 to 5 (innovators to laggards) and the average of the types of both participants is calculated in this step. Based on the assumption that two innovators interacting are more likely to mention the innovation compared to two laggards, the higher the average score is, the more improbable the mentioning of the topic is. On the contrary, a lower score indicates that the innovation is more likely to be discussed. The final mentioning probability is then adjusted by the influence of the adopter types in relation to the value of **max-influence-adopter-type**. To illustrate this with an example: if **max-influence-adopter-type** is set to 10%, **avg-mention-percentage** is 50% and two innovators interact with each other (maximum mentioning likelihood), the final mentioning probability becomes 55% (50% + 10% * 50%). Thus, the maximum influence is in relation to the base mentioning probability.

Another possible influence on the mentioning frequency is the number of previous points of contact of the participants with the innovation. We implement the model in a way that an agent who talked more about an innovation is more likely to mention it again. Nevertheless, if an agent talked a lot about an innovation and has decided not to adopt it yet, he may become annoyed and stop mentioning the innovation. Hence, we model the influence according to a bell-shaped probability distribution which is centered around the parameterized value **opt-number-previous-interactions**. Finally, the influence on the mentioning probability is determined by the obtained value in relation to the **max-influence-prev-interactions** parameter.

Similarly, the adoption state of the participants influences the mentioning probability. We assume that people who adopted an innovation tend to address it more in conversations. Regarding the decision process proposed by Rogers, agents which adopted the innovation are either in the *implementation* or *confirmation* stage [10]. In both stages, individuals try to actively reduce their uncertainty or justify their adoption and thus tend to communicate more with others. Hence, if both agents involved in an interaction are adopters, the final mentioning probability is adjusted by the value of **max-influence-adoption-state**.

Based on the final mentioning probability including the various influences, it is decided whether or not the innovation is mentioned in the current interaction. The probability whether the innovation is mentioned during a farm group meeting is equally calculated. The only difference is that the base probability is set by the parameter **avg-farmgroup-mention-percentage**.

If the innovation is mentioned, the impact of the conversation on the agents attitude towards the innovation is calculated in the next step. Otherwise, the current interaction process ends and the simulation continues.

Impact on Participants

After it is determined that the innovation will be brought up in the current interaction, the impact on the attitude of both participating agents is calculated. Intuitively, two agents interacting can be impacted differently. As an example considering the adopter types, an innovator may be more heavily influenced by positive information about the innovation than a laggard who may be more sceptical. We define four influences which affect the base attitude change of an agent. Like the mentioning calculation process, this process is implemented in an modular manner so it is adjustable or extendable with ease.

Not all conversation about an innovation are positive. Maybe someone has found unfavourable information about the innovation, has had a bad experience trying it, or generally tends to be very sceptical towards novel ideas. This is referred to negative word-of-mouth (WoM) [16, 31]. Hence, the model includes a parameter which determines a percentage whether a conversation is unfavourable and during each interaction, a positive or negative change is chosen accordingly.

Secondly, depending on the other participant, one's attitude may change differently. This is implemented during interactions with a chief. As chiefs are considered normally as more experienced and trustworthy by other farmers, their opinion has an above average impact on the other participant in the interaction. This goes both ways, a bad report about the innovation has a more negative influence on the attitude as normal and vice-versa for positive information. The different levels of influence of the agents are controlled by parameter adjustable by users.

The change in an agents attitude is influenced by assigned adopter type as well. The base assumption behind the following process is that the attitude of an agent is positively impacted by a favourable conversation about the innovation and negatively by unfavourable one no matter which type of adopter it is. However, the magnitude of the attitude change in any of the two directions is influenced by the adopter type. This is based on adopter characteristics which state that in average laggards are more receptive of negative information while innovators tend to trust positive information more as it supports their already existing attitude [10, 22]. Hence, during negative WoM, the effect of the conversation on the attitude is increased for late adopter types and decreased for early adopter types. The same calculation is done for positive WoM but the other way around.

Lastly, the opinion of an adopter tends to have more weight in average as an adopter already has hands-on experience with the innovation. This effect is incorporated into the model by checking the adoption state of the other participant in an interaction and assign more weight to its opinion if it adopted the innovation already. There is no negative impact if the counterpart has not adopted the innovation yet. The absolute impact of this criteria is controlled by an adjustable parameter.

4.2.5 Adoption Decision

Up until now, it was shown how agents interact, and their attitude towards the innovation changes over time. However, the actual adoption decision making of an agent was not discussed. Research shows that this decision is not very straightforward but based on a variety of social, economical and personal factors, some of which are not graspable in theoretical models [10–12,20–22]. Hence, we try to incorporate variables of different fields on which the decision is finally based.

Similar to the previous interaction related processes, this decision is based on a base adoption probability which is influenced by various aspects of an agent. Another important aspect is the timing of the decision. We implemented this decision as a repetitive action of an agent over time. The average interval between the decision makings of an agent is defined by the parameter **avg-adoption-decision-interval**. However, similar to the reasoning explained in the interaction section, this decision is not made strictly after x days but follows a varying pattern. Hence, after each (negative) decision of an agent, the time interval until his next decision is determined by a zero-truncated Poisson distribution:

$$X \sim Pois_0(avg\text{-}adoption\text{-}decision\text{-}interval) \tag{4.5}$$

When the point in time of the decision is reached, the agent decides whether or not to adopt the innovation. Trivially, this decision is more likely to turn out positive the more an agent knows about an innovation or as defined by Rogers, the smaller the uncertainty of an agent becomes [10]. We determine three types of influence on an agent's adoption decision: its attitude towards the innovation, its adopter type and the adoption rate in his direct social network. A set of parameter is available with which the adoption decision of an agent may be influenced.

avg-adoption-decision-interval Defines the average interval length between agent's adoption decision making.

base-adoption-percentage Defines the base probability of an agent adopting an innovation at a given time.

max-influence-adopter-type Defines the maximal influence of the agents adopter type on the adoption probability in percent.

max-influence-adoption-rate Defines the maximal influence of the adoption rate of the agent's social network on the adoption probability in percent.

max-influence-attitude Defines the maximal influence of the agent's attitude on the adoption probability in percent.

