KIMONO, a descriptive agent-based modelling method for the exploration of complex systems: an application to epidemiology.

Edouard Amouroux

▶ **To cite this version:**

Edouard Amouroux. KIMONO, a descriptive agent-based modelling method for the exploration of complex systems: an application to epidemiology.. Modeling and Simulation. Université Pierre et Marie Curie - Paris VI, 2011. English. NNT : . tel-00630779

HAL Id: tel-00630779
https://theses.hal.science/tel-00630779

Submitted on 5 Dec 2011

**THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité: Informatique

(Ecole doctorale: EDITE)

Présentée par
Edouard AMOUROUX

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

KIMONO: une méthode de modélisation descriptive centrée agent pour l'explication des systèmes complexes, une application en épidémiologie

soutenue le 30/09/2011

devant le jury composé de :

M. Directeur de thèse:	**Alexis Drogoul**,	Directeur de Recherche, IRD - UMMISCO / IFI - MSI
Rapporteurs:	**David Hill**,	Professeur des universités, Université Blaise Pascal, LIMOS UMR CNRS 6158
	François Roger,	Directeur de recherche, CIRAD, AGIRs Faculty of Veterinary Medicine, Kasetsart University
Examinateurs:	**Jean-François Perrot**	Professeur émérite, Université Pierre et Marie Curie
	Éric Ramat	Professeur Université du littoral Côte d'Opale
	Ho Tuong Vinh,	Professeur, Institut de la Francophonie pour l'Informatique (IFI) - Équipe de recherche MSI
	Nicolas Marilleau	Ingénieur de recherche, IRD - UMMISCO

75270-PARIS CEDEX 06 Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Abstract

Since a few years, in many different domains (e.g. sociology, ecology or economy), there is a growing trend in designing models that can be used for exploratory purposes rather than purely predictive ones. Loosely following a «complex systems» paradigm, exploratory models are often based on explicit and detailed representations of the components of the systems studied, and offer a large degree of freedom in the parameter or structural adjustments available to their final users (researchers, decision-makers or stakeholders).

They are intended to be used as a support of «as-if experiments» as they allow, through these adjustments, for the formulation of detailed hypotheses at various levels of description of the system. These hypotheses then lead to the generation of scenarios whose outcomes are explored and compared by way of repeated simulations.

Epidemiology is an interesting example of this situation. Its long modelling history can be characterised as a search for simple predictive models, but recent examples like the outbreaks of avian influenza in South-East Asia have shown their limits: without properly taking into account the interplays between social, ecological and biological dynamics to understand how this pandemic evolves, these models become useless as far as prediction goes. And when some of these dynamics are tentatively taken into account, the resulting models often become dependent on incomplete or qualitative data (for instance, the decision-making processes of social actors or the behaviours of birds), which prevents them to be used for any serious predicting purposes. As a consequence, there has been a recent shift of focus of the community on the design of exploratory models, which are meant to allow understanding the links between these dynamics, generating and studying various hypotheses, and measuring, with respect to these hypotheses, the impact of local or global policies in complex scenarios.

However, designing and using such models gives rise to serious methodological issues and existing modelling and simulation methods do not cope very well with this new type of models. When they are adapted to take their peculiarities into account, this often results in ad hoc solutions, which can barely be reused for other models in the same domain, let alone in different domains.

The objective of this thesis is to propose a domain-independent method (KIMONO) to facilitate the design and use of such exploratory models. Based on a series of examples from various domains (road traffic, social segregation, soil dynamics and more extensively epidemiology), I proceed from an account of the design requirements (taking conflicting and evolving hypotheses into account during the modelling process, producing highly-modular models, enabling an iterative modelling cycle, allowing for the collaboration of different experts and the combination of different formalisms, etc.) to a concrete proposal involving dedicated computer tools and a common accessible formalism, both aimed at facilitating the collaboration, communication and the implementation of «world models» (the name given in this proposal to open exploratory models).

