Toward complex models of complex systems

One step further in the art of Agent-Based Modelling

Habilitation à Diriger des Recherches
defended on 18th November, 2016
at the University Toulouse 1 Capitole

by

Benoit Gaudou

Jury:

François Bousquet, Research director, CIRAD - Reviewer
Guillaume Deffuant, Research director, IRSTEA - Reviewer
Bruce Edmonds, Professor, Manchester Metropolitan University Business School - Reviewer
Alexis Drogoul, Research director, IRD - Member
Amal El Fallah Seghrouchni, Professor, University Pierre et Marie Curie - Member
Frédérick Garcia, Research director, INRA - Member
Chihab Hanachi, Professor, University Toulouse 1 Capitole - Member
Laurent Vercouter, Professor, INSA Rouen - Member

En mémoire de Pierre Simon Iwao

Remerciements

Je souhaiterais tout d'abord exprimer mes profonds remerciements à François Bousquet, Guillaume Deffuant et Bruce Edmonds pour avoir accepté de lire et de rapporter ce travail et d'avoir pu se libérer pour assister à la soutenance. Leurs qualités professionnelles et humaines et les échanges lors de la soutenance et en dehors ont rendu cette journée incroyablement enrichissante.
Je remercie tout particulièrement Amal El Fallah Seghrouchni et Laurent Vercouter pour leur participation au jury. Il était très important pour moi qu'ils soient impliqués dans le passage de cette étape qu'est l'HDR car ils ont été présents à des moments importants de mon parcours et que nos chemins n'ont cessé de se croiser depuis. J'espère que ce sera également le cas avec Frédérick Garcia que je remercie tout particulièrement d'avoir accepté d'être membre du jury et d'avoir assisté à la soutenance.
Enfin un grand merci à Alexis Drogoul et Chihab Hanachi pour avoir tenu le rôle de directeurs d'HDR en m'accompagnant avec une très grande disponibilité tout au long de ce long et difficile processus, depuis la rédaction du manuscrit jusqu'à sa soutenance. Un grand merci également à Carole Adam pour les corrections du manuscrit.
Depuis ma thèse, mon parcours dans l'enseignement et la recherche s'apparente pour moi à un grand voyage, à la fois géographique, thématique et disciplinaire. Qui dit voyage, dit naturellement rencontres. J'adresse un immense merci à toutes celles et tous ceux avec que j'ai eu la chance de côtoyer, à tous ceux qui m'ont montré la voie, ceux qui ont accepté de suivre la mienne et à tous ceux qui ont voyagé et voyage encore à mes côtés. Je me sens extrêmement chanceux de toutes les personnes avec qui j'ai pu travailler ou simplement échanger. Je ne m'aventurerais pas à toutes et tous les nommer, mais j'ai une immense gratitude pour chacune et chacun d'entre eux. Une pensée toute particulière pour tous les collègues de mon port d'attache toulousain ainsi qu'aux communautés GAMA et MAPS qui jouent le rôle de phares et parfois de bouées dans cette aventure.
Pour finir, les mots manquent pour exprimer tous mes remerciements pour ma famille et mes amis pour me soutenir inconditionnellement. Et enfin à celle qui a accepté de m'accompagner et de me suivre dans les différentes étapes de ce voyage et dans les suivantes...

Abstract

Agent-Based Modelling has now become a relevant and recognized paradigm to design integrated models of complex systems such as socio-environmental systems. The MAELIA modelling project, that aims at assessing various water withdrawal policies, is a typical example of such complex models: it couples physical dynamics (*e.g.* water flow and plant growth) with agricultural activities (*e.g.* cropping plan decision-making) to provide a Decision-Support System about water management policies. Working on such a complex model has highlighted the limits of tools and methods currently used in modelling projects. This dissertation aims at investigating more particularly three research axes that appear necessary to improve the way we design and use agent-based models.

First, the dissertation focuses on the **integration of complex and cognitive agents in agent-based models**. Agent-Based Models are usually designed with very simple agents and these models are generally abstract and focused on a specific process (*e.g.* opinion diffusion). But it has appeared necessary to integrate, in socio-environmental system models, agents able to make complex decisions (such as cropping plan decision by farmer agents in the MAELIA model) and to reason about others in large-scale artificial societies. To this purpose, a BDI architecture coupled with a multi-criteria decision-making process has been proposed and integrated in the GAMA platform. In addition, models of agents with complex social cognitive capabilities (*e.g.* trust and social emotions) are presented.

The second research axis deals with **the integration of models using different paradigms into an agent-based model**, and more specifically with the coupling of Agent-Based Models with Ordinary Differential Equation models. This coupling is illustrated with the abstract MicMac model and more recently with a model of Dengue spread investigating the causal relationship between the opening of an economic corridor in South-East Asia and the number of Dengue fever cases. These models highlight (i) the benefits that the coupling of agent-based models (generative model at the microscopic level) with equation-based models (descriptive model at the macroscopic level) can bring to modellers, but also (ii) the methodological and technical difficulties of this coupling.

Finally, the last research axis focuses on issues related to **data and data management in agent-based models**. Agent-based models in general and socio-environmental models in particular require a huge amount of input data; they also produce a lot of data that needs to be analysed for calibration purposes or even to support decisions. To deal with these challenges, an integrated framework combining simulator, database management system and Business Intelligence tools is presented; its global architecture, implementation and application to a

case study (rice pests invasion monitoring in the Mekong delta) are also detailed.
One of the main characteristics of the research activity presented in this dissertation is that all the works have been implemented in one single agent-based platform, GAMA (developed in collaboration between the IRD and several French and Vietnamese universities), that is used in numerous training sessions every year (MAPS, MISS-ABMS, JTD).
After the description of the three previous axes, the last section of this dissertation focuses on research perspectives concerning the use of **qualitative data** (inquiries results, testimonies, interview...) to build, feed (at initialisation) and inform (during the simulation runtime) agent-based models and simulations.

Keywords: Agent-Based Modelling, GAMA platform, BDI, trust, emotion, ODE, coupling, data management.

Contents

1 Introduction

This *Habiliaton à Diriger des Recherches* thesis attempts to present a synthesis of my first 11 years of research with a particular focus on my[1] post doctorate work. I would like to start this thesis by a brief overview of my academic journey. It provides background elements helping to understand orientations taken in the work presented in the following sections.

1.1 Academic path overview

This journey has started with a PhD at the IRIT[2] lab under the supervision of Andreas Herzig (CNRS senior researcher) and Dominique Longin (CNRS researcher) in the LILaC[3] team. The PhD work comes within the field of Multi-Agent Systems and more specifically in the logical formalisation of agents' representation and reasoning process in the BDI (Belief, Desire, Intention) [42] framework. I worked on extending this framework with primitive group attitudes and more specifically group belief and group acceptance: the idea was to model the particular attitude that appears after a public expression and/or agreement on a proposition. This work was fully linked to what is expressed by agents and naturally I applied this formal concept to define a new public semantic for Agent Communication Languages such as the FIPA-ACL [76]. In addition, I had the opportunity to work with Carole Adam (PhD student at that time and now associate professor at the Grenoble-Alpes University) on the logical formalisation of emotions, again in the BDI framework. Although this work was fully abstract and formal, it was in fact my first step in multidisciplinary modelling as a lot of inspiration came from (analytic) philosophy, sociology and psychology.

Once my PhD achieved, I chose to extend my knowledge in Multi-Agent Systems beyond logical formalisation. My post-doctoral position, supervised by Nicolas Marilleau (IRD Research

[1] Of course, all this work is far from being only mine, as every single work or article presented here exists only thanks to collaborations with colleagues and students.

[2] IRIT stands for *Institut de Recherche en Informatique de Toulouse, i.e.* Toulouse Institute of Computer Science Research.

[3] LILaC stands for *Logic, Interaction, Language, and Calculus.*

Engineer) and Ho Tuong Vinh (Associate Professor at the IFI) at the IFI[4], Hanoi, Vietnam, introduced me to agent-based modelling and simulation of complex systems through the design of the PAMS collaborative modelling and simulation tool [133]. It was also my first steps in the new-born GAMA[5] community and the beginning of my long-standing collaboration with the UMI[6] 209 UMMISCO[7] of the IRD[8].

After 2 years in Vietnam, I got a position of Associate Professor in the University Toulouse 1 Capitole (UT1C), in the SMAC[9] team of the IRIT, to work more particularly with Frédéric Amblard (Associate Professor at UT1C) and Christophe Sibertin-Blanc (Professor at UT1C). I had the opportunity to be involved in several modelling and simulation research projects and in particular in the MAELIA[10] project that I present below. In addition I have joined several research and training networks such as SimTools and MAPS[11] networks associated with the RNSC[12] or the researcher network around the MISS-ABMS[13] and JTD (*Journées de Tam Dao*) training sessions. The next chapters of this thesis are dedicated to provide more details about these works.

1.2 Agent-based modelling and simulation

Both the tools I manipulate and design (Agent-Based modelling) and the application cases I worked on have anchored my work in the field of the modelling and simulation of complex systems [129]. I present in this section my positioning with respect to the state of the art and illustrate the challenges that have emerged from modern uses of models, on the example of the real-case MAELIA model.

[4] IFI stands for *Institut de la Francophonie pour l'Informatique, i.e.* Francophone Institute for Computer Science.
[5] GAMA stands for *GIS Agent-based Modelling Architecture.* GIS is *Geographical Information System.*
[6] UMI stands for *Unité Mixte Internationale, i.e.* international joint research centre.
[7] UMMISCO stands for *Unité Mixte Internationnale de Modélisation Mathématique et Informatiques des Systèmes Complèxes, i.e.* international joint research centre about mathematical and computer modelling of complex systems.
[8] IRD stands for *Institut de Recherche pour le Développement, i.e.* French Research Institute for Development.
[9] SMAC stands for *Systèmes Multi-Agents Coopératifs, i.e.* Cooperative Multi-Agents Systems.
[10] MAELIA stands for *Multi-Agents for EnvironmentaL norm Impact Assessment.*
[11] MAPS stands for *Modélisation multi-agents appliquée aux phénomènes spatialisés, i.e.* Agents Based Modelling applied to Spatial Phenomena.
[12] RNSC stands for *Réseau National des Systèmes Complexes, i.e.* National Network of Complex Systems
[13] MISS-ABMS stands for *Multi-platform International Summer School on Agent-Based Modelling & Simulation for Renewable Resources Management.*

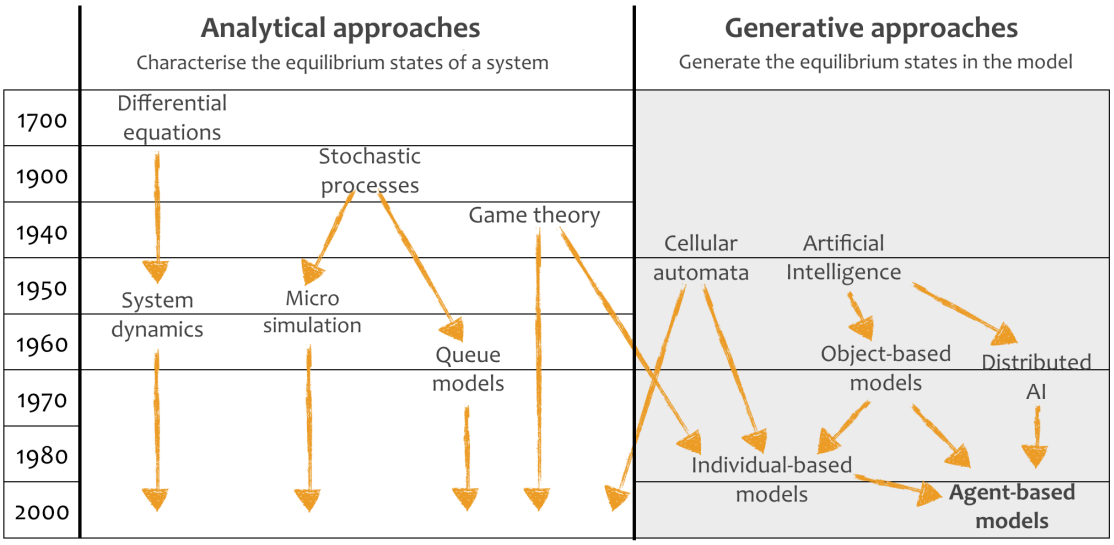

Figure 1.1: Dynamic model approaches over times in Social Science. Adapted by A. Drogoul in [65], from [86, 7].

1.2.1 State of the art in a nutshell and positioning

Modelling[14] is almost as old as science [79] and is one of the basic approaches in any research field and more particularly to study Complex Systems. Some detailed and recent states of the art about modelling complex systems in general and social science (more specifically geography) in particular can be found in [61, 163]. Agent-Based Model (ABM) approach is only one very recent paradigm among numerous other dynamic model paradigms[15]. Alexis Drogoul has summarised in Figure 1.1 [65], adapting the figure published in [86, 7], the main dynamic model approaches, drawing a line between Analytical and Generative approaches.

Most of the **Analytical approaches** (except queuing models) are equation-based approaches. They represent and reproduce thanks to equations (for example Ordinary Differential Equations in Systems Dynamics) the evolution of a system at the macroscopic level, such as in epidemiology to represent a disease spread in a population [109] or in ecology to reproduce population dynamics [119, 204]. On the contrary, **Generative Approaches** (Cellular Automata [80], Individual-Based Models [92] or Agent-Based Models [191]) attempt to model the system

[14]For the definition of models, I stick to Minsky's one: "*If a creature can answer a question about a hypothetical experiment without actually performing it, then it has demonstrated some knowledge about the world. For, his answer to the question must be an encoded description of the behaviour (inside the creature) of some sub-machine or "model" responding to an encoded description of the world situation described by the question. We use the term "model" in the following sense: To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A.*" [128]. Models are written in a particular "language" that is defined in a *metamodel*. In this dissertation, I will introduce many models, and focus most of the time on the implemented ones. In particular, I do not consider the translation between various kinds of models (*e.g.* from conceptual modelling to implemented models), as it can be the case in Model-Driven Architecture [125].

[15]With [191], I differentiate *static* (describing the structure of a system) from *dynamic* (describing its dynamics) models. A *simulation* is then simply defined as the execution of a dynamic model.

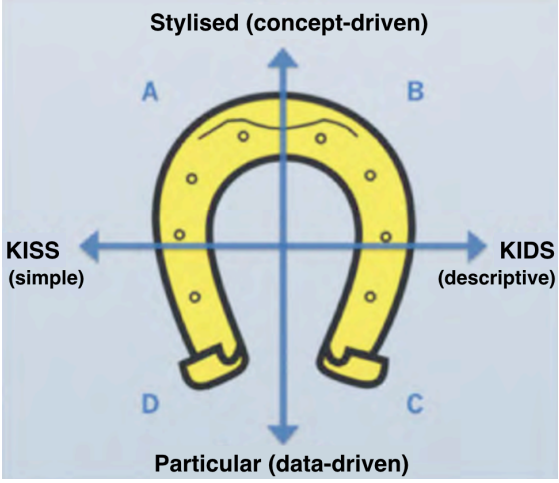

Figure 1.2: "Horseshoe" reading template proposed by [29] to classify models in geography.

at the individual level and to generate by simulation the observed macroscopic behaviour[16].

As [163] shows in his history of agent-based modelling in Social Science, the ABM approach appeared almost simultaneously in the 90's in Europa [86] and in the USA [72] with the same aim of reproducing artificial societies. From that time and with the development of computation power and dedicated tools, it spreads in most of social science fields, such as in economy [189], geography [43] or archaeology [21] and even in ecology [59]. Agent-based models take benefits from Individual-based models (in particular for the representation of the environment and its dynamics), Artificial Intelligence (AI) (aiming at developing an highly intelligent agent) [166] and Distributed Artificial Intelligence (DAI) (considering that intelligence is not fully in the single entities of the system but in the entities in interaction; intelligent behaviours will emerge from these interactions)[17] [37].

In parallel to the development of more and more models often implemented from scratch, a lot of tools and in particular agent-based modelling and simulation platforms have been developed to open this approach to non computer-scientists and to improve the sharing of models and discussions around them. [112] presents a recent survey of some of these platforms. Among the platforms that are open-source and generic enough to model any kind of system and phenomenon, 2 main tendencies have emerged, illustrated by the 2 main platforms: NetLogo [206] and Repast S [145]. NetLogo provides a simple domain-specific language dedicated to implement models. This platform is non computer scientist-oriented and is well-adapted to quickly develop small and simple models but remains limited for more complex ones. Repast S is dedicated to modellers with programming skills in Java. It provides

[16] Instead of considering these approaches as incompatible, the Chapter 3 of this thesis focuses on bridging the gap between these two approaches into a single model.

[17] DAI has later become Multi-Agent Systems (MAS). In the following sections, "MAS" will be used to denote the field concerning the design of **applications** based on a set of artificial software agents interacting together. It is highly influenced by Artificial Intelligence and DAI. "ABM" will denote **models** designed to represent a system and answer a question on it. Chapter 2 is focused on integrating MAS concepts into ABM.

a Java API, with many additional libraries helping to develop models. As a consequence it can deal with bigger models but is barely used by non computer scientists. The GAMA platform [181], in which I am involved, chose an intermediate approach providing a dedicated language (GAML for GAMA Modelling Language) but having the capabilities to deal with big and complex models (see Section 1.3.1 for details about GAMA).

Despite the use of these platforms easing the standardisation of models, a huge diversity exists among all the existing agent-based models, in particular because even the main concepts do not have a unique definition (*e.g.* agent or environment) or because the agent-based approach does not impose many modelling constraints. In addition models can be designed with very different purposes[18], which makes them even more diverse. [29] proposed the "Horseshoe" reading template of the models on 2 axes (Figure 1.2): simplicity and abstraction level. The **simplicity** refers to the opposition between the KISS (*Keep It Simple, Stupid* [17]) and the KIDS (*Keep It Descriptive, Stupid* [70]) approaches. The KISS approach favours simple models, with very simple agents' behaviour to reproduce complex systems: the model is simple, but the simulation results are complex. In contrarily, the KIDS approach militates for descriptive models, which remain explicative. The second axis is related to the **abstraction** level of the model: does the model represent a stylised fact or a particular given phenomenon? These two axes define four quadrants, the horseshoe illustrating the easy and natural path between these quadrants. For example, from a KISS-stylised fact model, it is often natural to move to a KIDS-stylised fact by complicating the model to improve its explanatory power. The path from KIDS-stylised fact to KIDS-particular case is typically the attempt to apply a theoretical phenomenon on a particular case-study whereas the converse is the attempt to generalise a particular phenomenon. The models I present in the next sections are definitely in the KIDS half (navigating between KIDS-stylised fact and particular phenomenon), mainly because they are often modelling systems including many ecological dynamics coupled with anthropogenic ones.

Such systems, named Socio-Ecological (or Socio-Environmental) Systems (SES) [148], have recently been the objects of many studies by agent-based modelling. These systems are really complex as they integrate actors, environment and institution dimensions in the same model, taking into account interactions between these three elements. These models are generally designed to be realistic and to provide tools for prospective and Decision-Support Systems. They thus bring new challenges for both modellers and computer scientists to design new advanced theoretic, conceptual and computational tools. These models are by essence descriptive (in the sense of the KIDS approach) and dedicated to particular cases. The MAELIA model is a typical example. It is presented in the next section in order to illustrate challenges raised by such models.

[18] Axelrod [16] lists seven purposes of simulation: prediction, performance of a task, training, entertainment, education, proof and discovery.

1.2.2 The MAELIA model: a prototypical model of Socio-Ecological Systems

Since 2010, I have been involved in the MAELIA project[19] [85]. It is a typical modelling and simulation project of socio-environmental systems with numerous kinds of agents (from very simple to very complex ones) and a huge amount of data needed. It has been implemented with the GAMA platform. It has been a great experience of interdisciplinary modelling work along with a great technical challenge. Here I present this model and use it as a starting point of this thesis because it has raised a lot of challenges to modellers and computer scientists: it is a great illustration of the improvements and new features that should (from my point of view) be integrated in modern modelling and simulation toolboxes.

The MAELIA project is dedicated to the development of an agent-based modelling and simulation platform to study the environmental, economic and social impacts of various regulations regarding water use and water management in combination with climate change. It is applied to the case of the French Adour-Garonne Basin, which is the most affected in France by water scarcity during the low-water period. Its ultimate aim is to become a Decision-Support System (DSS) providing information usable by institutions in charge of designing and implementing sustainable management strategies of water resources at the water basin level.

This model is a typical example of a large-scale integrated model of socio-ecological systems [148]: it combines spatio-temporal models of ecological (*e.g.* rainfall and temperature changes, water flow and plant growth), socio-economic (*e.g.* farmer decision-making process, management of low-water flow, demography, land use and land cover changes) and institutional (*e.g.* drought decree issuing) processes. The agriculture part (and in particular farmers' cropping decisions) is extremely important as it has been shown that in the considered basin, 75% of the water consumption during the low-level water period (which is the most critical) is due to irrigation (in particular for corn). The farmers' decision to sow such or such culture can thus have a huge influence on the whole model and must be taken into account very carefully.

During the modelling process, first a meta-model of socio-ecological systems has been designed [172]. It has then been instantiated in dozens of entities and dynamics. Soon in the implementation process two main principles have emerged: modularity and reusability of existing models. Modularity means that for any process or entity in the model, it should be possible to use several implementations. For example, a simulation can use a complex farmer decision-making process based on multi-criteria decision whereas another one can use a much simpler decision-making process. The second driving idea was to reuse existing models when we can (even when we need to reimplement them in GAML). New models (such as farmer decision-making process [183]) have been designed only when the existing ones were not appropriate. For example, to represent water flow we use the mathematical formalism of hydrological cycle from the SWAT (Soil and Water Assessment Tool) model [15], and the "AqYield" model for plant growth [143].

[19] The MAELIA project has been funded by the French "Sciences & Technologies for Aeronautics and Space" Foundation from 2009 to 2014. It is now led by the INRA (French National Institute for Agricultural Research) and is being transferred to agencies and companies.

Despite its complexity and simulation execution time, the MAELIA model behaviour has been explored. In particular, a dedicated sensitivity analysis method has been developed [114] and it is being calibrated too, first on the water flow process only, then on the farmer decision-making process only and finally on the whole integrated model. Its use as a Decision-Support System is also being developed: a recent work has connected its execution with a database of weather scenarios (weather conditions for the next six months) updated weekly. Every week, the MAELIA model will be executed with the historical weather data until the current day and will use one by one each weather scenario to get a bundle of predictions for the next six months.

1.2.3 Challenges from the MAELIA model

As far as I am aware, MAELIA is the biggest and heaviest model developed using the GAMA platform. It is also one of the most ambitious models I know. The GAMA platform and the MAELIA model have jointly evolved: the design of the MAELIA model has been influenced by the GAMA platform but it has also shaped and improved new features of the platform. The need for modularity, the number of model files to process and the integration of data have requested improvement and optimisation of the platform that benefit all the other projects. I focused mainly on three requirements (and new features) that clearly emerged from the development and execution of the model.

Agents with advanced cognitive capabilities. In the MAELIA model, the agriculture component and more specifically the cropping plan decision made by farmers has a large influence on the water consumption. As a consequence, the farmers' behaviour model and their decision-making process in particular have been the object of a particular attention and a dedicated model has been developed.

This example illustrates (from my point of view) the necessity to improve the possibility to give high-level cognitive capabilities to agents when needed. The path I chose to follow is an attempt to bridge the gap between traditional Artificial Intelligence models of human reasoning or decision-making (in particular using logic [161] and BDI architecture [42]) and computational agent-based models. This objective brings a lot of very interesting challenges from a modelling point of view. In particular, it requires to design an agent architecture dealing with all the steps from the perception of the environment, its interpretation and representation and the reasoning on these representations, to the decision-making process and the action performance. In addition, agents are inserted into large artificial societies and thus need advanced social capabilities (based on concepts of trust or emotions for example). Finally, one of my objectives is also to integrate all these additional capabilities into a generic platform (GAMA) which adds new constraints: the new tools should be generic enough to be integrated in any agent-based model, fast enough to support thousands of agents in the same simulation, and intuitive enough to be used by non-computer scientists, who are the main targeted users of such a platform. Tackling all the related issues is far beyond the scope of this

thesis. Proposed architectures and models are presented in Chapter 2.

Equation-based models inside agent-based models. The MAELIA model highlights the importance for an agent-based model to be able to manage very different agents in terms of paradigm describing their behaviour, and to let them interact. The model has needed in particular some agents (watersheds[20]) to have a behaviour (in a very wide sense, *i.e.* a dynamics) driven by a set of equations (coming from a particular hydrologic model SWAT [15]). From my point of view, this capability of permitting interactions of agents driven by very different kinds of model is one of the main strengths of ABM to ease interdisciplinary modelling collaboration. In particular, this is necessary to represent systems as complex as socio-environmental systems.

The sets of equations in MAELIA watershed agents are simple and linear equations. But this opened the door to more advanced coupling between equation-based models and agent-based models. Applications in epidemiology [26, 154] showed me the importance of integrating Ordinary Differential Equations (ODE) models into ABM. Indeed, they can be a very simple, intuitive and efficient way to deal with multi-scale models: ODE models represent macroscopic dynamics (*e.g.* a disease spread at the scale of a city or a part of a province) whereas ABM deal with individual dynamics (*e.g.* mobility of some individuals between cities or provinces). Similarly, an application in ecology [48] showed the interest of integrating Partial Differential Equations (PDE) in ABM. This work is presented in Chapter 3.

Management and integration of data. The MAELIA model has been designed from its very beginning as a model that requires a huge amount of input data and that produces many outputs. One of the ultimate objectives of the project is to build a Decision-Support System based on the simulations. Despite these two observations and similarly to other modelling projects in which I have been involved, all the data is manipulated manually as a set of files.

This raises some general questions about the management of data into simulations and more generally during the whole modelling process. Advanced tools exist to efficiently manage data in a business context, with in particular solutions for Decision-Support Systems and Business Intelligence. It thus appeared necessary to me to improve the management of data in agent-based platforms, first by simple connection to Databases and then by integrating dedicated Decision-Support System tools [193]. Chapter 4 is dedicated to this work.

[20] A watershed is a land area where all the water entering or falling in it is drained to a single outlet. In the MAELIA model, the SWAT hydrologic model is used to compute water flows and the watersheds are the basic spatial unit for this computation.

1.3 Research method

Due to the focus of my research on the improvement of agent-based tools to model complex systems, I cannot limit my work to pure theoretic work in collaboration with only computer scientists. I have thus to collaborate closely with numerous other researchers from very various fields, provide them with the developed tools and get them to adopt these new tools. To this purpose, from a methodological point of view, I can distinguish three aspects in my work, that enrich and "validate" each other continuously.

1.3.1 Make concepts operational: implementation in the GAMA platform

I argue that it is very important that everything I have designed (such as the integration of cognitive agents into agent-based models, the integration of data management tools into agent-based models, or the coupling of agent-based and equation-based models) is implemented and becomes operational concepts.

The aim is multiple. First it is a way to validate the design described in articles. In addition, everything is distributed as open-source code, so it can be provided to the scientific community and used by others for their own projects. Finally I like the idea that everything is integrated into a single agent-based platform. It allows to incrementally develop a tool more and more powerful and expressive; it thus eases interdisciplinary collaborations. It also gives a certain homogeneity to the whole work as each part is based on the same conceptual meta-model.

To this purpose, I chose for more than six years to take part in the design and development of the GAMA platform [181, 91, 63]. The GAMA platform[21] is a generic (in the sense that any kind of agent-based model in any kind of research field can be implemented in it) agent-based modelling and simulation platform, designed and developed initially in the UMI UMMISCO lab by Alexis Drogoul. At its beginning, it was built as a framework on top of the Repast J [145] agent-based platform, in order to provide a meta-model able to support the requirements of three research projects (avian flu propagation, maintenance in the environment and resurgence [11], urban emergencies [52] and household daily activities [13]). The Repast J platform was chosen as it integrates a powerful GIS library.

Later GAMA quickly became an independent modelling and simulation platform, providing a dedicated modelling language. The first key idea of the platform was to deeply link agent-based models and GIS data. The assumption behind the development of GAMA was that the available platforms at that time were not relevant to the needs of some users (in particular in geography): NetLogo did not integrate GIS data efficiently, while Repast J or Symphony [145] required knowledge in the object-oriented Java language to develop the model. The approach behind the design of GAMA was to take the best of other platforms, *i.e.* to provide a dedicated modelling language and advanced GIS data management capabilities.

[21] http://gama-platform.org

Another specificity of the GAMA platform is that its meta-model is fully agent-oriented and natively multi-level. This means that all top-level entities are agents: in-simulation agents, the environment, simulations and experiments are all agents. In addition all agents are embedded and scheduled by a higher-level agent. As a consequence, running an experiment required to create an experiment agent and to execute its behaviour (that is the scheduling of all its internal agents) similarly to executing any in-simulation agent. In addition, the environment of the agents is itself an agent, with its own attributes and dynamics[22].

The GAMA platform is now used in several training sessions and teaching units about agent-based modelling as discussed in the next paragraph, even outside of the GAMA community kernel.

1.3.2 Confront tools with modellers: training sessions

A second very important part of my research work is dedicated to training sessions in which I am involved in addition to my university teaching. They are very important as they give the opportunity to spread the GAMA platform to new modellers, but more importantly, they are the perfect occasion to interact with experts in various fields, build new potential collaborations and provide new ideas to improve the tools. From a pedagogical point of view, the following training sessions present two different ways to introduce new modellers to ABM: either by providing general lectures and letting trainees develop their own model; or by providing students with an existing implemented model, illustrating the lectures with this model and letting trainees manipulate the model and extend it, to answer a given modelling question.

All the training sessions presented here are very similar in terms of organisation and trainees profiles. They are long and intensive training sessions (from 1 to 2 weeks long). In terms of organisation they start with some general lectures (about modelling in general, conceptual modelling...), but at least half of the time is dedicated to group work on personal projects closely supervised by trainers.

The Tam Dao Summer University ("Les Journées de Tam Dao", JTD). The JTD[23] are a Regional Social Sciences Summer University organised since 2007 in Vietnam. It gathers every year between 80 to 100 trainees (mainly university lecturers from various social science fields but also stakeholders) and around 20 trainers (mainly French professors or researchers). Around an overall subject that is different every year, the two first days are dedicated to plenary sessions from international experts. They are followed by 5 days of interdisciplinary workshops in small groups. Since 2012, one of the workshops is dedicated to an introduction to agent-based modelling to face an issue related to the overall subject. The specificity of this training session is that the workshop is organised around a model we develop by ourselves with a

[22] This can be used in particular to manage exogenous dynamics, *i.e.* dynamics that are not generate by the model (*e.g.* weather).

[23] http://www.tamdaoconf.com/

thematician[24]. Lectures thus start with the presentation of the subject from the thematician point of view. Then the agent-based modelling methodology and the GAMA platform are introduced through the presentation of the model. The group work is dedicated to the extension of the model, by introducing new dynamics, new entities... The pedagogical ambition is to initiate trainees to a new way of thinking complex problems and to the possibilities of such an approach. This approach of training is particularly relevant for a very first introduction to the agent-based approach and an user-oriented position with relation to the model. But it prevents trainees from appropriating deeply the model and developing autonomously a model from scratch.

The two following training sessions adopt the opposite approach.

MAPS schools. The MAPS schools[25] is by far the most intensive training session. Trainees and trainers work and live together in a closed and remote location during 5 days. In addition to theoretic lectures, trainees mainly work on their own projects and experiment the whole modelling process from the modelling question definition and the conceptual modelling to the experiment design. There is a huge emulation among people, and modelling sessions can last long into the night.

MISS ABMS training sessions. MISS-ABMS[26] training sessions were originally organised around the CORMAS platform community to provide an agent-based training session oriented to CORMAS users. Since several years, it is open to other modelling communities and has integrated the NetLogo and GAMA platforms. Longer than other sessions (2 weeks), it takes more time to introduce concepts and associate exercises to lectures. A particular focus is made on conceptual modelling (using various static and dynamic UML diagrams) with lectures and exercises. At the end of the first week, the three modelling platforms are presented and compared (on the implementation of a same model). Trainees can then choose the most adapted one for their own needs, and thus implement their group project during the second week. It becomes very interesting when several platforms are used in a single group. Trainees should thus interact enough to design and keep the same conceptual model, and implement it on several platforms (which of course induces some implementation choices). This session is very rich and brings a lot due to discussion with other platform designers (CORMAS).