An agent decides according the the value of **avg-adoption-decision-interval** in a set interval whether it wants to adopt or reject an innovation for the time being. The base probability of adopting the innovation is determined by the value of **base-adoption-percentage** parameter. This probability is relatively influenced by a set of variables.

The adopter type of a potential adopter influences the base adoption probability positively for adopter types *innovator, early adopter, early majority* and negatively for the type *late majority, laggards*. The relative impact is determined by the parameter **max-influence-adopter-type**.

Furthermore, as many studies showed, the social network of an individual has a significant impact on the adoption likelihood [10, 12, 33–35]. Thus, during a decision, the proportion of friends and neighbors who adopted the innovation is determined and impacts the adoption probability to an extend specified in the **max-influence-adoption-rate-friends** parameter.

Lastly, the attitude of an agent influences the adoption decision. Attitude can either be positive, negative or indifferent towards the innovation. In our model, we approximate the direction of the attitude by comparing it with the number of previous innovation related interactions of an agent. If there were no influences on the attitude change during an interaction as explained in section 4.2.4, these two variables would grow evenly except to negative WoM. However, the various influences allow these two variables to diverge to a certain extend. Hence, these variables are compared during the adoption decision. When the attitude is significantly lower than the number of previous interactions (suggesting that in average interactions were not very convincing), the base adoption probability is negatively impacted. Likewise, if the attitude is significantly higher, the adoption probability is positively impacted. When both value are approximately equal, the base probability is not affected.

During a simulation with our model, the user interface gives visual feedback on the adoption rate in the modelled system. Agents who adopt the innovation are marked green. Villages in which all inhabitants adopted the innovation, also are coloured green as shown in figure 4.2. Lastly, there are UI elements which indicate to the user how the diffusion is progressing in the model. This feature should be disabled for performance reasons during larger simulations.

4.2.6 Intervention Strategies Implementation

According to the defined intervention strategies in the conceptual model in section 4.1.7, two intervention types are initially needed. Figure 4.2 shows how interventions can be manually triggered by a user on the user interface. At least one intervention must happen during a simulation as otherwise no knowledge of the innovation is present in the modelled system. Both types of interventions can be triggered at any point in time during a simulation.

For the direct advertisement of the innovation, a user can define the number of villages and the percentage of the population in these villages which should be addressed in the advertisement. The villages and its inhabitants are chosen at random by the model. Afterwards, their attitude is changed positively by a fixed amount defined by a user-adjustable parameter.

Likewise, in the training of trainers, a user can define the number of chiefs which will be addressed by the next intervention. These chiefs then are selected at random and their attitude is positively modified by a fixed amount defined by user-adjustable parameter.

Naturally, triggering interventions manually is not feasible for large simulations or simulations which are executed multiple times. The functionality provided on the user interface is intended for demonstration and trial purposes. In order to run intervention strategies over an interval of time, the source code can be easily extended with intervention plans which are automatically executed. However, this requires the user to have basic knowledge of the NetLogo scripting language.

Chapter 5

Analysis

In the previous chapter, we introduced an ABM which simulates innovation diffusion among smallholder farmer households in rural Tanzania. Now, we simulate and analyse the innovation diffusion when different intervention strategies are applied. This should provide a guideline for future applicants of this model on how intervention strategies could be compared in terms of effectiveness and efficiency. First, in order to retrieve significant results from the model, we perform a validation and verification. Then, we define five exemplary intervention strategies and compare the adoption rates in the respective simulations. Furthermore, we investigate how results from the model could be used in order to predict performance of intervention strategies.

5.1 Verification and Validation

As explained in section 3.3.2, verifying a model means to check its correct implementation. Validation, on the other hand, is the process of reassuring that the model actually corresponds to the real-world.

Verification The model was tested extensively to check that its behaviour complies with the conceptual model as well as the implementation decisions. The model's source code contains plausibility tests and further checks in all critical methods for the simulation. They check whether variables and parameter become invalid due to wrong behaviour during a simulation or due to implementation errors. In addition, the model was tested qualitatively in detail where variables were monitored throughout simulations. Furthermore, as many parameter are based on averages rather than absolute and unique values, variables were recorded over several runs in order to test whether they approximate the desired average value. One of these checks is shown in figure 5.1. It shows the distribution of adopter types of farmers which should approximately follow the adopter distribution proposed by Rogers [10]. By the frequency of the different types, we can assume adopter types are set correctly at the start of a simulation.

Validation Validation was performed by continuously challenging the conceptual design choices with feedback from the supervisor and with findings from the ongoing field study. However, the field study must progress first to find more precise values for parameter and thus a second, more detailed validation must be performed at a later stage. Figure 5.2 shows the averaged adoption rate of multiple simulations. The s-shaped curve suggests that the adoption rate in the model correlates with the theoretical adoption rate reviewed in section 3.2. Hence, the macro behaviour of the model seems plausible.

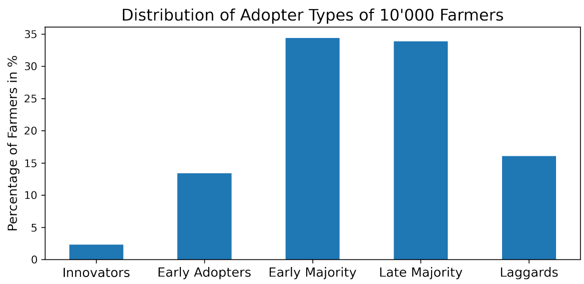

Figure 5.1: Distribution of assigned adopter types to 10'000 farmers with our model as an example of the verification process.

