The importance of being accurate in agent-based models - an illustration with agent aging

Julia Schindler

Center for Development Research, Bonn University,
Walter-Flex-Str. 3, 53113 Bonn, Germany
julia.schindler@uni-bonn.de
http://www.julia-schindler.de

Abstract. Sometimes, descriptive and empirical agent-based models contain sub-models of variables whose change over time cannot be easily modelled. Due to a frequent lack of long-term data, especially on demographic change, often logical design principles are used for modeling such changes. The problem with these modeling principles is that many of them exist and the choice is a difficult one. We would like to illustrate in this article that this choice has consequences, using the example of farm agent aging sub-models in an existing land-use change model. We will show that different design principles lead to different long-term changes in age pattern. The problem here is that one cannot be sure where the change in age structure comes from, i.e. whether it is an account of real-world change and/or whether it is caused by model assumptions. Then, we will show that simulated land-use change is highly sensitive to the type of aging sub-model. As a solution, we develop an alternative aging sub-model that exactly maintains an age pattern externally prescribed, but still allows individual aging.

Keywords: agent-based modelling, accuracy, critique, agent age, demography

1 Introduction

Although empirical agent-based models have often been rigorously built on snapshot survey data (e.g. [10], [13], [2]), a difficulty that often appears is how to model agent variables whose changes over time cannot be easily modelled using such data. Especially demographic variables, such as agent age, household size, household labor, but also animal stocks, etc., are difficult to model without a long-term database. Many approaches to circumvent this difficulty exist, and nearly every model has its own approach. However, we believe that the way we modellers circumvent the problem of unknown data can have consequences for model credibility and accuracy. In this article, we aim to demonstrate that different approaches to model change of age of farm agents can have far-reaching consequences for simulation results. We will show that different existing models of agent aging converge to different age structures using the same database, where it is not clear whether real-world change was simulated, and/or whether the change was caused by model assumptions. Further, we will illustrate in a case study that the change of aging model in an existing empirical agent-based model of land-use change can have drastic consequences for simulation results. We offer a solution for this problem, via developing and implementing an aging model that is able to control assumptions made.

In section 2, we review models of aging for farm agents, and the problem is described. In section 4, we offer an improved model to deal with the described problem, and in section 4, we illustrate with a case study what consequences the choice of model can have for simulation results. Finally, we conclude that modellers should be more self-critical when developing algorithms for unknown variable change.

2 Problem Analysis

We reviewed existing algorithms to model the change of age of farming agents. Doing so, we constricted ourselves to simple algorithms whose dynamics consist of agents' aging, death and replacement, not including algorithms that explicitly model mating and reproduction during lifetime. We excluded the latter because these can be regarded as sophisticated enough to replicate human reproduction explicitly - a quality that we think the former does not have. For each reviewed algorithm class we identified the modeling principle in order to be able to adapt the corresponding algorithm to our case study. Since in peer-reviewed articles the aging algorithm is often not explained in sufficient detail, the review was mainly limited to dissertations and master theses.

We start with the simplest idea of modeling the aging process of farming agents. None of the reviewed models in fact implemented it, but we found it worth examining, because it sounds straightforward but has an oblique rub in it. The algorithm consists of the replacement of agents that reach a certain age by young agents (equation 1). For simplicity, we only reset the age of the "dying" agent instead of deleting and recreating an agent. This applies to the entire remainder of the text.

$$\text{If } H_{age} > Age_{old} \text{ then } H_{age} \mapsto Age_{young} \tag{1}$$

where Age_{old} is the death age and Age_{young} the age of agents starting farming.

The rationale behind this algorithm is that in many real-world systems the role an agent has assumed during lifetime does not vanish with agent death, but is often only refilled with a younger agent. The question seems to be just where to set the death age and the age of the succeeding agent. But the more subtle problem is that this algorithm is not able to maintain the age distribution of the agent population when the distribution is non-uniform. What happens to non-uniform distributions is that the distribution form as such is maintained but simply shifts with age over time, while disappearing at old age and reappearing at young age (Figure 1).

