Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung Anwendungen 1

Christian Hüning, christian.huening@haw-hamburg.de

Socio-ecological modelling with Multi-Agent-Systems on Landscape-level

Fakultät Technik und Informatik
Studiendepartment Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Christian Hüning, christian.huening@haw-hamburg.de

Socio-ecological modelling with Multi-Agent-Systems on Landscape-level

Ausarbeitung Anwendungen 1 eingereicht im Rahmen der Ausarbeitung Anwendungen 1

im Studiengang Bachelor of Science Angewandte Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Thiel-Clemen

Eingereicht am: 20. September 2013

Christian Hüning, christian.huening@haw-hamburg.de

Thema der Arbeit

Socio-ecological modelling with Multi-Agent-Systems on Landscape-level

Stichworte

Multiagentensystem, Ökologie, Verteilte Systeme, Systemmodellierung, Lastverteilung

Kurzzusammenfassung

Christian Hüning, christian.huening@haw-hamburg.de

Title of the paper

Socio-ecological modelling with Multi-Agent-System on Landscape-level

Keywords

Multi-Agent-System, Ecology, Distributed Systems, Systemmodelling, Load Balancing

Abstract

Contents

1 Introduction **1**

2 Historical Overview & Background **2**

2.1 First Steps 2
2.2 Ten years of ecological modelling - A review 2
2.3 Integrating Models 4
2.4 Summary 5

3 Overview of MAS Frameworks **6**

3.1 Simulation Frameworks 6
3.1.1 JADE 6
3.1.2 GAMA 6
3.1.3 WALK 7
3.1.4 Vigueras 7
3.2 Case Specific Implementations 8
3.2.1 LUDAS 8
3.2.2 MASE 8

4 Area of research & Outlook **9**

1 Introduction

The usage of Individual Based Modelling (IBM) or Agent Based Modelling (ABM) in ecological science is well accepted for 25 years now. However a lot of experience on when and how to use IBM has been collected over time (Filatova *et al.*, 2013; Grimm, 1999; Huston *et al.*, 1988a) as well as new approaches, methods and technologies in computer science have emerged (Bellifemine *et al.*, 2008, 2007; Grimm *et al.*, 2006, 2010; Le *et al.*, 2008; Ralha *et al.*, 2013; Vigueras *et al.*, 2013).

Together with these improvements, new challenges and problems have arisen in the various domains of ecological science. Some of the major challenges are the integration of different models and almost arbitrary data into combined simulation models (Thiel-Clemen, 2013), execution performance of simulations and the need to significantly increase the scale of models, while at the same time be able to visualize the simulation. While the latter challenge is partly caused by globalization and the upcoming comprehension, that we must understand ecological systems as being coupled with their neighboring systems, the first two directly emerge from that requirement.

The coupling of ecological, social, economical and political systems as mentioned above, creates a huge complexity to the overall model and simulation. Since the creation and usage of multi-agent based simulation systems has proven to be a great tool to explore and investigate such models (Le *et al.*, 2008; Ralha *et al.*, 2013), the relevance to computer science arises from the massively increased requirements regarding execution performance, distributed computing, handling of vast amounts of simulation data and validation combined with a decent analysis of the simulated results.

The MARS research group at Hamburg University of Applied Sciences attempts to create a next-gen simulation system which satisfies these requirements. This paper shall give an overview of the history and background of ecological modeling (Chapter 2), provide a little evaluation of today's state of the art simulation frameworks and systems (Chapter 3) and finally define my own area of research and provide an outlook (Chapter 4).

2 Historical Overview & Background

2.1 First Steps

For more than 25 years, ecologists and social scientists among others are busy researching, how the usage of multi-agent-systems may help them to understand their areas of research more deeply and completely.

Although occasionally explored by others before, Huston *et al.* (1988b) where among the first to make a profound statement for the benefits of Individual Based Modeling (IBM) in ecological science. They point out essential rules of biology that were violated by models used so far, with the two major rules being, that every individual is different and that actions between two entities are inherently local.

They further find that those phenomenons cannot be depicted correctly by state-variable models and thus make their point towards IBMs.