[24] Previous edition models have been dedicated to past crisis reproduction [81] (with Olivier Tessier, historian), urban spread [179] (with Arnaud Banos, geographer), correlation between economic corridor and dengue fever spread [154] (with Marc Choisy, epidemiologist) and support to the design of urban energy transition policies (with Javier Gil Quijano, modeller at the CEA, the French Alternative Energies and Atomic Energy Commission)

[25] The MAPS network is the thematic network about multi-agent modelling applied to spatial phenomena. It is one of the RNSC (National Network of Complex System) networks and is dedicated to organise interdisciplinary schools and researcher workshops.

[26] MISS-ABMS stands for Multi-platform International Summer School on Agent-Based Modelling & Simulation for Renewable Resources Management.

Other non-recurrent training sessions. In addition, I was involved in many other non-recurrent one-week training sessions: *e.g.* at Can Tho university (Viet Nam) in 2013, at Manilla (Philippines) in 2015, and Siem Reap (Cambodia) in 2015. They are often very inspired by the JTD training sessions.

These training sessions (in addition to research networks such as MAPS or SimTools Network) are unique opportunities to meet researchers from very different research fields and to discover new applications.

1.3.3 Tend towards genericity: multiply the number of case-studies

Having the opportunities to work with many different people from various fields is a great inspiration source as well as a way to check the genericity and the applicability of tools designed for a given application or purpose. It is thus a way to validate them, especially since these applications are real case-studies with real questions. As an example, the MISS-ABMS training sessions allowed me to meet new people and start new collaborations: *e.g.* with hydrologists and geographers, on the study of draining basin in the area near Phnom Penh (Cambodia) or with agronomists on the Nitrogen cycle with application on villages in Senegal, always using the GAMA platform.

To illustrate the virtuous circle of the interactions between the three aspects presented in this section, I can cite the example of the MicMac project. This work about the coupling between agent-based and equation-based (Ordinary Differential Equation, ODE) models comes from the MAPS 4 researcher workshop, where I had the opportunity to collaborate with mathematicians and geographers. It has induced the implementation of the mathematical extension in GAMA (*c.f.* Appendix A.3) and led to several publications [25, 26, 24, 23]. It has then been used as support for later occurrences of the MAPS school and for additional applications.

1.4 Contributions and organisation of the manuscript

I have attempted to sum up my global contribution in the thesis title: “Toward complex models of complex systems - One step further in the art of Agent-Based Modelling”. As illustrated with the MAELIA model, models of socio-environmental systems are *de facto* complex because they reflect the multidisciplinary nature of the modelling project, inducing for example the necessity to build hybrid models combining various paradigms[27]. Building complex models raises new challenges and the following chapters describe in details the

[27] In addition I wanted to make a parallel between **complex systems** (that can be defined, in a very simple way, as a set of entities in interaction and whose global behaviour emerges from these interactions) and **complex models** (that can be described as a set of models in interaction whose global behaviour emerges from the interactions between these models).

contributions I proposed to the various challenges pointed out in this introduction[28].

Chapter 2 describes my contributions to the increase of agents' cognitive capabilities in agent-based models of complex systems. In particular, I detail the GAML extension introducing a BDI architecture in GAMA. In addition, I present models of trust and emotions for agents in large artificial societies. All these theoretic models are illustrated with specific implemented models (*e.g.* land-use change, information sharing or evacuation models).

In Chapter 3, I discuss from a methodological point of view the coupling of Equation-Based Models and Agent-Based Models. I illustrate the discussion with examples from two epidemic models coupling agent-based individual mobility and equation-based disease spread.

Chapter 4 presents a logical framework coupling agent-based modelling and simulation tools with data management and Business Intelligence tools. I then detail its implementation with the GAMA platform and illustrate its application to a rice pest invasion model. I demonstrate the various possible uses of such advanced tools and the benefits it could bring to the agent-based modelling and simulation field. I conclude this chapter with a perspective on the issue of missing data, and the possible solution of generating synthetic populations to feed simulations.

Finally Chapter 5 concludes and presents new perspectives and future work. In particular, I focus on the development of models driven by quantitative data that can be available, such as after crisis interviews, reports, inquiries or even tweets. The typical considered application is crisis management models.

I made the choice to not say much about GAML in these 4 chapters for the sake of clarity and genericity. Appendix A therefore gives a technical taste of the GAML language and of the extensions associated with each chapter.

Finally Appendix B provides a synthesis of my various scientific productions, including publications, supervisions, collaborations and projects.

[28]This thesis is expected to be a coherent synthesis of my work along three main axes. It is not a chronological nor an exhaustive picture of all my work. In particular, I have chosen to omit everything related to collaborative simulations. Similarly several models I worked on are not presented here.

2 Complex agents and agent-based models

2.1 Why do we need complex agents in simulation?

In socio-environmental models and in the MAELIA model in particular, the most challenging part to model is definitely the human behaviour component including knowledge representation of the environment state and of other agents or decision-making process.

But we can notice that, although most of the very classical models of computational social sciences are focused on human-related phenomena (*e.g.* [60] focused on opinion dynamics, [18] on culture diffusion, [168] on segregation, [102] on crowd move), human beings models remain very simple. As an example [102] have considered human beings in a crowd as particles in a (social) force field. These models are really interesting and their simplicity is a big part of their strength. They can be the basic brick of more complex behaviours. But all of them consider a very specific and isolated phenomenon. The study of socio-environmental systems, where human beings are plunged into a complex dynamic environment, requires to go further and to have more complex architectures to describe their behaviour. Human beings' decisions should take into account their environment and its past, present and expected states, their own past experience, current state or preferences. In addition, they are included in an artificial societies and should be able to integrate other human beings' choices (*e.g.* [192] consider that farmers are highly influenced by their spatial neighbours' decisions to make their own decisions in terms of land use changes), or information transmitted by others and be able to identify right and wrong information. Finally, studies [68] in real cases of evacuations have shown that people are not only submitted to social forces to evacuate but their decisions take also into account others (*e.g.* parents will not evacuate without their children, people tend to help each other's...) and more complex cognitive components, such as their emotions (influenced by many factors including others' emotions).

The chosen approach presented in the sequel is an attempt to start bridging the gap between traditional Artificial Intelligence models of human reasoning (in particular using logic [53, 161, 8] and BDI architecture [45, 39]) and computational agent-based models. This concerns both the whole reasoning architecture (in particular the BDI architecture), but also all the advanced

cognitive components that have been deeply studied in Artificial Intelligence. My work has only focused on the impact of emotions on behaviour, emotional contagion and reasoning about agents' and information reliability using trust concepts. In the following, I will thus use the expression "complex agents" in this sense of agents whose behaviour is influenced by various high-level cognitive components[1], such as emotions, trust, desires or intentions...

To this purpose, we kept in mind several characteristics of agent-based modelling and simulation. The proposed solutions should be light enough to be used in simulations running with thousands of agents and generic enough to adapt to many kinds of models. In addition, my goal is that models can be designed and implemented by non-computer scientists. As a consequence, the provided concepts should be simple and intuitive.

I first present in Section 2.2 the integration of BDI (Belief-Desire-Intention) agents into agent-based simulations. In the Section 2.3, I discuss the use of the concept of trust and how it could be computed and used to reason about other agents and their information. Section 2.4 discusses how to extend agent architectures to integrate emotions in a social context (social emotions and emotional contagion) and how they impact agents behaviour. Finally Section 2.5 concludes the chapter and opens new perspectives.

2.2 BDI architecture for agents in simulations

Proposals to make agents more complex in agent-based simulations are numerous. In this section I focus only on introducing BDI agents, choice I justify in the next section.

2.2.1 BDI agents in social simulations

The BDI (Belief, Desire, Intention) approach is a classical paradigm in Multi-Agent Systems domain to describe the way human beings are representing knowledge and reasoning on it. It is mainly based on Bratman's philosophical work [42] and attempts to capture the common understanding of how humans reason through: **beliefs** which represent the individual's knowledge about its physical and social environment and about its own internal state; **desires** or more specifically goals (non conflicting desires) which contain basically the way individuals want the world to be; and **intentions** which are the set of desires the agent has chosen and committed to itself to achieve; each intention is associated with the plan or sequence of actions the individual intends to follow in order to achieve its goals.

This general framework has been developed both theoretically, mainly using modal logic [53, 161, 8] and practically with numerous various implementations (and extensions) [45, 39].

[1] As discussed in [3, 7], I will not consider more complex and low-level cognitive architectures such as [177], as it appears that the chosen abstraction level is the most relevant one to map with the way modellers think in the human beings representation of the world. In addition, even if there exists a huge number of BDI architectures, they all share the same basic components and common focus, which is not the case of cognitive architectures, based on various physiological and neurological points of view.

While the specific way in which this is done varies among implementations, it is generally guided by folk psychology principles [144]. But the core functionalities that BDI systems should implement are: a representation of beliefs, desires and intentions; a perception module transforming percepts into beliefs; a rational process for selecting intentions; a flexible and adaptable commitment to the current set of intentions. From the initial BDI trinity, the framework (both logics and implemented architecture) has been extended to take into account various other cognitive components, such as emotions, obligations, norms, trust...

In [7, 3], we have provided, based on a survey of simulations integrating BDI agents, a classification of cases where and how BDI agents are particularly well-adapted. In particular, we have discussed needed agents' characteristics, application fields, simulation goals and the observation level. The BDI architecture is particularly suitable when we need individual agents with high-level representation and reasoning capabilities and can bring high benefits when we need agents with self-explanation capabilities as BDI architecture is based on folk psychology which allows the agent to provide intuitive explanations. In addition, high-level communication capabilities provided by Agent Communication Languages (such as the FIPA ACL [76]) can perfectly be used in such an architecture. We have also investigated the impact of the simulation application field on the use of the BDI architecture: among all the application fields listed by [120], we have observed that this paradigm can and has been used in all the fields. In fact, it appears that it is more the level of observation of the system rather the field that has an influence in the choice: when we are interested at simulation at a fine-grained scale, the choice of the BDI architecture can be particularly relevant, whereas it is not at all when the focus is at a higher scale. Finally, we followed the classification of [19] and investigated the influence of the goal of the simulation on the suitability of using such an architecture. We have shown that it fits more with simulations aiming at performing a task and training people, in particular when we need to have individual agents realistic enough for human beings to interact with them.

Finally, from a technical point of view, it appears that in social and socio-environmental models this architecture is in fact barely used. We can observe a split between work coming from Multi-Agent Systems which provide various implemented architectures and adapted methodologies and Multi-Agent-Based community with very fewer tools. In particular, we have observed that among all the agent platforms listed in [112], most of the classical agent-based modelling and simulation platforms (NetLogo, Repast S, Mason or GAMA) have only a limited support of the BDI. We can mention the exception of Sesame which can embed BDI agents through a coupling with Jade.

In the next sections, I present an attempt to integrate a BDI architecture inside the GAMA platform.

2.2.2 A situated BDI agent architecture for the GAMA platform

A first attempt to integrate a BDI architecture inside the GAMA platform has been done by Le Van Minh during his master internship [116] by developing in the GAML language a toy model (forest fire and its management by firemen) integrating BDI agents for firemen. All the BDI architecture components have been implemented in GAML language. This first work has provided promising results but also limitations in terms of the architecture reusability. To overtake this limitation, we had started to implement a BDI architecture in Java and provide GAML primitives in order that the architecture can be used in any model[2].

Presentation of the architecture.

The proposed BDI architecture [47, 180] has been implemented in an additional plugin (named simpleBDI): it provides to the modeller all the required primitives to create and manipulate beliefs, desires and intentions and to define the behaviour of agents controlled by the BDI architecture.

Data structure. As detailed in the Appendix A.2, our BDI agents are characterised by three bases of beliefs (what they know about themselves or the world), desires (what they want) and intentions (what they are doing to fulfil their desires). The content of each belief, desire or intention is defined as a `predicate`, a data structure containing the name of the predicate, a map of attached values, a priority and a truth value. A predicate can contain any type of information existing in the GAML language (quantity, location, Boolean value...). In addition, desires can have hierarchical links (sub/super desires) in order to create intermediary objectives. Similarly we can have intentions and sub-intentions. A stack of intentions has been introduced: an intention can be put on hold to execute sub-intention necessary to achieve intermediary objectives.

Behaviours. The general execution flow of the BDI architecture is described in Figure 2.1. It is based on three new behaviour structures introduced by the BDI extension: perception, rule and plan.

A **perception** is a command executed at each simulation step to update the agent's Belief base given the state of the environment. The general aim is to perceive the modifications in the environment and to map them to a predicate that will be integrated into the belief base. The relevant information to be perceived in the environment is defined by the modeller in the perception. It attempts to take advantage of space management provided by the platform to ease and automate the link between perceptions and belief bases.

[2] This is mainly a collaborative work with Carole Adam, Mathieu Bourgais, Philippe Caillou, Patrick Taillandier, Truong Chi Quang. This work is for a part funded by the ANR research project ACTEUR (led by Patrick Taillandier) and is part of Truong Chi Quang's and Mathieu Bourgais' PhD thesis.

A **rule** is a command executed at each simulation step to infer new attitudes (beliefs or desires) from existing ones. It is typically called after the perceptions in order to update the beliefs and desires depending on the new perceived information. No automatic coherence has been implemented, so it remains the modeller's responsibility to maintain (if needed) the coherence in what is added to the bases.

A **plan** defines a set of actions to be executed to fulfil an intention (several plans can be defined for the same intention). A plan will be executed if the specific intention exists and is not on hold. A plan can be defined as instantaneous, which means that another plan can be executed in the same simulation step: this has been introduced to deal with intentions related to cognitive processes that can be much faster than any physical actions. When no instantaneous plan is executed, only one plan is performed each simulation step.

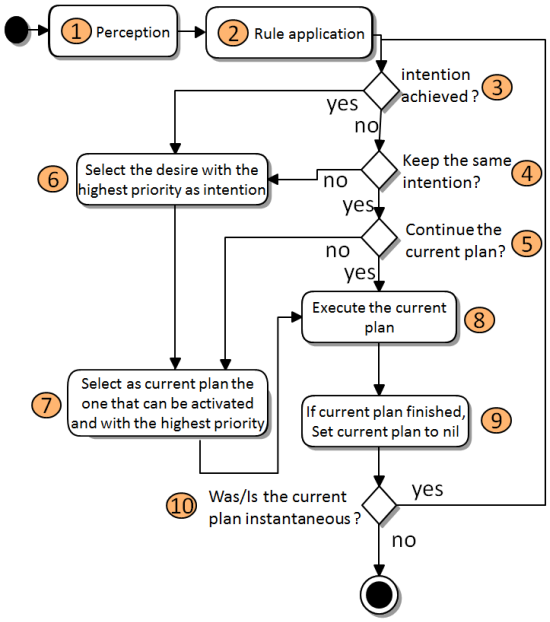

Figure 2.1: Activity diagram of 1 step of a simpleBDI agent [180].

Execution process. In addition to primitives easing the management of beliefs, desires and intentions attitudes and defining additional behaviour structures, the simpleBDI plugin also provides a dedicated way to execute the agent (illustrated in Figure 2.1). Following the classical perception-decision-action workflow [209], a simpleBDI agent first executes its perceptions (step 1), which updates its belief bases, and then its rules (step 2) to update its bases according to its new beliefs. Then it focuses on its intention: if the previous intention is achieved (3) (which can occur by previous actions of the agent, by other agents' actions or by environment modifications) or if it decides to drop it (4), it will choose as the new intention the desire with the highest priority (6).

Finally, it will execute its plan(s). It first chooses whether it will persist in its current plan (5) or

select a new one (7). It then executes its current plan (8). Once the execution of this plan is over, it checks whether it has finished its plan (9), for example when the `finished_when` facet of the plan is fulfilled. In the simpleBDI architecture, we have considered that some plans can be instantaneous, this is in particular the case when the actions to perform are mental (management of the bases). In this case, several plans can be performed in a single simulation step. Thus, the execution flow goes back to the step (3) and the agent restarts its intention elicitation and plan performance.

Main results.

The proposed architecture aims at being easy to use by modellers and scalable to deal with simulations including thousands of agents. I thus present in the two next sections results on these two points. In addition, I present how the architecture has been used in a real-case application.

Feedback from modellers. In order to evaluate the architecture relatively to its easiness to be used by modellers (being computer scientist or not), we made a small experiment with 6 modellers (3 computer scientists and 3 geographers, all of them knowing at least the bases of the GAML language) [180]. Only the 3 computer scientists are familiar with the BDI paradigm. This experiment has consisted in a short lesson (45 minutes) about the simpleBDI architecture, followed by an exercise which required to use this architecture (2 hours). At the end of the exercise, each participant answered a short survey to assess the architecture (with both open questions and closed-ended assessments). The exercise theme was the evacuation of the city of Rouen (France) in case of a technological hazard in one of the buildings of the city centre. Drivers can perceive the hazard at a given distance and can have a chance to understand the existence of the hazard by seeing another driver tying to evacuate. Those who are aware of the hazard try to reach one of the evacuation sites (shelters). The exercise was to improve this model following some given instructions and using the BDI architecture.

After the exercise, each participant answered a short survey about the BDI architecture. A first analysis of the results shows that all the participants have found the proposed architecture clear. Furthermore, the three participants that have a background in BDI architectures consider that our architecture translates the BDI paradigm well. Concerning the simplicity of use of the architecture inside GAMA, three of the participants assess it good or very good, and two pretty good. Concerning the comparison to other agent architectures available in GAMA, one of the modellers (the BDI expert) preferred simpleBDI to the other ones, while two of them found simpleBDI to be complementary to the existing ones, and he has mentioned that it allows to define the behaviour in a simpler way, avoiding to write many complex reflexes. Only one modeller mentioned that he has preferred the basic reflex-based architecture as he was more used to it.

An interesting remark is that some of the participants mixed the simpleBDI and reflex archi-

tectures, using the BDI architecture to define perceptions and objectives (especially the agents target), and reflexes for the repetitive operational behaviours (moving).

Scalability. All the following tests have been performed on a Macbook pro (2011) with an i7 processor and 4Go of RAM. We have computed the simulation time for two models [180]. First a classical Goldminer model has been tested with 10 000 BDI miners, 1000 golds and a square environment of 10 x 10 kilometres. The simulation was stopped when all the gold nuggets had been returned to the base. The average duration of a simulation step (without any graphical display) was 140ms.

The evacuation model used in the previous section was tested with 1000 drivers (due to the road network used and how the capacity of roads was defined, it was not possible to test the model with more driver agents because all the road would have been blocked) and a capacity for each of the evacuation sites of 200 driver agents. We stopped the simulation when all the drivers reached an evacuation site. The average duration of a simulation step (with no graphical display) was 70ms.

The first results obtained in terms of performance are promising and show that the architecture can already be used with medium-scale real-world problems. However, the architecture will still be continuously optimised and we plan to compare the results with the other GAMA architectures (especially the reflex one). The examples presented here are still quite simple in terms of reasoning process and of number of attitudes manipulated; the next step will be to test the architecture with more complex agents with many possible desires, sub-desires and plans.

Application to a real-case land-use change model. The simpleBDI has been applied in [47] to simulate land-use change in the Mekong delta (Vietnam). The Mekong delta area has to face directly the effects of climate change and in particular of the sea level increase and the intrusion of salt water in fields. The Vietnamese authorities design every 10 years a plan of infrastructure buildings and expected land-use. As shown in [146], we can observe a huge difference between what has been planned and the actual land-use, which can make built infrastructure and investments useless. The difference is mainly due to the fact that farmers' individual decisions are not taken into account in the plan design. The aim of this model is to better understand farmers' decision-making process, in order to be able to improve the authorities' plans. Interviews have revealed that the main decision drivers are the land type of the parcel, the expected profit and neighbours decisions.

In the model, we consider each parcel independently and we assimilate each of them with their owner (a farmer). A parcel is characterised by a land-unit type (close to a soil type). Each farmer agent has in its attitudes beliefs about the expected profit for each land-use and each land-unit type. The desires include the type of production (land-use) the farmer desires to do and possibly the desire to give information about its own profits to its neighbours.

Farmer agents have also two plans: `do_production` to install a land-use on their parcel and `inform_people` to transmit information about their own cultures to their neighbours.

This model shows that the simpleBDI architecture can also be applied in real-case models; but it also highlights a limit of the BDI architecture: this architecture is dedicated to organise the various components related to the agent reasoning process, but not the decision-making process itself. In the land-use model, for example, how will the farmers make their decision between the various possible land-uses. Similarly to the MAELIA model [183], we chose for the land-use change model to use a multi-criteria decision-making process. I present in the next section how it is coupled with the BDI architecture.

2.2.3 Multi-criteria decision-making and BDI agents

In the Mekong delta land-use change model introduced above [47], farmer agents have the installation of all the possible land-use types in their desire base; they thus have to decide which one will be selected. To this purpose the priority of each desire is computed using 3 criteria: the profit, the cost and the transition difficulty. Indeed, it is generally accepted that farmers tend to choose a production that maximises their profit, that minimises the cost and that are easy to implement. Detailed formula are presented in [47]. In this model a simple weighted mean is used to aggregate the various criteria. These weights represent the relative importance farmers associate to each of the criteria; they are defined as parameters of the simulation and evaluated in a batch experiment using genetic algorithm.

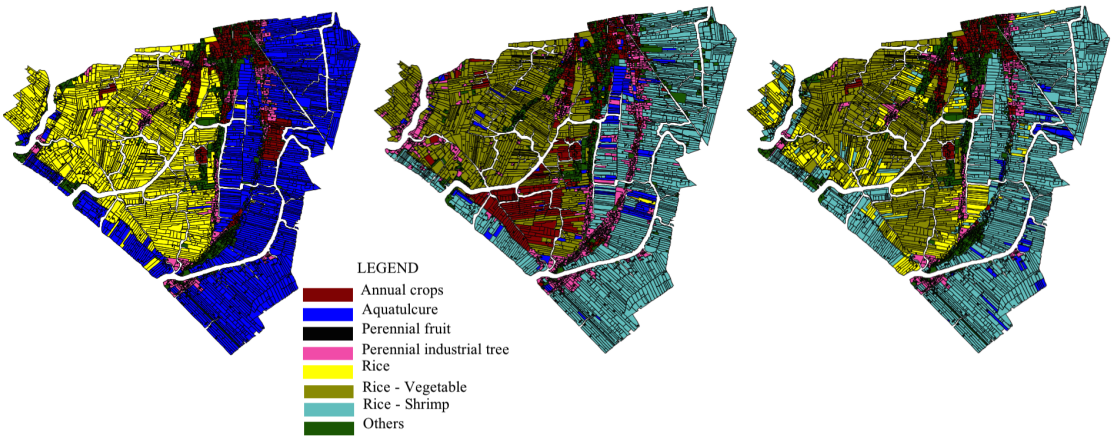

Figure 2.2: Application of the land-use change on the Binh Thanh village (Ben Tre province). Land-use for 2005 (left); land-use obtained for 2010 with the simulation (middle); observed land-use for 2010 (right) [47].

To evaluate the best values of these parameters we use the real data (of the land-use in the studied area) in 2005 and 2010. We use data of 2005 to initialise the simulation. The genetic algorithm will find the weights that minimise the distance between the actual and simulated data in 2010 (*c.f.* Figure 2.2). This distance is computed using the Fuzzy-Kappa coefficient

[203] that is well-adapted to compute distance between vector maps[3]. Simulations provide finally a similarity of 54.5%, which is rather good and shows that the model is able to reproduce in a relevant way the real dynamics. The mean duration of a simulation step was less than 0.6 seconds. This result is quite promising considering that we have more than 5700 BDI agents that can have many desires and that can activate many plans during the same simulation step (install production or information diffusion).

In addition we have shown on this same case study that numerous different decision-making algorithms can be used to model farmers' decision-making. In [192] we had compared three various architectures: a probabilistic model, a multi-criteria decision-making process and the previously presented BDI architecture with the multi-criteria decision-making process. The two last models provide quite close results (significantly better than the first one). The model using the BDI architecture is slightly better on the Fuzzy Kappa coefficient[4].

It thus appears that models of farmers using BDI or not BDI architecture with multi-criteria decision-making process provide pretty good and close results. It is very encouraging in the use of these complex models of reasoning and decision-making together. I argue that the model with BDI agents is more powerful and promising because it will allow us to integrate heterogeneity in farmers through imperfect knowledge and perceptions (providing to farmers wrong beliefs) they can get and heterogeneous social relationships with others. A solution to the latter issue of unreliable information is presented in the next section.

2.3 Trust to improve reasoning about information and other agents' reliability

As soon as agents can get unreliable information, coming either from communication with others or from their own perception, they should have the capability to reason about information reliability (and even about the source of these information). In this section I present an approach to deal with this issue based on the notion of trust[5].

2.3.1 Problematics

The issue of integrating heterogeneous agents, *i.e.* with different capabilities and (possibly contradictory) objectives and developed by different labs or company, in a single system is very old in the domain of Multi-Agent Systems. The development and standardisation of Agent Communication Languages (*e.g.* FIPA-ACL [76]) was a solution to make these various agents

[3] We first use it in [28, 179] to the same purpose in a urban growth model.

[4] I can also notice that, in preliminary works on the integration of BDI agents in the MAELIA model [183, 182], we went deeper by integrating a more complex multi-criteria decision-making algorithm using belief functions [171] for farmers' cropping plan decisions; we get also pretty good results.

[5] This work is based on Nguyen Vu Quang Anh's thesis [134] I supervised with Salima Hassas (Professor at Claude Bernard - Lyon 1 University), Richard Canal (Associate Professor and head of the IFI) and Frédéric Armetta (Associate Professor at Claude Bernard - Lyon 1 University).

interoperable. But in this context, nothing can ensure that transmitted information is right or accurate, that the agents are honest or that they are reliable in what they transmit[6]. To this purpose, various approaches have been proposed to make agents capable to reason on others' information, on their reliability or trustworthiness and to help them to decide with which agents to communicate or not and which information to consider reliable and appropriate to be used in their behaviour. Trust-based approaches [169, 201] give the capability to agents to compute a trust value about other agents, given information they have transmitted previously and/or their reputation.

This approach is classical in Multi-Agent Systems, where we can have systems with competition between heterogeneous agents. But even in Agent-Based Models, the introduction of such cognitive concepts in agents could help them to reason about information transmitted by other agents, that could be wrong or inaccurate, in particular if their perception can be flawed. As a consequence we developed a trust-based approach [74] allowing agents in a simulation to reason about other agents and information transmitted. In the sequel, I consider the trust in its classical quantitative sense in MAS: it will denote the probability with which an agent believes that another agent will enter in a beneficial interaction with it; so trust have a float value in $[0, 1]$.

2.3.2 Introduction of the TrustSet model

Hypotheses.

In this section we consider that agents can gather information about the environment (physical environment in the sequel, but it could be applied to the social environment too), store and communicate them. Some agents can be deceitful: we consider only that agents either gather wrong information (they are defective) or communicate wrong information they have gathered themselves (they are liars about gathered information). Agents will also communicate to others their trust about others: we consider that trust is built using its own assessment of gathered information and the trust values computed by others. The model is able to deal with agents that do not communicate their trust on others; in this case, agents can compute the trust only thanks to gathered information exchanges. But we argue that exchanging trust information will increase the speed to reach an accurate trust values of agents.

The TrustSet model.

The proposed model [135] is based on a data structure, named **TrustSet**, that is incrementally computed and updated by each agent given its interactions with other agents. The TrustSet is composed of the pair TrustGraph and TrustTable. The **TrustGraph** is a directed graph of agents (nodes), the edges carrying (direct or indirect) trust values. The *direct trust* is the trust

[6] These hypotheses are the bases of some so-called mentalist semantics of ACLs, that have been widely criticised (*c.f.* [82] for more details and discussions), in particular in systems with heterogeneous and concurrent agents.

the agent has computed on another agent, thanks to direct interactions with the agent by comparing its own information with information this other agent has communicated; *indirect trust* is a trust communicated by another agent. The graph is centred around the agent that is building the graph. It is the public part of the TrustSet as it will be communicated to other agents. The private part, named the **TrustTable**, is a simple table which contains the intrinsic trust value in other agents: the *intrinsic trust* is the trust value the agent will use to evaluate another agent and decide to communicate with it or not. It is computed from direct and indirect trusts. It is not communicated as we consider that each agent can have its own way of computing this intrinsic trust. For example, Figure 2.3 represents the TrustSet of the agent A. A has in its TrustTable the intrinsic trusts in all the agents of the TrustTable (T_{AA}, T_{AB} and T_{AC}). We can notice that in this model the agent is aware that it can be defective and thus computes its trust in itself too. In addition, A has a direct trust in B (DT_{AB}) and an indirect trust in C that has been transmitted by B (IT_{BC}).

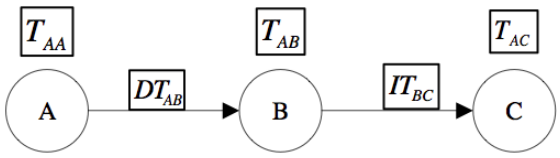

Figure 2.3: Example of TrustSet built by agent A.

Initially each agent creates its own TrustGraph with only itself as node. In the TrustTable it associates the value 1 to the trust in itself. During each communication with an agent, both agents exchange their knowledge about the environment (they have gathered or received from others) and their TrustGraph. They add these pieces of information to their knowledge bases and merge their TrustGraph. Each agent then computes again its TrustTable according to these new information.

For example (*c.f.* Figure 2.3), the agent A (with a TrustGraph limited to itself) meets the agent B (which has already encountered the agent C). We focus on A, but B will follow exactly the same steps. A stores in its knowledge base all the information transmitted by B. In addition A will integrate trust-related information coming from B. As the node associated to agent B does not exist in A's TrustGraph, it is added and linked to the A node[7]. As there is no value yet on the edge, the direct trust is computed by comparing the information A and B have collected on the same area of the environment[8] [137]. Finally, the TrustGraphs are merged [137]: from a topological point of view, all nodes and edges of B's TrustGraph are added to A's TrustGraph, the indirect trust values on edges are then merged: when an edge exists in both TrustGraphs, the value of the indirect trusts are merged using a weighted mean taken as weights the intrinsic trust that agent A has in the agents having provided the trust on the

[7] If agent A meets an agent whose associated node already exists in the TrustGraph, but but no edge between its node and A node exists, an edge is added.

[8] The direct trust is computed on the base of the gathered information by both agents: if they gathered the same or close information, the direct trust will be high, otherwise it will be low.

edges to merge [137].

The final step of the TrustSet dynamics is the updates of the TrustTable given the TrustGraph updates. First, the agent A computes its intrinsic trust in all the new nodes transmitted by agent B and then updates all the other intrinsic trusts. A's intrinsic trust T_{AX} in any other agent X is the mean of the direct and all the indirect trusts of any agent Y in X weighted by the intrinsic trusts of A in Y[9].

2.3.3 Use of the TrustSet by agents to improve their performance

The use of this intrinsic trust by agents is twofold: it influences their behaviour and helps them to compute information reliability. In [139], we propose various behaviour strategies taken into account the trust values. It allows agents to decide whether they will communicate with agents they meet (if they do not trust at all an encountered agent, the probability to communicate with it will be very low because it is a waste of time and resources or even harmful for the agent to exchange information with untrusted agents). In addition, physical behaviours can also be influenced by the computed trust: we consider that it could be positive for agents to build a cluster of trusted agents in which a self-organisation can emerge with various roles for agents (some can choose to stay far from trusted agents in order to increase the number of different information gathered and meet later to share) [139].

Each agent stores its own information about the world with for each information item a value gathered by an agent. As a consequence, for each information item, the agent has possibly several values, provided by various agents. To determine which is the right value for an information item, and what is the reliability of this value, the agent will use the trust it has in the agents that have gathered this value and the number of agents having gathered the same value. The idea is to balance the trust in an agent with the number of sources given an information item value[10].

2.3.4 Presentation of the application model

To illustrate the TrustSet model and how it can be used, we have developed an abstract model named "danger mapping". We consider an environment, unknown for the agents, with some danger areas. The objective of the agents is to build the more complete, precise and reliable map of the land using as little resources as possible. They can patrol in the environment and detect themselves the danger level around them thanks to their sensors and communicate (information about the land and other agents) with others. We consider that these sensors can be defective and provide wrong information about the environment. We assume that each robot agent has limited perception and communication ranges[11].

[9] See [134] for more details about the computation.

[10] See [134] for more details and precise formula about the computation of this reliability.

[11] This example was part of the AROUND project. This project has addressed the issue of developing a search and rescue multi-robot systems taking into account specific constraints of developing countries, such as for example

The model has been implemented using the GAMA platform (version 1.3): in a homogeneous space are located 200 danger areas (characterised by a danger level) and 50 robot agents. In a simulation step, they (1) collect data about the environment in their perception radius, (2) communicate (or not) with agents in their neighbourhood (defined by their communication radius), (3) update their data (TrustSet) if necessary and (4) move. To assess the trust model, three other trust models have been implemented: a probabilistic model [210], a fuzzy logic-based model [111] and the TrustNet model [169].