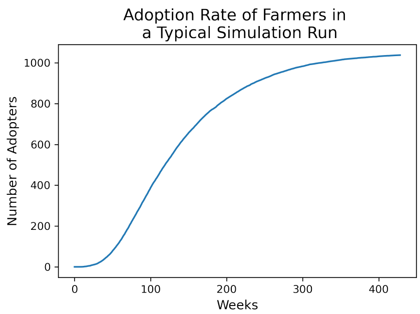

Figure 5.2: S-Shaped curve of the adoption rate of farmers in a random simulation with our model.

5.2 Comparison of Intervention Strategies

The conceptual model defines two types of intervention types: direct interventions and training of trainers. In this section, we define five possible intervention strategies which contain a combination of both intervention types which are executed repetitively.

These five intervention strategies are defined in table 5.1. Each intervention type is defined by a set of variable parameter. Direct interventions have a frequency, which defines the interval between interactions. During each interaction, the farmers of multiple villages are targeted. The number of villages is defined by the parameter **Nr. of Villages**. During each interaction, these villages are selected randomly. In addition, **% of Inhabitants** defines the percentage of the inhabitants which are addressed in the interaction. For the training of trainers intervention type, one variable controls the interval between trainings and the other one controls the number of chiefs which are trained during a single interval. Hence, strategy 2 is defined as followed: each week, 70% of the farmers of ten villages are targeted and one chief is trained.

Strategies 1-5 define possible intervention strategies each with a different specifications. These strategies are now compared with the help of our model. Due to their differences, the strategies should lead to varying rates of adoptions. Naturally, the model can be easily extended with other types of intervention strategies which are based on different variables. However, the process of their simulation and subsequent effectiveness analysis can be done in a similar manner.

	Direct Intervention			Training of Trainers	
	Frequency	**Nr. of Villages**	**% of Inhabitants**	**Frequency**	**Nr. of Trainings**
Strategy 1	weekly	15	70%	-	-
Strategy 2	weekly	10	70%	weekly	1
Strategy 3	weekly	5	70%	weekly	3
Strategy 4	weekly	5	70%	weekly	5
Strategy 5	-	-	-	weekly	7

Table 5.1: Definition of five different intervention strategies composed of different combinations of direct intervention and training of trainers.

5.2.1 Simulation Setup

As only preliminary results from the ongoing field study are available on the time of writing this thesis, the parameter values for the following simulations must be approximated. The values of all parameter are listed in table A.2 in the appendix. They were determined by a qualitative analysis of simulation runs in which the plausibility of the results was considered. Furthermore, results from previous field studies were used to compare these results and adjust parameter accordingly. As a consequence, the obtained results from this simulations have to be interpreted with caution and are rather intended as guidelines for later simulations with our model.

We use the tool *BehaviourSpace* of NetLogo, with which it is possible to automate simulations and vary parameter between runs [23]. In addition, each simulation run uses a different random-seed so that the independent decision of agents differs between runs. To minimize the total simulation duration, the execution is parallelized where each simulation is run on a different core of the processor.

For all strategies and all separate simulations of each strategy, the same environment of farmers and villages was used. Although the environment generation is implemented as a random process, using different environments in each run would make a meaningful analysis of the results impossible as the number of changing variables would be too high. A feature of NetLogo allows to export and import an environment which allows to run all simulations in the same one. The generated environment for all upcoming experiments contains 100 villages with 1051 farmer and chiefs in total.

5.2.2 Results

Each intervention strategy defined in table 5.1 is run 100 times with a different random-seed. Each run simulated the innovation diffusion for one year.

In each simulation, the number of farmers with the states *adopted*, *in consideration* and *unaware* as well as the number of distinct villages with at least one adopter are determined each day. The runs are then grouped by the applied intervention strategy and an average rate of adoption is determined. The results are shown in figure 5.3.

Figure 5.3 shows that there are significant differences between the selected intervention strategies in terms of adopters after one year. In subplot **(1)**, it is visible that strategy **2** outperforms the other ones. In general, strategies with more direct interventions tend to perform better. However, note that strategy **2** intersects with strategy **1** only after almost 300 days as the adoption curve is steeper. While strategies with more direct intervention reach people faster at the beginning, training of trainers interventions impacts are only perceptible after some time which is visible in subplot **(2)** in figure 5.3. This is valid for the number of villages with adopters as well, as con-

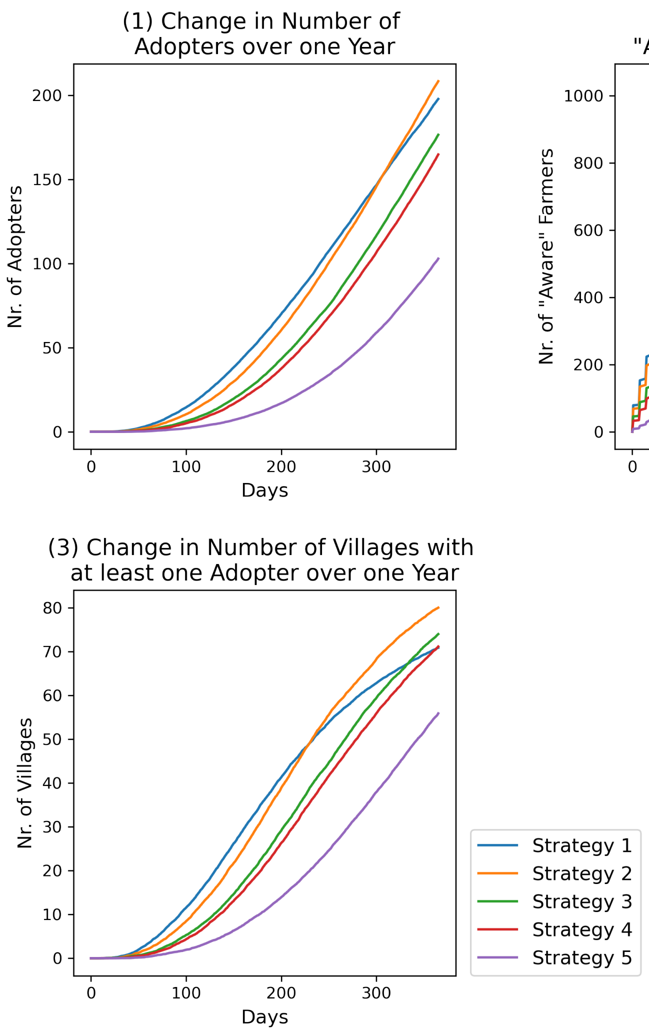

Figure 5.3: Plots comparing the performance of five intervention strategies based on different metrics. The values are the averaged results of one hundred simulations per intervention strategy.