In reviewed models on farming agents, this issue has been approached differently. We identified three different principles that suggest how to model aging more realistically. The first has been applied by the LUDAS (Land Use DynAmic Simulator) family (e.g. [10], [6]), where agents, when reaching the maximum age, are assigned an age randomly drawn from the range consisting of the mean age plus/minus the standard deviation (equation 2).

$$\text{If } H_{age} > Age_{max} \text{ then } H_{age} \mapsto Age_{mean} - Age_{std} + rand(2 \cdot Age_{std}) \qquad (2)$$

where H_{age} is an agent's age, Age_{max}, Age_{mean}, and Age_{std} are maximum, mean, and standard deviation of the original empirical age dataset, respectively, and $rand()$ is a function that draws a random floating number between zero and the assigned number.

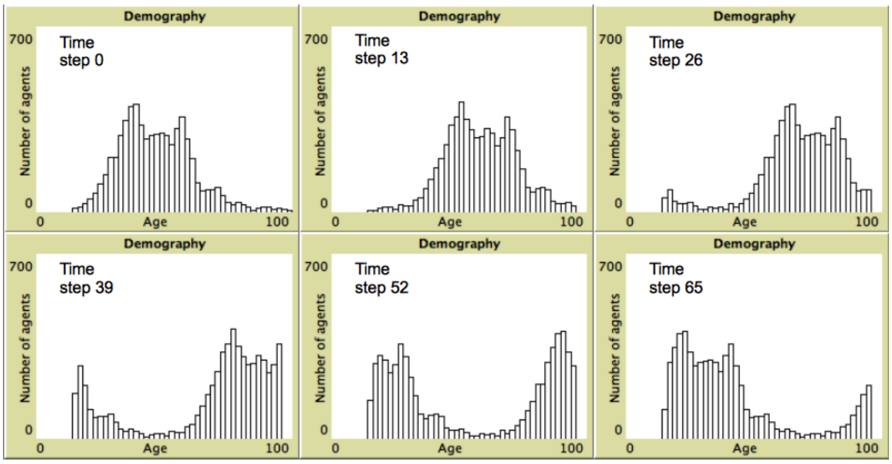

Fig. 1. Shift of age structure over time in terms of number of agents per age class for simple replacement algorithm (death age = 95, young age = 15)

The next modeling principle that we identified is drawn from a series of publications on regional structural change in Canada (e.g. [3], [15], [1]). It consists of letting agents die with death probabilities derived from empirical life tables and replacing them with agents with age one generation younger (equation 3). The idea of replacing agents by agents of age minus the generation-cycle interval is a widespread strategy. The widely used AgriPoliS model for simulating agricultural policies (e.g. [4], [8]) also uses this idea, although in none of the reviewed AgriPoliS publications this value was specifically given. In all three reviewed publications on regional structural change, the generation-cycle interval was set to 30 years, which the authors concede is an assumption.

$$\text{If } rand(100) < D_{age} \text{ then } H_{age} \mapsto H_{age} - G_{int} \tag{3}$$

Table 1. Parameter values used for implementing the selected principles for the Ghanaian case study

Parameter	Value (years)	Explanation and Source
H_{old}	75	mean death age calculated from D_{age} over all ages
H_{young}	21	mean of minimum age Age_{min} and reproduction age G_{init}
Age_{max}	95	maximum age derived from [12]
Age_{min}	15	minimum age derived from [12]
H_{mean}	46.3	mean age calculated from [12]
H_{std}	14.1	standard deviation of age calculated from [12]
G_{int}	27	rural marriage age of males by [7]
D_{age}	—	death probability per age class using [5]

where D_{age} is the death probability for the agent's age class, and G_{int} the generation-cycle interval.

The third principle we derived from a publication by Millington et al. [11], who also use life-table probabilities to model death, but replace the agent only by a new agent with younger age if the deceasing agent is old enough to have offspring. Millington et al. define this age as the minimum age an agent can become farmer plus the empirical generation-cycle interval [11]. Accordingly, the age of the new agent is assigned a random age between this minimum age and the minimum age plus generation-cycle interval. Millington et al. reason that if the "dying" agent is not old enough to have offspring, it is replaced by another random agent. Thus, in our terms, it simply obtains the age of another random agent. Although Millington et al. also implemented other reasons to stop being a farmer, we use an adapted version of this principle that only focuses on death as a cause, which is sufficient for our purpose (equation 4).

$$\text{If } rand(100) < D_{age} \begin{cases} \text{If } H_{age} > Age_{min} + G_{init} \text{ then } H_{age} \mapsto Age_{min} + rand(G_{init}) \\ \text{else } H_{age} \mapsto H_{age} \text{ of another random agent} \end{cases} \tag{4}$$

where Age_{min} is the minimum age at which agents can be farmer.