2.3.5 Results

The main parameters of the simulation are the ratio of defective agents, the trust algorithm (among the four ones that have been implemented) and various communication strategies (no communication, communication without taken into account trust...).

In order to evaluate the model, several indicators have been implemented:

- indicators about the performance of the system: the maximum (and the average) time for all the reliable agents to collect a complete and correct map of the environment;
- indicators about information gathering: at each simulation step, the known and unknown areas...;
- indicators about communication: messages sent/received by reliable/defective agents...

Results [135] was quite positive. For example, communication with trust computed using the TrustSet model was more efficient than with other tested strategies (848 step in average for reliable agents to get the real map, versus 2431 without communication, or 1553 with the probability trust model [210]). More precisely, it provides better results than other strategies for every percentage of unreliable agents from 30% to 90%. Finally, and it is correlated with previous results, the amount of information directly gathered by the agents themselves is the lowest when they use the TrustSet approach.

This trust model has been built independently from other works presented in this chapter as an independent and *ad hoc* GAMA model. It would thus be very interesting to implement in Java this trust model and provides GAML primitives to reuse the trust model in other models. In addition, as it has been done theoretically in [107], it could be integrated into the BDI architecture to provide a more complete model for cognitive agents. Finally, as it has been done in [131], it could also be linked to other cognitive concepts, such as emotions.

cheap robots that thus have a higher probability of being defective[40].

2.4 Emotion in artificial societies

Emotions, these reflexes that push human beings to make decisions quickly and without a deep and clear reasoning process, have been considered for a long time contrary to any rational reasoning process [174, 62]. Only recently, Damasio [58] has shown the key role of emotions in the decision-making process. From that time, computer scientists have started to be interested in taking into account emotions in decision-making algorithms [32, 155]. During my PhD, I have worked[12] on the modelling of emotions in a BDI logic [6]: we have formalised the triggering conditions of emotions classified by Ortony, Clore and Collins [147] and the links between these emotions.

I focus here on agents' emotions in a social context, as it is the case in agent-based simulations. As an example, in evacuation situations (and more generally in crisis management) stress and emotions have a huge influence on people behaviour [68]. As another example, in strategic interactions between people, even in a quiet and safe environment, emotions have a huge impact in the decision-making process: people want for example to avoid regret or have often an idea of fair interactions, driven by some bad feeling (*e.g.* guilt) when they do not respect their idea of ideality [98, 162].

In this section, I focus on 2 particular aspects of agents' emotions in a social context. First Section 2.4.1 presents the impact of **social emotions** on agent behaviour and choices: the respect of an ideality about social interactions is often driven by social emotions like guilt, relief, regret... Secondly, I focus on emotions in crisis situation and more specifically on **emotional contagion** in Section 2.4.2: particularly during crisis situation, interactions between people is not only physical or verbal; emotions are "exchanged" or more precisely the intensity of some emotions (*e.g.* fear) can be increased or decreased by agents' interactions.

2.4.1 Social emotions

This work takes place in the EmoTES research project[13] aiming at analysing and formalising social emotions and their impact on decision from a psychological, formal (game theory) and computational (social simulation) triple point of view. In this section I focus more specifically on a particular social emotion, the guilt, and how it could have led human beings to cooperative behaviours[14].

Contrarily to economic theory based on self-regarding assumption which assumes that cooperation appears only in indefinitely repeated interactions due to reciprocity benefits it

[12] This work was made in the context of Carole Adam's PhD [1] supervised by Andreas Herzig, Dominique Longin and Fabrice Evrard.

[13] The EmoTES ("Emotions in strategic interaction : Theory, Experiments, logical and computational Studies") research project has been funded by the French National Research Agency (ANR). Emiliano Lorini was the Principal Investigator.

[14] Computational work on a wider range of social emotions has been done in the context of the SocLab platform (https://soclabproject.wordpress.com/)[187, 186].

can bring, more and more evidence has been pointed out in favour of cooperation even in non-repeated or infrequently interactions with complete strangers [20, 75]. Following [41], we assume that acting selfishly, *i.e.* violating a fairness norm, has some negative effect on individuals, that is not only the interaction gain, but also a feeling of discomfort coming from *prosocial emotions*, such as guilt.

Mathematical model of guilt.

We consider the theoretical case study of an artificial society in which each individual has repeated one-to-one interactions with others agents and can learn from their past interactions. They have a moral value of fairness, and can evaluate the current or the expected situations after an interaction with relation to an ideal fair situation. The transgression of this moral value will trigger a guilt feeling, to which agents can have a different sensibility. Agents internalise these fairness norms in an utilitarian way: the expected utility of their actions will take into account the actual payoff, but also a possible loss due to fairness norm violation.

The general framework for the interaction is game theory (more particularly we consider in the simulation that agents interact following an iterated Prisoner Dilemma [157]). We first introduced formally the notion of normal form game with moral values (which extend the classical normal form game with a moral component) and then the concepts of guilt and guilt-dependent utility.

Definition 1 (Normal form game with moral values [84])**.** *A normal form game with moral values is a tuple* $\Gamma^+ = (N, \{S_i\}_{i\in N}, \{U_i\}_{i\in N}, \{I_i\}_{i\in N})$ *where:*

- $N = \{1, \dots, n\}$ *is a set of players;*
- S_i *is player i's set of strategies (*i.e. *all the possible alternative moves);*
- $U_i : \prod_{i\in N} S_i \longrightarrow \mathbb{R}$ *is agent i's personal utility function mapping every strategy profile in* $\prod_{i\in N} S_i$ *to a real number (*i.e. *, personal utility of the strategy profile for player i). This utility is often defined by the payoff matrix of the game;*
- $I_i : \prod_{i\in N} S_i \longrightarrow \mathbb{R}$ *is agent i's ideality function mapping every strategy profile in* $\prod_{i\in N} S_i$ *to a real number (*i.e. *, the ideality of the strategy profile for player i).*

In this context, we introduce the guilt emotion and defines it as the emotion arising from a deviation from an ideal behaviour: it can be expressed as the distance between the current situation and the counterfactual situation that could have occurred if the agent would have chosen the most ideal strategy, *i.e.* the strategy that maximises the ideality of the situation. We formally define it as follows.

Definition 2 (Guilt)**.** *Given a normal form game with moral values, player i will feel, after the*

strategy profile s is played, a guilt denoted by Guilt(i,s) and defined as follows:

$$Guilt(i,s) = I_i(s) - \max_{a_i \in S_i} I_i\big(a_i, s_{-i}\big) \tag{2.1}$$

where $I_i\big(a_i, s_{-i}\big)$ is the ideality of a strategy profile defined by the strategies of all the agents but i of the profile strategy s and the strategy a_i of agent i. The guilt is thus a measure of the effect of i's choices on the deviation from the ideal situation.

The guilt is supposed to provide a feeling of discomfort that the agent will try to avoid. To this purpose we assume that it will integrate it into its utility computation used to drive its choice. We thus define the guilt-dependent utility.

Definition 3 (Guilt-dependent utility)**.** *Given a normal form game with moral values, the guilt-dependent utility of the strategy profile s for agent i is defined as follows*[15]

$$U_i^*(s) = U_i(s) + c_i \times Guilt(i,s) \tag{2.2}$$

where $c_i \in \mathbb{R}^+ = \{x \in \mathbb{R} | x \geq 0\}$ is a constant measuring player i's degree of guilt aversion.

Definition of moral values. The utility $U_i(s)$ is determined by the kind of game and more specifically by its pay-off matrix. In these general definitions, the ideality $I_i(s)$ can be *a priori* any function. To define it, we choose to follow the utilitarian views of morality provided by Harsanyi [98, 99] and Rawls [162].

First Harsanyi [98, 99] has linked the morality to individual's utilities: he has argued that the moral motivation of individuals is to maximise the collective utility, that was defined as the sum of all the utilities[16]:

$$I_i(s) = \sum_{j \in N} U_j(s) \tag{2.3}$$

Later Rawls in [162] has proposed an alternative approach, arguing that the maximisation of the general well-being is linked to the maximisation of the utility of the agent with the lowest utility (it is a *maximin* approach). As a consequence, the ideality *à la* Rawls can be defined as:

$$I_i(s) = \min_{j \in N} U_j(s) \tag{2.4}$$

[15] To simplify the presentation, I consider here that the guilt sensibility is linear, but we could more generally consider a no decreasing function $\delta_i : \mathbb{R} \longrightarrow \mathbb{R}$ such that $\delta_i(0) = 0$.

[16] Once again I have simplified here the presentation: the collective utility is defined by the weighed sum of the utilities: $I_i(s) = \sum_{j \in N} k_{i,j} \times U_j(s)$, with $k_{i,j}$ is the empathy of i toward j. Here we suppose that every agent has the same and maximum empathy toward all the other agents, so $k_{i,j} = 1$ for all $i, j \in N$.

Computational model.

Given the proposed model of guilt and guilt-dependent utility, we have implemented an agent-based model (using the GAMA platform) in order to test various hypotheses, in particular the influence of the ideality (*à la* Rawls vs *à la* Harsanyi) and of the guilt aversion on agent behaviours. To this purpose, we choose to let agents interact one-to-one at each simulation step and play the Prisoner Dilemma[17]. Each agent can thus choose between only 2 strategies: Defect (D) or Cooperate (C).

Each agent is characterised by its *guilt aversion level* (a positive float number) and a *history* of past interactions, containing for each other agent, the total number of interactions and the number of interactions in which the other agent has cooperated. We provide agents with a way to reason about others' and learn their behaviour. To this purpose, we introduce in the agent a simple belief learning process known as the "fictitious play" or the Brown-Robinson learning process [44, 164]: an agent will assign to each possible strategy of its partner a probability to be chosen equals to the occurrence rate of this strategy in the past interactions. Of course the more both agents have interacted, the more precise the probabilities will be.

Initially, the simulation creates one agent for each guilt aversion value from 0 to 5 (a given maximum value), with a step of 0.1. This allows us to explore all the possible interactions between agents with different guilt aversions. At each step, agents are paired and each agent has thus to choose its own strategy, knowing its partner. To this purpose, it will compute the expected (guilt-dependent) utility of all possible situations and choose the one that maximises its expected utility. This expected utility will also depend on the fictitious play learning algorithm, which will associate a probability to each partner's possible strategies. Once every agent has chosen its strategy, the game is resolved: each agent is informed of the strategy of the other one, gets its payoff and updates its history.

The results are plotted in Figure 2.4 (with an ideality *à la* Harsanyi) and Figure 2.5 (with ideality *à la* Rawls). Axes contain the agents' guilt aversion value. Both figures thus plot the strategy pairs of both partners (depending on their guilt aversion). First of all we can observe a threshold in the guilt aversion value in both figures (1 for the ideality *à la* Harsanyi and 2 for the ideality *à la* Rawls). When the guilt aversion is lower for both players, they will both choose to defect, and this from the first interaction (the red square in both figures). Contrarily, when both guilt aversions are higher than this threshold, they will both cooperate (the dark green space in the figures).

We observe a difference between the two idealities in the cases where there is an asymmetry between both partners *w.r.t.* their guilt aversion. With the ideality *à la* Harsanyi, the agent with the lower guilt aversion will always defect and the other one will always cooperate. In contrarily, with the ideality *à la* Rawls, both agents will learn from previous interactions and converge either to a Defect-Defect or to a Cooperate-Cooperate strategy profile. To sum up,

[17] We choose the following payoff repartition: R = 2, T = 3, S = 0 and P = 1.

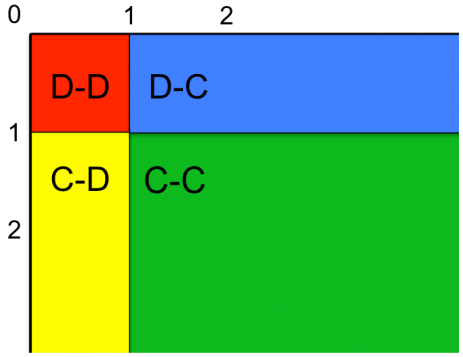

Figure 2.4: Fairness norm *à la* Harsanyi.

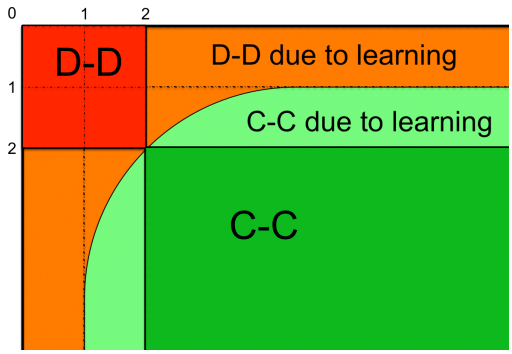

Figure 2.5: Fairness norm *à la* Rawls.

agents thus learn to not cheat or be cheated by the other ones.

We can conclude that agents with a higher guilt aversion will tend to cooperate and thus to have a higher payoff. In addition, from an evolutionary point of view, guilt aversion does play an important role in the suitability over time of fairness.

2.4.2 Emotional contagion and social behaviour in evacuation model

In addition to emotions linked to the reasoning about others and in particular to the impact of agents' actions on others, I present here a complementary work on the second aspect of emotions in a social context, that is the impact of agents' emotions on other agents and in particular the emotional contagion.

Basically in all the works presented below, models used as case study are quite similar. We have considered a case of a people evacuation from a closed space (this can be a shopping centre, a night club, or any building) in which a hazard has occurred. A lot of works already exist in the literature about similar problem of evacuation (from the social force-based model [103]). In the sequel, I present works to improve the agent behaviour model by integrating two essential components: emotions, with its impact on individual behaviour and on others' emotions, and social behaviours, in particular the influence of groups. Some researchers have indeed highlighted the fact that contrarily to what is often considered in evacuation models people do not evacuate individually and act selfishly, but have a social behaviour [69].

The environment of the model thus contains obstacles (walls all around the rooms but also inside the room), exits and hazards (*e.g.* fire, that can be dynamic and propagate in the environment). Agents are having a random behaviour until they are aware of the danger, which will make them try to evacuate the building by trying to reach an exit.

Emotion model.

Contrarily to previous work on the logic formalisation of emotion triggering conditions [6], I consider here only a single emotion that can be assimilated to the fear. We have considered that, in crisis situation, it is the most important and relevant emotion to take into account in the agent behaviour. In addition, we assimilate the emotion to its intensity (or fear level). We are interested not only to its triggering (due to danger perception or information transmitted by another agent), but also to its dynamics: (i) the emotion intensity has a tendency to decrease over time [202], (ii) direct stimuli (*e.g.* danger perception) or indirect stimuli (*e.g.* perception of high-level fear in others agents) can increase its intensity [115][18].

In addition in [115, 141], we have considered that agents behaviour is influenced by their emotion. As soon as they are aware that there is a danger in the building, agents will attempt to evacuate, moving toward an exit. Following [160], we have considered that people speed to move will increase with the fear intensity until a given limit. From this threshold, agents will tend to wander with a high speed in the space but without any objective, to simulate a panic behaviour.

As a result, we have shown that the evacuation with emotions is more efficient than without (in terms of number of people saved) [115]. This is obviously due to the fact that agents move faster when their fear level increases. But this is also due to the fact that this emotion contagion comes with an information transfer: observing an agent evacuating provides information of a danger to other agents. More interesting we can observe that evacuation is still more efficient even with the panic phenomenon, which induces a disordered fast move without objective and thus a less efficient individual behaviour.

In addition, we have added to this panic behaviour the fact that some agents can follow other agents to help them to evacuate [141]. As a consequence, in [178], we started to investigate the impact of groups in the evacuation in general (and not only in the case of panicked agents).

Impact of groups on the evacuation.

In [178], we made a first attempt to introduce groups into an evacuation process of a store. We consider that people agents in the model are either customers or leaders. Leaders are particular agents that have additional knowledge about the environment, in particular they know where is located the closest exit and the shortest path to reach it. They are representing typically security agents. We consider here that this role is determined before the start of the simulation and cannot change during it. Leaders will be the first ones to know that they have to evacuate and how to do it efficiently. Customers, that have detected a leader and do not know how to evacuate, will try to follow them.

We assimilate the leader and its followers to a group. But we consider that a group is more

[18] More details and in particular the complete formalisation can be found in [115].

than just a set of agents and that it has a feedback on agents constituting it: a group is a set of people helping each other's. As a consequence, the fastest ones will make the slowest agents move faster, but conversely the slowest agents will slow down the speed of fastest agents. We thus compute the speed of the group as the average speed of all of its members. The speed of each group member will thus become the speed of the group. The groups are also dynamic: new members can join at any time but others can leave too. It is particularly the case when the maximum speed of an agent is lower than the group speed. In this case, the agent's speed becomes its maximum speed; as it is moving slower than the group, it can thus be lost by the group and leave it.

As results, we have evaluated the survival rate and the average time of total evacuation. First of all we have noticed that the introduction of the groups is beneficial for agents' evacuation as it increases the survival rate, despite the fact that it has a negative impact on some agents by reducing their speed.

This preliminary work can definitely be improved in two main directions. First more flexibility can be introduced in the notion of group. For example, in an evacuation, groups are not limited to people following a leader: we can also have preformed groups (such as family of friends) that can thus merge with bigger spontaneous groups. Second, the social structure emerging from these groups should be used as the support of emotions contagion. Groups can have a smoothing effect on their members' emotions; but it can also be a catalyst of the emotion level increase.

To sum up, we can notice that, in all these case-studies about emotions in artificial societies, introducing emotions and social awareness in agents can, at the individual level, decrease individual efficiency (slow down evacuation speed (in groups), decrease pay-off in the Prisoner Dilemma-based interactions and possible panic behaviours) but this provides benefits at the global scale.

2.5 Conclusion and perspectives

This chapter has presented an attempt to go further in the integration of complex agents in Agent-Based Models. I have highlighted the benefits of a better integration of BDI agents and provided bases for a generic implementation under the GAMA platform. In addition, I have shown how to integrate other cognitive attitudes, such as trust or emotions and that they could be beneficial to represent human beings.

One of the most difficult issues with models embedding human being complex decision-making process is the issue of the validation, as it is often very difficult to get data about people decisions and reasons for their decisions. It has been partially possible in the case of farmers' cropping decision in the MAELIA model, as some interviews had been done before the development of the model. But they concern in general some strategic decision, *i.e.* decision

requesting a huge reasoning process. But as far as tactical or instantaneous decisions are concerned, in particular in evacuation, interviews are much rarer. We have started a modelling work on one of these dataset, the interviews after the Australian brushfires in the Victoria state [4]. We hope to be able to improve the quality of the models thanks to analysis of this data.

An important flaw (and thus a perspective) of the work presented in this chapter is that, although all the models have been developed on the GAMA platform, I have not been able yet to integrate them into a single framework. A possibility could be to consider the BDI architecture as the basic architecture for complex agents and to add to this architecture emotions and trust components[19]. But I argue that the challenge would be to have emotion and trust as separate components that could be used independently from any agent architecture and then plugged on the BDI architecture. For example we could imagine that the emotion module could be based on the generic notion of stimulus to trigger the emotion, which can match to new beliefs in case of a BDI-based agent or to any other agents' attribute otherwise.

[19] Mathieu Bourgais (PhD student supervised by Patrick Taillandier and Laurent Vercouter) has already started to integrate emotions and emotion contagion into the simpleBDI GAMA plugin.

3 Dynamical mathematical models and agent-based models

3.1 Introduction

A very interesting feature of Agent-Based Models is their capability to integrate agents representing entities at very different scales in a single model, *e.g.* models with both people and cities (containing people) as agents. It can be very challenging to manage their very different dynamics together and the interactions between agents at the various scales. In this chapter, I present an approach based on the coupling of models in different paradigms (in particular I focus on coupling Equation-Based and Agent-Based Models) to deal with multi-scale models.

As presented in Section 1.2.1, Equation-Based Modelling and Agent-Based Modelling are two among the most popular approaches to model complex systems. These two paradigms are very different in terms of objectives (descriptive vs generative approaches), formalisms (equations vs algorithms) and representation levels of the system (macro vs micro level). Whereas ABM aim at reproducing a system behaviour from interactions of its constituting entities, EBM describe the evolution of macroscopic variables of the system using equations. Among EBM approaches, I focus on differential equations (and more specifically on Ordinary Differential Equations models[1]) as it is a very classical way to model Complex Systems, in particular in epidemiology and ecology. I argue that, to deal with multi-scale agent-based models, coupling these two approaches can be very promising.

3.1.1 An example of EBM: the SIR model

As an example of EBM model based on ODE, I present here one of the simplest epidemic spread model, the SIR (Susceptible-Infected-Recovered) model [110]. This model was designed to reproduce the spread of diseases for which entities can be in three different infectious states: they can be *Susceptible* (*i.e.* they have not been infected yet and they could be), *Infected*

[1] Ordinary Differential Equations (ODE) refer to differential equations of only one independent variable (usually time t in ODE representing dynamics of systems), whereas Partial Differential Equations (PDE) are differentials of several independent variables (usually spatial variables x (y and z, depending on the number of dimensions in space) and time t).

(*i.e.* they have been infected and they can infect Susceptible people) and *Recovered* (*i.e.* they have recovered from the disease or died because of it; more generally they cannot infect people or be infected any more). The model formalises the disease spread at the population level by describing the evolution of the number of Susceptible (S), Infected (I) and Recovered (R) people over time with the equation system presented in (Equation 3.1).

$$\begin{cases} \dfrac{dS}{dt} &= -\dfrac{\beta}{N} IS \\ \dfrac{dI}{dt} &= \dfrac{\beta}{N} IS - \alpha I \\ \dfrac{dR}{dt} &= \alpha I \end{cases} \tag{3.1}$$

This kind of model is also often named a *compartment model*, as it describes the population as a set of compartments (here S, I and R) or stocks and the flows between these compartments (Figure 3.1). In the SIR model (Equation 3.1), each equation describes the evolution of the number of people in one state/compartment. For example, the instantaneous variation of R (in the third equation) is equal to a rate α of the number of Infected people (αI). We can observe here a flow of people from I to R stocks as the number of Infected people is decreased by this same term αI. The α parameter is the recovery rate (*i.e.* the rate of individuals flowing from I to R stocks) whereas $\frac{\beta}{N}$ is the infection rate (the rate of Susceptible people having encountered an Infected people that will become infected). The Figure 3.2 shows an example of the integration of the SIR system.

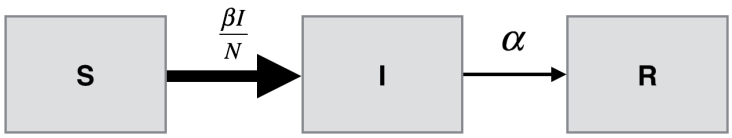

Figure 3.1: Stock-flow representation of the SIR equations.

It is important to notice that these ODE systems cannot be, in the general case, analytically solved, *i.e.* it is not possible to find the equations of $S(t)$, $I(t)$ and $R(t)$. As a consequence, to plot the solutions we have to use a numerical integration method (the classical ones are Euler or Runge-Kutta 4 [159] methods). The main idea of these methods is to provide an approximation of the value of $S(t+\Delta t)$ given $S(t)$ and the derivative dS/dt. The justification is that when Δt is very small, the function S between t and $t+\Delta t$ can be approximated by the derivation of S at t. This is the basic principle of the Euler method. Other methods are more precise, *e.g.* Runge-Kutta 4 computes 3 derivatives between t and $t+\Delta t$ [56].

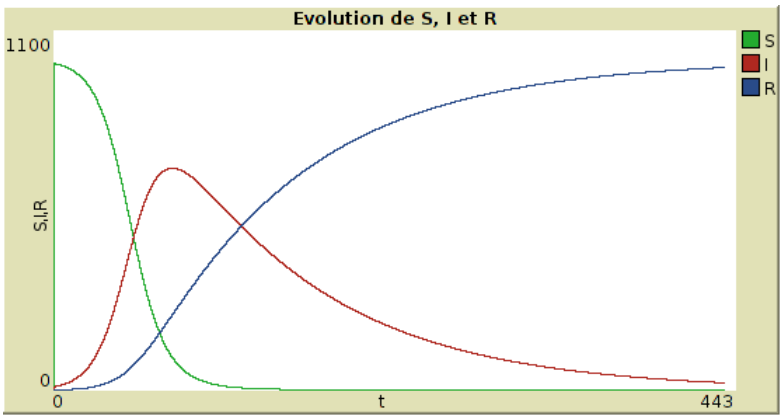

Figure 3.2: Evolution of the number of Susceptible, Infected and Recovered people over time with parameter values: $\alpha = 0.2$, $\frac{\beta}{N} = 0.5$ and with initial conditions: $S_{init} = 1000$, $I_{init} = 10$ and $R_{init} = 1$. The integration method is the Runge-Kutta 4 method with an integration step of 10^{-3}.

3.1.2 ABM and EBM: complementary or opposite approaches?

These two modelling approaches are by many aspects very different and they even appear opposite. As presented above, they do not consider the same representation level and have different purposes. But more importantly they do not make the same modelling hypotheses on the system. The Agent-Based approach makes almost no hypothesis on the system to study except the fact that its evolution can be represented and reproduced only by the behaviours of the entities constituting it and by their interactions [95]. In contrary, the Equation-Based approach is based on very strong hypotheses: the mathematical paradigm considers in particular a population with a huge number of individuals (in order that the dynamics can be seen as deterministic over the population). It is also seen as homogeneous, fully mixed in space with possible interactions between any entity of the system at any time step.

As a consequence, EBM and ABM seem to be two incompatible ways of modelling a system. Nevertheless, in this chapter, I argue that it is possible and could be very fruitful to couple these 2 paradigms in a single model. In particular, this coupling could help to deal with models integrating dynamics at very different scales: this is typically the case when we want to model a population with millions of people in which a disease spreads over the whole population and only few individuals move in the environment (*e.g.* this is typically the case of a population located in several cities with mobility between cities and disease spread in cities). To deal with individual mobility, it is necessary to have agents representing individuals; but as far as the disease spread over the millions of city inhabitants is concerned, this representation scale of 1 agent for 1 individual is not relevant. This dynamics could thus be represented using an equation-based system.

The chapter is organised as follows. In Section 3.2, I present quickly previous works tending

to combine and compare both approaches and in Section 3.3 a way to couple these two approaches practically and illustrate it on two examples. Section 3.4 presents a methodological guide for this coupling. Finally Section 3.5 concludes and opens new perspectives.

3.2 Existing works

3.2.1 Agent-based models

In the multi-agent-based modelling community, very few works were interested in equation-based models and their coupling with ABM. [151] or [49] have investigated the link between agent-based and equation-based models, but their point was to compare the two approaches in their representation of the same phenomenon. They highlight the difference of paradigm, scale of representation and way of thinking the phenomenon.

Similarly, [132] has shown that these two approaches are complementary to represent a complex system: agent-based models are well-adapted to understand "local causes" of the global evolution whereas equation-based approaches can deal with long-term evolution of the system. In addition, he went one step further by providing a methodology to induce an agent-based model from an equation-based model (and conversely), one of the main issues being to infer the parameters values to calibrate one model with the other.

3.2.2 Metapopulation models

Conversely in the EBM community, I am not aware of work integrating ABM. Nevertheless, in ecology the so-called *metapopulation* approach [97] has overcome one of the strongest limitation of the EBM approach, that is to consider the population as a whole: the population is split into several sub-populations into the nodes of a graph, the edges representing the possible migrations (*i.e.* flows of individuals) between nodes. The dynamics inside each sub-population is described by an ODE model whereas migrations are managed as instantaneous streams between nodes. This approach has also been applied in epidemiology [126, 55] to deal with the spread of a disease over a network of cities. It allows, for example, modellers to study how to control an epidemic by testing some mitigation strategies at the city level (such as the quarantine) or at the individual level (such as the influence of the risk culture). Nevertheless, such strategies remain limited as they cannot take into account heterogeneous individual behaviours (*e.g.* individual travel choices): individuals are not and cannot be represented in such models. This is the main feature of the agent-based modelling approach.

In all the examples presented above, the approach was to use either one or the other paradigm to study a system, but very few articles investigate deeply the coupling of these two approaches into one single model as we do in the following section.

3.3 Coupling EBM and ABM

3.3.1 General coupling principle

Due to the strong hypotheses of the EBM approach, the coupling between EBM and ABM can only be considered in cases of models integrating dynamics at different scales. This is typically the case when the model is composed of dynamics at the individual scale (*e.g.* mobility) and at the population scale (*e.g.* disease, rumour or opinion spread).

From a modelling point of view, I consider only the case of **an agent-based model integrating agents whose dynamics is controlled by an equation system**. This means that the global framework is the agent-based model running with its own discrete time scheduling. Individual agents have typically individual behaviours in the agent-based part of the model and can be aggregated in higher-level agents (*e.g.* cities containing a large population of individuals) in which a dynamics is applied at the population level thanks to equations. The connection between these two models is thus indirect in the sense of [27]: the two models affect different entities that are not at the same scale. Transformation functions are necessary to aggregate or conversely to extract entities in case of shift between scales. In order that some individual agents perform individual behaviours, such as move, they should be extracted from the city agent. This "extraction" should be coherent with the whole macro agent population: for example, if a population is composed of 70% of Susceptible, 20% of Infected and 10% of Recovered people, individuals extracted from the population should have the same probability distribution for the choice of their infectious state. Conversely, when an individual agent has to be integrated into the macro agent, in general 1 unit is added to the stock corresponding to its state and the individual agent disappears.

In addition, the coupling approach is integrative (again in the sense of [27]). The model contains agents at various scales, with extraction and integration rules. In the macro agents, the dynamics is described given an equation system. At each (agent) simulation step the equation system should be integrated by a given integration step, using a chosen integration method (*e.g.* Euler, Runge Kutta 4...). One very important part of the coupling will thus be to synchronise these two steps.

I present in the next sections two examples of models coupling EBM and ABM paradigms. The MicMac model is an abstract model allowing us to study the coupling and in particular to assess the divergence between an EBM and an ABM due to the release of the strong hypotheses of EBM. The second model is a practical application coupling dengue fever spread, mobility and control policies.

3.3.2 The MicMac model

General presentation of the model.

The MicMac model has been developed [24] and explored [26, 25] by an interdisciplinary researchers' team[2]. It started during a MAPS workshop. The MicMac model investigates the ABM-EBM coupling on a disease epidemic spreading over a network of cities connected by plane flights carrying people. The macroscopic dynamics of the disease spread follows a SIR model. The model contains two kinds of agents: `node` (representing cities) and `mobileGroup` (that can represent flights or any kind of transportation means). A `node` agent i contains a population of inhabitants split into the 3 compartments: S_i, I_i and R_i. The flows of inhabitants between these compartments is described by the SIR model.

Each city is linked to some other cities[3]. We consider that these links will be the support of the inhabitant mobility between cities: `mobileGroup` will land a given number of people from one city and bring them to a chosen target city. Each `mobileGroup` is thus characterised by the number of human beings in each infectious state it is carrying. Each `node` agent has a mobility rate g determining the number and frequency of the `mobileGroup` creation. The duration of the flight depends on the distance between the 2 cities. To this purpose a preliminary calibration step is necessary to calibrate the (SIR model) integration step, the `mobileGroup` speed and the size of the environment. As the flights are not instantaneous, an infectious dynamics also occurs in the plane during the flight. To this purpose we investigate two ways of integration of the SIR model: using continuous and step-based integration method (Runge-Kutta 4 in our case) or a discrete event-based Gillespie's method [87, 88].

The population conservation is fulfilled: the global population (in `node` and in `mobileGroup`) remain constant all over the simulation. The model interface is presented in Figure 3.3.

Coupling principle.

An overview of one simulation step of the MicMac model is described in the activity diagram of the Figure 3.4. A simulation cycle can be split into four main steps, each of them presenting a way of coupling EBM and ABM. The simulation runs until the end of the epidemic.

First the infectious states of the population in cities and planes is updated. This dynamics is managed by an SIR equation system. Each simulation step corresponds to one integration step of the ODE system: the numerical integration of the equation system and the agent-based model are thus synchronised. In order that this synchronisation makes sense, a preliminary calibration step is performed before each simulation: given data on an epidemic for a given

[2]This work is a collaborative work with Arnaud Banos, Nathalie Corson, Vincent Laperrière and Sébastien Rey Coyrehourcq. It has for a part be funded by the CNRS through the PICS MicMac project. I was the Principal Investigator of this project.

[3]The number of neighbours and more generally the topology of the connection graph is a parameter of the model.