2.2 Ten years of ecological modelling - A review

Eleven years later Grimm (1999) reviews the evolution of IBM since Huston *et al.* (1988b) stated their usefulness and develops "heuristic rules of individual-bases modeling". Grimm (1999) shifts the question from if and why someone should use IBM to how one should develop IBM based simulations. His rules are derived from a review of ecological models created and published between 1988 an 1999 and thus aim to provide a general approach for everyone who wants to create such a model of his own.

His first rule sounds as simple as it is important: "Individual-based modelling is modelling". He wants to make sure, that albeit IBM promises emergent effects from a group of more or less independent agents, this effect does never happen without proper modelling, as this activity means to really understand the problem at hand.

Another rule says that one should "Change the level of aggregation". This refers to the need of using the right level of detail when modelling reality. A too general level could ignore important facts in an underlying level, while a too detailed level could loose focus of the pattern one wants to model. Thinking about these matters, leads to the question whether a scale-down or scale-up approach is better suited when designing a new model. Grimm (1999) states that starting with a detailed model and then moving up to a more general one, would produce some quite interesting insights on the different levels of aggregation, though no one would really want to abandon a working detailed model, which is why this scaling-up approach is not really used too often.

Instead the scale-down approach is far more useful as it allows to recreate a general pattern within the simulation and then add detail to it by going down while continuously assuring that the overall pattern is still valid. This also allows to leave each aspect of the model at a different, but suitable level of aggregation, while going down the hierarchy.

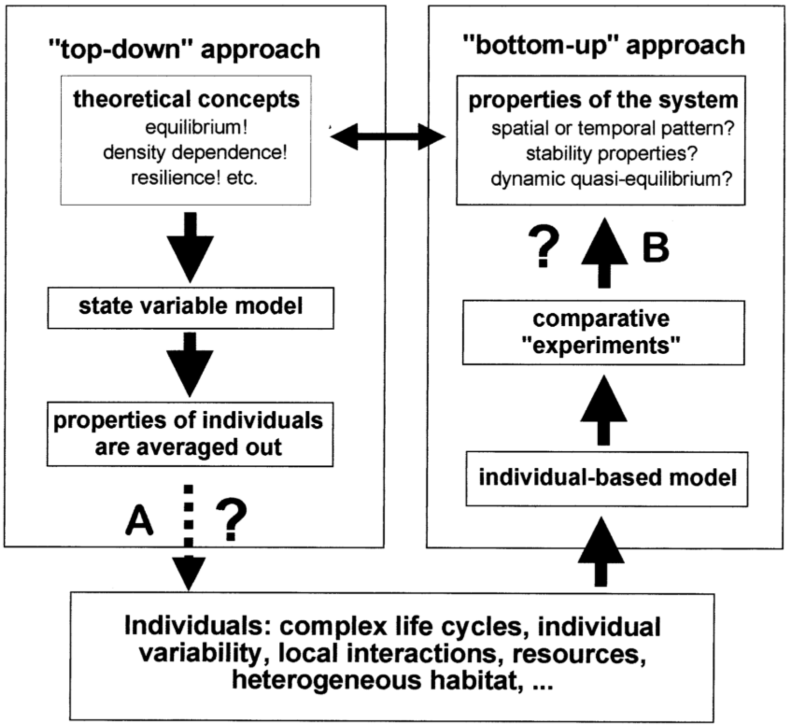

Figure 2.1: Mutual relationship of top-down and bottom-up approaches in ecological modelling. Source: (Grimm, 1999)

Comparing IBM with state-variable modelling means comparing a bottom-up approach with a top-down one. Figure 2.1 shows the mutual relationship between these two complementary approaches as Grimm (1999) sees them. They are complementary because each on its own does not lead to sufficient results. In the top-down approach you make a general assumption on the top level that then will be applied to each individual entity you find on the bottom end. This clearly might lead to wrong results in a multitude of cases. The bottom-up approach on the other hand, emerges from individual models for each different entity, but might run into trouble, if it is not clear what you want to know at the top end. So to avoid both problems, it turns out to be a good idea to combine the two approaches, by validating the results from bottom-up IBM against the general patterns of top-down state-variable models.