In order to test these modeling principles, we implemented each of them and calibrated them for a Ghanaian case study, namely the Atankwidi Catchment in Upper East Ghana (Table 1). Own data collected in the field were used [12], and life tables from the district from the same year [5]. The death probabilities were derived from the life tables via extrapolation, since they covered only ages above

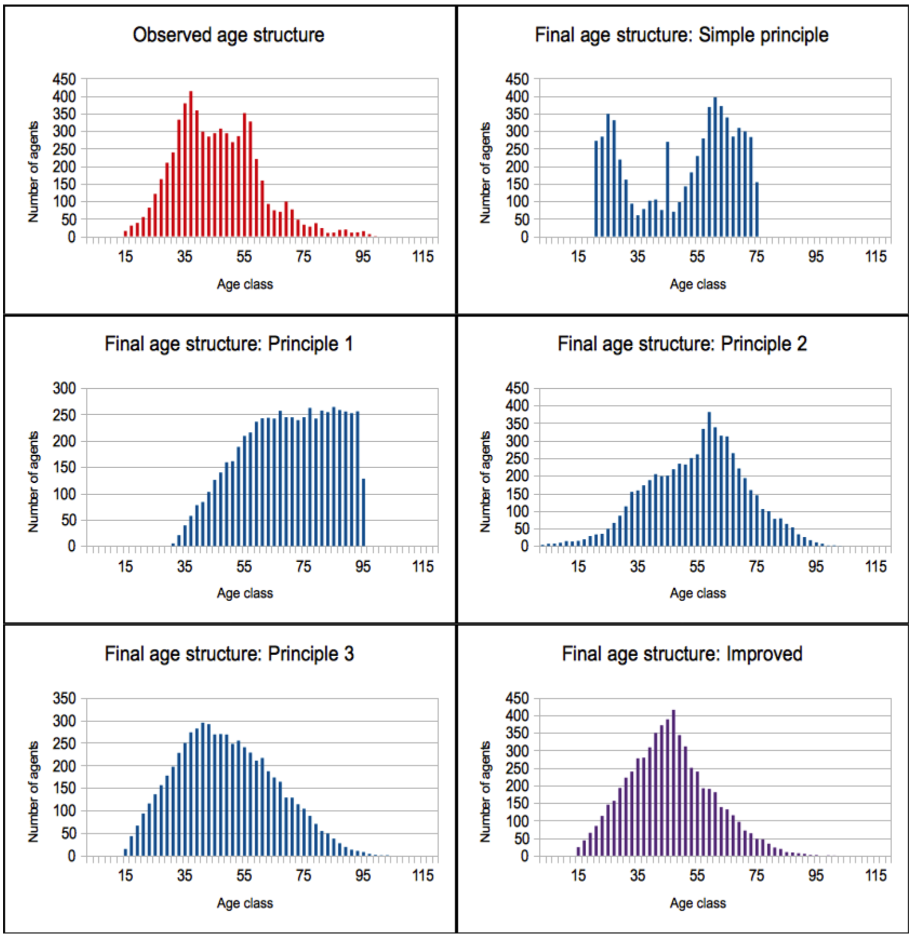

Fig. 2. Initial observed age structure and simulated age structures after 300 time steps

sixty years. The source for the generation-cycle interval for males in rural Ghana is unfortunately somewhat outdated (see [7]), but given the fact that cultural change was slow during the last decades in the study area, and since we believe that generation cycles of males are still high in rural West Africa, we decided to stick to the reported value (Table 1).