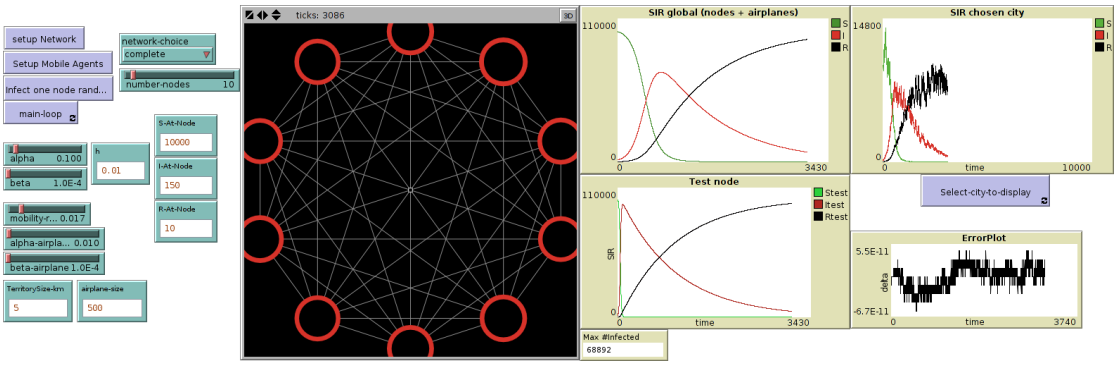

Figure 3.3: Graphical interface of the MicMac model. The user can choose the network topology, disease parameters and set up or run the model. He can monitor the simulation through plots of the global population repartition, *i.e.* the number of people in each compartment S, I, R. We can also observe the solution of a test node. Finally, an error measure between the initial population and the current total population is displayed in order to control the error in the simulation.

disease, we run an SIR model on a test node with the total population of the model merged into one single node. This integration runs until the epidemic ends (*i.e.* when the number of infected people is very small, lower than an epsilon threshold). We thus get the number of steps of the integration method and as we know the duration of the epidemics, we can compute the length of an integration step. This will have a consequence on the speed of the flights and thus on the transportation time.

Second, new `mobileGroup` agents are created. At each step, each `node` agent computes its population likely to move to another `node` (proportionally to its population and its mobility rate) and add it to its (float) stock of inhabitants who can move. As soon as this stock is greater than a plane capacity, the `node` creates a `mobileGroup` with a random target city in its neighbourhood. The population of the `mobileGroup` agent is picked from the city population. Given the capacity of an aircraft (hundreds of people), we consider that we have to move to the microscopic level and thus consider people individually. We thus extract a population of individuals representative of the `node` population. To sum up the algorithm, for each people in the plane, we determine its infectious state randomly, each state having a probability equals to its rate in the `node` population. Once the `mobileGroup` has been filled, the number of passengers is removed from the likely to move people stock of the `node`. The `mobileGroup` process is repeated if the stock number is high enough to fill another aircraft.

Finally, all the existing flights will move toward their target with a speed (computed during the calibration step). When a flight reaches its target city, it releases its population that is integrated in the city one and disappears from the simulation. If it contained some infected people, this can possibly make the disease appear in the city. This allows the disease to spread from city to city.

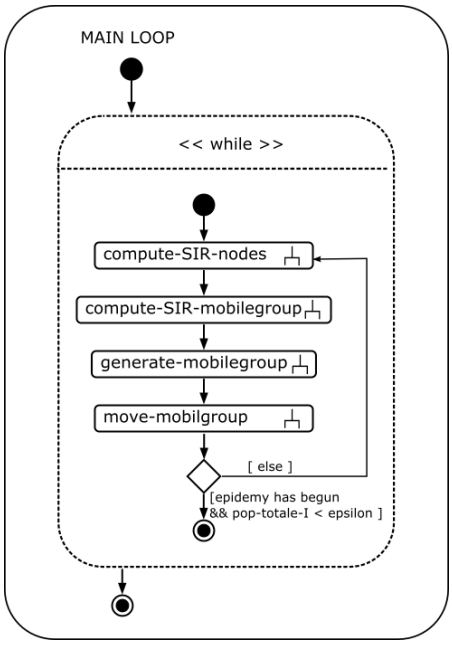

Figure 3.4: General activity diagram of the MicMac model.

Sketch of results.

[25] present an analysis of the MicMac model. We have focused on analysis the impact of the heterogeneity introduced by the agents on the results. To this purpose, we have compared the MicMac model with 2 models. The reference model is the SIR equivalent model: an ODE dynamics is applied on the whole population, mobility is thus not taken into account. The second model is a meta-population model (named MetaPop). This model contains the same `node` agents with the same dynamics as in the MicMac model. The difference is on the mobility part: the mobility is managed as instantaneous flows between neighbour cities. The number of inhabitants moving from one city to another one is the product of the mobility rate by the city stock in each infectious state and thus may not be an integer number. As a consequence, in the meta-population model, as soon as there is a touch of infected people in a city, all its neighbours will be infected in the next step. Contrarily, in the MicMac model the probability of contamination depends on the rate of infected people in the city.

The indicators used to analyse the simulations are: (i) $MaxI$, the maximum number of infected people in the population, (ii) $TimeofMaxI$, the simulation step corresponding to the maximum number of infected people and (iii) $Duration$, the duration of the epidemic. These indicators have been chosen because they allow us to characterise the evolution of a diffusion phenomenon (disease contamination) in the population.

The first interesting result is that both MicMac and MetaPop models can reproduce the reference SIR model on the whole population in the case of a complete network, homogeneous population and mobility rate repartitions over nodes, and instantaneous trips in the MicMac

model. From this state, we release one by one homogeneity conditions [25]. When we introduce heterogeneity in the initial population (and in particular with infected people only in one node initially) and trip duration in the MicMac model, both models of course differ from the SIR model, and have the classical characteristics of a progressive diffusion. The MicMac model presents a slower diffusion due to the fact that it is possible for an infected city to not infect its neighbours. In particular, we observe the following results:

- $MaxI$ of MetaPop > $MaxI$ of MicMac ;
- $TimeofMaxI$ of MetaPop < $TimeofMaxI$ of MicMac ;
- $Duration$ of MetaPop < $Duration$ of MicMac.

Concerning the node network topology influence of results, it has a greater impact on the MicMac model for the same reason as above. But results appear to be mainly influenced by the Average Path Length feature, independently of the kind of network topology.

3.3.3 Coupling in a model of dengue spread

I present in this section a second example of model coupling EBM and ABM. Contrarily to the MicMac model, this model has been built to tackle an actual issue: the link between the increase of economic trades and the dengue fever number of cases in the ASEAN[4] area. A more detailed description of the model and of its results can be found in [154][5].

Context.

The ASEAN is an association of 10 countries of the South-East Asia including Vietnam, Laos, Thailand or Myanmar among others. It aims for example at creating a unified economic area of free exchange and at increasing the economic trades. To this purpose, some economic corridor with improved infrastructures and simpler administrative procedures at the borders have been created. We focus on the East-West Economic Corridor (EWEC) crossing Myanmar, Thailand, Laos and Vietnam. In this corridor, we can observe a global increase of the intra-Asia countries trades (as expected) but we also observe a correlated increase of the dengue cases.

The dengue fever is a vector-borne disease: it is transmitted to human beings through infected mosquitoes (the vector) bites. Conversely, an infected human being can infect susceptible mosquitoes that have bitten her/him. The symptoms of the dengue fever are high fever, headache, pain behind the eyes, muscle and joint pains, nausea, vomiting, rash, and last for 2 to 7 days. But the real problems come with the severe dengue which can develop dengue shock

[4] ASEAN stands for Association of South-East Asian Nations

[5] This model has mainly been developed by Damien Philipon during his master internship [153] supervised by Marc Choisy and Alexis Drogoul. It has been used during the Journées de Tam Dao summer university in 2015 [51] and has been presented during the MABS 2016 workshop [154].

syndrome in 30% of cases and is lethal in 20% of the cases (but only 1% with hospital treatment) [154]. Dengue fever can be controlled by several means: guppy fishes to kill larvae, pesticide to kill adults, nets or repellent to prevent bites or vaccine to prevent human infection... It is important to notice that mosquitoes' life cycle (and thus dengue epidemic spread) is highly influenced by weather (in particular temperature and rains).

The aim of this model is thus to reproduce the dengue spread at the scale of the ASEAN East-West Economic Corridor given the increase of trades between countries and provinces thanks to the corridor opening. We aim at showing that there is a causality relationship (and not only a simple correlation link) between economic exchanges and dengue fever cases increases. We also aim at investigating the impact of the countries control policies.

General presentation of the model.

The considered reference system of our model is an area of approximately 1500 km by 400 km, which groups a selected number of districts and provinces in Myanmar, Thailand, Laos and Vietnam along the East-West corridor with a population of millions of inhabitants and much more mosquitoes [51]. The model starts in February 2004, with a time step duration of 12 hours. No time limit is fixed.

Given this huge area and the fact that we do not have data to locate spatially each individual human being and animal and that case data is available at the province scale, it is not relevant to model each individual (human being or mosquito) as a single agent as far as dengue spread is concerned. As a consequence, we have chosen to discretise the environment on a regular grid. Each `cell` contains the stocks of human beings and mosquitoes in each infectious state and the parameters of the epidemic dynamics. The infection process will thus be performed at the cell scale using an EBM.

Trades in the corridor will be simply represented by economic exchanges between big cities through truck flows. Each individual `truck` agent will be able to carry (infected) people or mosquitoes and release them at its target city. We have also added to this system `country` agents to manage health policies. In addition, the model contains several passive agents dedicated to integrating data in the model, such as `meteo station` dealing with temperature and rainfall data in the surrounding area, `district` and `province` agents to provide data about dengue case number and population, and `city` and `road` agents to support truck mobility.

At each simulation step, the daily data is updated (in particular for the weather station). Then countries apply their control policies on each cell. The epidemic dynamics is also computed in each cell: it will update the (mosquito and human) population in each infection states. As cell edges are not physical boundaries able to block the local dengue spread, an exchange of individuals is made by neighbour cells in order to allow the local dengue diffusion. Finally, the mobility process is executed to create new trucks, make them move to their target and

come back to their source. There is a small probability of epidemic interaction between trucks and cells. In addition, every year, countries update their global policy concerning dengue mitigation and the growth of country export is updated.

Coupling principle.

As presented above, we have three very different scales in the model: mobility at the individual level, epidemic spread at the cell level and control policies at the country scale. We focus here only on the coupling between the two first dynamics. As the dengue fever is a vector-borne disease, its spread is highly dependent of the vector (mosquito) population. Given the time scale (1 simulation step represents 12 hours, which is quiet small compared to the mosquito lifespan), we need to represent explicitly mosquito population and its life-cycle because, depending on the season, the number of mosquitoes can be extremely different and so is the number of new dengue cases[6].

At the scale of the cell we thus choose to use an ODE model for the disease dynamics. It is based on the model proposed by [123]. We consider 2 populations: the mosquitoes and the human beings (cf. Figure 3.5). Human beings can be in 4 states (Susceptible, Exposed, Infected and Recovered). Once a human being is recovered from the dengue, it cannot be infected by the same serotype. But as there are 4 serotypes of dengue, we consider that a recovered human can become susceptible again. Mosquitoes can be only in 3 states (Susceptible, Exposed and Infected) as they cannot recover. As shown in the Figure 3.5, the evolution of both populations are linked, as a human beings becomes Exposed because of a bite by an Infected mosquito, and conversely. For both populations, an individual is in the Exposed state when it has been infected, but cannot infect yet.

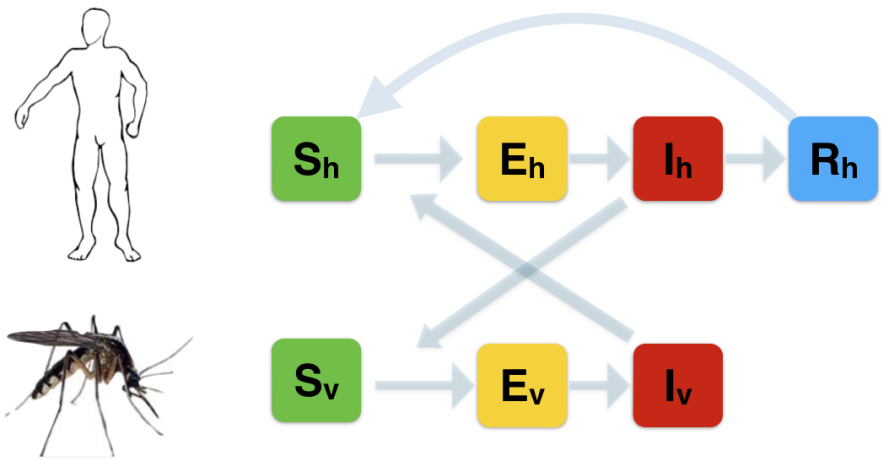

Figure 3.5: SEIRS-SEI compartment model for dengue contamination.

The evolution of this model is described using the following ODE systems, left column for the

[6] In addition, country policies will also have an influence on these populations.

mosquito dynamics (v is used for vector, here the mosquito), right column for the human beings one:

$$\frac{dS_v}{dt} = h_v(S_v + E_v, t) - \lambda_v(t)S_v - \mu_v S_v \qquad \frac{dS_h}{dt} = -\lambda_h(t)S_h + \omega_h R_h$$
$$\frac{dE_v}{dt} = \lambda_v(t)S_v - \upsilon_v E_v - \mu_v E_v \qquad \frac{dE_h}{dt} = \lambda_h(t)S_h - \upsilon_h E_h$$
$$\frac{dI_v}{dt} = h_v(I_v, t) + \upsilon_v E_v - \mu_v I_v \qquad \frac{dI_h}{dt} = \upsilon_h E_h - \gamma_h I_h$$
$$\frac{dR_h}{dt} = \gamma_h I_h - \omega_h R_h yc$$

with: $\lambda_v(t)$ (resp. $\lambda_h(t)$) denotes the transfer rate from Susceptible to Exposed for the vector (resp. human beings), υ_v (resp. υ_h) the transfer rate from Exposed to Infected for vectors (resp. human beings). γ_h (resp. ω_h) is the transfer rate from Infected to Recovered (resp. Recovered to Susceptible) for human beings. Given the time scale of the simulation, we take into account demography only for the vector: h_v is the emergence rate given a population and the time t and μ_v is the vector mortality rate. This emergence function comes from [149]: the population increase rate is a function of the temperature and rain given a double Poisson distribution.

At each simulation step, the evolution of mosquitoes and human being population is computed at the cell level, using the ODE system presented above. In addition, the mobility sub-model will control the move of individual trucks between cities: it creates new trucks to produce the expected trade exchanges, move along the corridor to their target city and when they reach their target they come back to their source. At the creation of the truck in a city, there is a probability (depending on the infected state of the cell on which the source city is located) that the truck brings infected mosquitoes that will be released in its target city.

Finally, every country has an annual budget to apply some control policies to fight again dengue spread [51].

Sketch of results.

The main result of this model is that we have been able to show that there is a causality relationship between the evolution of the economic exchanges and the evolution of Dengue cases (in the West-East ASEAN corridor area). Figure 3.6 shows the results in four different cases: with or without mobility and with or without control policies. The red and blue curves show unrealistic cases where policies would not be applied. The only goal of these scenarios is to show that without policies, the mobility has only an impact on the spreading speed of the disease, allowing the disease to reach its maximal incidence faster. The green and yellow curves are more realistic. The yellow one shows the incidence of the disease without mobility (only the diffusion in the neighbourhood is computed), which corresponds more or less to

the region before the setting up of the corridor. The green one represents the incidence of the disease with the setting up of the corridor, facilitating those outbreaks. We can observe the huge impact of policies application on Dengue cases number. We can also observe a speeding up effect of the mobility on the case number evolution.

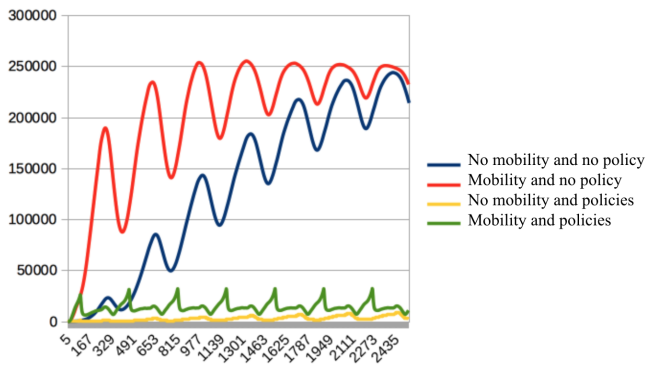

Figure 3.6: Results of different experiments: the figure shows the number of Dengue cases given the step number (one step is one day).

We also have tested the impact of the health policies synchronisation among ASEAN countries. In addition, this model has shown to be robust and flexible enough to be manipulated by non-computer scientist researchers[7]). They were able to implement and assess various coordination policy scenarios.

3.4 Methodological feedback

After having presented the general principle of the coupling and illustrated it on two model examples in the previous section, I now detail more particularly the methodological and technical issues we had to face in order to build models coupling these two modelling approaches. Indeed, such a model coupling brings out a number of fundamental problems, such as: (1) the articulation of different scales of time and space; (2) the preservation of the populations during transfers between scales; and (3) the interactions between a population considered as homogeneous (in the aggregated model components) and heterogeneous individuals (in its disaggregated components).

Finally by coupling ABM and EBM models, we do not have a deterministic model any more (as the EBM model is) but a stochastic one. The global model and its properties (that cannot be reduced to the properties of each of its components) should thus be carefully investigated by simulation through experiment designs, just as it is the case for any agent-based model (but not for ODE models).

[7] during an initiation to agent-based modelling in The Tam Dao Summer School in Social Sciences (JTD) www.tamdaoconf.com/ [51].

3.4.1 Articulation of different scales of time and space

Maintaining the spatial and temporal coherence in a model is always a challenge. Even in a "simple" agent-based model, as far as agents' mobility is concerned, the modeller should pay attention to the coherence between the expected duration of 1 simulation step (does it simulate 1 s, 1 mn, 1 hour or 1 day?) and the travelled distance in one step. When introducing a multi-scale aspect in the model by coupling EBM and ABM, this point becomes even more important. Indeed, in the chosen approach, the agent-based simulation controls the execution of the the ODE integration. We thus have to adapt either the integration of the ODE to the agent-based model execution, or the meaning attached to one simulation step.

In the MicMac model, we have considered as given the disease (and thus the attached infection and recovering parameters and the duration of the epidemics), the initial population in each city and each infectious state, the size of the space and the characteristics of the planes (in particular their speed). We thus have to compute the duration corresponding to a simulation step. To this purpose, we run (before each simulation) on an additional reference node containing the population of the whole network an SIR model until the epidemic ends. This gives us the simulation step corresponding to the end of the epidemic and thus the duration corresponding to an integration step h (and thus to the step of an agent-based simulation); this allows us to compute the distance a plane can cross during a simulation step.

In contrary, in the dengue model, the simulation step duration is an input of the simulation. As the consequence, the model has been developed such that the infectious coefficients are computed depending on this duration: all the infection and recovery rates have thus to be defined as a function of the integration step.

In both cases, the flexibility of the model relatively to its inputs is limited by validity bounds. Outside of these boundaries, the behaviour of the model is meaningless. For example, trainees of the Summer University of Tam Dao have have wanted to introduce the effects of climate change. The model had been designed to work from minutes to hours scale. Taking into account climate change would request to have a least a step duration of week, month or year. Such a duration for the integration step will lead to very inaccurate results (as the integration methods lose precision with the increase of the step duration). This case shows the limit of the approach where 1 simulation step = 1 integration step. A way to deal with this issue is to do several integration steps (that can thus be as precise as needed) during 1 simulation step.

3.4.2 Preservation of a population during transfers between macroscale and microscale agents

As we couple several scales in a single model, we thus need to introduce some transfer functions between macro scale and micro scale agents. The issues related to these functions are at least twofold: (1) the total population should be preserved; (2) some properties at one scale should be maintained at the other scale.

In the two previous examples, the models were quite similar *w.r.t.* this point: macroscopic agents (city or cell) contain a population as stocks (float numbers of people in each state) whereas microscopic agents are characterised by their own (infectious) state. The transformation functions should be very careful designed in the shift between integer and float values. The models require two functions: one extraction function from macro scale to micro scale and an aggregating function from the micro scale to the macro scale.

The transformation from micro scale to macro scale is immediate: each individual that should be aggregated in a macro scale agent will increase the stock value corresponding to its own state in the macro scale agent and disappear.

Conversely an individual agent can be extracted from the macro scale agent to act autonomously. To this purpose, the aim is to extract a single (or a set) of agents that are representative of the whole macro scale agent by using a proportional draw: for each individual to be extracted from the macroscopic agent, we randomly choose its epidemic state with a probability depending on the rate of individuals in the macroscopic agent with this state (for the MicMac model, the algorithm is detailed in [22]). An extracted agent will be removed from the macroscopic agent (the stock with its state is decreased by 1). For this step, it is very important to be careful in the management of the float values in order to preserve the total people populations (and avoid round errors for example).

3.4.3 Inclusion of a population considered as homogeneous and heterogeneous individuals

As far as we aim at letting interact such different entities as the ones manipulated in agent-based and equation-based models and integrating them in a single model, we have first to wonder in which case this is applicable. For example, we could consider that individual people and household have the same inclusion relationship as an individual people and a city. But it does not appear intuitive to apply an EBM on the household as it could be on the city[8].

As presented previously the hypotheses made by EBM are very strong and cannot be fulfilled in general in agent-based models. But in some multi-scale models, some entities can be close. For example, if we want to implement a model describing the worldwide epidemic spread and the impact of air traffic on it, we cannot simulate the 7 billions of people. But we can represent only cities with airports and planes as agents. In this case, cities are entities with a population of millions inhabitants, that will not be spatially located. As we are only interested in the disease spread, we are only interested in the number of infected people in the cities (and susceptible and recovered too). As a consequence, it appears particularly relevant to describe the evolution of the disease in the city using an ODE system.

I have discussed cases where it is possible to couple both approaches, now I focus on cases where it is particularly relevant to use it. Such an aggregated model can be very interesting

[8] We can link this fact with the classical problems of "small populations" related to the fact that small populations are highly sensitive to random hazards and have a high probability to disappear [118]

when we do not have (and/or do not want to gather) data at a microscopic scale, but only at the macroscopic one: we can simulate the phenomenon at a high level using an EBM[9]. In addition, EBM models have the advantage to not be sensible to population size in the integration process: millions or billions people does not bring a computation time increase, contrarily to agent-based models.

3.5 Conclusion and perspectives

To deal with models that should integrate dynamics at very different scales this chapter has presented a way of coupling Equation-Based Models and Agent-Based Models. The chosen coupling approach introduces some agents in the model that represent macro scale entities and describe their dynamics using an ODE system. I have illustrated this approach with two models before pointing out some methodological issues of this coupling. I have shown the benefits of this approach that allows the modeller to develop multi-scale models, by coupling ABM and EBM and thus combining the benefits of each of these approaches in a single model: EBM allows to describe a dynamics at a higher level, on large-scale population whereas the ABM is more precise and can deal with dynamics applied at individual scales.

As a perspective, I propose to go deeper in the combination of ABM and EBM. In the presented models, the use of ABM or EBM to describe a dynamics is static and defined *a priori* by the modeller: the mobility dynamics is always described using an ABM and the disease spread by an EBM. I argue that the choice of the paradigm could be dynamic during the simulation and can change on runtime depending on the observation scale of the user. For example, in the MicMac model, we can imagine being able to zoom in a city to observe more precisely the disease spread among individual people. This requires to be able to dynamically switch between EBM (used to describe the spread at the macroscopic scale) and ABM (used at the microscopic scale).

We[10] have started to develop a first prototype on a toy example: we have considered a model of fishery in Senegal with fish life-cycle and mobility between 3 spots (3 populations in a meta-population approach) and fishing boat mobility. The fish dynamics is described as follows: in the meta-population macroscopic part of the model, the evolution of the number of fish in each stock is described using a classical logistic function. At the micro scale level, the ABM model is used when boats have reached a fishing area: individual fishes are thus created in the fish area to be more precise on this part. We started to develop a generic aggregation/disaggregation approach that can be applied to any other model.

A second perspective is related to the dengue spread model: it has highlighted a limitation in the EBM approach, in the sense that it is aspatial: the dengue dynamics in a cell has no effect on neighbour cells. As a consequence, an additional diffusion process among cells has been introduced. In perspective I argue that agent-based can get a lot by being able to

[9] Chapter 4 will discuss solutions to this issue by providing tools to generate synthetic data at the needed scale.

[10] The model has been developed by Vincent Laperrière, Nicolas Marilleau, David Sheeren, Sébastien Rey Coyrehourcq and myself during the MAPS7 researcher school [83].

be coupled with not only ODE systems but also Partial Differential Equation (PDE) systems. These kinds of equations are typically used to model spatio-temporal phenomena involving diffusion (one of the most classical equation is the heat equation). We[11] have developed a first model of bark beetle (an insect communicating using pheromones) invasion using ABM for the beetle mobility part and PDE to model the pheromone diffusion (described as a PDE). The integration of the PDE is done on a discretization of the space (a grid) by applying (through a convolution product) a diffusion matrix defined by the modeller depending on the diffusions conditions (*e.g.* of the wind...).

[11] This work has been done through the GeoDiff project (funded by the CNRS), with Aymeric Histace (Professor at ENSEA Cergy-Pontoise), David Picard (Associate Professor at ENSEA Cergy-Pontoise), Nicolas Cazin (Master 2 Intern supervised by A. Histace and D. Picard) and Huynh Quang Nghi (PhD student, UMI UMMISCO, IRD) and has been published in [48].

4 Data and Agent-Based Models

One of the main advantages of Agent-Based Models, in particular to model socio-environmental systems, is their capability to integrate real data to make models much more realistic. This allows modellers to use them to tackle real-world issues and even to become an actual Decision-Support System, as it is the case for the MAELIA model.

The management of huge amounts of data in models still raises a lot of interesting questions (Section 4.1). As an improvement, I present in Section 4.2 a framework coupling agent-based modelling and simulation and Business Intelligence tools and in particular databases. Section 4.3 details the application of the framework on a real-case application. Finally I question the interest and benefits of the framework and show interesting new questions it brings in Section 4.4. I conclude this chapter by mentioning the issue of missing data and the solution of the generation of synthetic data and in particular of synthetic populations (Section 4.5).

4.1 Data and Agent-based models: challenges

Agent-based models, simulations and data are really closely related. I present in this section the challenges that are risen by the management of data in simulations, before introducing a way to address them.

4.1.1 Challenges

It is important to notice with [101] that **data is almost everywhere in the modelling process** as they are objective and precise views of a part of the reference system that is being modelled. In all the modelling cycles provided in the literature I am aware of (*e.g.* [64, 77, 86]), data plays an important role in most of the steps. Proposing the data-driven modelling approach, [101] have modified the logic of simulation of [86] (Figure 4.1) to highlight the importance of data, even in the modelling step. Data is really important in the model design in order to allow modellers to identify dynamics or patterns and to extract stylised facts that will be

implemented in the model. Data is also required at the initialisation of the simulation in order to create agents, locate them, give them attribute values... Finally, data about the reference system is also used for the validation (check of the similarity between the reference system and the simulation results) and calibration (tuning of parameters to fit with the reference system) of the simulations.

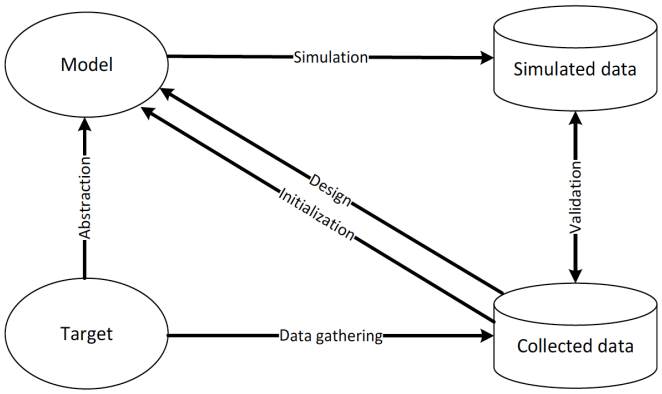

Figure 4.1: Data-driven approach [101] (modification of the logic of simulation of [86]).

As detailed in the previous chapter, the capability of ABM to integrate, in a single model, agents at very different spatial scales and acting at distinct time scales is a major strength of this modelling approach. As an example, the Dengue fever spread model (Section 3.3.3) mixes agents representing individuals with agents representing various administrative entities (provinces or countries). Similarly, in the MAELIA model, farmers perform each step daily activities (sowing, irrigation, harvesting...) but only once a year they make cropping plan decisions. These remarks highlight the necessity of having available **data at several scales** (or the ability to shift between these scales by upscaling or downscaling data).

In addition, many models, such as MAELIA, are designed for a huge area (*e.g.* the whole Adour-Garonne draining basin, covering all the south-west of France) with various kinds of entity and all the gathered data. But simulations are often only launched on a sub area. Thus when modellers want to **switch between sub areas**, they have to prepare at hand a dedicated dataset for each sub area.

Another specificity of Agent-Based simulations is their stochastic nature and the fact they often have a lot of parameters. As a consequence, their exploration requires a huge number of simulations and the output data produced by the simulations is very big. We can conclude that the **amount of data that has to be manipulated can be really huge**.

Finally, more and more modelling works have for objective to build not only agent-based simulations but also a **Decision-Support System** based on these simulations. This means that data has to be easily accessed, retrieved, aggregated or used for statistical analysis or data mining.

All these remarks make clear that a new approach in the way data is managed in an agent-based

modelling and simulation project is required.

4.1.2 Toward the integration of advanced data management tools

Despite the increasing need and use of data in agent-based simulations, a deep gap can be observed between the way data is managed in actual modelling and simulation projects and the way they could be with advanced tools. This has been the starting motivation to investigate the integration of data management tools and agent-based modelling and simulation platforms. This work has started during the MAELIA project and has been continued with Truong Minh Thai's PhD thesis [193]. First the need for management and organisation of a huge amount of data in big modelling projects drove us to the idea of using classical Database Management Systems to store input data and to retrieve it in an easy way. The initial idea was not only to store all the data in the databases but also to allow agents to interact directly with the databases from the simulations[1].

This solution addresses basically issues about the storage and organisation of data. It has also the advantage of splitting the responsibility of data preprocessing: the data provider is thus responsible of the database update and of the cleaning of the data.

But with Truong Minh Thai's thesis, we noticed that the characteristics and needs of agent-based simulations go beyond the scope of Database Management System and require advanced tools such as the ones provided by Business Intelligence (BI) approach. BI solutions are classically composed of a Data Warehouse, integrated data tools (ETL, Extract-Transform-Load tools) and Online Analytical Processing tools (OLAP tools) [105]. These tools are dedicated to store huge amounts of data, to organise them, to allow users to retrieve efficiently needed subset of data and to support the decisions by easing analyse (such as data mining, statistical or prediction analysis...).

Data Warehouse and OLAP tools have already shown their capabilities to deal with huge amount of data coming from simulations, such as in [173, 121]. As an example, in [122], authors have used such tools to store and analysis simulation results of a complex model including the coupling of biological and meteorological processes. In addition, they have also been used to build decision-support or forecast systems from simulations in various fields, *e.g.* logistics [71] or patient flow in hospitals [200].

The next section is dedicated to the presentation of the logical framework proposed (during Truong Minh Thai's thesis [193]) and its implementation in the GAMA platform.

[1] To this purpose we have extended the GAML language so that agents can directly retrieve data from the databases to initialise the simulation or insert output data, as presented in Appendix A.4.

4.2 Toward a framework to couple Agent-Based Simulations and BI tools

4.2.1 Architecture

The Combination Framework of BI solution and Multi-agent platform (CFBM) has been proposed in [194]. It aims at improving agent-based modelling and simulation platforms with their data management by coupling such a platform with BI solutions. The architecture is presented in Figure 4.2; it is important to notice that the figure illustrates the logical architecture of the framework that can thus be implemented with any agent-based modelling and simulation platform, database management system or BI solution. An implementation using the GAMA platform is freely available with the platform.

The framework is composed of the three systems detailed below: simulation system, data warehouse system and decision-support system.

Simulation system.

The **Simulation system** is the part of the framework dedicated to the modelling and simulation. The core component is thus the **Multi-agent simulation models**. With the help of the **Simulation interface** (responsible of everything related to the visualisation of the models and simulations), it allows the modeller to design and implement the models and to run simulations. These two components are classically implemented in any agent-based modelling and simulation platforms.

In order to link the simulations with the data sources, we made the choice of integrating directly in the models data-providers connected to databases, through particular agents (**SQL-agent**). Agents used as data proxies in a model can be found in many models, such as in the MAELIA model or in the Dengue fever spread model (Section 3.3.3) for weather data read from a file. During simulations, they can provide data to other agents through interactions or communications. The SQL-agents have the capability to connect to databases, interact with them to retrieve data, insert data in the databases or even manage the databases. More generally they can send any kind of SQL query to databases[2].