	Values				Unit
Direct Intervention Interval	7	14	21		days
Number of Villages	5	10	15	20	villages
Percentage of Inhabitants	100				%
Training of Trainers Interval	7	14	21		days
Number of Trainings	1	3	5		trainings

Table 5.2: Different parameter values which lead to a total of 108 varying intervention strategies used for the effectiveness prediction experiment.

tinuous direct interventions tend to spread knowledge about the innovation more. Moreover, all curves in subplot **(2)**, except for strategy **5**, follow a zigzag pattern which reflects the repetitive direct interventions targeting unaware farmers. Note that analyzing the absolute numbers of adopters and other metrics would not be meaningful as the parameter of the model are not fine-tuned according to the real world and such results would not provide any added value.

5.2.3 Effectiveness Prediction

In table 5.1, we define five intervention strategies which are defined by values of five variables. Compared to testing such strategies in the real world during a field study, determining their effectiveness by running simulation on our model is economically inexpensive. However, finding optimal values for all variables of a strategy requires non-negligible computational power as multiple runs for each variation must be made. Therefore, we have investigated whether the creation of a predictive model is possible based on results obtained by simulations.

First, we created a sample data set by simulating innovation diffusion repetitively with our model. For this reason, we defined for all parameter of an intervention strategy a set of values as shown in table 5.2. The percentage of the inhabitants is fixed to 100% because to its correlation to the number of villages parameter. On average 70% of 100 villages equals to 100% of 7 villages. This leads to a total of 108 different strategies each differing by at least on parameter value.

The values of the parameter (listed in table A.2) as well as the environment are the same as in the previous section. For each strategy, 50 runs were executed which resulted in 5400 separate simulations. Additionally, the data was averaged by strategy which resulted in second data set with 108 data points.

Linear Regression In a first step, we tested whether a linear regression model can be used to predict the effectiveness of an intervention strategy based on the parameter determined in table 5.2. The resulting coefficients and the R^2 from a linear regression based on the simulated data are shown in table 5.3. The ordinary least squares method was used for the linear regression [48]. Furthermore, we introduced a constant to the regression as the number of adopters never was zero. The dependent variable is the number of adopters while the parameter of the strategies are independent variables.

The coefficients obtained by the linear regression seem plausible on the first glance and all coefficients are significant. Both coefficients of the frequencies are negative as longer intervals lead to overall less interventions and thus less adopters. Moreover, the positive values of the remaining two parameter suggest that an increase in addressed villagers and trainings result in more adopters. Intuitively, the standard deviation is on average higher in the regression on averaged data as less observations can be used for the regression. Furthermore, both R^2 would suggest that the model predicts values reasonably well. Nevertheless, further statistical tests on the data and

	All		Averaged	
R^2	0.895		0.950	
	Coefficient	*SD*	*Coefficient*	*SD*
Direct Intervention Interval	-6.432	1.255	-6.432	0.231
Number of Villages	7.941	0.047	7.941	0.236
Number of Trainings	4.237	0.049	4.237	0.808
Training of Trainers Interval	-0.888	0.166	-0.888	0.231
Constant	200.6	0.047	200.6	6.102

Table 5.3: Table showing coefficients of a linear regression fitted to the simulation results of our model. Even though, the R^2 values are indicating a good fit, the underlying data does not comply to necessary requirements for a linear regression approach.

the results reveal that a linear regression may not be a suited modelling technique to represent this data. The residual values of the linear regression model should be distributed randomly for all independent variables [48]. However, this is not the case with this data suggesting that the variables are not linearly related. Mathematical transformations of dependant as well as independent variables do not resolve this problem.

Linearity would suggest that, for example, each additional included village would lead to a linear increase in the number of adopters. However, considering that after each intervention, agents start interaction with each other and innovation diffusion is happening through WoM, suspecting non-linear exponential relation is more reasonable. In this case, using a linear regression to predict effectiveness would over- or underestimate outcomes depending on the parameter combination.

Non-linear Regression The variables in the linear regression do not comply with the underlying assumptions in order that the predicted results are reliable. Hence, we applied two different types of regression on the data set in order to obtain more reliable predictions and demonstrate how machine-learning algorithms could be applied on simulation results.

Random Forest Regression is a supervised learning algorithm which consists of a set of uncorrelated decision trees [49,50]. When predicting a value, each decision tree proposes a result based on its previous training on data. The output of the algorithm is the average of these preliminary results. As it combines multiple decision trees, a random forest regression tends to outperform single decision trees. Furthermore, the random forest algorithm is able to predict values based on non-linear correlations. For our experiments, we use the *RandomForestRegressor* of the Python-based machine-learning library *scikit-learn* [51]. We use a set of 50 estimators in our random forest and train the *RandomForestRegressor* on the data set introduced at the beginning of this sections with 5400 data points.

The second machine-learning algorithm used for the following predictions is gradient boosting of decision trees [52,53]. "Boosting" is the technique of continuously improving the used estimators by adapting to errors of previous estimators. In contrast to the Random Forest algorithm, in which all decision trees are built independently, the estimators are generated subsequently with this approach. Furthermore, gradient boosting analyses the residuals of the estimated values for possible patterns and applies these findings to the next estimator which should outperform the previous one. For this experiment, we use the *XGBoost* library in Python which provides all needed functionality and has good performance records [52]. The *XGBRegressor* is set to 300 estimators, a maximal depth of 4, and a learning rate of 0.01. It is trained on the same data set as the *RandomForestRegressor*.

