Kennedy, W.G., Cotla, C.R., Gulden, T., Coletti, M, Cioffi-Revilla, C. (submitted) Towards Validating a Model of Households and Societies in East Africa. Proceedings of the 4th World Congress on Social Simulation. Springer (available from http://www.mllab.com)

Towards Validating a Model of Households and Societies in East Africa

William G. Kennedy [1], Chenna Reddy Cotla [1], Tim Gulden [1], Mark Coletti [1], and Claudio Cioffi-Revilla [1]

[1] Center for Social Complexity, Krasnow Institute for Advanced Study, George Mason University, 4400 University Drive, Fairfax, Virginia 22030, USA
{wkennedy, ccotla, tgulden, mcoletti, ccioffi}@GMU.edu

Abstract. Model validation is one of the major challenges of social simulations. In particular, experimentation is not practical when modeling societies, which makes model validation inherently difficult. However, one strength of an Agent-Based Model (ABM) is its implementation of a theory of behavior for low-level agents from which higher-level behavior emerges. Our ABM of societies is based on low-level decision making of rural households in a 1,600 km (1000 mile) square around Lake Victoria in East Africa. We report on the first validation of this model by discussing the development of its layers and comparing the emergent high-level activities against actual societal data collected by anthropologists.

Keywords: Agent-Based Modeling (ABM), household decision-making, East Africa.

1 Introduction

One of the major challenges of Agent-Based Modeling (ABM) is the validation of a model. When modeling societies where experimentation is not practical or ethical, validation of models is inherently difficult. However, one of the significant strengths of the ABM approach is the faithful implementation of a theory of behavior for relatively low-level agents and their associated environmental dynamics and then observing high-level behaviors emerging from the low-level theory's implementation. That is the approach we have taken toward validating our model.

A team of scientists at George Mason University and at Human Relations Area Files (HRAF) at Yale University has been working a few years on an agent-based model of a large area of East Africa, including validation-related fieldwork in 2010. The purpose of this project is to answer research questions on social dynamics, such as internal conflict and responses to natural disasters and humanitarian relief. This follows earlier validation efforts in the same project on other components [1], [2], [3], and [4]. We will discuss the available validation of the layers of our model up to the model of household decision-making.

Our model includes detailed representations of the environment, specifically land types, water supplies, and weather. From these we have modeled vegetation growth for grazing of domestic herds and farming activities. We have modeled the people in the region at the household level, and households manage herds, farming, and labor activities. These subsistence activities are modeled down to actions taken on a daily basis, such as deciding when to plant, when to harvest, and where to move the herd each day.

It is relatively easy to compare modeled vegetation to actual data on vegetation. To test the macro-level performance of our model of household behavior, we use anthropological data from local ethnographies, such as primary sources from HRAF and secondary sources from extant literature (e.g, [5], [6], and [7]). Anthropologists have catalogued approximately 135 different ethnic groups in the modeled area, including data on how each the people of each culture makes their living. Therefore, by implementing theoretical decisions at the household level driven by the environmental conditions, we can see if the simulated results match anthropological data. We report on the validation of our household model against the anthropological data for the region, which will provide insights into model validation for models of societies.

2 RiftLand Model Overview

Our model of the Rift Valley of East Africa surrounds Lake Victoria. Our model, called “RiftLand” [8], was developed using the MASON system [9] and [10], and represents the area and the modeled actors at a scale appropriate for our research questions. The subject area is show in Figure 1. As an agent-based model, we attempt to model the environment and agents at a relatively low level and allow their interactions to produce macro-level results. The environment is represented at the one square kilometer level throughout the 1,600 by 1,600 km area. The time scale has one step of the model representing one day. People are modeled at the household level, keeping track of the number of individuals in each household. Our model runs are typically over several years of simulated time, or thousands of days. Major model components and behaviors are detailed below.