Finally, the simulation system is composed of various databases which can be logically split into two categories: **reality data** and **simulation data**. The former databases contain empirical data gathered from the reference system. They are typically used to initialise the simulation, but can also be retrieved at any step of the simulation to update some agent data. They will also be the reference data for the calibration and validation of the models. The latter databases will be typically used by agents to store simulation results; but they will also contain all the data about simulations (*i.e.* the metadata), such as possible alternative models, scenarios, parameter values... The main distinction between these two kinds of database is that Reality

[2]Technical details about the implementation under the GAMA platform can be found in Section A.4.

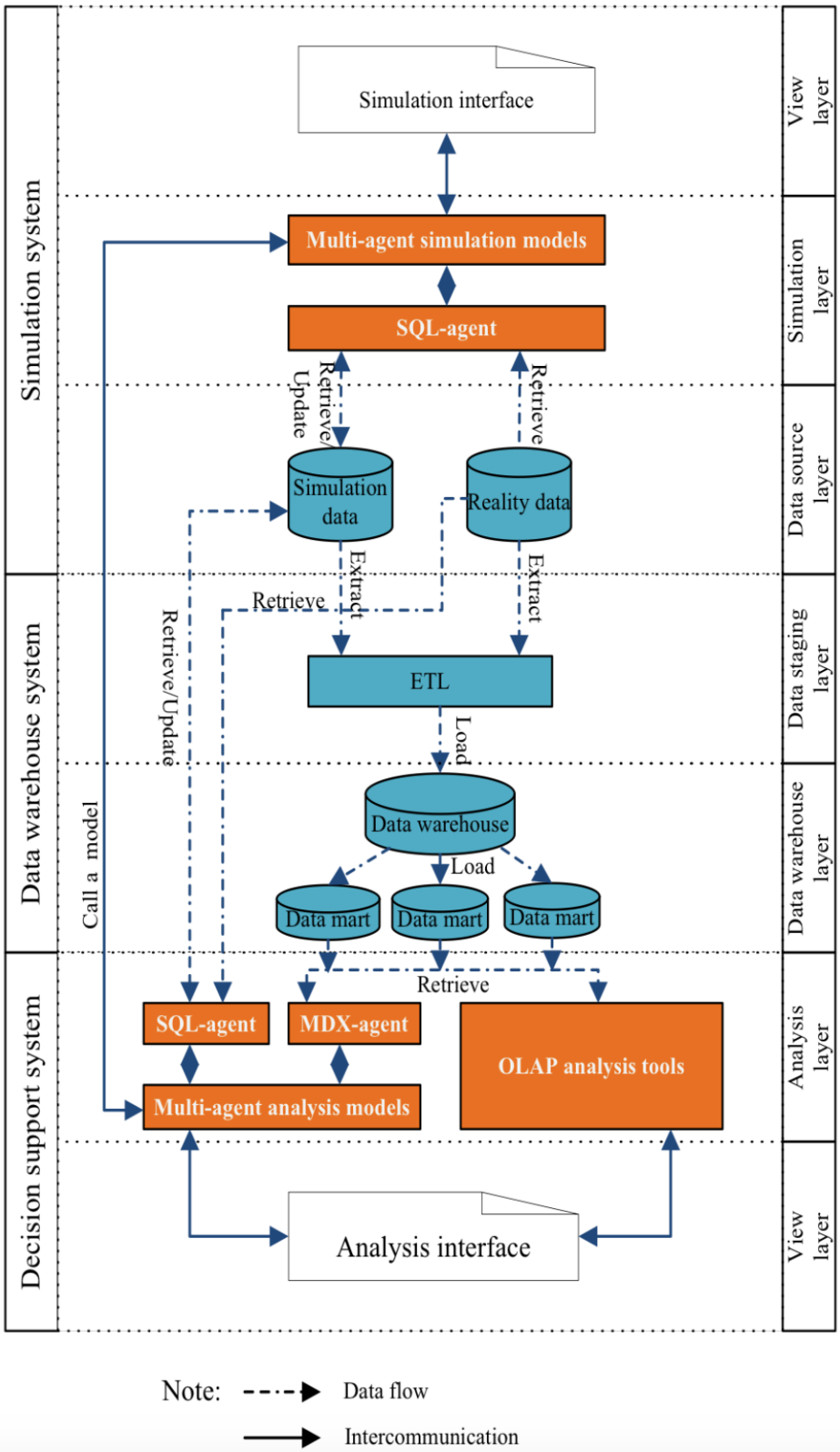

Figure 4.2: CFBM logical architecture [193].

data cannot be updated by simulations. In contrarily, Simulation data can be both retrieved and updated by agents in simulations.

This data source layer will be used as an interface between simulations and the other parts of the framework: data extracted from them will be inserted into the data warehouse and used for Decision-Support part.

Data Warehouse system.

The **Data Warehouse System** is important to make the link between the simulation system and the Decision-Support System. To improve the efficiently of analysis, it integrates all the data and provides tools to retrieve them efficiently (through Data marts). The Data Warehouse itself is fed by an **ETL** (Extract, Transform, Load) tool. The ETL extracts the data from all the various data sources, it transforms them, in particular to be compatible in terms of formats, and loads them into the **Data Warehouse**. A Data Warehouse (DW) is a huge relational database storing all the historical data loaded by the ETL, DWs have been designed to organise all the data of a company for example. Following [105, 193], a Data Warehouse is characterised by the following features: "it is subject-oriented" (it is a hinge between the key business concepts, such as customers, sales, products...), "it is integrated and consistent" (a DW is designed to integrate in a single location all the data) and "it shows data evolution over time and it is not volatile" (data in a DW are only growing and thus data has a time stamp to discriminate them in time). Finally **Data Marts** are defined to store subsets of the Data Warehouse data relevant for a given analysis in order to improve its performance.

Data from the various databases or from these data marts is finally used for analysis in the third system of the framework.

Decision-Support system.

The **Decision-Support system** provides to the modellers and even to users of the whole framework tools to support their thinking and decision-making processes.

The **Analysis interface** will thus allow human beings to visualise and interact with the results. In the CFBM framework, we have provided two kinds of tool to get these results. First we can use any **OLAP analysis tools**, to retrieve data from the Data Warehouse and provide and compute results. The CFBM framework has been designed to be modular and as a consequence it lets the users free to use any chosen OLAP tool. Second **Multi-agent analysis models** can be used to produce analysis results. The key ideas here are that, on the one hand it can be efficient and intuitive to design analysis tools in the same formalism as the one used to produce the data. On the other hand, [90, 89] have shown that agent-based models can be a very interesting tool to visualise data in a flexible manner. As a consequence analysis models will have to access required data. To this purpose they will typically have to integrate **SQL-agents**, that can get data from both Simulation data and Reality data of the Simulation System.

But these models can also take benefits from the whole framework and get pre-aggregated data from the data marts. Data marts are multidimensional databases and data can thus be retrieved only using multidimensional expressions-based queries. So we propose to integrate a new kind of agents, **MDX-agent**, able to communicate with data mart and retrieve data using dedicated expressions. MDX-agents will thus play the same role of data proxies as SQL-agents, but managing queries on multidimensional databases.

The CFBM framework as it has been presented in this section is completely generic and modular. In particular, modelling and simulation-related tools, databases, data warehouse or OLAP analysis tools can be implemented in a lot of different manners. In the next section, I detail an implementation of the framework using the GAMA platform for all the components related to agent-based modelling and simulation. But it remains independent of the chosen database and data mart management systems. This implementation is released with the platform.

4.2.2 Implementation of CFBM using the GAMA platform

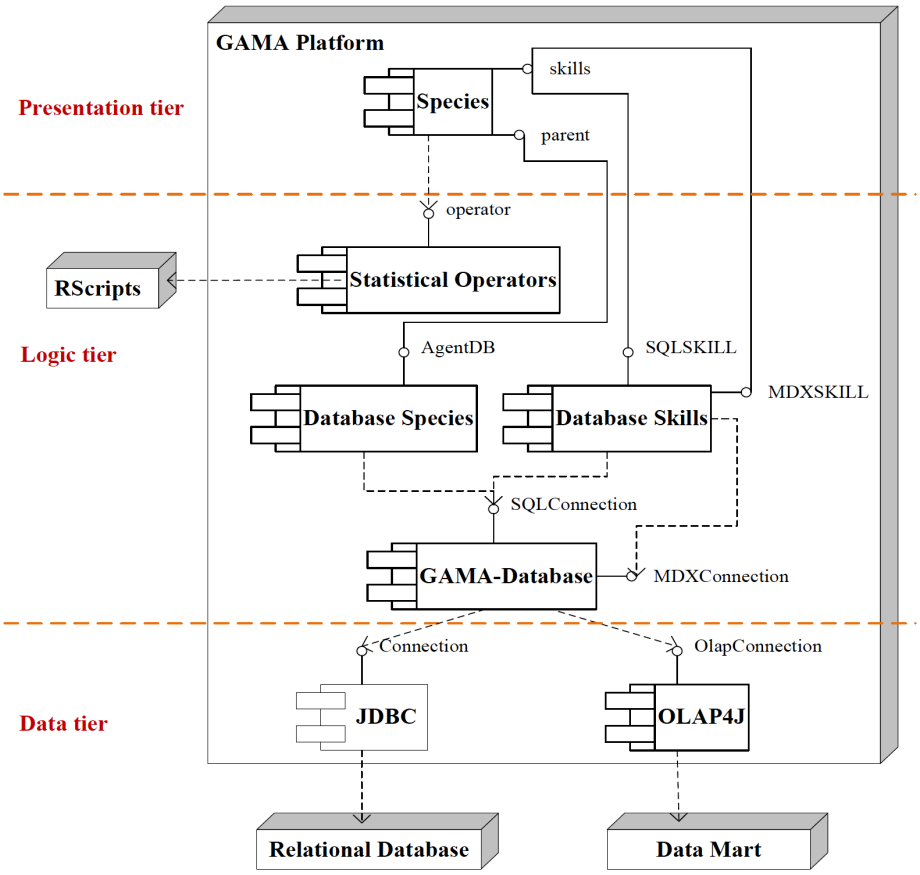

Figure 4.3: Software architecture of CFBM in GAMA [193].

First of all, the GAMA platform is perfectly adequate to implement the CFBM as it already provides both visualisation and simulation tools: models can be implemented, run and visualised in the platform. As a first attempt, we can also use its graphical user interface to display various simulation results (2D or 3D displays, charts of indicators...).

The general software architecture is presented in the Figure 4.3 as a UML component diagram. As detailed in the brief summary of the GAML language (Appendix A.1), the kinds of agent in a GAMA models are named `species` and built-in functions they can use `operators`. In order to integrate SQL or MDX-agents in models, the GAML language has been extended to allow the modeller to give to some species additional features in order that agents of this species can connect and make queries to databases using SQL (SQLSKILL) or multidimensional expressions (MDXSKILL). Modellers can attach these new features to any species of the model. It is very important as it allows modellers to focus in a first step on the design of their model and on the choice of the kinds of agent in it, only for modelling reasons relatively to the modelling questions on the reference system. The addition of the possibility to get data from databases can thus be added later, without modifying the logic of the model.

The so-called **Database Skills** is only a proxy linking keywords of the GAML language to Java methods. The **GAMA-Database** component is thus responsible of the interactions with the data sources. It relies on Java APIs (such as JDBC or OLAP4J) which provide the low-level primitives to connect with any kind of Database Management System (using dedicated connectors).

Finally, to improve the capability of the GAMA platform in terms of analysis tools, a bridge between GAMA and the statistical software R[3] has been implemented. R is the open-source software of reference for statistics with numerous and up-to-date extensions. A generic GAML operator has thus been added, allowing any agent to call any R functions or scripts with parameter values from the current simulation. This solution is not very efficient in terms of performance; but this weakness is compensated by computation possibilities provided by the R functions.

4.3 Application to a real case study: modelling of BPH invasion

In order to test and evaluate the CFBM framework implemented in the GAMA platform, we chose to use it on the model developed in the DREAM research team (Can Tho University) [198, 197]. It is an integrated model composed of a Brown Plant Hopper (BPH) invasion model and a surveillance network model and requires a huge amount of data [196].

[3] https://www.r-project.org/

4.3.1 Model

Brown Plant Hopper (BPH) is a rice pest attacking rice-growing fields [124]. Every year it induces in Asia an economic loss of millions of dollars for farmers [197], is a threat for the food security and is thus a very serious burden for these countries and in particular for Vietnam. To monitor BPH invasions in the Mekong river delta, a surveillance network has been built in 13 provinces of the delta [198]. The key element to measure the number of insects in an area is a light trap, *i.e.* a light trapping insects; they are daily counted by people and this measure is sent in the network.

The DREAM team is being conducting research to develop a tool allowing to monitor BPH break-out, growth and invasion, and to provide forecast tools based on a multi-agent model[4]. In particular, [197] has implemented a model aiming at optimising the location of light traps in the surveillance network. To this purpose, his model (named BSM for Brown Plant Hopper Surveillance Model) combines a surveillance network (providing historical empirical or simulated data on the number of BPHs) model with a BPH model. The latter has to reproduce the invasions and the evolution of BPH populations on the area monitored by the network. We mainly focus on this model, but only light traps from the surveillance network model will be considered here to calibrate the model.

The Surveillance network.

The Surveillance Network Model is mainly composed of surveillance devices (light traps) agents able to capture insects; the real number of insects in the surrounding area can be computed from the number of trapped ones. [198] has also introduced a way of correcting measures based on the measures of the neighbour light traps in the surveillance network.

The BPH prediction model.

Every year, Brown Plant Hoppers migrate all over the South-East Asia damaging and destroying rice on fields. These insects move following dominant winds possibly on a very long distance to find food. They can install and reproduce on rice fields but also on grass. Their life-cycle is quite short (maximum duration of 32 days) with 3 steps: egg, nymph and adult. Only adults can reproduce or migrate when local conditions are not appropriate any more. When migrating, BPH will be more attracted by some given area with favourable temperature, humidity and state of the crop on fields.

Given the huge number of insects in the area, the low number of data on the location of the insects and the size of the reference system (from at least a Vietnamese province to the whole Mekong Delta), it has been chosen to discretise the environment on a grid (just as for the Dengue fever Spread model presented in Section 3.3.3). A cell contains (and computes)

[4] For example, the model has been used to investigate the impact of the synchronisation of rice sow period on the BPH population.

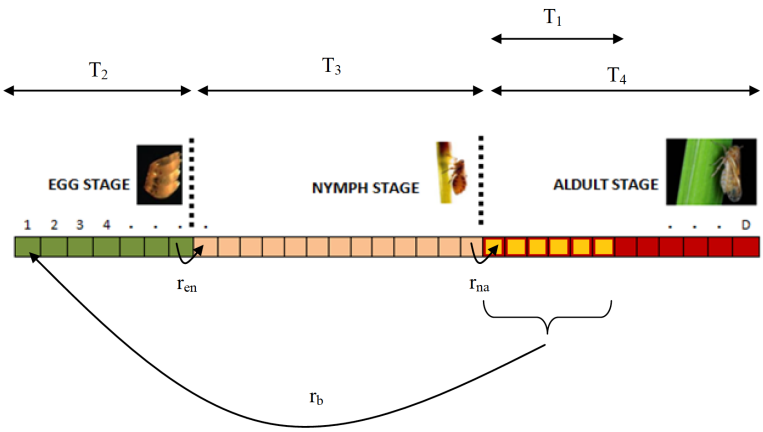

Figure 4.4: BPH density vector in cells and growth model [197].

information about its attractiveness and obstruction for insects, the number of adults and more importantly the density vector of BPH. The vector contains the population of BPH on the cell structured by age and updated thanks to the growth model.

The BPH density vector V (Figure 4.4) contains the number of BPH in the cell at each age: the ith element of the vector ($V[i]$) contains the number of BPHs that are i days old (a BPH can live until 32 days).

$$V[t] = \begin{cases} \sum_{k>(T_2+T_3)}^{k<=(T_2+T_3+T_4)} V[k] * r_b * (1-m) & t = 1 \\ V[t-1] * r_{en} * (1-m) & t = T_2 \\ V[t-1] * r_{na} * (1-m) & t = T_2 + T_3 \\ V[t-1] * (1-m) & otherwise \end{cases} \tag{4.1}$$

The density vector is updated using the growth model described by Equation 4.1. At each simulation step (1 day), each hopper gets 1 more day or dies (the mortality rate is denoted by m): so each element $V[t]$ gets the number of BPH in the previous element of V (*i.e.* in $V[t-1]$) that have survived. In addition, there are three special elements of the vector that are updated differently: they correspond to the shift between egg and nymph (only a rate r_{en} of the surviving eggs become nymphs), between nymph and adult (only a rate r_{na} of the surviving nymphs become adults) and to the laying time (each adult will lay r_b eggs). The other parameters of the model are the durations of each stage: a BPH remains egg during T_2 days, nymph during T_3 days and adult during T_4 days. The duration of the laying period is T_1. The growth model has thus 8 parameters whose value has to come either from data or be found by calibration.

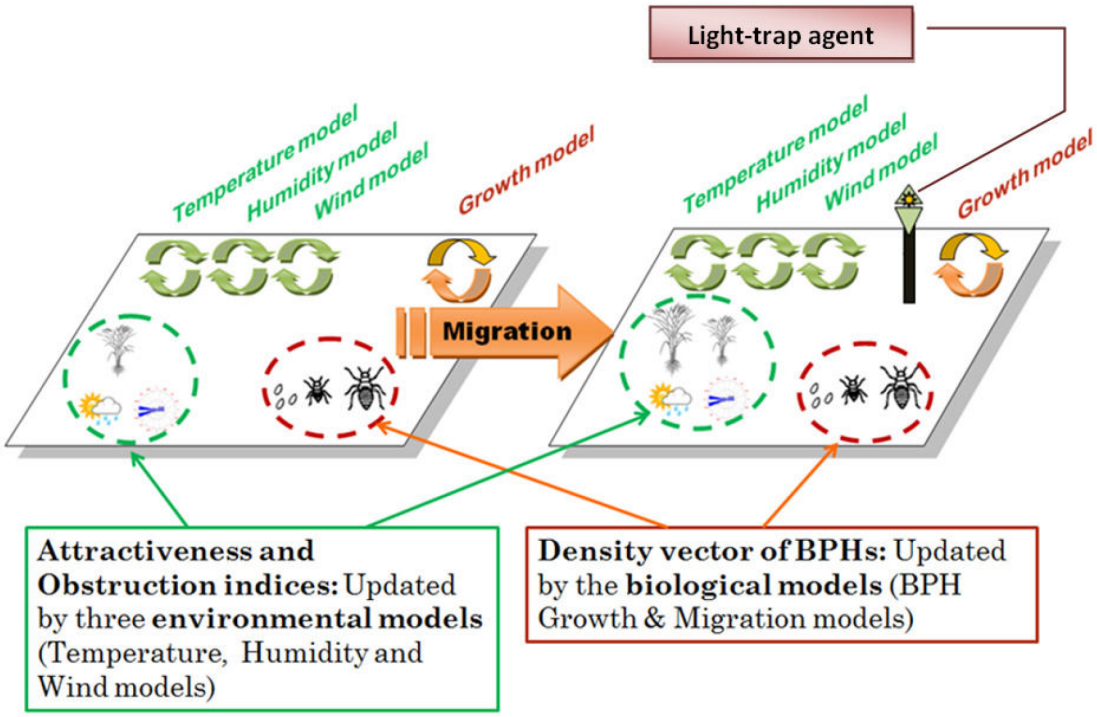

Figure 4.5: BPH prediction model and its interaction with the surveillance network model [197].

The global dynamics of the model is summarised in the Figure 4.5. At the scale of the cell, at each simulation step, the ecological models are computed: temperature, humidity or wind are updated and their new values are used to update the cell attractiveness and obstruction indices. In addition, the BPH growth model updates the BPH density vector. Given the cell attributes, the execution of the migration model will make move adult BPH from cells to cells, updating the density vectors of the source and target cells of the migration. The light-trap agent is able to monitor the BPH population in the cell.

Adequacy of the model with the CFBM framework.

The considered model has been designed and implemented in GAML independently of the CFBM. But we find out that it is typically the kind of model for which the framework has been designed: it is a spatial model including a lot of data of various types (maps, sampling data, ecological data...), it can be run on various spatial areas and scales (several districts or provinces of the Mekong river delta) and aims at being a Decision-Support tool (at various scales). We argue that CFBM can ease many tasks in this model from the data preparation to the data aggregation, analysis and processing [193].

4.3.2 Implementation of CFBM on the BPH model

As a generic implementation of CFBM in GAMA is already provided, using it on a particular GAML model requires only the few following steps. The considered model is thus the Simulation Model of the framework (Figure 4.2). The first step is to prepare the databases. The second one is to slightly modify the simulation models in order that it can get and store data in the databases. Finally analysis models have to be developed.

Database preparation.

The first step is to integrate data into the "Reality data" database and metadata in the "Simulation data" database. To this purpose, we first need to design the database schema, and in particular the various tables and the links between them (*c.f.* Figure 4.6 and Figure 4.7). In particular, the multi-level administration organisation (with regions, provinces, districts and small towns) is represented and spatial data is stored for each level. In addition, the data schema allows to easily associate data with their respective spatial level: wind data is available at the region scale, whereas land-use and light trap data is at the small town level.

The simulation data databases will include the simulation results (SIMULATIONDATA_SM, *i.e.* the average density of BPH at the scale of small towns, and SIMULATIONDATA_LT, *i.e.* the number of BPH in light traps). All the result data is linked to the simulation that produced them through the *replication* number of a scenario (SCENARIO_MODEL) associated with a model (*c.f.* Figure 4.7). A scenario is simply a set of parameter-value pairs. All the tables related to model and scenario are generic and could be reused for any model[5].

More generally, the database organisation that should be built has to include three main kinds of table: (i) the empirical data tables (containing all the data on the reference system), (ii) the metadata related to the model and scenario (it will in particular contain all the combinations of parameters values of an experiment design), and (iii) the result data (that will contain the result values of all the indicators). The result data table will be at the interface between the 2 other kinds of database: it is attached to a replication number (*i.e.* the identifier of the simulation that produced them) but also to empirical data tables (*e.g.* a light trap measure has been gathered in a given small town). This allows to address the issue of the level at which the data has been gathered and ease the shift between the scales.

Update of the simulation models.

The main and only adjustment to make in the model is related to the data access: the reading and writing in files have to be replaced by retrieval from and insertion in the databases. I can mention here two typical uses of data in an agent-based model.

[5]Related works can be found in [122] where authors propose a multidimensional data schema of a data warehouse for storing and analysing simulation results.

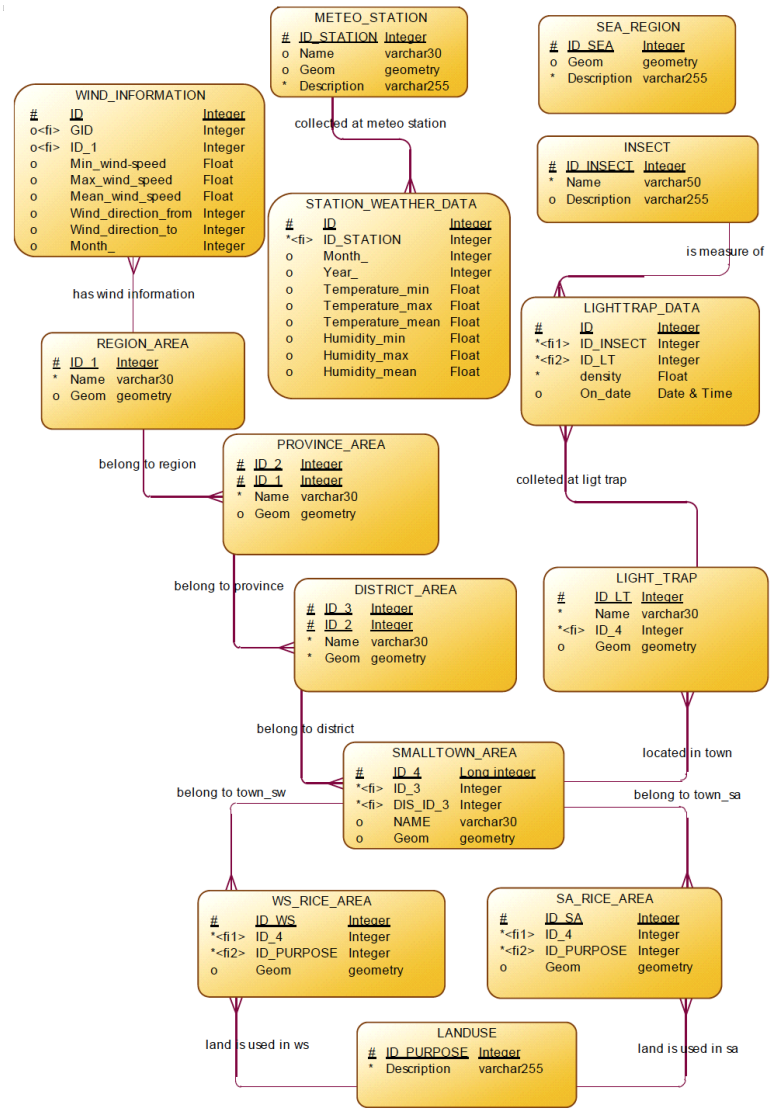

Figure 4.6: Empirical data databases schema [193].

First some specific agents read data from files to update their state or write data in files to save results of the simulations. We have used here to possibility of the GAML language to give new capabilities to agents, simply by adding to their species definition an additional skill (*c.f.* Appendix A.1 for more details). We thus have developed extensions (SQLSKILL and MDXSKILL) to the GAML language to have the possibility to attach database interaction capabilities to any agent. The data retrieval from the databases or their insertion will thus replace straightforwardly the data reading from a file or the data writing in a file.

Second, data files are used to initialise agents when they are created. Once again, the GAML language has been extended to give the possibility to create agents directly from a SQL request

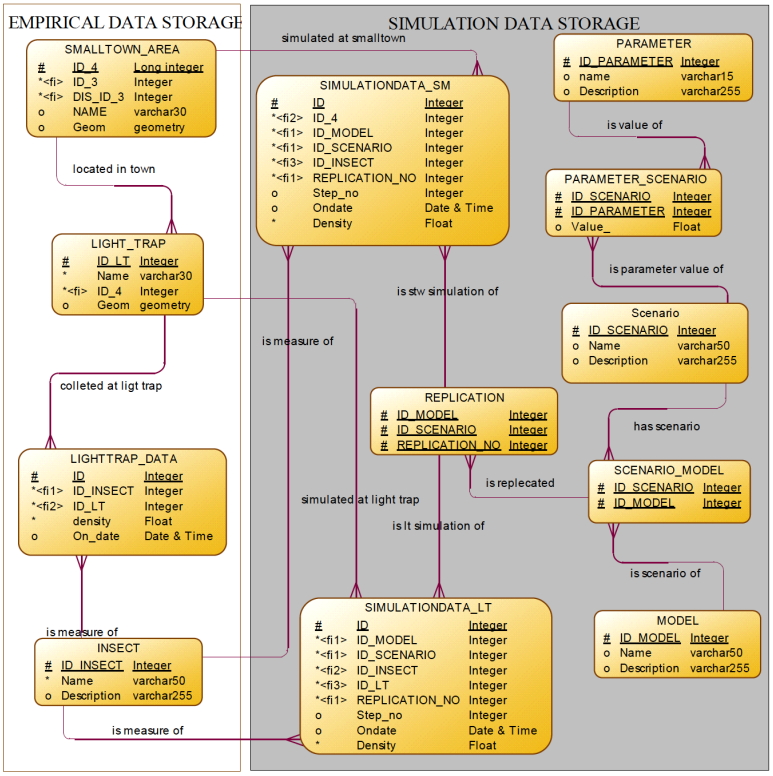

Figure 4.7: Simulation data database schema [193].

results. The modification of the initialisation of the simulation will thus be minimum[6].

Analysis models.

Finally, multi-agent analysis models can be implemented in order to explore the model behaviour or to validate and calibrate it. We can in particular take advantages of the databases to store the various scenarios, run simulations, store results and compare the results with empirical data for validation purpose. In the next section, I give an illustration on the BPH model.

4.3.3 Application to the calibration and validation of the BPH growth model

The calibration of the BPH model will be limited to the calibration of the parameters of the BPH growth model. The other processes are initialised thanks to input data. As detailed above, the growth model has 8 parameters ($T_1, T_2, T_3, T_4, r_{en}, r_{na}, r_b, m$). Among all these parameters, [130] provides the following values: $T_1 = 7$ (days), $r_{en} = 0.4$, $r_{na} = 0.4$ and $r_b = 360$ (eggs). We thus have to deal with the 4 remaining parameters. We know that the possible values for these parameters are the following ones: $T_2 \in [6,7]$, $T_3 \in [12,13]$, $T_4 \in [10,11,12]$ and

[6]Details of the modifications are provided in the Truong Minh Thai's thesis [193].

$m \in [0.15, 0.2, 0.25, 0.3, 0.35, 0.4]$. We thus have 72 possible combinations of parameters[7].

The simulation data database (*c.f.* Figure 4.7) thus contains the 4 parameters and each possible value. A scenario is defined as a set of parameters associated with a value. An experiment design thus corresponds to a set of selected scenarios. The execution of a set of scenarios starts by the load of the parameter values from databases, the launch of the simulation and the storage of results (BPH density and number) in the database, associated with the model, scenario and replication identifiers.

Once the replications of the simulation model have been executed, the analysis model can be executed to assess the (scenario applied on the) model, for example by comparing its results with real-data. The analysis model should thus retrieve data from the databases of empirical and simulated data and compare them using given indicators. To this purpose, various indicators can be implemented or reused. For example, to compare map, the fuzzy-kappa operators had been implemented (*c.f.* Section 2.2.2). For the BPH model, two indicators have been introduced: the Root-Mean Square Error (measuring the distance or error between 2 sets of values) [207] and the Jaccard index (measuring their similarity) [106, 142].

We can then define a fitness condition to characterise the acceptable scenarios. As an extension of the database schema we can also introduce a table to store the best or acceptable scenario. From these scenarios, we can also for example produce new scenarios (by inferring possible parameter values that could provide better results), that will be executed on the BPH model and assessed. For example it could be particularly interesting to be more precise on the mortality rate in some area of the parameter space. This use of the CFBM is thus quite promising as it allows modeller to design analysis model that can work not only on the models results but also on the scenario definition themselves.

4.4 Discussion about the CFBM framework and its use

4.4.1 Advantages and limits of the architecture

To conclude on the CFBM framework, I present in this section its many advantages [194]. It provides a logical architecture but also an existing implementation into an agent-based modelling and simulation platform. In addition, the architecture is fully modular: its implementation under GAMA allows modellers to use their own Database Management Systems, Data Warehouse or OLAP tools. But even if modellers have designed and implemented their model in GAMA and set up the whole data management system, the framework allows them to implement the same model on another platform (or even from scratch in a generic programming language). After having linked this platform with the framework, they can reuse all the databases. This

[7] We can notice that the set of values of m is only a discretization of the possible values of the parameter, chosen to pave the possible values space and to be able to define an exhaustive experiment plan. Nevertheless the framework introduces the possibility to define new scenario, which could be used to be more precise on some interesting areas if needed.

could be particularly interesting to compare results of various implementations of the same conceptual model.

In addition, its deep integration into a particular modelling platform allows modellers to develop their own model independently of their choice of using or not the framework, letting them design freely the model. The integration of the model in the framework, in a further step and without changing its design principles, is eased by the extensions of the GAML language.

Nevertheless, even if it is always possible to use this framework, it (but also any solution based on the management of data using databases) is not always the best solution to deal with data in a model. It requires additional work and skills in terms of database preparation, but also it is not really efficient for small amount of data. It is thus not relevant for small models, with few data, where benefits of the framework disappear under the execution time to get data.

But in addition to its benefits in terms of efficiency to manage data, it also raises some new interesting questions about the relationship between data and simulations.

4.4.2 New questions and challenges

Managing various versions of the models.

Another possible use of an integrated data storage is the possibility to manage, compare and analyse data from various versions of the same model. The modelling process often requires the development step by step of the model(s). It could be interesting to keep all these versions and their results. It could allow modellers to avoid for example regression bugs (by testing that an expected behaviour, represented by the value of an output, is fulfilled in later versions of the model). In addition, it could allow modellers to compare results of models with and without a given dynamics or with various implementations of the same process, in order to assess its impact on results or to compare various implementations of the same dynamics. Modellers could thus go further than the simple versioning tools that register only modifications and differences between versions. Here model versions could be attached with results giving a metrics of comparison between the versions.