	Mean Absolute Error	Root Mean Squared Error
Linear Regression	15.096	20.12
Random Forest	15.504	20.988
XGBoost	13.577	18.667

Table 5.4: Measured error values of the linear, Random Forest and XGBoost regressions showing that XGBoost has the best performance in predicting intervention strategy success.

Comparison In order to evaluate the performance of the linear, Random Forest and gradient boosting regressions, we generated a second data set with our model which contains **4050** entries. The parameter of the intervention strategies were altered in value to create new combinations which should be predicted by the regressions. The errors of these predictions are shown in table 5.4.

The *XGBoost* predictions are more accurate compared to the ones of the Random Forest and linear regression. The linear regression outperforms the Random Forest algorithm slightly. However, as the data does not follow a linear relation, the results of the linear regression are not reliable. Considering the possibility that the relation is exponential due to network effects, we assume that the linear regression underestimates results more heavily with larger values. The average number of adopters for all simulations in the test data lies at **221** farmers. Hence, an absolute error of **13.577** farmers suggests a relative error of about **6.1**% for predictions when using *XGBoost*. The root mean squared error of the regression is approximately **30**% higher than the mean absolute error for all regression suggesting that there is some variation in the magnitude of errors.

Furthermore, all regression types state that the parameter **Direct Intervention Interval** and **Number of Villages** are significantly more important for the predictions than Training of Trainers parameter. Both non-linear regressions rate the feature importance of the direct intervention parameter at around **95**% while the other parameter only are responsible for **5**%. This suggests that the predictions rely mostly on the direct intervention values.

Chapter 6

Discussion and Conclusion

After summarizing the necessary background information on innovation diffusion, agent-based modelling and Tanzania, we developed a ABM simulating innovation diffusion among smallholder farmer households. Afterwards, we used said model to analyse the effectiveness of different intervention strategies and their effect on innovation diffusion. In this chapter, we combine all these findings to answer the research questions of this thesis and challenge our findings. Eventually, we propose directions for future research on this topic.

6.1 Addressing the Research Questions

In the first chapter of this thesis, we proposed two research question which this thesis should answer. In this section, we answer both with findings from the previous chapters.

Answer First Research Question The first research question to be answered is:

RQ1 How can innovation diffusion among smallholder farmer households be modelled in Sub-Saharan regions?

To answer this question, we reviewed the latest research on related topics in section 2 and summarized the most important aspects in section 3. Furthermore, the ongoing field study of the University of Zurich in Tanzania, where they have direct contact with smallholder farmers, allowed to get first-hand insights into the local social system which lies at the base of our model. Based on these findings, we constructed a conceptual model for an agent-based modelling approach explained in chapter 4. We defined that the core driving forces behind diffusion of innovation are the farmers and especially the interactions between them [33–35]. Hence, we determined three types of interactions over which WoM recommendations happen in Tanzania: inter-village interactions, intra-village interactions and farm group meetings. All of them play a role in the overall diffusion as all interactions between agents eventually reduces uncertainty about the innovation [10]. This resulted in a conceptual model which captures all properties of the smallholder farmer households system in Tanzania in order to implement an ABM of it.

We chose an agent-based approach due to various reasons. Even though innovation diffusion being a macro-level process, each individual decides for itself whether or not to adopt an innovation [10]. Furthermore, very limited information of system processes in rural Tanzania is available and therefore, higher level simulations producing useful results are not possible to implement at this time. ABMs tend to be very flexible for adjustments which is critical when modelling a system of which many aspects are not known yet.

As a result, we mapped the generic ideas of the conceptual model to an agent-based implementation. Besides the sophisticated generation of the environment replicating the characteristics of rural Tanzania, the main focus of the model lies on the interaction logic between agents. All aspects concerning the decision process of an agent are built in a modular manner which gives users a wide range of possibilities to adjust the model to his demands. Furthermore, all decision are based on a parametric base likelihood which is influenced by various aspects of the agents involved. This design reflects the complexity and individuality of decisions of agents. In order to introduce some variety to the actions of the agents, most actions are happening based on distribution probabilities rather than on fixed, predefined values. We further critically validated and verified our model to ensure that obtained results are significant [14].

In section 5, we analysed how intervention strategies can be tested with our model. Various qualitative and quantitative experiments showed that simulations with our model produce plausible results. Hence, the model seems to be suitable representation of the smallholder farmer system in Tanzania. However, more data from the field study is needed in order to fine-tune the parameter of the model and accurately evaluate its results with observations from the real world. Due to the implementation of the model, such adjustments should not pose any issues and can be completed with minimal effort.

In conclusion, we find that innovation diffusion among in smallholder farmer households can be suitably modelled with an agent-based approach. This type of model allows for much flexibility which is needed for modelling the dynamic process of innovation diffusion in an region, which is not researched thoroughly. Furthermore, our tool provides researchers with a cost-effect option to simulate intervention processes in these regions compared to on-field trials.

Answer Second Research Question The second research question of this thesis is:

RQ2 How can different intervention strategies be evaluated and compared?

Our agent-based model developed in section 4 provides the grounds to test different approaches on how to accelerate the diffusion of innovation in a system. To test such approaches in the setting of Sub-Saharan smallholder farmer households, only field experiments or findings from other research areas could be used. On the one hand, field experiments certainly generate the most significant results but in return are also very costly and labour intensive. On the other hand, using research from a different setting is inexpensive however it may not be possible to map all findings reasonably well due to the discrepancies between the application areas. This gap is filled with our model. It is representative of the Sub-Saharan smallholder farmer households system, however, evaluating different intervention strategies is inexpensive. Hence, a user is able to test the performance of a planned strategy beforehand and already optimize it before testing it in the real world.