The method I propose focuses on two elements: the implication of the experts and a detailed representation of the system.

Experts are at the centre of the modelling process, which starts with extensive descriptions of their knowledge, possibly reusing their formalism, and further proceeds through iterative amendments (of increasing or decreasing complexity) that they are able to evaluate and validate in interaction with the modellers. The iterative process comes to an end when the experts estimate that they have a sufficient insight of the system or when further investigations require field experiments.

Regarding the kind of representation suitable for supporting this process, I propose an adaptable and modular combination of two implementation systems: Agent-Based Modelling (ABM) and Geographical

Information Systems (GIS). I show that this combination provides for an arbitrary level of description of the components of a system, that it allows both qualitative and quantitative knowledge to be equally represented and that it supports a high level of evolution of the hypotheses during the modelling process. The interactions between modellers and experts are based on two abstractions of these implementation details, using both the ODD (Overview, Design concepts, Details) protocol for communication purposes, and the GAML modelling language for the collaborative programming of the model.

The method proposed has been applied and validated in the context of a large study undertaken in South-East Asia (especially North-Vietnam) by epidemiologists and veterinarians to understand the role of various hypotheses in explaining the recurring outbreaks of the avian influenza epidemics among domestic poultry. During a four years long interdisciplinary collaboration, several «world models» have been co-designed, implemented on the GAMA platform and used as «virtual laboratories» by experts. This collaboration, and its unique outcomes, have allowed them to test a broad range of hypotheses (especially on the local conditions of persistence), better understand the role of various spatial, ecological or social factors in the survival and propagation of the virus and reorient some of their field studies in consequence.

Résumé

Depuis plusieurs années, on peut observer une tendance, dans de nombreux domaines (Sociologie, Ecologie, Economie, etc.), à construire des modèles vers l'exploration des systèmes qu'ils représentent que vers la prédiction ou l'explication. S'inscrivant dans le paradigme des "systèmes complexes", ces modèles exploratoires utilisent généralement une représentation explicite et détaillée des composants du système étudié. Ils offrent aussi aux utilisateurs finaux (chercheurs, décideurs, parties prenantes) une grande liberté d'adaptation en termes de paramètres et de structure du modèle. Ces modèles servent de support à des expériences "as-if" en permettant, au travers de ces ajustements, la formulation d'hypothèses détaillées et ce à différents niveaux de descriptions du système. A partir de ces hypothèses, des scénarios sont construits, dont les résultats sont explorés et analysés grâce à des simulations répétées.

L'épidémiologie est exemplaire de cette situation. Elle a une longue histoire de modélisation qui peut être caractérisée comme la recherche de modèles essentiellement prédictifs. Cependant, certaines situations, comme l'émergence de foyers épidémiques de grippe aviaire en Asie du Sud Est, montre la limite d'une telle approche: sans une prise en compte adéquate des interactions entre les dynamiques sociales, écologiques et biologiques, l'utilisation de modèles prédictifs est sans fondement. De plus, dès lors que l'on tente de prendre en compte ces dynamiques, les modèles deviennent dépendants de données incomplètes ou qualitatives (le processus de décision des acteurs sociaux ou bien le comportement des oiseaux, par exemple). En conséquence, on assiste actuellement à un changement d'orientation de la communauté épidémiologique vers la conception de modèles plus exploratoires, mieux adaptés à la génération et à l'étude d'hypothèses variées, et mieux à même d'aider à mesurer, par rapport à ces hypothèses, l'impact des politiques locales et globales de lutte contre les épidémies dans le cadre de scénarios complexes.

Cependant, concevoir et utiliser de tels modèles souligne les sérieux problèmes méthodologiques auxquels ne peuvent réellement répondre les méthodologies de modélisation et simulation existantes. Et quand celles-ci ont été adaptées pour prendre ces spécificités en compte, il en résulte des solutions ad hoc qui ne peuvent être réutilisées ni dans le domaine en général, ni dans d'autres domaines.