Relationship between data.

A critical advantage of using databases (and more generally BI solutions) is that they allow modellers to store all the kinds of data related to their models and simulations in a common place and thus compel them to think about the nature of the link between their data. Obviously these data storages will contain all the required data about the reference system (*e.g.* weather data, administrative vector map, number of disease case per country over time...). But modellers will also store all the results of the simulations (for example the number of simulated cases of a disease over time computed by the simulations). These simulation results should be linked to input parameters values applied on a given model (or even a version of a model in

the sense of the previous paragraph). Thus storing metadata (models, with their versions or input parameters values...) in addition to (empirical and simulated) data can obviously help modellers to organise their modelling process.

But it is also interesting to note that the distinction between empirical real data and simulated data in the role they play *w.r.t.* a simulation can become fuzzy. As mentioned in the introduction, modellers have often to face a lack of data at the right time or space scale. This issue could be addressed by using an appropriate model to produce the needed data. The results of the simulations, stored in the Simulated Data databases, can thus be used to initialise another model, if they have been assessed suitable. Integrating all the data and metadata in a single structure allows the modellers to switch easily between real and simulated data (even as input of some simulations).

4.5 Conclusion and perspectives

In this chapter, I have discussed a way to manage data efficiently in agent-based models and simulations using the CFBM framework. I have detailed its logical architecture, described an implementation in the GAMA platform et illustrated its use on a real-case model.

The close relationship between (agent-based) models and data is also often one of the main drawbacks of the agent-based approach, as it requires a lot of data at a given time and space scale to be used. Even if we have now more data that we can manage, we are always lacking the right data for a particular model. A way to deal with this issue is to generate synthetic data, statistically coherent, to initialise our models.

Introduction to the GenStar project. We have started in 2014 the research project named GenStar (standing for Generation of spatially and socially structured synthetic populations for social simulation) and funded by the ANR. It aims at generating data to initialise the populations of agents in the models, *i.e.* at generating synthetic populations. We thus aim at providing a generic library able to generate a synthetic population that could be used to initialise the population of agents in any model. This population will be generated based on the statistical and macroscopic data (*e.g.* census) available in the interesting research area. One of the motivation to start the project was that such a kind of data does not have the same resolution and quality in all the countries. When we want to build a model in a country like Vietnam, data is much more sparse than the ones in USA. We thus have a need for such a generator.

In addition, the available data is often very macroscopic, whereas we can need them at a more or less microscopic scale. Moreover, we can have either a sample of the population, that we have to generalise, or only statistical data, from whose we have to produce the population. The provided library should thus be designed to manage upscaling or downscaling of data, with or without sampling generation methods.

Finally, when we want to generate a synthetic population of agents representing human beings, several kinds of attribute can be generated. First, we generate their individual characteristics (*e.g.* age, sex, incomes...). Second, such a population is in general spatially located: the second step of the generation is thus to generate a location for the agents (that should be coherent with the previous attributes: members of a household live generally together). Finally, agents are in a social environment. This is often represented by the various social networks of their family, friends or colleagues... These three components should be taken into account to generate a good synthetic population. We[8] focus mainly on the social network generation part.

Generation of synthetic social networks. The need for synthetic data generation methods is even higher for the social network component of the population, as we have less data on this part. Indeed, it is really difficult to get actual data about the social aspect. Many network generators have already been proposed in the literature. In [10], we made a review of the algorithms that have been used to generate synthetic social networks in agent-based models. The considered dataset includes all the articles having been published in the JASSS journal[9].

From this review, we can observe that five main kinds of generator are used more than others: regular lattices [208] (*i.e.* a grid, such as in cellular automata), random networks [73] (networks where nodes are connected randomly to a given number of other nodes), small-world networks [205] (networks with low average path length among nodes), scale-free networks [30] (networks with a power law degree distribution) and spatial networks [31] (networks connecting nodes that are spatially close). We can notice that all these network generators are really efficient to reproduce one particular feature of real social networks (*e.g.* the power law node degree distribution), but these network generators are far from being realistic. [54] has shown that any real social network is often closer to the target social network than any of those created by such generators.

In addition, we have shown in [10] that the use of these simple algorithms does not even decrease over time in favour of advanced generators than produces more realistic networks. For some ones (*e.g.* random network), the number of articles using it even increases over time. It appeared that in recent works, networks produced by such simple generators are no more use for their realistic character but rather as classical testbed to explore the model behaviour (*e.g.* in sensitivity analysis). Another explanation is that these algorithms are really simple to implement, understand and use (which is not the case for more recent generators). All these arguments drive us both (i) to the development of more precise network generators, and (ii) to their integration in library easy to use in an agent-based platform.

To this purpose, Audren Bouadjio Boulic's (supervised by Chihab Hanachi, Professor at the UT1C, and Frédéric Amblard) and Carlos Sureda Gutierrez's (supervised by Pascale Zaraté, Professor at the UT1C, and myself) theses has started with the project. Audren Bouadjio Boulic

[8] with Frédéric Amblard, Audren Bouadjio Boulic and Carlos Sureda Gutierrez
[9] Journal of Artificial Societies and Social Simulation: http:\jasss.soc.surrey.ac.uk.

is working on a network generator using the diffusion of attachment rules between agents during the evolution of the network. Complementary, Carlos Sureda Gutierrez is developing a model generating social network under the combined influence of individual attitudes and social spots (with an application to the radicalisation network generation).

5 Conclusion and Perspectives

5.1 Conclusion

In this manuscript, I have presented an organised synthesis of my research work, aiming at going one step further in the art of the agent-based modelling and simulation by providing advanced methodological and operational tools. We observe nowadays the ambition to build more and more complex models of complex systems (in particular to represent socio-environmental systems). This comes with a need for more powerful tools to design and implement these models. I chose to focus in this synthesis on three particular aspects to be improved. First, in Chapter 2, socio-environmental systems integrate human beings agents that should be able to make a representation of their environment, reason on it and make complex decisions. To this purpose, I have presented the integration of a BDI architecture in the GAMA platform, which can involve in particular multi-criteria decision-making process. In addition, I have described two models of cognitive attitudes that agents in artificial societies should be endowed with (trust and emotion). Second, as socio-environmental systems embed in general various kinds of agents at various scales and designed with various modelling approaches, I have explored the coupling of two traditionally opposite paradigms in a single model: equation-based and agent-based modelling paradigms (Chapter 3). Finally, such models require huge amounts of data, that are often managed in agent-based modelling projects in an inefficient way. To this purpose, I have proposed a framework integrating agent-based simulations with advanced tools to manage data and tools dedicated to Business Intelligence (Chapter 4).

All this work was done with in mind the constraints that all these tools should be operational, freely available, usable by non-computer scientists and generic to be applied to any kind of agent-based models. From my point of view, the best way to fulfil most of these constraints was to integrate all of them in a single tool: the GAMA platform.

5.2 Research project and perspectives

My research project inspiration comes mainly from the recent observations in the crisis management and response community (and in particular the ISCRAM[1] community). Some research groups have the capability to get mobilised quickly in order to efficiently use modern IT tools to help rescue or to provide insights or predictions about future crisis evolution. For example, the OpenStreetMap community is able to gather and share precise data in major crisis areas as it was the case during the 2015 Nepal earthquake [14]. Another example is the Hackathons organised during the Ebola 2014 crisis, taking advantage of innovative analysis of data that was gathered and shared as open-data. Such events had the aim of "assisting the global response in an informed capacity: to link together various pieces of innovation and expertise from different departments of the University in order to find solutions for problems encountered in the field" [100].

The following research project is centred around the idea to provide the ABM community with tools allowing researchers to bring insights into such situations evolution and possible managements thanks to models. I argue that tackling this question can not only bring many societal benefits, but also can improve, ease and speed up the way we build any kind of model. In Section 5.2.1, I detail the new scientific and technical issues risen by building models in a crisis situation. I focus on two particular ones: the design of a built-in model library (Section 5.2.2) and the use of quantitative data to design agent behaviours (Section 5.2.3).

5.2.1 Building agent-based models in crisis situation

Designing, implementing and using an agent-based model in a situation of crisis brings additional constraints in many steps of the modelling and simulation process, in particular with relation to the model design, the initialisation of the simulation with data, or the assimilation of new data during the simulation.

First, it introduces an additional constraint of time to the traditional modelling cycle (see [28] for an example of modelling cycle we use during training sessions) as first models should be used quickly after the crises. As a consequence, this imposes to reuse as much as possible existing and available models, knowledge and data. It is not possible to go on the ground to gather new data or spend months to design and implement new specific models.

We can notice that even if one question should drive the design of one particular model, and two different questions on the same reference system can lead to two very different models, it appears that often several processes are quite similar between models. As a consequence, it would be a great benefits for modellers to have a set of sub-models that could be directly used and combined to build first versions of the model. These off-the-shell models could be made more complex or simply modified later, but they will help modellers to start quickly in the implementation of a first working model. It is interesting to notice that this idea is

[1] ISCRAM stands for Information Systems for Crisis Response and Management (http://www.iscram.org/)

neither specific to quick model design in crisis situations nor new in the ABM community but remains poorly addressed. This question was for example deeply discussed in the last ABMUS[2] workshop about urban modelling: although there exist lot of crowd models and traffic models, there does not exist a single unified model in which human agents can have a specific behaviour and be driven by a built-in traffic model when they move using their vehicle, or by a built-in crowd model when they are pedestrians. This reveals the need for building a library of models and a way to combine them (see details in Section 5.2.2). But the combination should not be limited to plug the outputs of one model on the inputs of another one: it should be able to provide tools to semantically match entities in both models (the human beings in a crowd can be the drivers of cars).

Second, once a model has been built, the next step is to initialise the simulation with appropriate data. With the Open Data trend, a lot of quantitative or spatial data is now available: *e.g.* national statistical agencies (such as INSEE[3] in France) provide demographic data about the population and websites such as OpenStreetMap[4] or the recent but very promising Google Earth Engine[5] provide continuously updated spatial data. The GAMA community has made a great effort to support various kinds of data in the models, being tabular data, shapefiles or even data stored in remote (geographical) databases. When data is not available at all, or not at the right scale, we need to produce statistically correct synthetic data. It is particularly the case when dealing with a human population to be integrated into a model. The GenStar project presented previously deals in particular with the generation of a synthetic, localised and socially structured population from statistical macro data or data at a higher than needed scale. I will not focus on this point in the following but rely on existing work in the GAMA community.

Finally, an additional challenge is the fact that modellers design an initial model of a system that is not in a stable state, *i.e.* the situation and the system will evolve structurally (in the real-world) during the simulation. As a consequence, the simulation should be able to assimilate real-time data from the actual system. This can be done by directly getting new information from the real-world through physical sensors or using connections to databases updated in real-time by external sources. Another very promising data source, in particular in crises involving many human beings, is the use of information from the Twitter social network.

This interaction between simulations and real-time data creates new challenges in terms of data assimilation. In particular, new external data can be simply new values of a data series (such as in the case of weather data) that will only provide a new input data for the next step. But this data can also be an update of data on which the simulations are built (for example the destruction of a building or of a road...). Finally, the simulation can get new data that only gives more precisions to the existing data already used in the simulation. Assimilation of this

[2] ABMUS stands for Agent-Based Modelling of Urban Systems

[3] INSEE stands for *Institut National de la Statistique et des Etudes Economiques, i.e.* National Institute of Statistics and Economic Studies.

[4] http://www.openstreetmap.org/

[5] https://earthengine.google.com/

data raises the question of continuing the simulation with this more precise data, or even to step back in the simulation, restart from a past time point and relaunch simulations with this new more precise data.

The use of Twitter data is promising to monitor crisis situations. I argue that it could be a great source of information on populations in the area of a crisis, providing for example spatial locations or additional information. In addition to this quantitative information, Tweets can also provide more qualitative information, for example on the individual behaviours and on social interactions or self-organisation behaviours. It could thus be very interesting to mine this data source in order to detect new behaviours or patterns that should thus be integrated in the model. This point particularly interests me and I detail it in Section 5.2.3.

Another challenge on which I will not focus is the necessity to run simulations (and experiment designs) in real-time to get useful insights and information from the model. To this purpose, the use of High-Performance Computing will be necessary. On this point, I will also rely on work in the GAMA community, such as the EPIS project [35] or collaborations with the OpenMole[6] community [152].

5.2.2 Toward a library of models

Models representing a crisis situation and its management (similarly to models of socio-environmental systems) are typically complex models integrating multiple processes or sub-models. ABM has become the paradigm of choice to this purpose by allowing modellers to integrate models from different domains in a single complex model. As shown previously in Section 1.2.2, the MAELIA model couples various processes (or sub-models) in different formalisms and at different scales. It could thus be very interesting for the whole GAMA community that the sub-models developed for MAELIA are easily usable in other models. However this is not the case so far for two main reasons, that I will detail in the following sections:

- the need for a high-level semantic description of models: all the traditional coupling models solutions (such as HLA[7] or DEVS[8]) [57, 104] do not address the *semantic* problems that arise in order to ensure that a composition of models makes sense,
- the lack of a models library, *i.e.* a set of models that has been designed, implemented and semantically described to be then coupled together. For example, even if it is very modular (in the sense that one sub-model implementation can be easily replaced by another one), the MAELIA model has been implemented as a single model without paying attention to the reusability of the various parts.

[6] OpenMole is an open-source application easing the execution of simulations on remote grids or clusters http://www.openmole.org.
[7] HLA stands for High Level Architecture.
[8] DEVS stands for Discrete Event System Specification.

High-level semantic description of models.

From a technical point of view, coupling models is a question that has been widely addressed, but in general outside of the ABM community. The standard in the domain are the technical solutions HLA [57] or DEVS [104]. More recently [67] have addressed this issue in an agent-oriented way (and implemented it in the GAMA platform) through a paradigm called *comodelling*. A comodel is an agent-based model in which some agent behaviours are computed by the execution of another external model. This is a way to reuse an existing models in a new model. In addition, this approach has the interest of merging the notions of data sources and models: both are wrapped in agents and it is not possible from an external point of view to differentiate whether they compute or simply read the data they make available for other agents.

But these approaches are purely technical and do not address the semantic problems that arise in order to ensure that an composition of models makes sense. One reason is the lack of standard high-level descriptions of models. Some description protocols are based on natural language, like ODD[9] [92], or graphical languages, like the one proposed in the MAGEO platform [113]. Some others use ontologies [94] like OWL[10] or Mr. Potatohead [156, 150], or more elaborate "semantic" metadata, such as CityGML [93]. Even initiatives like OpenABM[11], Modeling4All [108], NetLogo [206], or Record [34] do not go beyond simple taxonomies of simple models. Thus the lack of standard high-level model descriptions accounts for the difficulty in defining a library of models that could be reused.

The availability of a semantic description for models will allow developers to choose which models to combine thanks to their description. This description should provide the models main points, document their limits and hypotheses, and should support some degree of automated processing, to allow design experiments that rely on the substitution of models. The starting point will be to rely on the *de facto* standard ODD protocol and replace its ambiguous parts [12] with a graphical and unambiguous description. One advantage of coupling ODD with an approach supporting a graphical representation is that it is easy to understand even for non-computer scientists. The description will be completed by the specification of how to execute a model so it can be integrated in other (co)models. This last part can rely on either FMI[12] [36] or WSDL[13] [38].

Models library.

A set of models should thus be defined, implemented and indexed in a library using the description formalism defined in the previous section. Defining this library of models will be a huge task, that should be done incrementally given the chosen applications. Ideally

[9] ODD stands for Overview, Design concepts, and Details.
[10] OWL stands for Web Ontology Language.
[11] http://www.openabm.org/
[12] FMI stands for Functional Mock-up Interface.
[13] WSDL stands for Web Services Description Language.

this library should be a shared repository in which all the GAMA community can share their atomic models. Since nowadays the population is more and more concentrated in cities, a good starting point to build this library should be by integrating urban simulation models (*e.g.* traffic or pedestrian mobility...) and a set of crisis situation models that apply on a city (earthquake, tsunami, epidemics, fire...). Several instances of models will compose the different categories in order to fulfil the need for models at various scales and/or with different formalisms, input parameters, goals, etc. Finally, more process-focused models will also be indexed, especially population generation methods.

In addition to this model library, it could be very interesting to integrate a library of indicators that can be reused in several models. These indicators can be economic, social or environmental. They can measure the resilience or the sustainability of a system or the performance in the evaluation of a management response to a crisis or in the assessment of a policy. Specific indicators can provide an advanced understanding of the simulation or domain-specific information on a simulation. They will help stakeholders or domain experts to understand its dynamics and results. Another interesting benefit is the fact that it could allow to more systematically compare several models with exactly the same indicators.

5.2.3 Behaviour from qualitative data

In this section I am interested in the questions of building, informing and feeding a model with qualitative data, and of designing the behaviour of (human) agents from texts, reports, testimony, Tweets or more generally any kind of qualitative data. I have to mention that I had to face this question in two different modelling projects (ARCHIVES and SWIFT). In both models, human behaviour description is provided in reports or inquiry reviews. To build the models, I had no choice but to arbitrarily choose a small piece of statements and to manually propose a way to implement human behaviours reproducing this subset. It is of course neither satisfactory nor generic, and totally unsuitable in case of a dynamic environment (*i.e.* when simulations dynamically get new information about people behaviours). As a consequence there is a need for providing new tools to deal with this issue. This problem is by the way quite recent: the JASSS[14] journal has devoted a special session (Volume 18, Issue 1 of January 2015) to the subject, and an ESSA[15] Special Interest Group named *Qual2Rule – Using qualitative data to inform behavioural rules* has been created recently.

Qualitative data is far more heterogeneous than quantitative data. As a consequence, I propose to consider three different datasets on three specific applications to investigate the question of the design of an agent-based model from qualitative data. These case-studies are presented from most structured to least structured.

- ARCHIVES project dataset [81]. The aim of the model developed in the ARCHIVES

[14] JASSS stands for The Journal of Artificial Societies and Social Simulation, http://jasss.soc.surrey.ac.uk/
[15] ESSA stands for European Social Simulation Association.

project is to reproduce the 1926 floods of Hanoi (Vietnam), from geo-historical data on the environment and the management of the crisis by authorities. The dataset contains a set of reports, information notes and orders related to the crisis management.

- SWIFT project dataset [4]. The model aims at reproducing Australian bushfires in 2009 and human behaviours when facing them. The dataset is a set of interviews from survivors, reports and police statements about victims available on-line. They describe the population activities but also their preparedness state before bushfires and their feelings during them. This is particularly interesting in order to build a model with complex agents.

- Tweets can now be used as a real-time information source. A challenging work would be to design new ABM from this qualitative data in real-time. We have a dataset of Tweets of the day of the terrorist attacks in Paris (the 13th of November 2015) as a case study.

These three datasets are very interesting because they are very rich and different, in particular in terms of the point of view, of actors involved and the link with formal official organisations. I present in the following sections the way I plan to tackle these case studies, before concluding on how the tools and methods developed on these case studies could also be applied to extract agents behaviours from participative simulations.

ARCHIVES dataset.

The ARCHIVES project[16] is dedicated to the study, documentation and reproduction of past events, with a case study on the floods in Hanoi (Vietnam) in 1926. One of the aims of the ARCHIVES project is to show the benefits of agent-based models for Humanities and in particular History: it has for goal to propose a methodology that would enable the support ot historians' work, in a systematic and automated way, from the analysis of documents to the design of realistic geo-historical computer models. It is part of the recent trend of *Digital Humanities* that consists in integrating computer tools into the activities of humanities scholars: we argue that using the developed models, users can both visualise what happened and explore what could have happened in alternative "what-if" scenarios. Our claim is that this tangible, albeit virtual, approach to historical "fictions" will provide researchers with a novel methodology for synthesising large corpora of documents and, at the same time, become a vector for transmitting lessons from past disasters to a contemporary audience.

A first model for this project has been built for the JTD training sessions [66] and preliminary results have been published in [81]. We have focused mainly on the reproduction of the crisis (river water level evolution and the effects on dykes) and on its management by authorities. This model has been built by gathering data from several different sources: for example,

[16]The ARCHIVES project is led by the UMI 209 IRD, UMMISCO and funded by the Université des Sciences et Technologies de Hanoi (USTH), Hanoi, Vietnam.

historical maps of Hanoi buildings, dykes and rivers have been found in the IGN[17] archives and in the Vietnamese National Archives. As far as crisis management data is concerned, it has been gathered and organised by Olivier Tessier (Historian at the EFEO[18], Ho Chi Minh City, Vietnam) and Nasser Gasmi (Master 2 intern at the USTH in 2014). These archives describe hour by hour human activities to deal with the crisis. They contain documents (information letters, orders, activities reports...) sent between the various actors (French and Vietnamese province, district or commune authorities or technical services...). In addition O. Tessier was able to formalise the formal organisation at that time with hierarchical relationships between various roles [81][19].

The main issue I am interested in this case study is, given a model of the floods and other physical processes, to provide a way to generate the crisis management process thanks to the reports in archives. It is important to notice that the dataset is very particular in the sense that documents have two main origins: they come either from information reports, orders or help requests sent between actors during the events, or from an *a posteriori* inquiry to find responsibilities in the crisis management. Documents are thus mainly focused on describing precisely what the actors did and how they interacted. Since we have access to the theoretical hierarchical organisation, I argue that the simplest starting point is to focus on discovering interaction protocols among these actors.

A preliminary study [46] started to consider all these archives as log files and to use Protocol Mining (PM) [199] on them to extract some processes (in this case interaction protocols) that can then be integrated in the model. This approach is widely used in crisis management to formalise rescue intervention protocols from multiple crisis cases. Its application to this case study will have to face several difficulties[20]. Indeed, we only have a single crisis case, which will make the generalisation process difficult. I argue that it is possible to balance this lack of cases with additional information we have gathered. For example we know for each action performed its location and that similar actions (or sequences of actions) have been performed in several locations. The generalisation will thus rely on this spatial information to distinguish various cases for generalisation purpose.

In addition, traditional uses of PM are dedicated to find the scheduling of activities in a log file. This does not take into account the richness of the exchanged messages, in particular in terms of the illocutionnary force [170, 76] of statements. We will therefore follow [96] to enrich PM algorithms in order to discover dynamic organisation, which will give us the various

[17] IGN stands for *Institut national de l'Information Géographique et forestière, i.e.* the French national institute of geography.

[18] EFEO stands for *Ecole Française d'Extrême-Orient, i.e.* French School of Asian Studies

[19] In 1926 Vietnam was a French protectorate which means that both French and Vietnamese authority organisations remain. The hierarchical relationships also describe the kinds of interaction that are theoretically possible between actors. An exhaustive analysis of the archives showing the cases where the hierarchy is not followed has brought very interesting information to historians too.

[20] I should mention here quickly the issue of automatic extraction of semi-formal logs from the archives in natural language. On this point, I rely on other members of the ARCHIVES project who are currently working on the document image analysis and recognition.

interactions between actors (or more generally roles) involved in the crisis management.

This tool should provide a formal representation of the interactions among actors in the system. It will also help the modeller to elicit the agent species that should be defined and express the atomic actions and interactions they can have with other agents (and the interaction protocol). This will be the basic element to build the model. Of course all the atomic actions will have to be implemented manually by the modeller or be taken from an existing sub-models library, but I am convinced that this approach can provide the skeleton of the model.

This first case study is particular because the qualitative data are well-structured. In this case, I argue that the Protocol Mining approach would be very well appropriate to build the models.

Bushfires.

The second dataset is composed of (*a posteriori*) testimonies of witnesses[21] [184] of the 7 February 2009 bushfires in Victoria (Australia), also known as the Black Saturday bushfires. It is the worst bushfire disaster in Australia history with the death of 173 people. In addition 414 people have been injured and the cost of the disaster is estimated to $4 billions [185]. All data comes from the 2009 Victorian Bushfires Royal Commission final report[22].

The policy (and the recommendation) in the state of Victoria (Australia) related to population behaviours during a bushfire is either to leave their house and attempt to escape the fire or to stay and defend it. In both cases, people should be prepared in advance. The inquiry in response to this massive disaster has tried to understand the high number of casualties during the Black Saturday. Reports contain information about the behaviour of residents during the fire (what they have actually done, what were their feelings, the reason for their choices...), their *a priori* choice (between leave or defend), their preparedness and the way they got and perceived information (*e.g.* alert...).

A preliminary analysis of the document [2] has shown subjective reasons of dangerous behaviours such as an under-estimation of danger, an over-estimation of capabilities, individual differences in the danger perception (even in a single family or couple)... From this analysis, we recently built a first model integrating this discrepancy between objective and subjective danger and objective and subjective capabilities [4].

This case study is particularly interesting first because it is well-documented with extensive reports. Contrarily to the ARCHIVES project, the dataset is here focused on describing the disaster victims' behaviours (and not the crisis management). It describes in details the activities during the bushfire. But even more interesting, it contains the subjective part of the reasoning of the victims, providing insights into the actual decision-making process of victims (that appears to be quite different from the decision-making process expected by authorities). Finally, it contains details about the use of communication means, and the

[21] The list of testimonies can been found at the address: http://vol4.royalcommission.vic.gov.au/index3037.html.
[22] http://www.royalcommission.vic.gov.au/Commission-Reports/Final-Report.html

impact and perception of information provided by authorities by different means.

This provides us with a very good dataset to investigate the generation of human agents with a complex and cognitive behaviour thanks to this subjective description of their behaviour. The challenge will thus be to use the proposed cognitive improvements provided in this manuscript and in particular to identify the key concepts (*e.g.* beliefs, desires or emotions) that drove the inhabitants' behaviours. In addition, this will be the occasion to integrate the impact of communication on agent behaviour on a real-case example. This final part should be very useful when we build a model from the last dataset (Tweets detailed in the next section) and we want to deal with the impact of authorities information communication.

Twitter.

Compared to the two previous sources of data, Twitter has the specificity to provide real-time datasets with all the features of Big Data (Volume, Variety and Velocity). It is nowadays a rich data source to monitor actual situations in real-time, interesting in particular a lot of researchers in the ISCRAM community (a track of the conference is dedicated to *Socia Media Studies* and includes many papers dealing with Tweets to better manage crisis).

Twitter is now used monthly by more than 300 millions people[23] and in particular in case of crisis or disaster [176]. It has thus become a media of choice for rescuers and the population. Indeed, it provides rescuers with real-time information about about population facing crisis and thus about the crisis itself. It is also important for the impacted population as it allows self-organisation among people for help purpose, as it was the case with the *#PorteOuverte* trend during the Paris terrorist attack.

Given the big issues induced by the huge number of exchanged Tweets and by their unstructured nature, several works propose methodologies and tools to mine them. [188] apply an information extraction tool (Twitcident) on Tweets before, during and after a huge storm in Belgium. This tool identifies hazards from an external source[24] and then creates automatic queries to old and current Twitter messages to retrieve associated Tweets. Using automated filtering, they were able to extract trending topics and valuable information for crisis management. Similarly, other works focus on the automatic detection of events only through Tweets. As an example, [167] can detect in real-time particular events (*e.g.* an earthquake) only by monitoring Tweets thanks to a classifier, and are able to locate it using a particle filter on Twitter users. They apply it on the detection of Japan earthquakes with a magnitude higher than 3. There exist several other quite similar applications aiming at detecting events (and in particular crises) from Tweets such as Tweet4act [165] or Emergency Situation Awareness [158].

As a real and interesting case-study we have gathered a dataset of Tweets during the 24 hours

[23] Source: https://about.twitter.com/company (may 2016).
[24] The Netherlands real-time public paging messages sent to emergency serviceshttp://www.p2000.nl/

after the Paris terrorist attacks of the 13 November 2015, which will be used as testbed. [175] studied, on a similar datasets, how and why some volunteers self-organised during the 2010 Haiti earthquake, and what were their activities and the benefits of their actions. Of course as any uncontrolled communication, information communicated on Twitter is far from being flawless, and the question of the trust in this information is pregnant [127].

Despite its flaws, Twitter is now a valuable (and often the only) source of information about people behaviours during a crisis. Combined with other open-data sources (*e.g.* OpenStreetMap...) and synthetic population generation, it could become a powerful tool to create and/or feed simulations on the fly. One of the main challenges with Twitter will be to be able to extract relevant data and knowledge from this huge source of data.

Building a model in crisis situations based on Twitter data will require several improvements of existing tools. An insight into the envisioned steps to reach this goal is presented below.

First, we should be able to create a preliminary simulation from a combination of data from Tweets and statistical macroscopic data (*e.g.* INSEE data) thanks to synthetic population tools (such as the one developed in the GenStar project). This simulation will not be much more than a visualisation of the situation (human beings agents will be given no behaviour), but it should integrate a way to assimilate new Tweets, in particular to update the population of human agents. As far as the spatial data is concerned, either the space limitation is chosen *a priori* and Tweets are filtered given this chosen location, or the spacial area can be computed from a selection of Tweets and spatial data retrieved from spacial databases.

Second, dedicated behaviours should be added to the model. At first simple behaviour models from the model library could be used (*e.g.* a generic model of people mobility in a crowd). At runtime, given mined data, these models can be updated during the simulation through data assimilation. But we can also mine the Tweets in order to find emergent behaviours and (self-)organisations. This mining should be able to formalise these behaviours to integrate them into the models at runtime, through a kind of behaviour assimilation.

Finally, this also opens many other research tracks. In particular, work developed in the ARCHIVES project about the mining and generation of management processes can be reused here to deal with rescue intervention processes. But as the situation is highly dynamic (with for example self-organisations among people), these processes should be regularly adapted (or can even be self-adaptive). An interesting addition could also be to mine sentiments in Tweets in order to add an emotional dimension in the model. The model could thus reuse the emotion model presented in Section 2.4.2 in order to take into account people emotions in their behaviour, or simulate potential emotion contagions.

Application: behaviour elicitation from participative simulations.

This manuscript has mainly presented simulations used to explore behaviours of the model or to support decisions. Another very interesting way of using the (agent-based) simulations is to

let human beings play agents' roles and decide of their actions. These participative simulations [117, 33] allow to build serious games in which stakeholders or even decision-makers can play their own role (or a different one, depending of the purpose of the game). Here the simulation is a support for discussions, interactions and reasoning. It can even lead to decision-making. The GAMA platform provides all the necessary tools to do such participative games and has already been used to create a serious game about managing coastal floods on the Oleron Island [9].

In addition to its purpose to support human discussion, I argue that we could use participative simulation to extract knowledge from actors: we could record logs of the actors' actions during a simulation. The idea is to apply the methods developed on the qualitative datasets on these logs to extract agents' behaviours. Similar ideas have been investigated in [52]. Following the methodology of [78], in modelling projects of socio-environmental systems, it appears very efficient to develop two distinct models. The first one (the exploration model) is very simple, light and suitable for participative simulations. The knowledge extracted from this first model (and in particular agents' behaviours) can be used to develop a second model (the prediction model), that aims at reproducing the system very precisely.

To sum up, from a global point of view, this research project attends to build methodological and operational tools to design and implement models in crisis situations. To this purpose, I argue that it is necessary (i) to build a library of models, with a semantic description, that can be coupled, and (ii) to give them the capability to assimilate real-time quantitative and qualitative data. This requires to develop tools to mine quantitative data (reports, testimonies, narratives or even Tweets) and extract new behaviours that can be assimilated by simulations. Beyond the particular case of building models in crisis situation, this could definitely improve the way we build models.

A Integration in the GAMA platform

As mentioned in Section 1, most of the work presented in this thesis have been done using the GAMA platform and had required extensions of the modelling language GAML. In this appendix, I just aim at giving a taste of the improvements that have been added to the language. To this purpose, I first present the basic elements of the GAML language in Section A.1. The three following sections refer each to a chapter of the thesis: Section A.2 presents basic elements of the simpleBDI architecture, Section A.3 presents how ODE systems can be integrated in a GAMA model and finally I present the way agents in models can interact with databases in Section A.4. This chapter is very synthetic on purpose; interested readers can refer to the official GAML documentation for more details[1].

A.1 GAML in a nutshell

Skeleton of a GAMA model. A GAMA model (written in the GAML language) is composed of a header defining the name of the model (`model`) and three kinds of blocks:

```
model my_model

global {
  int a_variable_name <- 0;
}

species my_species {
  ...
}

experiment my_exp {
  ...
}
```

[1] An exhaustive documentation of the GAMA platform and of elements of the GAML language can be found on the website: http://gama-platform.org/.

The `global` block defines the global environment in which the agents will be defined. In this block will be defined the global variables and global dynamics. It will also define the boundary of the physical space. Finally the initialisation of the global will be the place to create all the agents of the simulation.

Each `species` blocks (we can define as many `species` block as needed) will define a particular kind of agents, with its own attributes, behaviours and ways of display. Each agent in the simulation will be an instance of a species.