For this thesis, we analyzed two different intervention types which potentially can influence the attitude of farmers towards an innovation. Direct intervention and training of trainers were selected as they both are used in the ongoing field study of the University of Zurich. For both types of intervention, we defined certain parameter which together describe the operation principle of a strategy in detail in section 5. Moreover, we defined five exemplary intervention strategies which are composed of both intervention types in various configurations. Even though these strategies are potentially too rudimentary for real life deployment, they illustrate how potential strategies may be defined for future users. Parametrization of potential intervention types is important as it allows further analysis of the results and the comparison between different types and strategies.

A first qualitative comparison of these five intervention strategies revealed that the configuration of the contained intervention types significantly impacts the outcome in terms of the number of adopters. While direct intervention reaches a large number of farmers fast, training of trainers seems to increase the rate of adoption after some time. We found that the combination of both intervention types in a strategy produces the best results and highest rates of adoption. Nonetheless, these first results must be interpreted with caution as the model must be fitted adequately to the real world before drawing final conclusions.

In a larger experiment, we used three different approaches to compare the performance of different intervention strategies. A multi-variant linear regression on the intervention parameter and the number of adopters proofed not to be adequate as the correlation seems to be non-linear. Hence, we applied two supervised learning algorithms to predict the performance of the intervention strategies which both are used frequently in machine learning. While the Random Forest approach performed similarly to the linear regression, the gradient boosting algorithm performed reasonably well with an average relative error of **6.1%**. However, not all parameter of the interventions were of the same importance. It showed that direct interventions were significantly more important than the training of trainers approach. Note that this may also imply the parameter configuration of the model may need adjustments. Nevertheless, we showed that the application of machine learning on such results is feasible and has potential to be researched in more detail.

All things considered, intervention strategies can be efficiently compared with the help of our model. Implementing our ABM in NetLogo allows users to run simulations and test different strategies with minimal effort. Furthermore, we provide guidance how a potential evaluation of strategies can be executed by applying different regression techniques as well as a more in-depth qualitative analysis.

6.2 Future Work

The possibility for future research based on this thesis in manifold. We propose suggestions in two directions: extension of our ABM and the evaluation of the effectiveness of intervention strategies. In a first step, the model should be adjusted by parameter values obtained from the ongoing field study to accurately represent the underlying social system. This is key for all future work based on this model as it allows to draw representative conclusion from performed simulations. Moreover, the model is built in a very modular manner in order that various features may be dynamically added or removed. Therefore, new concepts may be easily added to the model without effort. An interesting addition to the model would be to incorporate geographical data of the underlying environment. The random generation of villages could then be replaced with precise data which paves the way to model the social system more in detail. This data was not available at the time of writing this thesis. Modelling the economy in rural Tanzania could be another interesting addition to our model. This could particularly influence the adoption decision as economic limitations of farmers could be simulated more accurately. Considering the seasonality of food insecurity caused by the time of harvest, the adoption rate is certainly influenced by the time of the year.

Furthermore, we implemented two intervention types to accelerate diffusion of innovation and generated different intervention strategies based on them to demonstrate how their effectiveness may be evaluated and compared. Especially using more sophisticated machine-learning algorithms in order to analyse the performance of intervention strategies should be examined in more detail. A common problem for machine learning is the limiting amount of data available for training a model. However, as our model is able to produce unlimited data, various machine learning approaches could be used. Moreover, new intervention types can be added to the model to simulate more elaborate intervention strategies. For example, it could be tested whether focused targeting of opinion leaders in the social system leads to faster diffusion.

Appendix A

Parameter Definition

Type	Paramter Name	Group	Description
Setup	avg-nr-of-farmers-per-village	Network	Defines the average number of smallholder households inhabiting a village.
	nr-of-villages	Network	Defines the number of villages which form the environment of the modelled system.
	avg-nr-inter-village-friends	Network	Defines the average number of inter-village friends of a farmer.
	max-influence-adopter-type-mention	Interaction - Mentioning	Defines the maximum influence of the participants adopter types on the mentioning probability in percent.
	max-influence-prev-interactions	Interaction - Mentioning	Defines the maximum influence of participants previous, innovation related interactions on the mentioning probability in percent.
	opt-number-previous-interactions	Interaction - Mentioning	Defines the optimal number of previous innovation related interactions for an agent, which did not adopt yet, to mention the innovation again.
	max-influence-adoption-state	Interaction - Mentioning	Defines the maximum influence of the participants adoption state on the mentioning probability in percent.
	avg-adoption-decision-interval	Adoption	Defines the average interval length between agent's adoption decision making.
	max-influence-adopter-type	Adoption	Defines the maximal influence of the agents adopter type on the adoption probability in percent.
	max-influence-adoption-rate-friends	Adoption	Defines the maximal influence of the adoption rate of the agent's inter-/intra-village network on the adoption probability in percent.
	max-influence-attitude	Adoption	Defines the maximal influence of the agent's attitude on the adoption probability in percent.
	adopter-influence	Adoption	Defines the influence of an adopter during an interaction.
	attitude-decrease-per-tick	Adoption	Defines by which amount the attitude of an agent declines if no topic related interaction happens.
	base-attitude-change	Interaction - Attitude	Defines by which amount the attitude of an agent declines if no topic related interaction happens.
	base-influence-counterpart	Interaction - Attitude	Defines the base influence of an agent on another.
	chief-influence	Interaction - Attitude	Defines the influence of a chief on the members during an farm group meeting.