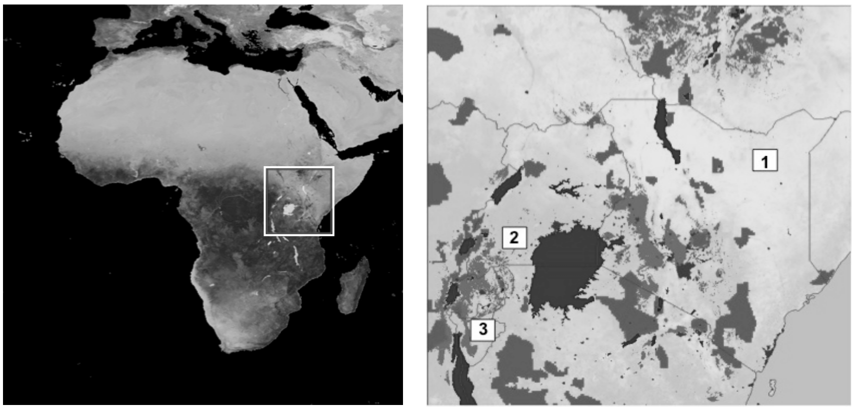

Fig. 1. The RiftLand model area and three test regions identified. Sources: NASA (left) and RiftLand model output with different land/water types (right).

1.1 Reference Data for Validation

An atlas collecting the results of many anthropologists who have lived and studied the people of the Rift Valley provides anthropological data concerning how the people of this region make their living [11]. The atlas was published by G.P. Murdock as a series of journal articles (e.g., [12]) in the journal *Ethnology* in the late 1960's. These data have been updated by HRAF. The 135 HRAF-coded ethnographic cultures of the Rift Valley region are shown in Figure 2.

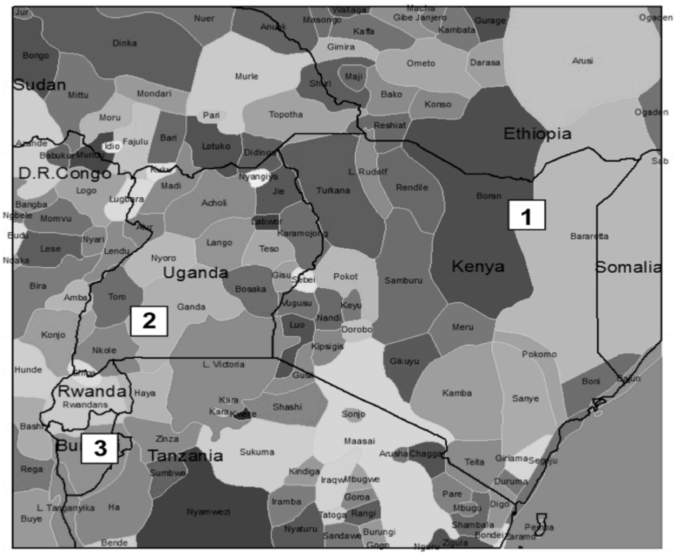

Fig. 2. Cultural Map of Rift Valley Region, based on HRAF-coded ethnographic cultures. Source: HRAF.

1.2 Modeling the Environment

The subject area includes several biomes, cultures, and polities. The 1,600 by 1,600 km area is represented by 1-km^2 parcels of land or water. This level of resolution supports reasonably accurate representation of the landscape, political borders, and locations of the people. We did not model every rock and blade of grass, but did model major factors, actors, and interactions affecting the macro-level behavior of interest. We represent differences in land use by different types of parcels. Parcels can be all water, either saltwater (in the Indian Ocean) or freshwater (km^2 parcels of Lake Victoria, other named lakes, and some major rivers). On land, we differentiate urban areas, forests, parks, and open parcels available for grazing or farming. The major difference is the impact on vegetation and people.

Given the importance of water in this region of the world, we carefully modeled water sources in the area. The easy part is placing water sources along large rivers and around freshwater lakes. Using the elevation information for the region, we also placed water sources in the lowest areas to represent seasonal water sources and man-made watering holes. Using available data on the number of water sources in some areas, we are able to appropriately populate a given region with water sources.

1.3 Modeling Vegetation Growth

Vegetation growth is very important to local populations through its impact on both agricultural and pastoral activities. We modeled vegetation based on three factors: local rainfall, land fertility, and remotely sensed data on the resulting vegetation over time. Using rainfall data collected frequently over large parts of this area, we developed approximate 50 weather cells covering approximately 30 km^2 regions. For each region and covered parcels, we generate a rainfall amount each day, which is used to determine the vegetation produced on each parcel.