Finally `experiment` blocks (we can also define as many `experiment` blocks as needed) define the ways a model will be run: launching a simulation is executing a particular experiment. It contains the parameters (`parameter`) and the outputs of the simulation, defined as a set of `display`. The GAML language includes two kinds of experiment: `gui` mode (experiments with a graphical user interface) and `batch` mode (these experiments define an experiment design and run simulation in batch mode).

In GAML, every line ends either by a semi-colunm character `;` or opens a new block, opened and closed by curly-brackets `{` and `}`. The only exception is the first line (with the `model` statement).

The GAML basics: statements and operators. The GAML language is a both functional and procedural language, composed of two basic components: **statement** and **operator**. Every line is a statement, often starting by the name of the statement: `global`, `species` or `experiment` in the above example, or other classical structures such as conditionals (`if` or `match`), loops (`loop`). The only exception is the declaration of variables and the affectation: the line starts with the type of the new variable (or only its name if it has already been declared). The statement symbol is here the arrow `<-`. Many statements can be given additional parameters using facets: `statement_name facet1: val1 facet2: val2 ... ;`. All the statements have a facet which name can be omitted to lighten the code.

`operator` are almost equivalent to functions. Every operator takes 1 or more operands and returns a computed value. Except for unary operators that can be written only using the prefix notation: `operator_name(operand)`, the binary or more operators can be written using either the prefix (`operator_name(op1,op2,...)`) or the infix (`op1 operator_name op2` or `op1 operator_name(op2, op3, ...)`) notations.

As an example, the following lines combines the statements `loop`, `write` (writing a value in the console) and the affectation with the two operators `length` (returning the number of elements in a list) and `at` (returning the element of the list at the index).

```
list<int> l <- [0,1,3,6,8];
loop i from: 0 to: length(l) - 1 {
  write "" + l at i;
}
```

Structure of a species. Any species can define the following elements. It declares a set of attributes (with a type, a name and an initial value). Each species has already built-in attributes, such as `name`, `shape` or `location`... It also defines a set of `aspect`, that define the ways agents of this species can be displayed.

```
species my_species {
  int attribute1 <- 0;
  string attribute2 <- "";

  init { ... }

  reflex r1 { ... }

  action act1 { ... }

  aspect a1 { ... }
}
```

Finally, the basic architecture of an agent, which defines its behaviour, uses two kinds of statements: `init` and `reflex`. When an agent is created, the `init` block is called. It contains what should be executed at the creation of the agent[2]. All the `reflex` blocks are executed at each simulation step, when the agent is scheduled. It contains the set of code describing the behaviour of the agent. `action` blocks contain code that is executed only when called (in a `reflex` for example). They are very similar to methods defined in a class in Java.

Notes on species:

- The reflex-based architecture is the basic one for agents. Others can be attached to a species with the facet `control`: `species my_species control: fsm { ... }`. In this example, this species has a Finite State Machine architecture; it can thus define various `state` for the agent.
- GAML has an inheritance mechanism: a species can inherit from another with the facet `parent`.
- Additional built-in skills can be attached to a species using the `skills` facet to add new capabilities to agents of this species. For example, the skill `moving` adds additional attributes (*e.g.* `speed`) and built-in actions (*e.g.* `wander`, `goto`...) for species with this skill.

[2]The `init` block of the global is executed when the global agent is created, *i.e.* at the creation of the simulation. The other agents can be created in this block, using the `create` statement.

A.2 Cognitive agent extension: the simpleBDI architecture

Declaration of a species controled by the BDI architecture. First of all, the species of agents using the simpleBDI architecture should declare it:

```
species my_species control: simple_bdi{ ... }
```

This gives to all agents of this species additional attributes (such as its `belief_base`, `desire_base` or `plan_base`) and the capability to use additional behaviour structures. In particular, each agent has its own `intention_persistence` (its commitment to its intentions in $[0,1]$), `plan_persistence` (its commitment to its plans in $[0,1]$) and `probabilistic_choice` attributes (which enables the agent to use a probabilistic or a deterministic choice when trying to find a plan or an intention).

A new type: predicate. `predicate` is a new data structure that is used to define the content of any belief, desire or intention. A predicate is defined by a name, a map of values, a priority and a truth value.

```
predicate a_predicate <- new_predicate("my_prop", true)
  with_priority 3;
```

Management of bases. A set of actions have been introduced to manage the belief, desire and intention bases. Agents can for example add (`add_belief`), remove (`remove_belief`) or replace (`replace_belief`) a belief, check its existence (`has_belief`) or get the predicate value of a belief (`get_belief`). Similar actions exist for desires. We can also get the current intention (`get_current_intention`) and add subintentions (`add_subintentions`)...

```
do add_belief( new_predicate("my_prop", true) with_priority 3);
```

Perception. At the beginning of each step, the agent will first check its perceptions. The aim of these perception blocks is to update the belief base depending of the agent's perception of the environment. The perception can be focused on the agent itself or on other kinds of agent. In particular, when the agent attempts to perceive other agents, we can define what we want to store for each of the perceived agents in a given radius (using the `focus` statement).

```
perceive target:self{
  if(a_value>0){
    do add_belief(a_predicate);
    do remove_belief(another_predicate);
  }
}
```

```
perceive target: another_apecies in: 10 {
    focus predicate_name_to_add var: an_attribute priority: 10;
}
```

Rules. `rule` are a new statement introduced to produced automatically new beliefs and desires from beliefs or desires. They can be used in particular to represent inference rules between these attitudes.

```
rule belief: existing_belief new_belief: p_new_belief new_desire:
  p_new_desire;
```

Plan. Finally, `plan` statements define the sets of actions to perform when a given intention exists.

```
plan my_plan intention: intention_predicate {
  do an_action;
}
```

Additional facets can be defined for a plan such as `finished_when` (it defines the termination condition), `instantaneous` (if false, no other plan can be executed afterwards during the current simulation step).

A.3 Integration of Equation-Based Models in GAMA

A.3.1 Integration of ODE models

A stereotypical use of Equation-Based Models in a GAMA agent-based model is to describe some agents' attributes evolution using an ODE system. As a consequence, the GAML language has been increased[3] by two main concepts: (1) the definition of equation systems using the `equation` statement; (2) the numerical integration of an equation system using the `solve` statement.

Definition of an equation system.

The definition of an ODE system has to be done in the definition of a species: the equation system can thus manipulate species attributes. An equation system is defined using the `equation` statement. An `equation` block is composed of a set of `diff` statements describing the evolution of species attributes. In the following example, we define a species embedding an SIR system (*c.f.* (Equation 3.1)). The equation system named `SIR` will thus control the evolution of the `S`, `I` and `R` attributes of each agent of the species.

```
species agent_with_SIR_dynamic {
  float t;
  int N <- 1500 ;
  float S <- N - 1.0;
  float I <- 1.0;
  float R <- 0.0;

  float alpha <- 0.2 min: 0.0 max: 1.0;
  float beta <- 0.8 min: 0.0 max: 1.0;
  float h <- 0.01;

  equation SIR {
    diff(S,t) = (- beta * S * I / N);
    diff(I,t) = (beta * S * I / N) - (alpha * I);
    diff(R,t) = (alpha * I);
  }
  // Call to the integration method
}
```

In addition, an equation system can be split into several species and each part of the system is synchronized using the `simultaneously` facet of `equation`. The system split into several agents can be integrated using a single call to the `solve` statement. The interest is that the modeller can create several agents for each compartment, which different parameter

[3] This extension has been a collaborative work over years mainly between Hyunh Quang Nghi, Tri Nguyen-Huu, Alexis Drogoul and myself. It has also been partially funded by the CNRS through the IMEA PICS project, which Patrick Taillandier is the Principal Investigator.

values. For example in the SIR model, the modeller can choose to create 1 agent describing the evolution of the number of Susceptibles, 2 agents for the Infecteds and 1 agent for the Recovereds. The `beta` attribute will have different values in the two agents, to represent 2 different strains.

Integration of an equation system.

The `solve` statement has been added in order to integrate numerically the equation system. It should be called into a reflex. At each simulation step, a step of the integration is executed, the length of the integration step is defined in the `step` facet. The `solve` statement will update the variables used in the equation system. The integration method (here Runge-Kutta 4) can be chosen using the `method` facet.

```
    // Call to the integration method
    reflex solving {
        solve SIR method: "rk4" step: h ;
    }
```

A.3.2 Diffusion

GAMA provides the possibility to represent and simulate the diffusion of a variable through a grid topology. To this purpose, the `diffuse` statement has been introduced[4] to make an integration step of the diffusion following the diffusion matrix.

The following model is a very simple example illustrating the use of the `diffuse` statement to diffuse the value of the variable `phero` (defined using the facet `var`) on the grid of agents `cell` (defined using the facet `on`) following the diffusion matrix `mat_diff` (facet `matrix`). At each simulation steps, an integration step is performed[5].

```
global {
  matrix<float> mat_diff <- matrix([
        [1/9,1/9,1/9],
        [1/9,1/9,1/9],
        [1/9,1/9,1/9]]);

  reflex diffusion {
    diffuse var: phero on: cells matrix: mat_diff;
  }
}
grid cell height: 64 width: 64 {
    float phero <- (flip(0.2))? 1.0 : 0.0;
}
```

[4]This extension has been a collaborative work over years mainly between Hyunh Quang Nghi, Julien Mazars, Alexis Drogoul and myself.

[5]Many other facets have been defined in addition for more complex diffusions.

A.4 Data management in GAMA

From its very start, the GAMA platform has been designed to allow modellers to integrate and use various kinds of data in their models. It has initially been built over the Repast platform (for its early management of GIS data) with the aim of providing a modeller-friendly modelling language. With the development of the platform many data format has been supported such as: shapefiles, OpenStreetMap[6] but also raster data such as .csv files or spatialize .asc of .tif files[7]. In this section, I will focus only on the primitives dedicated to manage data using databases[8].

The GAML language has been extended with primitives to interact with Database Management Systems (DBMS) and Multi-Dimensional Databases. They allow agents to execute any kind of SQL query (create, insert, select...) to various kinds of DBMS and MDX (Multidimensional Expressions) queries to select multidimensional objects, such as cubes, and return multidimensional cellsets that contain the cube's data.

We made the choice that DataBase capabilities to be associated with agents in two different ways: (i) through a skill (`SQLSKILL`, `MDXSKILL`), or (ii) through the built-in species `AgentDB`. Thanks to the GAML inheritance, any species can inherit from the `AgentDB` species, thus modeller can use either the latter of the former solution in any species definition. The difference between the two approaches is that `AgentDB` agents create a connection to the DataBase when they are created and keep this connection open as long as they exist in the simulation. This saves time for each requests, but as the number of simultaneous connections to a database is limited, we cannot create as many `AgentDB` agents as needed in the simulation. Conversely, agents using the `SQLSKILL` open a new connection to the database for each query and close it after. This is obviously less efficient for each query but saves resources in terms of connection.

All the primitives of interaction with databases are independent of the DBMS (only the connection paramaters are DBMS-dependent). The following DBMS are currently supported: SQLite, MySQL Server, PostgreSQL Server, SQL Server, Mondrian OLAP Server and SQL Server Analysis Services.

A.4.1 SQLSKILL

Definition of a species using the SQLSKILL. First of all, the species should declare it uses the `SQLSKILL` in order to be able to use additional actions

```
species foo skills: [SQLSKILL] {
  //insert your description here
}
```

[6] http://www.openstreetmap.org/

[7] The support of many data formats is mainly due to Patrick Taillandier.

[8] I had started to develop these extension for the MAELIA model. It has then been extended by Truong Minh Thai, Viet Xuan Truong and Alexis Drogoul.

Maps of connection parameters for SQL queries. In all of the actions of the SQLSKILL we have to give the connection parameters to the database. It is fully dependent of the DBMS, I present below only the parameters for the MySQL DBMS.

```
// MySQL connection parameter
map <string, string>  MySQL <- [
  'host'::'localhost',
  'dbtype'::'MySQL',
  'database'::'table_name', // it may be a empty string
  'port'::'3306',
  'user'::'root', 'passwd'::'abc'];
```

Test a connection to database. The action `testConnection` tests whether the connection to a given database is possible.

```
if (self testConnection(MySQL)){
  write "The connection is possible." ;
}
```

Interaction with the database. The SQLSKILL provides 2 actions to interact with databases depending on the type of query: `select` to execute select statements (*i.e.* queries returning a result) and `executeUpdate` to execute all the SQL statements that do not return a set of values (*e.g.* create, drop, insert...). These 2 actions first create a connection to the database, execute the query and close the connection.

If the selection succeeds, the action `select` returns a list with three elements: (i) the list of column name, (ii) the list of column type, and (iii) a data set.

```
list<list> t <- list<list> (self select(MySQL, "SELECT * FROM
  points ;"));
```

If the action `executeUpdate` succeeds, it returns the number of records modified.

```
// Insert into
do executeUpdate( MYSQL, "INSERT INTO registration " + "VALUES
  (100, 'Zara', 'Ali', 18);");
```

A.4.2 AgentDB

Definition of a species from AgentDB. AgentDB is a built-in species. It can thus be used as it in any model. But it can also been used from any other species, using the inheritance mechanism.

```
species fooDB parent: AgentDB  {
  //insert your descriptions here
}
```

Connection to the database. Once these agents has been created, it should be connected manually to the database. To this purpose the action `connect` has been introduced. Similarly, the `close` action exists. In addition the action `isConnected` can check whether the connexion has succeeded.

```
create fooDB  {
  do connect (MYSQL);
  if (self isConnected() ){
    write "The connection has succeeded.";
  }
}
```

Interaction with the database. The same actions as previously have been introduced to interact with the database, but they do not need to get as parameters the connection parameters.

```
// Insert into
do executeUpdate( "INSERT INTO registration " + "VALUES(100, 'Zara
  ', 'Ali', 18);");

// SELECT query
list<list> t <- list<list> ( self select( "SELECT * FROM
  registration ;") );
```

A.4.3 MDXSKILL

`MDXSKILL` plays the role of an OLAP tool using select to query data from OLAP server to GAMA environment and then species can use the queried data for any analysis purposes. It is really similar to the SQLSKILL in its use (with the action `testConnection` and `select`). The main difference is in the maps of connection parameters and in the language for queries.

A.4.4 Using database features to define environment or create agents

In GAMA, we can use results of select query to create agents or define boundary of the environment in the same way we do with shape files. Further more, we can also save simulation data that are generated by simulation including geometry data into a database.

Initialisation of the environment bounds from database data.

To this purpose we have to define an extended connection map, including the request. The envelope operator will thus create itself the database connection, perform the query, close the connection and use results.

```
map<string,string> BOUNDS <- [
  'host'::'localhost',
  'dbtype'::'postgres',
  'database'::'spatial_DB',
  'port'::'5433',
  'user'::'postgres', 'passwd'::'tmt',
  'select'::'SELECT ST_AsBinary(geom) as geom FROM bounds;' ];

geometry shape <- envelope(BOUNDS);
```

Initialisation of agents from database data.

Agents of a given species can be created using data results from a select query. An agent to manage database should thus be created, run the select query and then the results will be used to create and initialize agents.

```
string LOCATIONS <- 'select ID_4, Name_4, ST_AsBinary(geometry) as
   geom from vnm_adm4 where id_2=38253 or id_2=38254;';

ask fooDB {
  create locations from: list(self select (  select: LOCATIONS))
  with:[ id:: "id_4", custom_name:: "name_4", shape::"geom"];
}
```

Other features are also available such as the storage of geometries in dedicated databases.

B Summary of the scientific production

This appendix summarises chapter by chapter my scientific production including collaborations, projects, students' supervision, publications...

B.1 Chapter 2

Supervision. The material of the Chapter 2 includes the work of 3 PhD students and 2 Master students I have supervised. I have supervised Nguyen Vu Quanh Anh (he has defended in 2012) with Salima Hassas (Professor at Claude Bernard - Lyon 1 University), Richard Canal (Associate Professor and head of the IFI) and Frédéric Armetta (Associate Professor at Claude Bernard - Lyon 1 University).

I am also supervising Truong Chi Quang (from 2012) with Alexis Drogoul, Vo Quang Minh (Associate Professor at the Can Tho University) and Patrick Taillandier (Associate Professor at the Rouen University); and Ta Xuan Hien (from 2013) with Dominique Longin (CNRS Researcher at IRIT lab).

In addition, I have supervised Charles Berthaume's (in 2014) and Le Van Minh's (in 2011) master internships.

Collaborations and projects. The MAELIA project (led by Pierre Mazzega, CNRS Senior Researcher at the UMR GET) has been funded by the RTRA STAE from 2009 to 2014. It is now led by Olivier Thérond, Research Engineer at the INRA (French National Institute for Agricultural Research).

The integration of the BDI architecture has been partially funded by the ACTEUR ANR project (led by Patrick Taillandier). This project is a collaboration with the IRD (UMMISCO), the universities of Rouen and La Rochelle.

Works investigating social emotions have been funded by the EMOTES ANR project (led by

Emiliano Lorini, Researcher at the IRIT CNRS lab) from 2012 to 2014 and is a collaboration with the IRIT lab and the CLLE CNRS lab.

Publications. The material of this chapter has partially been published in 3 journal articles: KER [3], Kybernetes [187], JAIHC [135], and 18 international conference and workshop proceedings: MABS 2016 [180], MABS 2015, [192], SSC 2015 [47], SoICT 2015 [178], ICAART 2014 [186], ISCRAM-MED 2014 [141], iEMSs 2014 [190], MABS 2014 [85], KES-AMSTA 2013 [116], JFSMA 2013 [140], ESSA 2013 [84], PRIMA 2012 [115, 138], iEMSs 2012 [183], JFSMA [182], SASO 2011 [139], RIVF 2010 [137], RIVF 2009 [136].

B.2 Chapter 3

Collaborations and projects. Research on the MicMac model has been partially funded by the MicMac PEPS CNRS project (led by myself in 2013 and 2014) in collaboration with the universities of Le Havre and Aix-Marseille and the CNRS Géographie Cités lab. The study of the coupling between PDE models and ABM has been funded by the GeoDiff PEPS CNRS project (led by Aymeric Histace in 2013 and 2014) and is a collaboration with the ENSEA *grande école*.

The Dengue Fever model is the result of an informal collaboration with the universities of Rouen and Can Tho, the UMI UMMISCO, the UMR MIVEGEC, IRD and the Oxford University Clinical Research Unit (OUCRU).

Publications. The material of this chapter has partially been published in 1 journal article: Systems [25], 3 book chapters [22, 24, 51], and 4 international conference and workshop proceedings: MARAMI 2014 [23], MABS 2015 [26], PAAMS 2015 [48] and MABS 2016 [154].

B.3 Chapter 4

Supervision. The material presented in the Chapter 4 is related to the work of 2 PhD students and 4 Master students I have co-supervised. I have co-supervised Truong Minh Thai's PhD (he has defended in 2015) with Frédéric Amblard and Christophe Sibertin-Blanc and I am supervising Carlos Sureda Gutiérrez (from October 2014) with Pascale Zaraté (Professor at UT1C). I have also co-supervised, with Frédéric Amblard, Thomas Fumey and Paterne Chokki during their Master internships (about synthetic social network generation) in 2014 and William Chapotat and Lionel Houssou (about the simulation of oil pollution in Equateur, that can be used as a case study for the GenStar library) with Mehdi Saqalli (CNRS Researcher in the UMR GEODE (Environmental Geography) in the University Toulouse 2) and Audren Bouadjio Boulic (PhD student at the UT1C).

Collaborations and projects. Work about the synthetic population generation is being funded by the ANR research project named GenStar (led by Alexis Drogoul, from 2014 to 2018). This project involved the IRD, the universities of Rouen and Toulouse 1 Capitole and Airbus Defense and Space.

Truong Minh Thai' PhD has been done in informal collaboration with UMMISCO (IRD) and DREAM (IRD, Can Tho University) teams. Similarly, the supervision of William Chapotat and Lionel Houssou is an informal collaboration with the UMR GEODE.

Publications. The material of this chapter has partially been published in 5 international conference and workshop proceedings: AIDE 2013 [194], SoICT 2013 [196], ICAART 2014 [195], the Winter Simulation Conference 2015 [10] and iEMSs 2016 [50].

B.4 Chapter 5

Supervision. I have supervised 1 Master intern: Alan Benier, in 2015.

Collaborations and projects. The preliminary ARCHIVES model is one of the result ot the ARCHIVES project (led by Alexis Drogoul, in 2014-2015) funded by the USTH. The partners are: IRD, University of La Rochelle, University of Toulouse, USTH, Vietnamese Academy of Science and Technology, Vietnamese National University, Vietnamese National Archives and the EFEO.

The preliminary model about Australian bushfire is the result of an informal collaboration with Carole Adam (Associate Professor, University Grenoble-Alpes).

Publications. Preliminary results are about to be published in 1 book chapter [66], the 3 conference and workshop proceedings: MABS 2014 [81], SSC 2016 [4] and CNIA 2016 [5].

My exhaustive publication record can be found on the IRIT website[1]. It includes 4 conference/workshop proceedings, 9 national and international journal articles, 15 book chapters, 46 national and international conference papers (at the day of the 18th of November, 2016).

[1] https://www.irit.fr/-Publications-?code=3390&nom=Gaudou+Benoit

Bibliography

[1] C. Adam. *Emotions: from psychological theories to logical formalization and implementation in a BDI agent*. PhD thesis, University of Toulouse, 2007.

[2] C. Adam, E. Beck, and J. Dugdale. Modelling the tactical behaviour of the australian population in a bushfire. In *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*, pages 53–64. Springer, 2015.

[3] C. Adam and B. Gaudou. BDI agents in social simulations: a survey. *The Knowledge Engineering Review*, 31(3):207–238, 2016.

[4] C. Adam and B. Gaudou. Modelling human behaviours in disasters from interviews: application to melbourne bushfires. In *submitted to Social Simulation Conference*, 2016.

[5] C. Adam and B. Gaudou. Modélisation de comportements humains en situation de crise à partir d'entretiens : application aux incendies de forêt de melbourne. In *Proceedings of CNIA*, page to appear, Clermont-Ferrand, 2016.

[6] C. Adam, B. Gaudou, A. Herzig, and D. Longin. OCC's emotions: a formalization in a BDI logic. In J. Euzenat and J. Domingue, editors, *The Twelfth International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA'06), Varna, Bulgaria*, number 4183 in LNAI, pages 24–32. Springer-Verlag, 2006.

[7] C. Adam, B. Gaudou, S. Hickmott, and D. Scerri. Agents BDI et simulations sociales. Unis pour le meilleur et pour le pire. *Revue d'Intelligence Artificielle, Simulation Sociale Orientée Agent*, 25(1):11–42, 2011.

[8] C. Adam, A. Herzig, and D. Longin. A logical formalization of the OCC theory of emotions. *Synthese*, 168(2):201–248, 2009.

[9] C. Adam, F. Taillandier, E. Delay, O. Plattard, and M. Toumi. Sprite - participatory simulation for raising awareness about coastal flood risk on the oleron island. In *Proc. of ISCRAM-med*, 2016. to appear.

[10] F. Amblard, A. Bouadjio Boulic, C. Sureda Gutierrez, and B. Gaudou. Which models are used in social simulation to generate social networks? A review of 17 years of publications in JASSS. In L. Yilmaz, W. Chan, I. Moon, T. Roeder, and C. Macal, editors,

Winter Simulation Conference, pages 4021–4032, Huntington Beach, CA, USA, 2015. IEEE.

[11] E. Amouroux, S. Desvaux, and A. Drogoul. Towards virtual epidemiology: An agent-based approach to the modeling of h5n1 propagation and persistence in north-vietnam. In T. D. Bui, T. V. Ho, and Q. T. Ha, editors, *Proceedings of Intelligent Agents and Multi-Agent Systems: 11th Pacific Rim International Conference on Multi-Agents, PRIMA 2008*, pages 26–33, Hanoi, Vietnam, 2008. Springer Berlin Heidelberg.

[12] E. Amouroux, B. Gaudou, S. Desvaux, and A. Drogoul. O.D.D.: a Promising but Incomplete Formalism For Individual-Based Model Specification (short paper). In *IEEE International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, pages 1–4, Hanoi, Vietnam, 2010. IEEE.

[13] E. Amouroux, T. Huraux, F. Sempé, N. Sabouret, and Y. Haradji. SMACH: agent-based simulation investigation on human activities and household electrical consumption. In *Agents and Artificial Intelligence - 5th International Conference, ICAART. Revised Selected Papers*, pages 194–210, Barcelona, Spain, 2013.

[14] J. Anhorn, B. Herfort, and J. a. P. de Albuquerque. Crowdsourced validation and updating of dynamic features in openstreetmap an analysis of shelter mapping after the 2015 nepal earthquake. In *Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management ISCRAM 2016*, Rio de Janeiro, Brazil, 2016.

[15] J. Arnold, R. Srinivasan, R. Muttiah, and J. Williams. Large area hydrologic modeling and assessment. Part I: Model development. *Journal of the American Water Resources Association*, 34(1):73–89, 1998.

[16] R. Axelrod. Advancing the art of simulation in the social sciences. In *Simulating social phenomena*, pages 21–40. Springer, 1997.

[17] R. Axelrod. *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton University Press, 1997.

[18] R. Axelrod. Dissemination of culture: A model of local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226, 1997.

[19] R. Axelrod. Advancing the art of simulation in the social sciences. In J.-P. Rennard, editor, *Handbook of Research on Nature Inspired Computing for Economy and Management*. Hersey, PA: Idea Group, 2005.

[20] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[21] R. L. Axtell, J. M. Epstein, J. S. Dean, G. J. Gumerman, A. C. Swedlund, J. Harburger, S. Chakravarty, R. Hammond, J. Parker, and M. Parker. Population growth and collapse in a multiagent model of the kayenta anasazi in long house valley. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7275–7279, 2002.

[22] A. Banos, N. Corson, E. Daudé, B. Gaudou, and S. Rey Coyrehourcq. Macro Models, Micro Models and Network-based Coupling. In A. Banos, C. Lang, and N. Marilleau, editors, *Simulation spatiale à base d'agents avec NetLogo*, volume 2, chapter C, pages 63–84. ISTE editions – Elsevier, 2016.

[23] A. Banos, N. Corson, B. Gaudou, V. Laperrière, and S. Rey Coyrehourcq. Couplage de modèles dynamiques micro-macro sur réseaux : Application à la propagation de maladies (Modèles et Analyses Réseau : Approches Mathématiques et Informatiques (MARAMI), Paris, France). 2014.

[24] A. Banos, N. Corson, B. Gaudou, V. Laperrière, and S. Rey Coyrehourcq. MicMac, épidémie dans un réseau de villes, systèmes d'équations différentielles et réseaux. In C. MAPS, editor, *Recueil de fiches pédagogiques du réseau MAPS - Modélisation multi-Agents appliquée aux Phénomènes Spatialisés - 2009 - 2014*, pages 107–135. Collectif MAPS, juillet 2014.

[25] A. Banos, N. Corson, B. Gaudou, V. Laperrière, and S. Rey Coyrehourcq. The Importance of Being Hybrid for Spatial Epidemic Models: A Multi-Scale Approach. *Systems*, 3(4):309–329, 2015.

[26] A. Banos, N. Corson, B. Gaudou, V. Laperrière, and S. Rey Coyrehourcq. Coupling Micro and Macro Dynamics Models on Networks: Application to Disease Spread. In B. Gaudou and J. Sichman, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, number 9568 in Lecture Notes in Computer Science, pages 19–33, Istanbul, Turkey, mars 2016. Springer.

[27] A. Banos, N. Corson, C. Lang, N. Marilleau, and N. Taillandier. Multiscale Modeling: Application to Traffic Flow. In A. Banos, C. Lang, and N. Marilleau, editors, *Agent-Based Spatial Simulation with NetLogo*, volume 2, chapter B. ISTE editions – Elsevier, 2016.

[28] A. Banos, A. Drogoul, B. Gaudou, Q. N. Huynh, T. Q. Truong, and D. A. Vo. Tools and Models for Understanding and Exploring Urban Spatial Dynamics. In S. Lagrée, editor, *A Glance at Sustainable Urban Development - Methodological, Crosscutting and Operational Approaches*, Regional Social Sciences Summer University, chapter 2.2, pages 173–200. Journées de Tam Dao, july 2015.

[29] A. Banos and L. Sanders. Modéliser et simuler les systèmes spatiaux en géographie. In F. Varenne and M. Silberstein, editors, *Modéliser & simuler*, volume 2, page 2. Edition Matériologiques, 2013.

[30] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1–4):69–77, 2000.

[31] M. Barthélémy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[32] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[33] N. Becu, F. Amblard, N. Brax, B. Gaudou, and N. Marilleau. How to Involve Stakeholders in the Modeling Process,. In A. Banos, C. Lang, and N. Marilleau, editors, *Agent-Based Spatial Simulation with NetLogo*, volume 1, chapter 6, pages 223–252. ISTE editions – Elsevier, http://iste-editions.fr/, 2015.

[34] J. Bergez, H. Raynal, and F. Garcia. Record: an open platform to build, evaluate and simulate integrated models of farming and agroecosystems. In *Proceedings of IFSA*, 2012.

[35] E. Blanchart, C. Cambier, C. Canape, B. Gaudou, T. N. Ho, T. V. Ho, C. Lang, F. Michel, N. Marilleau, and L. Philippe. EPIS: a grid platform to ease and optimize multi-agent simulators runnning (short paper). In Y. Demazeau, M. Pechoucek, J. M. Corchado, and J. B. Pérez, editors, *International Conference on Practical Applications of Agents and Multiagent Systems (PAAMS), Salamanca, Spain*, volume 88, pages 129–134. Springer, 2011.

[36] T. Blochwitz, M. Otter, J. L. Åkesson, M. Arnold, C. Clauss, H. Elmqvist, M. Friedrich, A. Junghanns, J. Mauss, D. Neumerkel, H. Olsson, and A. Viel. Functional mockup interface 2.0: The standard for tool independent exchange of simulation models. In *9th International Modelica Conference In Proceedings of the 9th International Modelica Conference*, pages 173–184, 2012.

[37] A. Bond and L. Gasser. *Readings in Distributed Artificial Intelligence*. Morgan Kaufman, 1988.

[38] D. Booth and C. K. Liu. *Web services description language (wsdl) version 2.0*, 2007.

[39] R. Bordini, J. Hübner, and M. Wooldridge. *Programming multi-agent systems in AgentSpeak using Jason*. Wiley-Interscience, 2007.

[40] A. Boucher, R. Canal, T.-Q. Chu, A. Drogoul, B. Gaudou, V. T. Le, V. Moraru, N. V. Nguyen, V. Q. A. Nguyen, P. Taillandier, F. Sempé, and S. Stinckwich. The AROUND project: Adapting robotic disaster response to developing countries. In *IEEE International Workshop on Safety, Security & Rescue Robotics (SSRR), Denver, USA*, pages 1–6. IEEExplore digital library, 2009.

[41] S. Bowles and H. Gintis. Prosocial emotions. Working Paper 02-07-028, Santa Fe Institute, 2003.

[42] M. Bratman. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press, 1987.

[43] A. Bretagnolle and D. Pumain. Simulating urban networks through multiscalar space-time dynamics (Europe and United States, 17th–20th centuries). *Urban Studies*, 47(13):2819–2839, 2010.

[44] G. W. Brown. Iterative solutions of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York, 1951.

[45] P. Busetta, R. Ronnquist, A. Hodgson, and A. Lucas. Jack intelligent agents-components for intelligent agents in java. *AgentLink News Letter*, 2:2–5, 1999.

[46] A. Bénier. Génération de simulation multi-agents à partir de données géo-historiques. Master's thesis, Université Toulouse III Paul Sabatier, 2015.

[47] P. Caillou, B. Gaudou, A. Grignard, C. Q. Truong, and P. Taillandier. A Simple-to-use BDI architecture for Agent-based Modeling and Simulation. In *The Eleventh Conference of the European Social Simulation Association (ESSA 2015)*, Groningen, Netherlands, 2015.

[48] N. Cazin, A. Histace, D. Picard, and B. Gaudou. On The Joint Modeling of The Behavior of Social Insects and Their Interaction With Environment by Taking Into Account Physical Phenomena Like Anisotropic Diffusion. In J. Bajo, K. Hallenborg, P. Pawlewski, V. Botti, and N. Sánchez-Pi, editors, *Highlights of Practical Applications of Agents, Multi-Agent Systems, and Sustainability - The PAAMS Collection*, volume 524 of *Communications in Computer and Information Science*, pages 151–164, Salamanca, Spain, 2015. Springer International Publishing.