	direct-ad-influence	Intervention - Attitude	Defines the influence of a direct intervention on an agent’s attitude.
	train-chiefs-influence	Intervention - Attitude	Defines the influence of a ToT intervention on a schief’s attitude.
Simulation	avg-intra-village-interaction-frequency	Interaction	Defines the average time span which elapses between intra-village interactions started by an agent.
	avg-inter-village-interaction-frequency	Interaction	Defines the average time span which elapses between inter-village interactions started by an agent.
	avg-farmgroup-meeting-frequency	Interaction	Defines the average time span which elapses between farmgroup meetings.
	farmgroup-meeting-attendance-percentage	Interaction	Defines the percentage of farmgroup members which actually attend the meeting.
	avg-mention-percentage	Interaction - Mentioning	Defines the base probability of the innovation being mentioned in an interaction.
	avg-farmgroup-mention-percentage	Interaction - Mentioning	Defines the base probability of the innovation being mentioned during a farm group meeting.
	percentage-negative-WoM	Interaction - Attitude	Defines the percentage of the conversations which influence the participants negatively.
	base-adoption-probability	Adoption	Defines the base probability of an agent adopting an innovation at a given time.
	nr-of-chiefs	Intervention - ToT	Defines the number of targeted chiefs in a ToT intervention.
	nr-of-villages	Intervention - Direct	Defines the number of targeted villages in a direct intervention.
	percentage-of-villagers	Intervention - Direct	Defines the percentage of inhabitants targeted in selected villages in a direct intervention.
	direct-ad-frequency	Intervention - Direct	Defines the interval between direct interventions in a intervention strategy.
	train-chiefs-frequency	Intervention - ToT	Defines the interval between ToT interventions in a intervention strategy.

Table A.1: A complete list of all parameter, which are configurable in our model, with a description.

Parameter Name	**Value**
avg-nr-of-farmers-per-village	10
nr-of-villages	100
avg-nr-inter-village-friends	5
max-influence-adopter-type-mention	25%
max-influence-prev-interactions	20%
opt-number-previous-interactions	10
max-influence-adoption-state	10%
avg-adoption-decision-interval	5 days
max-influence-adopter-type-adoption	10%
max-influence-adoption-rate-friends	20%
max-influence-attitude	50%
attitude-decrease-per-tick	0.05
adopter-influence	120%
base-attitude-change	1
base-influence-counterpart	100%
chief-influence	175%
direct-ad-influence	150%
train-chiefs-influence	200%
avg-intra-village-interaction-frequency	4 days
avg-inter-village-interaction-frequency	5 days
avg-farmgroup-meeting-frequency	7 days
farmgroup-meeting-attendance-percentage	90%
avg-mention-percentage	1%
avg-farmgroup-mention-percentage	10%
percentage-negative-WoM	10%
base-adoption-probability	1%

Table A.2: Parameter Settings Used for all Experiments in this Thesis in Section 5.

Bibliography

[1] *Transforming our world: the 2030 Agenda for Sustainable Development*. 2015.

[2] *Transforming food systems for food security, improved nutrition and affordabel healthy diets for all*, volume 2021 of *The state of food security and nutrition in the world*. FAO, 2021.

[3] Lutz Goedde, Amandla Ooko-ombaka, and Gillian Pais. Winning in africa's agricultural market. `https://www.mckinsey.com/industries/agriculture/our-insights/winning-in-africas-agricultural-market`. Accessed: 2022-03-22.

[4] National bureau of statistics - population and housing census. `https://www.nbs.go.tz/index.php/en/census-surveys/population-and-housing-census`. Accessed: 2022-02-19.

[5] Michael Brander, Thomas Bernauer, and Matthias Huss. Improved on-farm storage reduces seasonal food insecurity of smallholder farmer households – evidence from a randomized control trial in tanzania. 98:101891, 2021. PII: S0306919220300932.

[6] Paul Christian and Brian Dillon. Growing and learning when consumption is seasonal: Long-term evidence from tanzania. 55(3):1091–1118, 2018. Journal Article Research Support, Non-U.S. Gov't.

[7] M. Huss, M. Brander, M. Kassie, U. Ehlert, and T. Bernauer. Improved storage mitigates vulnerability to food-supply shocks in smallholder agriculture during the covid-19 pandemic. 28:100468, 2021. PII: S221191242030122X.

[8] Hippolyte Affognon, Christopher Mutungi, Pascal Sanginga, and Christian Borgemeister. Unpacking postharvest losses in sub-saharan africa: A meta-analysis. 66:49–68, 2015. PII: S0305750X14002307.

[9] Adebayo B. Abass, Martin Fischler, Kurt Schneider, Shamim Daudi, Audifas Gaspar, Janine Rüst, Esther Kabula, Gabriel Ndunguru, Daniel Madulu, and David Msola. On-farm comparison of different postharvest storage technologies in a maize farming system of tanzania central corridor. 77:55–65, 2018. PII: S0022474X17304447.

[10] Everett M. Rogers. *Diffusion of Innovations*. Free Press, 5th ed. edition, 2003.

[11] Seline S. Meijer, Delia Catacutan, Oluyede C. Ajayi, Gudeta W. Sileshi, and Maarten Nieuwenhuis. The role of knowledge, attitudes and perceptions in the uptake of agricultural and agroforestry innovations among smallholder farmers in sub-saharan africa. 13(1):40–54, 2015.

[12] Miyuki Iiyama, Athanase Mukuralinda, Jean Damascene Ndayambaje, Bernard S. Musana, Alain Ndoli, Jeremias G. Mowo, Dennis Garrity, Stephen Ling, and Vicky Ruganzu. Addressing the paradox – the divergence between smallholders' preference and actual adoption of agricultural innovations. 16(6):472–485, 2018.

[13] Seth Tisue and Uri Wilensky. Netlogo: A simple environment for modeling complexity. pages 16–21, 2004.

[14] William Rand and Roland T. Rust. Agent-based modeling in marketing: Guidelines for rigor. 28(3):181–193, 2011. PII: S0167811611000504.

[15] Steven M. Manson, Nicholas R. Jordan, Kristen C. Nelson, and Rachel F. Brummel. Modeling the effect of social networks on adoption of multifunctional agriculture. 75:388–401, 2016. PII: S1364815214002795.

[16] Guillaume Deffuant, Frédéric Amblard, Sylvie Huet, Stéphan Bernard, and G. Weisbuch. Agent-based simulation of organic farming conversion in allier. In *Complexity and Ecosystem Management: the Theory and Practice of Multi-agent Systems*, page 19. 2002.