The amount of vegetation produced on a parcel is computed using a logistic equation with rainfall as input. We use Normalized Difference Vegetation Index (NDVI) [13], which provides daily vegetation data based on remote sensing, to validate results. We then determine appropriate land fertility values for each parcel as the residue of a linear regression model of rainfall and NDVI data, based on NDVI as reference vegetation, our vegetation growth model, and rainfall data. These factors together allow the simulation to grow vegetation based on the land's fertility and rainfall for both agriculture and pastoral modeling purposes.

1.4 Modeling Herds

Domesticated animals are modeled as generic tropical livestock units (TLUs), which represent 1.0 head of cattle, 0.7 camels, 11 sheep, 10 goats, and other numbers of other animals [6]. Using this abstract unit, we did not need to differentiate the types of animals in a herd. As a herd of TLUs, we modeled the herd's need for water and vegetation. These physiological parameters allowed us to model the daily intake needs, current levels, and general health of the herd. We also assumed a birth rate and death rate when these physiological needs were not met, resulting in each herd being modeled as a number of TLUs with daily needs and health status as a function of water and vegetation.

3 Household Modeling

The model keeps track of the number of people engaged in different daily subsistence activities, although people are modeled as households who make decisions. The primary activities of households are farming and herding. Family members are also modeled as engaged in wage labor, bringing cash to their household. Household activities are semi-independent subsistence activities. Ideally, each would be self-sufficient, but the model recognizes the activities as part of one household and, if necessary, the people of one activity can live on resources of another. After describing the activities that household members could engage in, we describe how they decide.

3.1 Farming Activity

Only the key parts of farming were modeled. These included planting a specified amount of land and harvesting its crop. Part of the household's population was assigned to farming. This was envisioned to include small children and to occur on the household's farm, one of the parcels suitable for grazing or farming. Those engaged in herding, which could be a relatively independent activity, were expected to be at the household's farm to provide labor from two weeks before planting until after harvest.

Farming involves deciding when to plant, how much land to plant, and how long the growing season will be. Vegetation growth was then modeled daily over a growing season of approximately 90 days. Planting (i.e., starting from zero vegetation), occurs on a specific day and, after the adjustable growing season, the crop is harvested. We included a farming productivity factor that increases the yield on tended farmland over what would grow wild. However, we did not specify the crop type. Vegetation was modeled as producing a yield in units of kilograms of dry matter, for which we had production and consumption data.

The history of rainfall on the farm indicated the best month to plant. Based on the region's weather pattern, there could be one or two best times to plant and the farm was modeled to have one or two growing seasons. A week before the scheduled harvest, the farmer makes an assessment of the crop's potential yield. The farm's next growing season is then extended or shortened based on comparing the potential harvest a week before the actual harvest.

At harvest, the crop's yield is added to the farm's grain store. The farm's store is the source of daily food for farmers and other household members, if needed. After harvest, those assigned to the herding activity take their TLUs and go to open land to graze the herd until next season.

3.2 Herding Activity

Herding is concerned with where to move a herd each day. To make that decision, we have employed a "fast and frugal" decision tree that prioritizes concerns for a need to change watering holes, avoid conflict, and move herd to graze or to water the herd. The decision-making process was described previously [3].

Herders make their daily living directly from animals. A set number of TLUs is necessary to support a herding household. If the number of TLUs in the herd decreases below the number necessary to support the herding household, then its subsistence depends on whether the herd is in the field or at the household's farm. If at the farm, the unsupported herders can subsist from farm store or cash reserves. If in the field, we presume they can find a living from other sources. If the herded TLUs increase beyond what the herding household can manage and the herd has returned to the farm, the animals are sold to increase the household's cash assets.

With the general practice of sending teenage boys with the herd when there is a household base on a farm, the people engaged in herding do not reproduce. However, if the household does not have a farm and the household is therefore moving with the herd, normal reproductive activity takes place.

3.3 Labor Activity

All households in the RiftLand area need cash for some obligations, such as paying taxes and paying for schooling. We address this need by including labor activities when households involve more than ten members. Household members involved in labor activity generate cash for the household and are presumed to not need to subsist through other household activities. Human resources in labor, however, can be re-assigned to support farming or herding activities, if needed. Also, if herding or farming activities are not self-sufficient, the unsupportable individuals will be transferred to labor activities. However, if the fraction of the household involved with labor becomes too great, those persons become internally displaced (so-called "IDPs") and are no longer part of the household.