[49] F. Cecconi, M. Campenni, G. Andrighetto, and R. Conte. What do agent-based and equation-based modelling tell us about social conventions: The clash between ABM and EBM in a congestion game framework. *Journal of Artificial Societies and Social Simulation*, 13(1):6, 2010.

[50] W. Chapotat, L. Houssou, A. Bouadjio Boulic, N. Maestripieri, E. Lerigoleur, B. Gaudou, and M. Saqalli. An Agent-Based Model of the Amazonian Forest Colonisation and Oil Exploitation: the Oriente Study Case (poster). In S. Sauvage, J.-M. Sanchez Perez, and A. E. Rizzoli, editors, *International Environmental Modelling and Software Society (iEMSs), Toulouse, France*, volume 5, pages 1335–1335. International Environmental Modelling & Software Society, juillet 2016.

[51] M. Choisy, A. Drogoul, B. Gaudou, N. Marilleau, D. Philippon, T. Q. Truong, and D. A. Vo. Epidemiological Risks and Integration of Health Policies on a Regional Scale: Modelling to Make Better Decisions. In S. Lagrée, editor, *Shared Challenges for Development within ASEAN - Applied and Analytical Methods*, Regional Social Sciences Summer University, chapter 2.2, pages 167–207. Knowledge Publishing House, july 2016.

[52] T. Q. Chu. *Using agent-based models and machine learning to enhance spatial decision support systems : application to resource allocation in situations of urban catastrophes*. PhD thesis, Université Pierre et Marie Curie, 2011.

[53] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2–3), 1990.

[54] J.-P. Cointet and C. Roth. How realistic should knowledge diffusion models be? *Journal of Artificial Societies and Social Simulation*, 10(3):online, 2007. <http://jasss.soc.surrey.ac.uk/10/3/5.html>.

[55] V. Colizza and A. Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *Journal of Theoretical Biology*, 251:450–467, 2008.

[56] N. Corson and D. Olivier. Dynamical Systems with NetLogo. In A. Banos, C. Lang, and N. Marilleau, editors, *Agent-Based Spatial Simulation with NetLogo*, volume 1, chapter E, pages 183–221. ISTE editions – Elsevier, 2015.

[57] J. S. Dahmann. High level architecture for simulation. In *Proceedings of the First International Workshop on Distributed Interactive Simulation and Real-Time Applications*, pages 9–14, Singapore, 1997.

[58] A. R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam Pub Group, 1994.

[59] D. DeAngelis and L. Gross. *Individual-Based Models and Approaches in Ecology*. Chapman and Hall, New York, 1992.

[60] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4), 2002.

[61] E. Delay. *Geographical investigations about the use of agent based model in the understanding of the evolution in steep slopes vineyard areas : Côte Vermeille and val di Cembra case*. PhD thesis, Université de Limoges, 2015.

[62] R. Descartes. *Les passions de l'âme*. Livre de Poche, LGF, 1649.

[63] A. Drogoul, E. Amouroux, P. Caillou, B. Gaudou, A. Grignard, N. Marilleau, P. Taillandier, M. Vavasseur, D. A. Vo, and J.-D. Zucker. GAMA: multi-level and complex environment for agent-based models and simulations (demonstration) . In T. Ito, C. Jonker, M. Gini, and O. Shehory, editors, *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1361–1362, Saint-Paul, MN, USA, 2013. IFAAMAS.

[64] A. Drogoul, V. Diane, and T. Meurisse. Multi-agent based simulation: Where are the agents? In *Proceedings of MABS'02 (Multi-Agent Based Simulation)*, LNCS, Bologna, Italy, 2002. Springer-Verlag.

[65] A. Drogoul and B. Gaudou. Methods for Agent-Based Computer Modelling. In S. Lagrée, editor, *Water and its Many Issues. Methods and Cross-cutting Analysis.*, Regional Social Sciences Summer University, chapter 1.6, pages 130–154. Journées de Tam Dao, 2013.

[66] A. Drogoul, B. Gaudou, N. Gasmi, A. Grignard, P. Taillandier, O. Tessier, and D. A. Vo. Understanding Past Crises to Better Manage the Present: Initiation to Geo-Historical Risk Modelling (The 1926 Red River Swelling). In S. Lagrée, editor, *Perception and Risk Management - Methodological approaches applied to development*, Regional Social Sciences Summer University, chapter 2.4, pages 299–336. Journées de Tam Dao, juillet 2014.

[67] A. Drogoul, N. Quang Huynh, and Q. C. Truong. Coupling Environmental, Social and Economic Models to Understand Land-Use Change Dynamics in the Mekong Delta. *Frontiers in Environmental Science*, 4:19, Mars 2016.

[68] J. Drury and C. Cocking. The mass psychology of disasters and emergency evacuations: A research report and implications for practice. Technical report, University of Sussex, 2007.

[69] J. Drury, C. Cocking, and S. Reicher. Everyone for themselves? a comparative study of crowd solidarity among emergency survivors. *British Journal of Social Psychology*, 48:487–506, 2009.

[70] B. Edmonds and S. Moss. From KISS to KIDS: an anti-simplistic modelling approach. In P. Davidsson, editor, *International Workshop on Multi-Agent-Based Simulation*, volume 3415 of *LNAI*, pages 130–144. Springer, 2005.

[71] J. F. Ehmke, D. Großhans, D. C. Mattfeld, and L. D. Smith. Interactive analysis of discrete-event logistics systems with support of a data warehouse. *Computers in Industry*, 62(6):578–586, 2011.

[72] J. M. Epstein and R. Axtell. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[73] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[74] R. Falcone and C. Castelfranchi. Social trust: cognitive anatomy, social importance, quantification and dynamics. In *Autonomous Agents '98 Workshop on "Deception, Fraud and Trust in Agent Societes"*, pages 35–49. Minneapolis, USA, 1998.

[75] E. Fehr and S. Gaechter. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181, 2000.

[76] FIPA. FIPA Communicative Act Library Specification. http://www.fipa.org/specs/fipa00037/, Foundation for Intelligent Physical Agents, 2002.

[77] P. A. Fishwick. Computer simulation: growth through extension. *Trans. Soc. Comput. Simul.*, 14:13–24, 1997.

[78] P. Fosset, A. Banos, E. Beck, S. Chardonnel, C. Lang, N. Marilleau, A. Piombini, T. Leysens, A. Conesa, and I. Andre-Poyaud. Exploring Intra-Urban Accessibility and Impacts of

Pollution Policies with an Agent-Based Simulation Platform: GaMiroD. *Systems*, 4(1):5, 2016.

[79] R. Frigg and S. Hartmann. Models in science. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information (CSLI), Stanford University, fall 2012 edition, 2012.

[80] M. Gardner. The fantastic combinations of John Conway's new solitaire game of Life. *Scientific American*, 223:120–123, 1970.

[81] N. Gasmi, A. Grignard, A. Drogoul, B. Gaudou, P. Taillandier, O. Tessier, and D. A. Vo. Reproducing and exploring past events using agent-based geo-historical models. In E. Norling and F. Grimaldo, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, volume 9002 of *Lecture Notes in Computer Science*, pages 151–163, Paris, France, 2014. Springer-Verlag.

[82] B. Gaudou. *Formalizing social attitudes in modal logic*. Thèse de doctorat, Université de Toulouse, juillet 2008.

[83] B. Gaudou, V. Laperrière, N. Marilleau, D. Sheeren, and S. Rey Coyrehourcq. Metafish – modélisation métapopulationnelle d'une ressource halieutique, 2014. MAPS7.

[84] B. Gaudou, E. Lorini, and E. Mayor. Moral Guilt: An Agent-Based Model Analysis. In *Conference of the European Social Simulation Association (ESSA), Warsaw*, volume 229 of *Advances in Intelligent Systems and Computing*, pages 95–106. Springer, 2013.

[85] B. Gaudou, C. Sibertin-Blanc, O. Thérond, F. Amblard, Y. Auda, J.-P. Arcangeli, M. Balestrat, M.-H. Charron-Moirez, E. Gondet, Y. Hong, R. Lardy, T. Louail, E. Mayor, D. Panzoli, S. Sauvage, J. Sanchez-Perez, P. Taillandier, V. B. Nguyen, M. Vavasseur, and P. Mazzega. The MAELIA multi-agent platform for integrated assessment of low-water management issues. In S. J. Alam and H. V. D. Parunak, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, volume 8235 of *Lecture Notes in Computer Science*, pages 85–110, Saint-Paul, MN, USA, 2014. Springer.

[86] N. Gilbert and K. Troitzsch. *Simulation for the social scientist*. McGraw-Hill Education (UK), 2005.

[87] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22:403–434, 1976.

[88] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.

[89] A. Grignard. *Modèles de visualisation à base d'agents*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2015.

[90] A. Grignard, A. Drogoul, and J.-D. Zucker. Online analysis and visualization of agent based models. In *ICCSA*, pages 662–672, Ho Chi Minh City, Vietnam, 2013.

[91] A. Grignard, P. Taillandier, B. Gaudou, D. A. Vo, N. Q. Hyunh, and A. Drogoul. GAMA 1.6: Advancing the Art of Complex Agent-Based Modeling and Simulation. In G. Boella, E. Elkind, B. T. R. Savarimuthu, F. Dignum, and M. K. Purvis, editors, *Pacific Rim international Conference on Multi-Agents (PRIMA), Dunedin, New Zealand*, volume 8291 of *Lecture Notes in Computer Science*, pages 117–131. Springer, 2013.

[92] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1—2):115–126, 2006.

[93] G. Gröger, T. H. Kolbe, C. Nagel, and K.-H. Häfele. OpenGIS city geography markup language (CityGML) encoding standard. Technical report, Open Geospatial Consortium, 2012.

[94] T. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

[95] T. Grüne-Yanoff and P. Weirich. The Philosophy and Epistemology of Simulation: A Review. *Simulation & Gaming*, 41(1):20–50, jan 2010.

[96] C. Hanachi and I. Khaloul. Découvertes de protocoles et de structures organisationnelles dans le workflow. In *international conference NOuvelles TEchnologie de la Repartition, NOTERE*, pages 23–27, Lyon, France, 2008.

[97] I. A. Hanski and M. E. Gilpin, editors. *Metapopulation Biology: Ecology, Genetics, and Evolution.* Academic Press, Waltham, Massachusetts, USA, 1997.

[98] J. C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63:309–321, 1955.

[99] J. C. Harsanyi. Morality and the theory of rational behaviour. In A. K. Sen and B. Williams, editors, *Utilitarianism and beyond*, pages 39–62. Cambridge University Press, Cambridge, 1982.

[100] P. Hartigan and Z. Anderson. Ebola crisis support - hackathon report. Technical report, Oxford Launchpad, Saïd Business School, University of Oxford, 2014.

[101] S. Hassan, J. Pavón, L. Antunes, and N. Gilbert. Injecting data into agent-based simulation. In *Simulating Interacting Agents and Social Phenomena*, pages 177–191. Springer Japan, 2010.

[102] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[103] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Physical Review E*, 51(5):4282–4286, 1995.

[104] D. R. Hild. *Discrete Event System Specification (DEVS) Distributed Object Computing (DOC) Modeling and Simulation.* PhD thesis, University of Arizona, 2000.

[105] W. Inmon. *Building the Data Warehouse.* Wiley Publishing Inc., 2005.

[106] P. Jaccard. Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 44(163):223–270, 1908.

[107] B. Jarvis and L. Jain. Trust in LORA: Towards a Formal Definition of Trust in BDI Agents. In B. Gabrys, R. J. Howlett, and L. C. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems: 10th International Conference, KES 2006, Bournemouth, UK, October 9-11, 2006. Proceedings, Part II*, pages 458–463. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[108] K. Kahn and H. Noble. The Modelling4All Project - A web-based modelling tool embedded in Web 2.0. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques (SIMUTools'09)*, 2009.

[109] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, 115(772):700–721, 1927.

[110] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. *The Journal of hygiene*, 39(3):271–288, 1939.

[111] T. K. Kim and H. S. Seo. A trust model using fuzzy logic in wireless sensor network. In *Proceedings of world academy of science, engineering and technology*, volume 32, 2008.

[112] K. Kravari and N. Bassiliades. A survey of agent platforms. *Journal of Artificial Societies and Social Simulation*, 18(1):11, 2015.

[113] P. Langlois, E. Daudé, B. Blanpain, and E. Sapin. AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie. In *Sageo, atelier OntoGéo*, 2010.

[114] R. Lardy, P. Mazzega, C. Sibertin-Blanc, Y. Auda, J.-M. Sanchez-Perez, S. Sauvage, and O. Therond. Calibration of simulation platforms including highly interweaved processes: the maelia multi-agent platform. In *Proceedings of the 7th International Congress on Environmental Modelling and Software, June*, pages 15–19, 2014.

[115] V. M. Le, C. Adam, R. Canal, B. Gaudou, T. V. Ho, and P. Taillandier. Simulation of the Emotion Dynamics in a Group of Agents in an Evacuation Situation. In N. Desai, A. Liu, and M. Winikoff, editors, *Pacific Rim international Conference on Multi-Agents (PRIMA), Kolkata, India*, volume 7057 of *Lecture Notes in Computer Science (LNCS)*, pages 604–619. Springer, 2012.

[116] V. M. Le, B. Gaudou, P. Taillandier, and D. A. Vo. A New BDI Architecture To Formalize Cognitive Agent Behaviors Into Simulations. In D. Barbucha, M. T. Le, R. J. Howlett, and C. J. Lakhmi, editors, *Advanced Methods and Technologies for Agent and Multi-Agent Systems (KES-AMSTA), Hue, Vietnam*, volume 252 of *Frontiers in Artificial Intelligence and Applications*, pages 395–403. IOS Press, 2013.

[117] C. Le Page, N. Becu, P. Bommel, and F. Bousquet. Participatory agent-based simulation for renewable resource management: The role of the cormas simulation platform to nurture a community of practice. *Journal of Artificial Societies and Social Simulation*, 15(1):10, 2012.

[118] S. Legendre, J. Clobert, A. P. Møller, and G. Sorci. Demographic stochasticity and social mating system in the process of extinction of small populations: the case of passerines introduced to new zealand. *The American Naturalist*, 153(5):449–463, 1999.

[119] A. J. Lotka. *Elements of Physical Biology*. Williams & Wilkins Company, 1925.

[120] C. M. Macal and M. J. North. Tutorial on agent-based modeling and simulation. In *37th Winter Simulation Conference. Introductory tutorials: agent-based modeling*, pages 2–15, 2005.

[121] H. Madeira, J. Costa, and M. Vieira. The OLAP and data warehousing approaches for analysis and sharing of results from dependability evaluation experiments. In *IEEE/IFIP International Conference on Dependable Systems and Networks, Dependable Computing and Communications, DSN-DCC*, pages 22–25, 2003.

[122] H. Mahboubi, T. Faure, S. Bimonte, G. Deffuant, J.-P. Chanet, and F. Pinet. A multidimensional model for data warehouses of simulation results. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, pages 1–19, 2010.

[123] C. A. Manore, K. S. Hickmann, J. M. Hyman, I. M. Foppa, J. K. Davis, D. M. Wesson, and C. N. Mores. A network-patch methodology for adapting agent-based models for directly transmitted disease to mosquito-borne disease. *Journal of Biological Dynamics*, 9(1):52–72, 2015.

[124] M. Matsumura, H. Takeuchi, M. Satoh, S. Sanada-Morimura, A. Otuka, T. Watanabe, and D. Thanh. Current status of insecticide resistance in rice planthoppers in asia. *Planthoppers: New threats to the sustainability of intensive rice production systems in Asia*, pages 233–244, 2009.

[125] S. J. Mellor. *MDA Distilled, Principles of Model Driven Architecture*. Addison-Wesley Professional, 2004.

[126] S. Meloni, N. Perra, A. Arenas, S. Gomez, Y. Moreno, and A. Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports*, 1, 62(7):1–7, 2011.

[127] M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: Can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 71–79, New York, NY, USA, 2010. ACM.

[128] M. Minsky. Matter, mind and models. In *IFIP Congress*, pages 45–50. Spartan Books, 1965.

[129] M. Mitchell. *Complexity: A guided tour*. Oxford University Press, 2009.

[130] N. H. Ngo. *Principles and Applications in Mathematical model for Biological Studies, Agriculture and Environment*. Vietnam: Agriculture Publishing House, 2008.

[131] M. H. Nguyen. *A Logical Framework for Trust-Related Emotions: Formal and Behavioral Results*. PhD thesis, University of Toulouse, 2010.

[132] N. D. Nguyen. *Coupling Equation-Based and Individual-Based Models in the Study of Complex Systems – A Case Study in Theoretical Population Ecology*. PhD thesis, University Pierre and Marie Curie, 2010.

[133] T. K. Nguyen, B. Gaudou, T. V. Ho, and N. Marilleau. Application of PAMS Collaboration Platform to Simulation-Based Researches in Soil Science: The Case of the MIcro-ORganism Project. In T. Cao, R.-D. Kutsche, and A. Demaille, editors, *International Conference on Computing and Communication Technologies, Da Nang, Vietnam*, pages 65–72. IEEExplore digital library, 2009.

[134] V. Q.-A. Nguyen. *Cohérence et robustesse dans un système multiagent perturbé : application à un système décentralisé de collecte d'information distribué*. PhD thesis, Université de Lyon 1, 2012.

[135] V. Q. A. Nguyen, R. Canal, B. Gaudou, S. Hassas, and F. Armetta. TrustSets - Using trust to detect deceitful agents in a distributed information collecting system. *Journal of Ambient Intelligence and Humanized Computing*, 3(4):251–263, 2012.

[136] V. Q. A. Nguyen, B. Gaudou, R. Canal, and S. Hassas. Coherence and Robustness in a Disturbed MAS (short paper). In T. Cao, R.-D. Kutsche, and A. Demaille, editors, *International Conference on Computing and Communication Technologies (RIVF), Da Nang, Vietnam*, pages 1–4. IEEExplore digital library, 2009.

[137] V. Q. A. Nguyen, B. Gaudou, R. Canal, S. Hassas, and F. Armetta. TrustSets - Using trust to detect deceitful agents in a distributed information collecting system. In *IEEE International Conference on Research, Innovation and Vision for the Futur (RIVF), Hanoi, Vietnam*, pages 1–6. IEEE, 2010.

[138] V. Q. A. Nguyen, B. Gaudou, R. Canal, S. Hassas, and F. Armetta. A Cluster-Based Approach for Disturbed, Spatialized, Distributed Information Gathering Systems. In N. Desai, A. Liu, and M. Winikoff, editors, *Pacific Rim international Conference on Multi-Agents (PRIMA), Kolkata, India*, volume 7057 of *Lecture Notes in Computer Science (LNCS)*, pages 588–603. Springer, 2012.

[139] V. Q. A. Nguyen, S. Hassas, F. Armetta, B. Gaudou, and R. Canal. Combining Trust and Self-Organization for Robust Maintaining of Information Coherence in Disturbed MAS. In *IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO), Ann Arbor, MI, USA*, pages 178–187. IEEE, 2011.

[140] V. Q. A. Nguyen, S. Hassas, B. Gaudou, R. Canal, F. Armetta, and M. H. Nguyen. Couplage de dynamiques, auto-organisation et confiance dans un système multi-agents perturbé. In S. Hassas, editor, *Journées Francophones sur les Systèmes Multi-Agents (JFSMA), Lille*, 2013.

[141] V. T. Nguyen, D. Longin, T. V. Ho, and B. Gaudou. Integration of emotion in evacuation simulation. In C. Hanachi, F. Benaben, and F. Charoy, editors, *Information Systems for Crisis Response and Management in Mediterranean countries (ISCRAM MED), Toulouse, France*, number 196 in Lecture Notes in Business Information Processing (LNBIP), pages 192–205. Springer, 2014.

[142] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu. Using of Jaccard Coefficient for Keywords Similarity. In *International MultiConference of Engineers and Computer Scientists*, pages 280–384, 2013.

[143] J.-M. Nolot and P. Debaeke. Principes et outils de conception, conduite et évaluation de systèmes de culture. *Cahiers Agricultures*, 12(6):387–400, 2003.

[144] E. Norling. Folk psychology for human modeling: extending the BDI paradigm. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent System (AAMAS)*, New York, 2004.

[145] M. J. North, N. T. Collier, J. Ozik, E. R. Tatara, C. M. Macal, M. Bragen, and P. Sydelko. Complex adaptive systems modeling with repast simphony. *Complex adaptive systems modeling*, 1(1):1–26, 2013.

[146] M. of Natural Resources and Environment. Detailing the establishment, regulation and evaluation planning, land-use planning, 2009.

[147] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions.* Cambridge University Press, Cambridge, MA, 1998.

[148] E. Ostrom. A general framework for analyzing sustainability of social-ecological systems. *Science*, 325(5939):419–422, 2009.

[149] P. Parham and E. Michael. Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, 118(5):620–626, 2010.

[150] D. Parker, D. G. Brown, J. G. Polhill, P. J. Deadman, and S. M. Manson. Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: the mr potatohead framework. In *Agent-based modelling in natural resource management*, pages 23–51, 2008.

[151] H. V. D. Parunak, R. Savit, and R. L. Riolo. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *First International Workshop on Multi-Agent Systems and Agent-Based Simulation (MABS)*, pages 10–25, 1998.

[152] J. Passerat-Palmbach, M. Leclaire, R. Reuillon, Z. Wang, and D. Rueckert. OpenMOLE: a Workflow Engine for Distributed Medical Image Analysis. In *International Workshop on High Performance Computing for Biomedical Image Analysis (part of MICCAI 2014)*, Boston, United States, 2014.

[153] D. Philippon. Réalisation d'un modèle épidémiologique avec prise en compte des politiques de santé entre différents pays. Master's thesis, Université Nationale du Vietnam, Institut Francophone International, 2015.

[154] D. Philippon, M. Choisy, A. Drogoul, B. Gaudou, N. Marilleau, P. Taillandier, and C. Q. Truong. Exploring trade and health policies influence on dengue spread with an agent-based model. In L. Antunes and L. G. Nardin, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, Lecture Notes in Computer Science, Singapore, 2016. Springer. to appear.

[155] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

[156] J. Polhill and N. Gotts. Ontologies for transparent integrated human-natural system modelling. *Landscape Ecology*, 24:1–5, 2006.

[157] W. Poundstone. *Prisoner's Dilemma*. Doubleday, NY NY, 1992.

[158] R. Power, B. Robinson, J. Colton, and M. Cameron. Emergency situation awareness: Twitter case studies. In *Information Systems for Crisis Response and Management in Mediterranean Countries (ISCRAM-Med)*, pages 218–231. Springer, 2014.

[159] W. H. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, 3rd edition, 2007.

[160] E. Quarantelli. The nature and conditions of panic. *American Journal of Sociology*, 60(3):267–275, 1954.

[161] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *KR'91*, pages 473–484. Morgan Kaufmann, 1991.

[162] J. Rawls. *A theory of Justice*. Harvard University Press, Cambridge, 1971.

[163] S. Rey-Coyrehourcq. *Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation multi-agents*. PhD thesis, Université Paris 1 Panthéon-Sorbonne, 2015.

[164] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54(2):296–301, 1951.

[165] S. Roy Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, and C. Castillo. Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *10th International ISCRAM Conference*, 2013.

[166] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach.* Upper Saddle River (New Jersey), 1995.

[167] T. Sakaki, M. Okazaki, and Y. Matsuo. Tweet analysis for real-time event detection and earthquake reporting system development. *Knowledge and Data Engineering, IEEE Transactions on*, 25(4):919–931, 2013.

[168] T. C. Schelling. Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1:143–186, 1971.

[169] M. Schillo, P. Funk, and M. Rovatsos. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence, Special Issue on Trust, Deception and Fraud in Agent Societies*, 14(8):825–848, September 2000.

[170] J. R. Searle. *Speech acts: An essay in the philosophy of language.* Cambridge University Press, New York, 1969.

[171] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, 1976.

[172] C. Sibertin-Blanc, O. Therond, C. Monteil, and P. Mazzega. Formal modeling of social-ecological systems. In *7th International Conference of the European Social Simulation Association (ESSA 2011)*, Montpellier, France, 2011.

[173] J. Sosnowski, P. Zygulski, and P. Gawkowski. Developing data warehouse for simulation experiments. In M. Kryszkiewicz, J. F. Peters, H. Rybinski, and A. Skowron, editors, *Proceedings of Rough Sets and Intelligent Systems Paradigms (RSEISP 2007)*, pages 543–552, Warsaw, Poland, 2007. Springer Berlin Heidelberg.

[174] Spinoza. *L'éthique.* Folio, Gallimard, 1994.

[175] K. Starbird and L. Palen. "voluntweeters": Self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1071–1080, New York, NY, USA, 2011. ACM.

[176] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 241–250, New York, NY, USA, 2010. ACM.

[177] R. Sun, E. Merrill, and T. Peterson. From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, 25(2):203–244, 2001.

[178] X. H. Ta, D. Longin, B. Gaudou, and T. V. Ho. Impact of group on the evacuation process - Theory and simulation. In L. De Raedt, Y. Deville, M. Bui, and T. T. D. Lin, editors, *Symposium on Information and Communication Technology (SoICT), Hue, Vietnam*, pages 350–357. ACM DL, 2015.

[179] P. Taillandier, A. Banos, A. Drogoul, B. Gaudou, N. Marilleau, and C. Q. Truong. Simulating urban growth with raster and vector models: A case study for the city of can tho, vietnam. In P. Perez, L. Padgham, K. Nagel, A. L. C. Bazzan, and M.-R. Namazi-Rad, editors, *Agent Based Modelling of Urban Systems (ABMUS 2016)*, pages 154–171, 2016.

[180] P. Taillandier, M. Bourgais, P. Caillou, C. Adam, and B. Gaudou. A situated BDI agent architecture for the GAMA modelling and simulation platform. In L. Antunes and L. G. Nardin, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, Lecture Notes in Computer Science, Singapore, 2016. Springer. to appear.

[181] P. Taillandier, A. Grignard, B. Gaudou, and A. Drogoul. Des données géographiques à la simulation à base d'agents: application de la plate-forme gama. *Cybergeo: European Journal of Geography*, Systems, Modelling, Geostatistics(671):(on line), mars 2014.

[182] P. Taillandier, O. Thérond, and B. Gaudou. Une architecture d'agent BDI basée sur la théorie des fonctions de croyance : application à la simulation du comportement des agriculteurs (short paper). In P. Chevaillier and B. Mermet, editors, *Journées Francophones sur les Systèmes Multi-Agents (JFSMA), Honfleur, France, 17/10/2012-19/10/2012*, pages 107–116. Cépaduès Editions, 2012.

[183] P. Taillandier, O. Thérond, and B. Gaudou. A new BDI agent architecture based on the belief theory. Application to the modelling of cropping plan decision-making. In R. Seppelt, A. A. Voinov, S. Lange, and D. Bankamp, editors, *International Environmental Modelling and Software Society (iEMSs), Leipzig, Germany*, pages 2463–2470. International Environmental Modelling and Software Society, 2012.

[184] B. Teague, R. McLeod, and S. Pascoe. Final report - volume 4: the statements of lay witnesses. Technical report, Victorian Bushfires Royal Commission, 2009.

[185] B. Teague, R. McLeod, and S. Pascoe. Final report, volume 1: The fires and the fire-related deaths. Technical report, Victorian Bushfires Royal Commission, 2009.

[186] O. R. Teran Villegas, C. Sibertin-Blanc, and B. Gaudou. Identifying Emotions in Organizational Settings: towards dealing with morality. In B. Duval and J. Van Den Herik, editors, *International Conference on Agents and Artificial Intelligence (ICAART), Angers (France)*, pages 284–292. SciTePress, 2014.

[187] O. R. Teran Villegas, C. Sibertin-Blanc, and B. Gaudou. The influence of moral sensitivity on organizational cooperation. *Kybernetes*, 44(6/7):1067–1081, septembre 2015.

[188] T. Terpstra, A. de Vries, G. Paradies, and R. Stronkman. Towards a realtime twitter analysis during crises for operational crisis management. In *Proceedings of 9th International*

Conference on Information Systems for Crisis Response and Management (ISCRAM), Vancouver, Canada, 2012.

[189] L. Tesfatsion. Agent-based computational economics: growing economies from the bottom up. *Artificial Life*, 8:55–82, 2002.

[190] O. Thérond, C. Sibertin-Blanc, R. Lardy, B. Gaudou, M. Balestrat, Y. Hong, T. Louail, V. B. Nguyen, D. Panzoli, J.-M. Sanchez-Perez, S. Sauvage, P. Taillandier, M. Vavasseur, and P. Mazzega. Integrated modelling of social-ecological systems: The MAELIA high-resolution multi-agent platform to deal with water scarcity problems. In D. P. Ames and N. Quinn, editors, *International Environmental Modelling and Software Society (iEMSs), San Diego, California, USA, 15/06/2014-19/06/2014*, page (electronic medium), http://www.iemss.org, 2014. International Environmental Modelling and Software Society.

[191] J.-P. Treuil, A. Drogoul, and J.-D. Zucker. *Modélisation et simulation à base d'agents: exemples commentés, outils informatiques et questions théoriques.* Dunod, 2008.

[192] C. Q. Truong, P. Taillandier, B. Gaudou, Q. M. Vo, T. H. Nguyen, and A. Drogoul. Exploring agent architectures for farmer behavior in land-use change. A case study in coastal area of the Vietnamese Mekong Delta. In B. Gaudou and J. Sichman, editors, *International Workshop on Multi-Agent-Based Simulation (MABS)*, number 9568 in Lecture Notes in Computer Science, pages 146–158, Istanbul, Turkey, mars 2016. Springer.

[193] M. T. Truong. *To Develop a Database Management Tool for Multi-Agent Simulation Platform.* PhD thesis, University of Toulouse, 2015.

[194] M. T. Truong, F. Amblard, and B. Gaudou. Combination Framework of BI solution & Multi-agent platform (CFBM) for multi-agent based simulations (short paper). In *Atelier aide à la Décision à tous les Etages (AIDE), Toulouse*, pages 35–42. IRIT, 2013.

[195] M. T. Truong, F. Amblard, B. Gaudou, and C. Sibertin-Blanc. To Calibrate & Validate an Agent-Based Simulation Model - An Application of the Combination Framework of BI solution & Multi-agent platform. In B. Duval and J. van den Herik, editors, *International Conference on Agents and Artificial Intelligence (ICAART), Angers, France*, pages 172–183. SciTePress, 2014.

[196] M. T. Truong, F. Amblard, B. Gaudou, C. Sibertin-Blanc, V. X. Truong, A. Drogoul, X. H. Hyunh, and M. N. Le. An implementation of framework of business intelligence for agent-based simulation. In Q. T. Huynh, T. B. Nguyen, V. T. Do, M. Bui, and H. S. Ngo, editors, *Symposium on Information and Communication Technology (SoICT), Da Nang, Viet Nam*, pages 35–44. ACM, 2013.

[197] V. Truong. *Optimization by simulation of an environmental surveillance network - Application to the fight against rice pests in the Mekong delta (Vietnam).* PhD thesis, University Pierre & Marie Curie-Paris, 2014.

[198] V. Truong, H. Huynh, M. Le, and A. Drogoul. Modeling a surveillance network based on unit disk graph technique - application for monitoring the invasion of insects in mekong delta region. In *PRIMA 2012: Principles and Practice of Multi-Agent Systems*, pages 228–242, 2012.

[199] W. Van Der Aalst. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media, 2011.

[200] C. Vasilakis, E. El-Darzi, and P. Chountas. A decision support system for measuring and modelling the multi-phase nature of patient flow in hospitals. In *Intelligent Techniques and Tools for Novel System Architectures*, pages 201–217. Springer, 2008.

[201] L. Vercouter and G. Muller. L.I.A.R.: Achieving Social Control in Open and Decentralized Multiagent Systems. *Applied Artificial Intelligence*, 24:723–768, September 2010.

[202] P. Verduyn, E. Delvaux, H. Van Coillie, F. Tuerlinckx, and I. van Mechelen. Predicting the duration of emotional experience: two experience sampling studies. *Emotion*, 9(1):83, 2009.

[203] H. Visser and T. de Nijs. The map comparison kit. *Environmental Modelling & Software*, 21(3):346–358, 2006.

[204] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[205] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[206] U. Wilensky and I. Evanston. Netlogo. center for connected learning and computer based modeling. Technical report, Northwestern University, 1999.

[207] H. Wolda. Similarity indices, sample size and diversity. *Oecologia*, 50(3):296–302, 1981.

[208] S. Wolfram. *Theory and applications of cellular automata*. Singapore: World Scientific, 1986.

[209] M. Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, 2009.

[210] T. A. Zia. Reputation-based trust management in wireless sensor networks. In *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 163–166. IEEE, Dec. 2008.