[17] Floor H.W. Ambrosius, Mark R. Kramer, Alisa Spiegel, Eddie A.M. Bokkers, Bettina B. Bock, and Gert Jan Hofstede. Diffusion of organic farming among dutch pig farmers: An agent-based model. 197:103336, 2022. PII: S0308521X21002894.

[18] Tanzania - the world factbook. `https://www.cia.gov/the-world-factbook/countries/tanzania/#people-and-society`. Accessed: 2021-11-15.

[19] Zhanli Sun, Iris Lorscheid, James D. Millington, Steffen Lauf, Nicholas R. Magliocca, Jürgen Groeneveld, Stefano Balbi, Henning Nolzen, Birgit Müller, Jule Schulze, and Carsten M. Buchmann. Simple or complicated agent-based models? a complicated issue. 86:56–67, 2016.

[20] G. Edwards-Jones. Modelling farmer decision-making: concepts, progress and challenges. 82(6):783–790, 2006. PII: S1357729806000944.

[21] Bartosz Bartkowski and Stephan Bartke. Leverage points for governing agricultural soils: A review of empirical studies of european farmers' decision-making. 10(9):3179, 2018. PII: su10093179.

[22] Floor H. W. Ambrosius, Gert Jan Hofstede, Bettina B. Bock, Eddie A.M. Bokkers, and Adrie J.M. Beulens. Modelling farmer decision-making: the case of the dutch pork sector. 117(10):2582–2597, 2015. Stefano Pascucci, Dr Liesbeth Dries, Dr (Editor).

[23] Uri Wilensky. Netlogo (and netlogo user manual). `http://ccl.northwestern.edu/netlogo/`. Accessed: 2022-02-19.

[24] Nick Collier. Repast: An extensible framework for agent simulation. 2001.

[25] Sebastian Rodriguez, Nicolas Gaud, and Stephane Galland. Sarl: A general-purpose agent-oriented programming language. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE, 2014.

[26] Haifeng Zhang and Yevgeniy Vorobeychik. Empirically grounded agent-based models of innovation diffusion: a critical review. 52(1):707–741, 2019. PII: 9577.

[27] Elmar Kiesling, Markus Günther, Christian Stummer, and Lea M. Wakolbinger. Agent-based simulation of innovation diffusion: a review. 20(2):183–230, 2012. PII: 210.

[28] Kristen C. Nelson, Rachel F. Brummel, Nicholas Jordan, and Steven Manson. Social networks in complex human and natural systems: the case of rotational grazing, weak ties, and eastern us dairy landscapes. 31(2):245–259, 2014. PII: 9462.

[29] Nikola Diemer, Philipp Staudacher, Aggrey Atuhaire, Samuel Fuhrimann, and Jennifer Inauen. Smallholder farmers' information behavior differs for organic versus conventional pest management strategies: A qualitative study in uganda. 257:120465, 2020. PII: S0959652620305126.

[30] Duncan J. Watts. Networks, dynamics, and the small–world phenomenon. 105(2):493–527, 1999.

[31] Guillaume Deffuant, Sylvie Huet, and Frédéric Amblard. An individual–based model of innovation diffusion mixing social value and individual benefit. 110(4):1041–1069, 2005.

[32] Mark Granovetter. Threshold models of collective behavior. 1978.

[33] Barak Libai, Ruth Bolton, Marnix S. Bügel, Ko de Ruyter, Oliver Götz, Hans Risselada, and Andrew T. Stephen. Customer-to-customer interactions: Broadening the scope of word of mouth research. 13(3):267–282, 2010.

[34] Jonah Berger. Word of mouth and interpersonal communication: A review and directions for future research. 24(4):586–607, 2014. PII: S1057740814000369.

[35] Andrew T. Stephen and Donald R. Lehmann. How word-of-mouth transmission encouragement affects consumers' transmission decisions, receiver selection, and diffusion speed. 33(4):755–766, 2016. PII: S016781161630088X.

[36] Chiteji, Frank Matthew, Bryceson, Deborah Fahy, Mascarenhas, Adolfo C. and Ingham, Kenneth. Tanzania. encyclopedia britannica. `https://www.britannica.com/place/Tanzania`. Accessed: 2021-11-15.

[37] World Bank. Tanzania mainland poverty assessment. `https://openknowledge.worldbank.org/handle/10986/33031`, 2019. Accessed: 2021-11-15.

[38] Sarah K. Lowder, Jakob Skoet, and Terri Raney. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. 87:16–29, 2016. PII: S0305750X15002703.

[39] R. Msigwa. Challenges facing urban transportation in tanzania. 2013.

[40] Emmanuel Fungo. Road infrastructure improvement for efficient utilisation of the agricultural potential : a case study of morogoro, tanzania. 2018.

[41] Uri Wilensky and William Rand. *Introduction to agent-based modeling*. The MIT Press, 2015.

[42] Mitchel Resnick. Beyond the centralized mindset. 5:1–22, 1996.

[43] The AnyLogic Company. `https://www.anylogic.com`. Accessed: 2022-02-10.

[44] Jayson W. Richardson. The adoption of technology training by teacher trainers in cambodia: A study of the diffusion of an ict innovation.

[45] Jose Fulgencio and Tutaleni Asino. Oer training of trainers model: A possible diffusion strategy. 16(2), 2021.

[46] Maru Mormina and Sophie Pinder. A conceptual framework for training of trainers (tot) interventions in global health. 14(1):100, 2018. Journal Article Research Support, Non-U.S. Gov't.

[47] Frank Avery Haight. *Handbook of the Poisson distribution*. Publications in operations research no. 11. J. Wiley, 1967.

[48] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd, fifth edition edition, 2012. Montgomery, Douglas C., (author.) Peck, Elizabeth A., (author.) Vining, G. Geoffrey, (author.).

[49] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, number 1, pages 278–282, 1995.

[50] Leo Breiman. Random forests. 45(1):5–32, 2001. PII: 354300.

[51] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. 12(85):2825–2830, 2011.

[52] Tianqi Chen and Carlos Guestrin. Xgboost, 2016.

[53] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. 29(5), 2001.