3.2 Displaced Persons

Displacement is generated as an emergent phenomenon in the model and displaced persons are part of the model's rural household decision-making. The intent, in a later version of the model, is that displaced people move to urban centers for subsistence and if their numbers exceed the capacity of the local urban center, they will move to larger urban centers, possibly also moving across national borders officially becoming refugees.

3.2 Household Decision Making

Each household makes decisions concerning human resources to apply to its subsistence activities. If all goes well, a household would have an operating farm, herd, and some members engaged in labor. On a daily basis when the herd is at the farm, all activities share the resources necessary to meet daily needs and excesses become accumulated wealth. If an activity does not do well enough to feed its assigned people for some period of time, some of the people change to another activity, as shown in Figure 3. The reason for these changes is based on the success of the activity in providing food for the household; changes in human activity amount to transfers of household human resources among activities. Farming activities are evaluated at each planting and harvest, while herding activities are evaluated every four months.

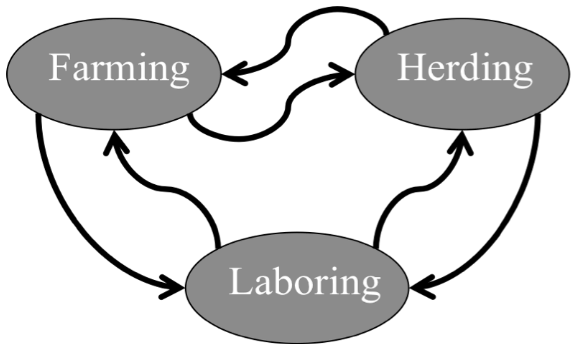

Fig. 3. Transfers of Household Human Resources among Activities. Source: the authors.

If farming or herding activities fail completely, the household will try the failed activity again after some time. On the positive side, successful farming or herding will result in starting a new household with approximately half the original household's assets. Successful households divide their resources creating a new farm and herd as appropriate. Our $1km^2$ parcels can support several co-located farms and herds. However, each parcel has a maximum carrying capacity.

4 Modeling Household Behavior: Preliminary Results

The RiftLand model has been run under a variety of conditions to demonstrate how households make a living throughout the region. The expectation is that environmental factors will cause different subsistence strategies in different regions. We have focused on three different regions to validate our model of household decision-making in the RiftLand area, since overall variability is exponentially (not just linearly) proportional to the number of individually validated regions: one valid region is good, two are better, three are far better. Three regions studied are shown in Figure 1 as white squares. In each of the three study areas, our simulations involved placing 1,000 households in the region and simulating their behavior over 3,650 days (i.e., 10 years). Households are initialized with a number of persons divided nearly evenly between farming and herding activities. If the household has greater than 10 persons, then approximately 10 percent are initialized as being engaged in labor to generate cash for paying taxes, educational, or other needs. Households start with both farming and herding and subsequently adjust human resources applied to each activity based on their success in feeding the household. Preliminary results are presented in the next three subsections.

4.1 Survival in Semi-Arid Regions of Northwest Kenya (Region 1)

The first area tested is in the relatively dry northwest region of the modeled area. We focused on an area 150 km by 150 km in northwest Kenya indicated as a white square numbered 1 in Figure 1 (right). This area, near the Bokhol Plain (500–1000 m elevation) in Wajir District is close to the Mandera region studied in previous work [2] and [4]. The area lacks major population centers (closest is Wajir, ca. 30,000 pop., 100 km away), water supplies, forests, or parklands. This is an arid, open area with low population density (< 10 inhab/km^2) inhabited primarily by Somali people. Figure 4 is a plot of the simulation's results. This plot presents the mix of subsistence activities as the average of 1,000 households and shows those results for 30 runs.