2.3 Integrating Models

As more and more ecological models were created and programmed over the years, more and more paradigms and ways of implementation of these models emerged. With that another interesting aspect of IBM came along, the integration of different models. The idea is simple. Connect and integrate domain specific models from domain specific experts to create a new super model of a certain domain. If for example you would want to create a large scale model of the ecosystem of a national park in south Africa, it would be very helpful, if you could use existing models of elephants, cheetahs etc.. Actually doing that, turns out to be much more difficult, since every group of scientists working on a model uses another, individual paradigm, architecture, programming language, data format and so forth. Villa (2001) proposes his Integrating Modelling Architecture (IMA) for hat purpose. He singles out three characterizing dimensions for connecting different models:

Representation A unified semantic relating to the depiction of space, time and behavior in every respective model is needed.

Domain A clear distinction between the domain spaces of each sub-model must be made. In particular this relates to the input and output parameters which are valid for each sub-model.

Scale Data, which is exchanged between models, must be compatible or translated in space and time dimensions.

A recent contribution to the Scale dimension has been made by Thiel-Clemen (2013), who proposes a data warehouse based information integration process on the simulation data.

These dimensions target the difficulty when technically connecting different models. A more functional view has been made by Liu *et al.* (2007) who take a look at the complexity of coupled human and natural systems. Their integration efforts aim at taking interdisciplinary research on a broader scale into account, as well as exceeding local and temporal boundaries when modelling certain ecological system. As shown by their findings, almost every ecosystem today is tightly coupled with its neighboring economic or social systems and thus these need to be taken into account when watching the evolution of that ecosystem. Filatova *et al.* (2013) move even further and demand that the corresponding aspects of ecological systems like economy, social systems and bio-physical dynamics need to be integrated into the representation of a heterogeneous landscape representation.

2.4 Summary

The discussion today circles around the fields of model re-usage (Holst (2013)), model integration (Filatova *et al.* (2013), Le *et al.* (2008), Liu *et al.* (2007), Villa (2001)), which makes distributed, parallel simulation execution (Cicirelli *et al.* (2010), Wang *et al.* (2009), Wang *et al.* (2012), Bellifemine *et al.* (2007), Thiel (2013), Vigueras *et al.* (2013)) necessary and the question of spatial-temporal information integration (Thiel-Clemen (2013), Filatova *et al.* (2013)) is raised.

Since the above mentioned ideas produce a lot of computing complexity, the need for appropriate simulation tools and frameworks arises. Over the past years there have been quite a lot approaches to this field, which will be further examined in detail in the next chapter.

3 Overview of MAS Frameworks

A huge number of MAS frameworks and domain specific implementations have been created over the past years. Since I strive to create yet another framework, it makes perfect sense to look at the previous work and evaluate their capabilities and usefulness.

3.1 Simulation Frameworks

3.1.1 JADE

One of the most famous frameworks is JADE (Bellifemine *et al.*, 2007) which allows to execute a simulation distributed across several JADE container processes or just locally in a single container. JADE was developed in Java to create a reference implementation of the FIPA agent specification (http://www.fipa.org). The performance of JADE has been extensively investigated by Mengistu *et al.* (2008). Their findings show that JADE has significant performance issues in the fields of communication and agent migration due to the usage of the LDAP protocol and slow message transport services. JADE's Lookup-Directory-Service also is measured to be slow, which is caused by not using local caching on the respective nodes. Mengistu *et al.* (2008) propose improvements to both mechanisms and present promising results from experiments they conducted. However a more recent investigation of JADE's performance seems appropriate, given that the paper is almost 6 years old.

3.1.2 GAMA

GAMA (Amouroux *et al.*, 2007) is a modeling and simulation framework which is based on RepastJ. It features a nice model description language, called GAML, which allows non-programmers to create complex models. GAMA is written in JAVA and thus executable on all java enabled systems. A very strong feature of GAMA is its visualization feature, especially when it comes to using GIS data. An easy import function allows to quickly create a scenario's environment and visualization from a GIS file and thus allows for a quick integration of that kind of data.