We simulated each principle five times for 300 time steps (years) (the simulation period was set this high to ensure convergence of the age structures (see Figures 2 and 3)). Henceforth, with principle 1 we will refer to the principle of LUDAS by [9] with principle 2 to that of the Canadian group (e.g. [3], and with principle 3 to that largely following Millington et al. [11]. The simple algorithm described above was also simulated, showing the behavior as discussed above, i.e. the distribution simply shifted and was not maintained, resulting in an up-and-down bounce of mean age and standard deviation (Figure 3). Imple-

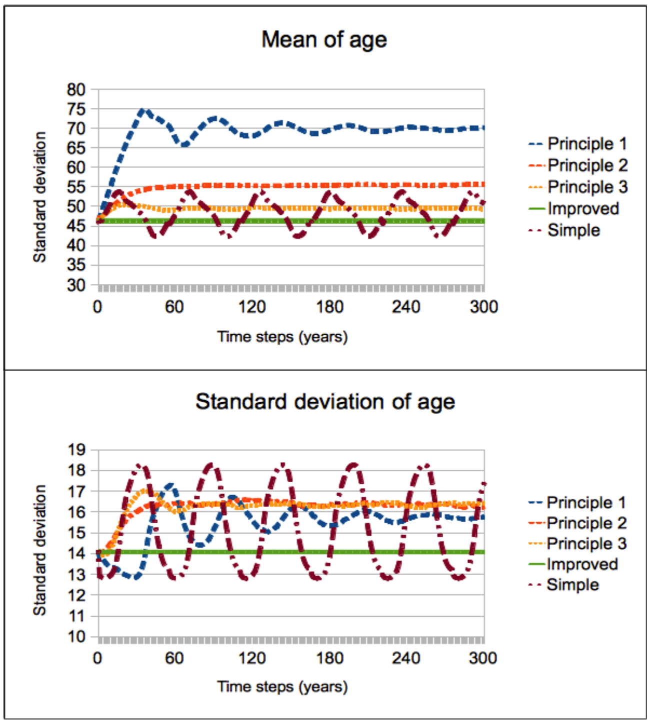

Fig. 3. Simulated age and standard deviation after 300 time steps

menting principle 1 on the other hand, which aimed at keeping the distribution more-or-less stable, resulted in a convergence to a significantly higher mean age and standard deviation over time (Figure 3). The reason is that the algorithm assigns to new agents only values around the initial mean, thus removing the low age values in the long term but maintaining high values, leading to a shift of mean age in the positive direction over time. Furthermore, the shape of the age structure converged to a significantly changed structure over time (Figure 2). One of the causes for this significant change in structure shape is that agent life was designed to end at the same age for all. Principles 2 and 3 solved this issue by using death probabilities for ages classes, resulting in a smooth curve of the right-hand side of the age structure (Figure 2). But both of them also showed an increase of mean age and standard deviation (Figure 3).

The problem with these modeling principles is that one cannot determine where the change in age structure comes from. Is it an account of real-world change or is it caused by model assumptions, or is it even a bit of both? As long as one does not model demographic change in detail, the model is prone to forcing the modeller to make (possibly weak) assumptions. We do not claim that these models are wrong, but the problem is that one cannot be sure that one is right. We neither claim that one is not allowed to make assumptions, but that one should keep control what happens with these assumptions. That is, whether the emergent age structure is the one that was indeed intended by assumption. The solution that we propose is to externalize assumptions in such as way that the shape of the age structure itself can be controlled. This implies for us to develop a model that enables individual aging of agents on the one hand, but on the other is able to adjust death and youth ages in such a way that the age structure changes the way prescribed, as e.g. by expert opinion. This way, we aim to offer a model that is cleaned from internal and potentially inaccurate assumptions.

```
If (mean [age] of turtles > mean-age) [
  while [mean [age] of turtles > mean-age] [
   ask one-of turtles with [any? turtles with [age < [age] of myself]] [
    set age [age] of one-of turtles with [age < [age] of myself]
    ask turtles with [age != nobody] [set age-1 age set age+1 age]
    set age-1 age - 1 set age+1 age + 1
    while [
     (abs(standard-deviation [age] of turtles - std) >
     abs(standard-deviation [age-1] of turtles - std) and age > 15) or
     (abs(standard-deviation [age] of turtles - std) >
     abs(standard-deviation [age+1] of turtles - std))
    ] [
     ifelse (abs(standard-deviation [age-1] of turtles - std) <
     abs(standard-deviation [age+1] of turtles - std)) [
      set age age - 1
     ]
     [
      set age age + 1
     ]
     set age-1 age - 1 set age+1 age + 1
    ]
   ]
  ]
]
```

Fig. 4. First part of algorithm extension for equation 4 for improved aging modelling

3 An Improved Model for Simulating Aging of Farm Agents

For developing such an improved model we decided to build on the model by principle 3, since it comes closest to the observed structure in shape. We extended the algorithm of principle 3 in such a way that the simulated mean and standard deviation coincide with the prescribed mean and standard deviation. Please note that the observed is **not** normally distributed (significant at 0.001 level with

Kolmogorov-Smirnov test). But in the absence of other simple measures that can describe the observed age structure, we went for the simplest available, although we must concede that other important characteristics such as kurtosis may not be maintained this way.