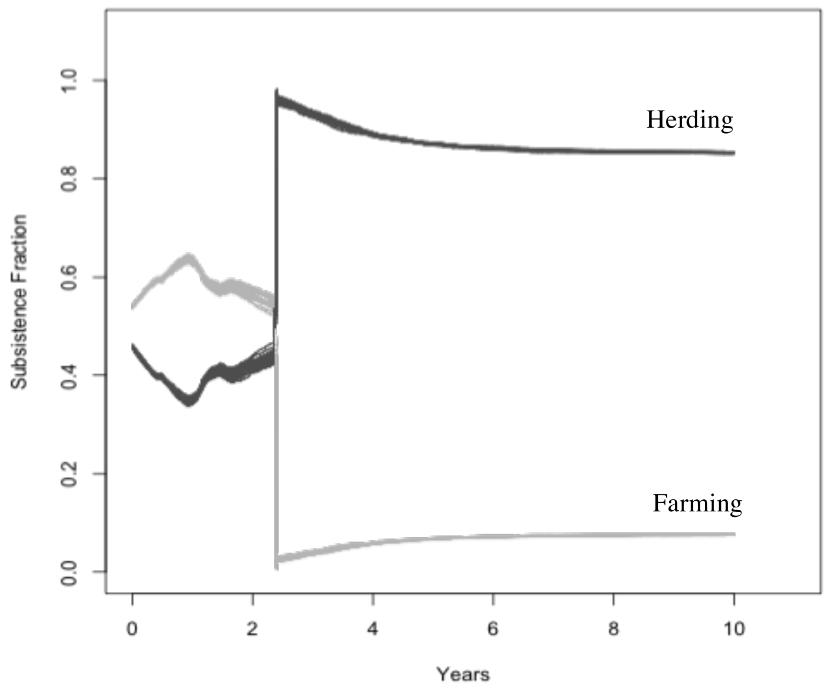

Fig. 4. Results of Household Decision-Making in Region South of Mandera.

Figure 4 shows average results of 1,000 households deciding how to make a living over 10 years. After the first several unsuccessful harvests, occupations changed dramatically toward favoring herding over farming. In the first three years, decisions are reconsidered and the strategy mix changed. The major changes shown occur at harvest times, but after three years, household farming activities are diverse enough that actions at harvest are no longer visible for the whole society. This region settles at an 80/20 preference for herding over farming, with only residual labor activity.

The reference anthropological data for this region is contained in the *Ethnographic Atlas* [11] and [12] updated by HRAF. With the modern-day updates, the people of this region, the Bararetta [5], are described as "PA," meaning they are both pastoralists (herders) and agriculturalists (farmers), consistent with our simulation results.

4.2 Survival in a Relatively Wet Area, Southwest Uganda (Region 2)

The second area tested is in the relatively wet region west of Lake Victoria. We focused on an area 100 km by 100 km in southwestern Uganda as indicated in Figure 1 (right). The area size was selected to be primarily rural and to avoid major population centers, forests, or parklands. This region consists of open areas, available for farming and grazing family herds, with relatively high population density (100–300 inhab./km^2). Figure 5 is a plot of the simulation's resulting population in this region for 30 runs.

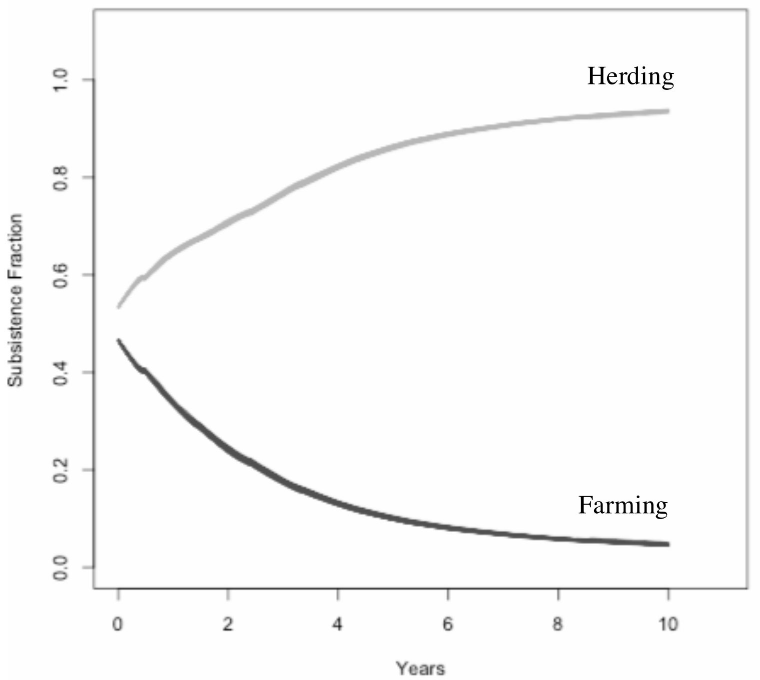

Fig. 5. Results of Household Decision-Making in a Southwest (Ganda) Region of Uganda.