The downside of GAMA is, that it's not possible to distribute the system and that it does not scale well across multiple CPU cores. In fact when testing GAMA, it actually used only just up to 4 cores while running on a 24 core machine. While testing I found GAMA to have a perfomance threshold around 80.000 agents, with one simulation step taking more than 800ms on the aforementioned machine.

3.1.3 WALK

Also from 2013 comes a solution with a strong focus on evacuation scenarios which has been developed here at the Hamburg University of Applied Sciences and is called WALK (Thiel, 2013). It features a dynamic (re)partitioning and distribution of agents across several compute nodes and is thus capable of running simulations with hundreds of thousands agents on commodity hardware. In fact Thiel (2013) showed in his final tests that WALK can run a 300.000 agent random walk simulation in near real time. Also remarkable about WALK is, that its agents pass the RiMEA tests and thus provide a pretty good behavior. As a recent addition Stefan Münchow added support for leadership models and social behavior to the agents implemented in WALK. These additions show very promising results and create a very high interest in re-using the agent implementation from WALK in the new system whenever human agents are explored.

3.1.4 Vigueras

Another interesting architecture (Vigueras *et al.*, 2013) proposes an almost completely asynchronous, distributed simulation execution to implement interactive simulations, that may be visualized in near real-time. The only time Vigueras *et al.* (2013) synchronize the execution of their agents is, when they happen to act or move beyond the boundaries of their respective environment patch.

When it comes to visualization of the simulation Vigueras *et al.* (2013) utilize visualization nodes (VS) that also act asynchronously on the distributed nodes. Each VS has a camera-style definition of its field of view and may thus only ask those nodes for information containing parts of the environment, which is in that field of view. This is very contrary to other visualization approaches (e.g. GAMA, NetLogo), since it does not attempt to visualize the whole simulation at once.
Considering the amount of agents and the sheer sice of simulated space in our upcoming scenarios, this approach might become very valuable.

3.2 Case Specific Implementations

3.2.1 LUDAS

LUDAS (Land-Use Dynamic Simulator) (Le *et al.*, 2008) implements a social-ecological, land-use/cover change (LUCC) model featuring four components, which implement human population including behavior, the environment, various policy factors with focus on land-use choices and lastly a decision making procedure which integrates the first three features. The model simulates "a watershed in Vietnam for integrated assessments of policy impacts on landscape and community dynamics". The implementation has been done in NetLogo and thus does not provide a very high performance, but showcases the scenario pretty nice.

It is not performance nor distribution which makes LUDAS interesting, but the great integration of LUCC components into a working simulation scenario. If that model can be translated into a larger, more capable software architecture, it could provide some very decent results in future, larger scale LUCC simulations.

3.2.2 MASE

MASE (Ralha *et al.*, 2013) is another LUCC simulation which targets the development of robust land-use strategies. The showcase features a region called Cerrado in Brazil. Whats remarkable about MASE is, that it utilizes a methodical, empirical parameterization process for human behavior, which has been developed by Smajgl *et al.* (2011). The implementation has been done with JADE (Bellifemine *et al.*, 2007) and Matlab.

4 Area of research & Outlook

In my bachelor thesis I designed an architecture by the name of RUN, which prototypical implemented a couple of possible solutions to the challenges mentioned earlier in this paper.

Integration and re-usability of models are taken care of by a layer-oriented approach, which results in treating simulation layers as plugins on an architectural level. Each layer realizes an aspect of the whole model and may be constructed and implemented in whatever way the modeler finds suitable as long as some lightweight requirements are met. A LayerContainer (LC) software has been written to take hold of layers and provide the needed execution environment.

Parallel execution of the simulation is treated in a new way by RUN, as it is runs synchronously, but without a central clock. This is achieved by an Erlang-based implementation of the SAPD algorithm (Synchronization Algorithm of Parametric Difference) presented by Fan *et al.* (2013). After a short calibration phase, the algorithm enables each LC to execute its layers on its own, without having to report back to a central instance as it was necessary in WALK (Thiel, 2013).