We extended the modelling principle 3 by an algorithm that adjusts deaths and newcomer ages to match the prescribed mean and standard deviation. It consists of two parts. The first part does adjustments when the mean is too high, and the second when the mean is too low. The first part is depicted in NetLogo 4.1.3 code (Figure 4), while the second part has the same code except that the "larger than" and "lower than" signs in the first four lines are to be reversed. In the first part, as long as the mean is lower than the prescribed mean (in the algorithm: mean-age), a random agent with non-minimum age is selected and its age is changed to the age of a random agent with lower age. Then, the age of this adjusted agent is again adjusted in steps of one year until a minimum difference between simulated standard deviation and prescribed standard deviation (in the algorithm: std) is achieved. Thereat, the age is not allowed to fall below 15 years, which is the minimum age boys can be household heads in the study area. The explanation of the second part is analogous, given the reversal of the signs in the first four lines.

We implemented this algorithm for the case-study data set (Table 1), and conducted five test simulations for 300 years like for the other modeling principles. Results verify that the simulated mean and the standard deviation remain constant at the prescribed mean and standard deviation, respectively (Figure 3), which we set to the initial observed values. Also, the shape of the age distribution seems to mimic the age structure better than the other models (Figure 2, bottom right).

4 The Importance of Being Accurate: A Case Study

In order to illustrate what consequences the choice of modeling principle for aging of farm agents can have, we implemented the selected algorithms in a fully-fledged model for simulating land-use change in the selected study area. The model is one of the model versions described in detail in [14] (version Ord-L) and follows the modeling framework of the LUDAS family, from which we derived the first modeling principle.

The LUDAS framework envisages the representation of the human-environment feedback loop via integrating both long-term change of farm livelihood strategies and short-term system changes in the human land-use choice process (for details see [10], [14]). This is implemented by household agents running a land-use choice phase, in which patches are selected for use, i.e. their location, land-use type, and management mode is selected, as long as the agent has sufficient spare labor for management. Thereat, first owned patches are cultivated, and then alien patches, as long as labor is available. The choice of land-use type is modelled via multinomial logistic regression using environmental and household variables, while the preference coefficients are dependent on the livelihood type