Simulation results show that the households took a longer time to settle on their mix of subsistence activities. During each of the first several years' harvests, there were dramatic shifts in activities before a few years of consistent trends settled on approximately an 80/15/5 percent division between farming, herding, and labor.

Murdock's 1967 atlas reported the Nyoro people [14] in this region are coded as "Ap", primarily agriculturalists with some pastoralist activity, specifically 56-65 percent subsisting on agriculture and 16-25 percent subsistence on animal husbandry.

4.3 Survival in Burundi (Region 3)

The third area tested was less arid (500–1000 mm annual precipitation), in Burundi, southwest of Lake Victoria. We again focused on an area 100km by 100km to be primarily rural and to avoid major population centers, forests, and parklands. This region of our model consists of some open areas available for farming and grazing family herds. Figure 6 is a plot of the simulation's resulting survival strategies in this region for 30 runs.

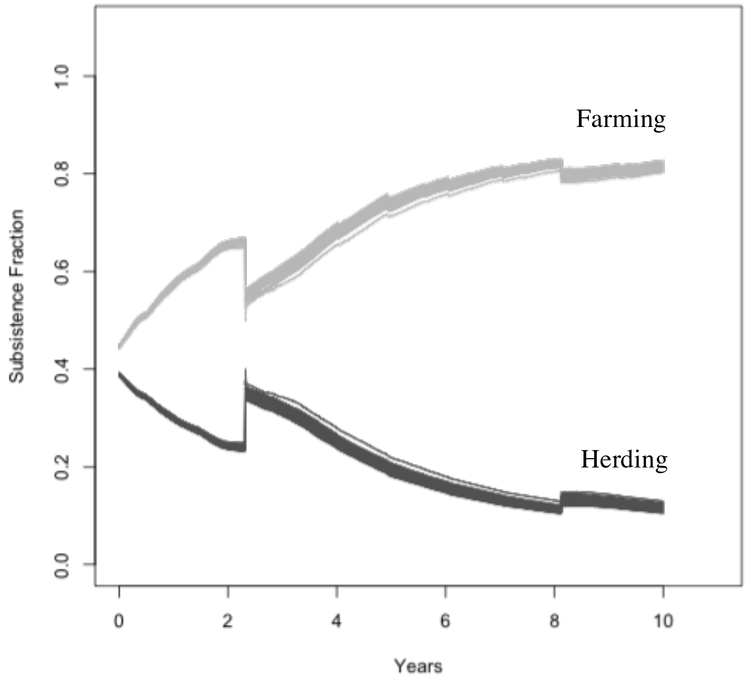

Fig. 6. Results of Household Decision-Making in Burundi.

In this region, it was very clear after the first harvest and definitely clear after the second harvest that herding was not a viable strategy for making a living. However, farming was very good and produced households large enough to support some labor activities. The resulting distribution was approximately 95/0/5 for farming, herding, and labor. The anthropological data for the Moru people of this region is coded as "Ap", but with largest subsistence fraction being 36-45 percent in agriculture and the second largest fraction being 16-25 percent in herding.

5 Progress in RiftLand Model Validation

Our approach to validating this agent-based model is to start with micro-level models from which emerge societal-level behaviors that match available empirical (in this case ethnographic) data. Our models of subsistence activities employed in East Africa are based on available data, weather, TLUs, vegetation, and other socio-natural features. Household agents adjust their human resources based on the performance of their subsistence activities. Based on these inputs and dynamics, the model generates distributions of farming, herding, and labor activities. Available anthropological data coded subsistence activities in five categories and we are attempting to have our model match the top two activities (farming and herding). So far, our model has matched the primary subsistence activity in each region—either as farming or herding—and does not yet match the approximate percentages reported for the three regions. However, data in the atlas are recorded assessments of one or more anthropologists based on local observations at some time in history. Some of these observations are over 100 years old. As such, these assessments can vary from current, "on-the-ground" truth by quiet a bit. For our purposes, the comparison of model results to anthropological data is primarily of ordinal value, not interval or ratio value.