However the central challenge of my work will deal with the architecture of MARS and how it has to be designed to fulfill the requirements of scaling out in an efficient, effective and desirably transparent way when more hardware is added, while at the same time produce valid results (Mengistu *et al.*, 2008). The aforementioned simulation systems (3.1) usually do scale up to a certain (rather low) extent (3.1.2) but do not scale out or at least not too well like JADE (Mengistu *et al.*, 2008).

The creation of the MARS SYSTEM is and will be an extremely collaborative task within the MARS GROUP (http://www.mars-group.org). As such my next milestone for Project 1 will be the creation of a showcase for the Abdoulaye Model by Hodabalo Pereki (Pereki *et al.*, 2013), which is scheduled to be finished next summer. The showcase shall feature a vertical slice through the whole MARS System and thereby provide valuable insights into the practicability of features and solutions depicted above. Hereby the Abdoulaye Model acts as my fundamental source of requirements, since each of the challenges MARS seeks to solve are present within Abdoulaye's scenarios.

Two of the major indicators in the Abdoulaye Model are the biomass and amount of carbon held in the forest at any given time. According to Hodabalo Pereki a reduction in size or further simplification of the Abdoulaye Model is not possible, since that would distort the results being computed. Furthermore in his model every tree is an agent and the landscape has to be captured as a large grid of 20m by 50m patches on the lowest level of aggregation. Given the size of Abdoulaye National Park and an estimation on the number of trees, this results in 4.000.000 tree agents and 1.700.000 environment patches which have to be simulated at once. Looking back at chapter 3 it becomes clear, that no present simulation framework or system can cope with these numbers in an efficient way. Especially not, because Hodabalo Pereki stated that a visualisation, which ideally runs in near real-time, is most desirable. So high performance and very good scalability are key issues to meet the requirements of the Abdoulaye Model. Therefore I will work on the following questions:

- How can the layer-oriented approach for the development of integrated simulation models be designed, so that it allows for a transparent distribution of each layer across multiple compute nodes?
- What are suitable metrics and algorithms to partition agents across multiple compute nodes?
- How can an efficient and effective communication solution for inter-layer, intra-layer and controlling communication be designed and implemented?
- How can the overall MARS architecture be integrated, e.g. what are suitable connectors?
- How does one use MARS to create, execute and evaluate a given domain model with MARS?

My AW2 Paper will describe and analyze the results and findings from the showcase and will derive the next steps of research and development for MARS, while taking new publications in my field of research into account. Since the paper from Mengistu *et al.* (2008) is now almost 6 years old, I will probably have to examine the recent version of JADE towards its scalability capabilities.

Bibliography

Amouroux, Edouard, Thanh-quang, C H U, Boucher, Alain, & Drogoul, Alexis. 2007. GAMA : an environment for implementing and running spatially explicit multi-agent simulations.

Bellifemine, Fabio, Caire, Giovanni, Poggi, Agostino, & Rimassa, Giovanni. 2008. JADE: A software framework for developing multi-agent applications. Lessons learned. *Information and Software Technology*, **50**(1-2), 10–21.

Bellifemine, FL, Caire, G, & Greenwood, D. 2007. *Developing multi-agent systems with JADE.*

Cicirelli, Franco, Furfaro, Angelo, Giordano, Andrea, & Nigro, Libero. 2010. Parallel Simulation of Multi-agent Systems Using Terracotta. *2010 IEEE/ACM 14th International Symposium on Distributed Simulation and Real Time Applications*, Oct., 219–222.

Fan, Linjun, Ling, Yunxiang, Wang, Tao, Zhu, Xiaomin, & Tang, Xiaoyong. 2013. Novel clock synchronization algorithm of parametric difference for parallel and distributed simulations. *Computer Networks*, **57**(6), 1474–1487.

Filatova, Tatiana, Verburg, Peter H., Parker, Dawn Cassandra, & Stannard, Carol Ann. 2013. Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental Modelling & Software*, Apr., 1–7.

Grimm, Volker. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling*, **115**(2-3), 129–148.