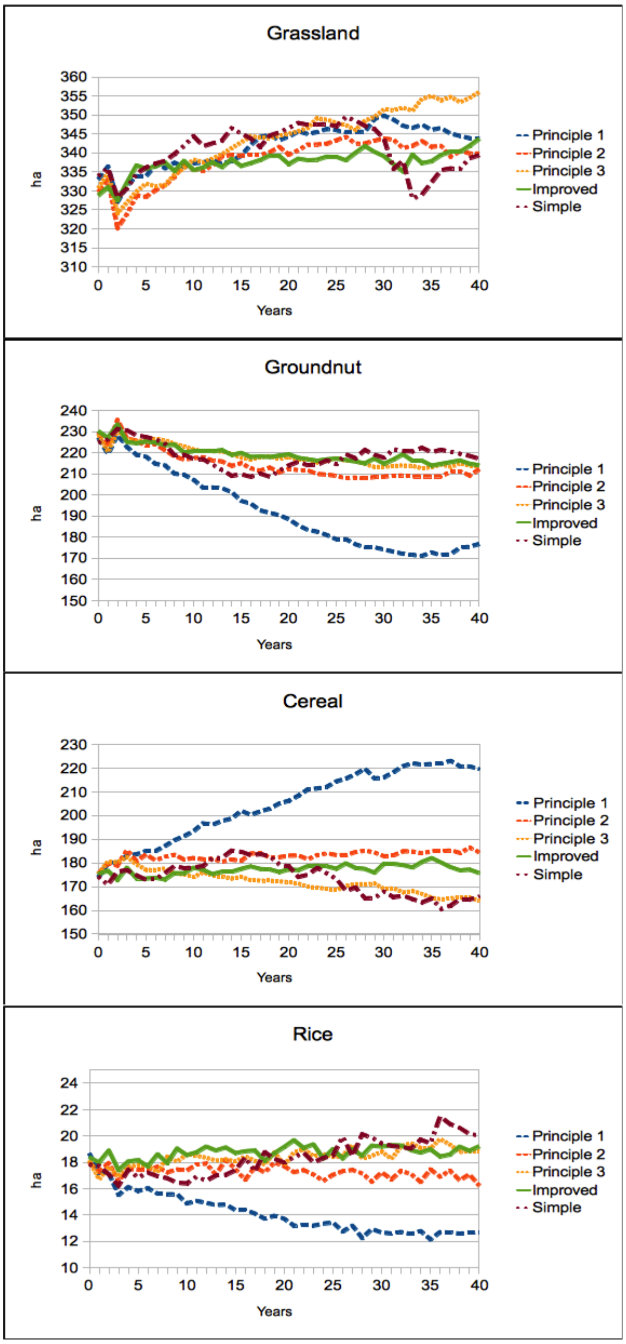

Fig. 5. Simulated land-use areas (ha) over time

of the agent. These livelihood types represent farm agent archetypes with typical livelihood strategies, which is assigned to agents anew each time step, dependent on the agent's (changed) variable values. In this version, the management-mode choice is simply modelled via assignment of categorical values dependent on other agent and landscape characteristics. Models of ecological returns to these decisions complete the model. These feed back into the agent's livelihood and agricultural choice routines. All used data and sub-models are based on data collected during two household surveys conducted in 2006 ([13], [12]). The simulated land-use types comprised grassland as grazing land and the main staples in the area, i.e. cereals, groundnut, and rice. The simulated area comprises an area of 918 ha located in the Atankwidi Catchment, which is the same study area from which age data were used above.

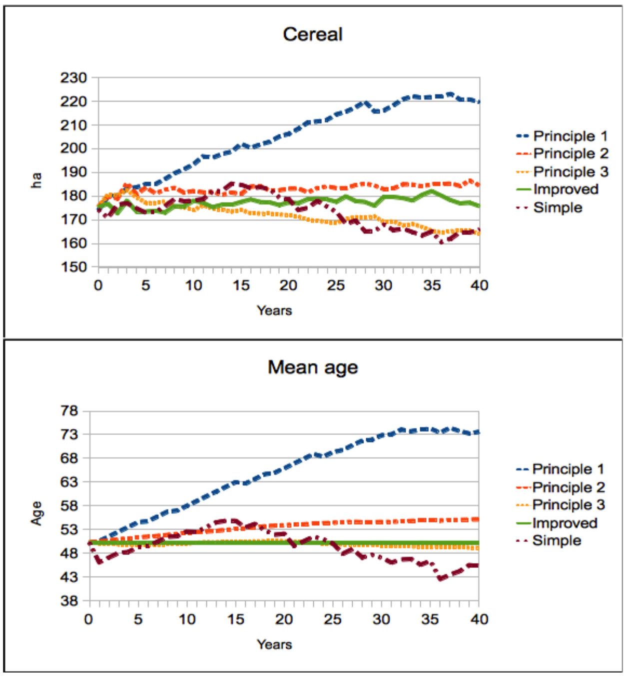

Fig. 6. Simulated mean age (years) and cereal area (ha) over time, rescaled

We simulated each principle 20 times for 40 years, which is the time period that we estimate the model can more-or-less reproduce. Results clearly show how sensitive the model is to the choice of aging sub-model (Figure 5 and 6), where simulated areas of land use, in specific the staples highly correlated with simulated mean age (Table 2). The reason for this correlation is that the land-use choice model contains age as an explanatory variable. The selection of explanatory variables for the land-use choice model was based on hypotheses collected

Table 2. Overall correlation between mean age (years) and land-use area (ha)

Land-use type	Correlation coefficient	R^2
Grassland	0.066	0.004
Groundnut	-0.678	0.460
Cereal	0.710	0.504
Rice	-0.601	0.361

in the field, which were (partially) confirmed statistically (see [14]). Age was insofar selected as the field study indicated that younger farmers tended to favor the more cash-oriented staples groundnut and rice, while older individuals rather opted for the traditional but subsistence-oriented cereal crops such as millet and sorghum. This is also indicated by the results, where the large increase in age under principle 1 is accompanied by an accordingly lower area of rice and groundnut, and an accordingly larger area of cereal (Figure 5).