6 Conclusions

We modeled the weather, parcel type, vegetation, and household subsistence activities in a 1,000 by 1,000 mile area around Lake Victoria in East Africa at the one square kilometer, one day, and household level. Results from the MASON RiftLand agent-based model match the types of primary and secondary subsistence activities, farming and herding, as reported in the anthropological data for the three sample regions. However, we do not yet match the exact percentages of those activities. We are continuing our modeling efforts, conducting and preparing additional model runs, and we will expand our model comparisons to all of the 135 cultures represented in the study area and catalogued by anthropologists. This initial "modal" or "ordinal" validation in behavioral patterns is a necessary step in the direction of more extensive validation via additional tests based on more stringent interval and ratio standards.

Acknowledgments. This work was supported by the Center for Social Complexity at George Mason University and by the Office of Naval Research (ONR) under a Multidisciplinary University Research Initiative (MURI) grant no. N00014-08-1-0921. The authors thank members of the Mason-HRAF Joint Project on Eastern

Africa (MURI Team), especially Jeffrey K. Bassett, Atesmachew B. Hailegiorgis, Joseph Harrison, and Eric Scott for comments and discussions. Carol Ember, Ian Skoggard, and Teferi Abate of HRAF provided assistance with anthropological data. The opinions, findings, and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the sponsors.

References

1. Cioffi-Revilla, C.: MASON RebeLand and Data Aspects of Agent-Based Simulation Models. In: Schmorrow, D. and Nicholson, D. (eds.), Advances in Cross-Cultural Decision Making. pp. 291-301. CRC Press/Taylor and Francis, Orlando (2010)
2. Hailegiorgis, A.B., Kennedy, W.G., Catalin Balan, G., Bassett, J.K., Gulden, T.: An Agent Based Model of Climate Change and Conflict among Pastoralists in East Africa. In: International Congress on Environmental Modeling and Software. Ottawa (2010)
3. Kennedy, W.G., Bassett, J.K.: Implementing a "Fast and Frugal" Cognitive Model within a Computational Social Simulation. In: Second Annual Conference of the Computational Social Science Society of the Americas. Santa Fe (2011)
4. Kennedy, W.G., Gulden, T., Hailegiorgis, A.B., Bassett, J.K., Coletti, M., Catalin Balan, G., Clark, M., Cioffi-Revilla, C.: An Agent-Based Model of Conflict in East Africa and the Effect of the Privatization of Land. In: Third World Congress on Social Simulation. Kassel (2010)
5. Lewis, I.M.: Peoples of the Horn of Africa. London: International African Institute. (1955)
6. Little, P.D., McPeak, J., Barrett, C.B., Kristjanson, P.: Challenging Orthodoxies: Understanding Poverty in Pastoral Areas of East Africa. Dev. Change, 39(4). pp. 587—611 (2008)
7. McCabe, J.T.: Cattle Bring Us to Our Enemies. University of Michigan Press, Ann Arbor (2004)
8. Cioffi-Revilla, C., et al.: MASON RiftLand: An Agent-Based Model for Analyzing Conflicts, Disasters, and Humanitarian Crises in East Africa. Working Paper. Mason-Yale Joint Project on Eastern Africa. Center for Social Complexity, George Mason University, Fairfax, Virginia 22030 (2012)
9. Luke S., Cioffi-Revilla, C., Sullivan, K., Catalin Balan. G.: MASON: A Multiagent Simulation Environment. Simulation 81(7). pp. 517—527 (2005)
10. Luke, S.: Multiagent Simulation and the MASON Library http://cs.gmu.edu/~eclab/projects/mason/
11. Murdock, G.P.: Ethnographic Atlas: a summary (Codes). Int J Cultural Soc. Anthro. 4. pp. 154—169 (1967)
12. Murdock, G.P.: Ethnographic Atlas. Ethnology 5(3). pp. 317–345 (1966)
13. Vrieling, A., de Beurs, K.M., Brown, M.E.: Variability of African farming systems from phenological analysis of NDVI time series. Climatic Change, 109(3-4). pp. 455--477 (2011)
14. Taylor, B.K.: The West Lacustrine Bantu. International African Institute: London (1962)