Grimm, Volker, Berger, Uta, Bastiansen, Finn, Eliassen, Sigrunn, Ginot, Vincent, Giske, Jarl, Goss-Custard, John, Grand, Tamara, Heinz, Simone K., Huse, Geir, Huth, Andreas, Jepsen, Jane U., Jø rgensen, Christian, Mooij, Wolf M., Müller, Birgit, Pe'er, Guy, Piou, Cyril, Railsback, Steven F., Robbins, Andrew M., Robbins, Martha M., Rossmanith, Eva, Rüger, Nadja, Strand, Espen, Souissi, Sami, Stillman, Richard a., Vabø, Rune, Visser, Ute, & DeAngelis, Donald L. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, **198**(1-2), 115–126.

Grimm, Volker, Berger, Uta, DeAngelis, Donald L., Polhill, J. Gary, Giske, Jarl, & Railsback, Steven F. 2010. The ODD protocol: A review and first update. *Ecological Modelling*, **221**(23), 2760–2768.

Holst, Niels. 2013. A universal simulator for ecological models. *Ecological Informatics*, **13**(Jan.), 70–76.

Huston, M, DeAngelis, D, & Post, W. 1988a. New computer models unify ecological theory. *BioScience*, **38**, 682–691.

Huston, M, DeAngelis, D, & Post, W. 1988b. New computer models unify ecological theory. *BioScience*, **38**, 682–691.

Le, Quang Bao, Park, Soo Jin, Vlek, Paul L G, & Cremers, Armin B. 2008. Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system. I. Structure and theoretical specification. *Ecological Informatics*, **3**(2), 135–153.

Liu, Jianguo, Dietz, Thomas, Carpenter, Stephen R, Alberti, Marina, Folke, Carl, Moran, Emilio, Pell, Alice N, Deadman, Peter, Kratz, Timothy, Lubchenco, Jane, Ostrom, Elinor, Ouyang, Zhiyun, Provencher, William, Redman, Charles L, Schneider, Stephen H, & Taylor, William W. 2007. Complexity of coupled human and natural systems. *Science (New York, N.Y.)*, **317**(5844), 1513–6.

Mengistu, Dawit, Tröger, Peter, Lundberg, Lars, & Davidsson, Paul. 2008. Scalability in Distributed Multi-Agent Based Simulations: The JADE Case. *2008 Second International Conference on Future Generation Communication and Networking Symposia*, Dec., 93–99.

Pereki, Hodabalo, Wala, Kperkouma, & Thiel-Clemen, T. 2013. Woody species diversity and important value indices in dense dry forests in Abdoulaye Wildlife Reserve (Togo, West Africa). *International Journal*, **5**(June), 358–366.

Ralha, Célia G., Abreu, Carolina G., Coelho, Cássio G.C., Zaghetto, Alexandre, Macchiavello, Bruno, & Machado, Ricardo B. 2013. A multi-agent model system for land-use change simulation. *Environmental Modelling & Software*, **42**(Apr.), 30–46.

Smajgl, Alex, Brown, Daniel G, Valbuena, Diego, & Huigen, Marco G A. 2011. Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling Software*, **26**, 837–844.

Thiel, Christian. 2013. *Analyse von Partitionierungen und partieller Synchronisation in stark verteilten multiagentenbasierten Fußgängersimulationen.* Master Thesis, Hamburg University of Applied Sciences.

Thiel-Clemen, Th. 2013. Information Integration in Ecological Informatics and Modelling. *Pages 89 – 96 of:* Wittmann, J., & Müller, M. (eds), *Simulation in Umwelt- und Geowissenschaften, Workshop Leipzig.* Leipzig: Shaker.

Vigueras, Guillermo, Orduña, Juan M., Lozano, Miguel, & Jégou, Yvon. 2013. A scalable multi-agent system architecture for interactive applications. *Science of Computer Programming*, **78**(6), 715–724.

Villa, Ferdinando. 2001. Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. *Ecological Modelling*, **137**(1), 23–42.

Wang, B, Yao, Y, & Himmelspach, Jan. 2009. Experimental analysis of logical process simulation algorithms in JAMES II. *Pages 1167–1179 of: Proceedings of the Winter 2009 Simulation Conference.*

Wang, Y, Lees, M, & Cai, W. 2012. Grid-based partitioning for large-scale distributed agent-based crowd simulation. *Proceedings of the Winter 2012 Simulation Conference.*