5 Summary

A review and test of aging models for farm agents showed that they, using the same database, converge to different age structures over time. Each of the simulated age structures differed from the observed age structure in a different way. The problem is that it is not clear whether simulated emergent change is an account of real-world change or a result of inaccurate assumptions, or a bit of both. And one cannot identify for which algorithm which of these causes apply. Further, it is not clear whether authors tested the assumptions made in their models, i.e. whether the resulting age structure is in line with their global understanding of how the structures would evolve. Therefore, we argue that assumptions for variables whose change over time is not or little known should be held under control at the emergent global level. In this article, we aimed to offer such an algorithm for aging of farm agents, and showed that it is able to match the age distribution prescribed over time, but also lets agents age by one year every annual time step. We finally illustrated what effects the choice among presented algorithms can have, using a case-study model for land-use change in Northern Ghana. Results suggest that the selection of aging model is one of the major factors that determines both the direction and extent of land-use change in the case-study model.

References

1. Arsenault A.M.: A Multi-Agent Simulation Approach to Farmland Auction Markets: Repeated Games with Agents that Learn. Master thesis. University of Saskatchewan (2007).

2. da Feitosa F.F.: Urban Segregation as a Complex System: An Agent-Based Simulation Approach. Dissertation. University of Bonn. Ecology and Development Series 70 (2010).
3. Freeman T.R.: From the ground up: An agent-based model of regional structural change. Master thesis. University of Saskatchewan (2005).
4. Happe K., Kellermann K. Balmann A.: Agent-based Analysis of Agricultural Policies: an Illustration of the Agricultural Policy Simulator AgriPoliS, its Adaptation and Behavior. Ecology and Society 11(1): 49 (2006).
5. Khagayi S., Debpuur C., Wak G., Odimegwu C.: Socioeconomic status and elderly adult mortality in rural Ghana: Evidence from the Navrongo DSS. African Population Studies Journal (2011).
6. Kaplan M.: Agent-based modeling of land-use changes and vulnerability assessment in a coupled socio-ecological system in the coastal zone of Sri Lanka. Dissertation. University of Bonn. Ecology and Development Series 77 (2011).
7. Lamptey P., Nicholas D.N., Ofosu-Amaah S., Lourie I.M.: An Evaluation of Male Contraceptive Acceptance in Rural Ghana. Studies in Family Planning 9(8): 222-226 (1978).
8. Lobianco A.: The Effects of Decoupling on Two Italian Regions. An Agent-based Model. Dissertation. Polytechnic University of Marche (2007).
9. Le Q.B.: Multi-agent system for simulation of land-use and land-cover change: A theoretical framework and its first implementation for an upland watershed in the Central Coast of Vietnam. Ecology and Development Series 29. Göttingen: Cuvillier Verlag (2005).
10. Le Q.B., Park S.J., Vlek P.L.G., Cremers A.B.: Land use dynamic simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. 1. Structure and theoretical specification. Ecological Informatics 3(3): 135-153 (2008).
11. Millington J., Romero-Calcerrada R., Wainwright R. Perry G.: An Agent-Based Model of Mediterranean Agricultural Land-Use/Cover Change for Examining Wildfire Risk Journal of Artificial Societies and Social Simulation 11(4): 4 (2008).
12. Schindler J.: Household surveys July and November 2006. For questionnaire see [13], (2006).
13. Schindler J.: A multi-agent system for simulating land-use and land-cover change in the Atankwidi catchment of Upper East Ghana. Dissertation. University of Bonn. Ecology and Development Series 68 (2010).
14. Schindler J.: Justification styles and their effects in agent-based models - An illustration with a land-use model. Submitted (2012).
15. Stolniuk P.C.: An Agent-Based Simulation Model of Structural Change in Agriculture. Master thesis. University of Saskatchewan (2008).