L'objectif de cette thèse est de proposer une méthodologie (KIMONO) qui, sans être spécifique à un domaine particulier, facilite la conception et l'utilisation de ces modèles exploratoires. En partant d'une série d'exemples tirés de différents domaines (trafic routier, ségrégation sociale, dynamique du sol, mais aussi, et de façon plus extensive, épidémiologie), je commence par une caractérisation des besoins de conception (prise en compte d'hypothèses contradictoires et évolutives lors du processus de modélisation, génération de modèles extrêmement modulaires, rendant possible un cycle de modélisation itératif, permettant la collaboration entre différents experts et la combinaison de différents formalismes, etc.) pour aboutir à une proposition concrète impliquant un outil informatique dédié et un formalisme commun et accessible, orienté aussi bien vers la facilitation de la collaboration, la communication et l'implémentation de "modèles monde" (le nom donné dans ce document à ces modèles exploratoires ouverts).

La méthodologie que je propose se concentre sur deux éléments: l'implication des experts et la représentation détaillée du système.

Les experts sont au coeur de processus de modélisation. Celui-ci s'appuie sur une description étendue de leurs connaissances, potentiellement exprimées dans leurs propres formalismes, description qui est ensuite amendée de façon itérative (soit pour la complexifier, soit pour la simplifier) dans un dialogue continu avec les modélisateurs et en utilisant le modèle pour support. Ce processus itératif s'arrête quand les experts estiment qu'ils ont obtenu suffisamment de précisions sur le système ou lorsque la poursuite de ces itérations nécessite des expériences ou données de terrain.

Concernant les types de représentations qui soient adaptées à un tel processus, je propose une combinaison modulaire et adaptable de deux systèmes d'implémentation: les modèles à base d'agent (MBA) et les Systèmes d'Information Géo-référencées (SIG). Je montre que cette combinaison offre une très grande souplesse de description des composants d'un système (réel), qu'elle permet de représenter de façon équivalente les

connaissances qualitatives et quantitatives des experts, et qu'elle supporte un haut niveau d'évolution des hypothèses au cours du processus de modélisation. Les protocoles d'interaction proposés entre modélisateurs et experts se basent sur deux abstractions de cette implémentation: ODD (Overview, Design concepts, Details, un protocole de communication de modèle) et GAML (un langage de modélisation pour la programmation collaborative du modèle).

La méthodologie proposée a été appliquée et validée dans le contexte d'une étude détaillée située en Asie du Sud Est (essentiellement au Nord Vietnam) par des épidémiologistes et des vétérinaires. Ceux-ci voulaient pouvoir évaluer, en l'absence de données de terrain ou de résultats d'expérimentation, l'effet de différentes hypothèses expliquant la réapparition récurrente de foyers épidémiques de grippe aviaire parmi la population domestique de volailles. Au cours de cette coopération interdisciplinaire, qui a duré quatre ans, plusieurs "modèles monde" ont été co-conçus et implémentés au sein de la plate-forme GAMA, et utilisés comme "laboratoires virtuels" par les experts. Cette collaboration, et ces résultats, ont permis de tester un large champ d'hypothèses (en particulier sur les conditions locales de persistance), d'avoir une meilleure compréhension du rôle de l'environnement spatial, des facteurs écologiques et sociaux dans la survie et la propagation du virus et ont également permis de réorienter certaines des études de terrains.

Acknowledgements

First of all, I would like to thank David Hill and François Roger whom accepted to review this thesis. Their analyses and critics are stimulating ones and give me directions for future works. I would like, as well, to thank my defence jury: Jean-François Perrot, Eric Ramat, Ho Thuong Vinh and Nicolas Marilleau whom contribute to give further very interesting directions and questionings. Finally, a special member of the jury, Alexis Drogoul deserves a special thanks as a PhD supervisor. It was a great adventure to pursue my phd with you in Vietnam. You made me discover 3 whole new worlds: Vietnam, research and inter-disciplinary works!

After, the formal (nonetheless friendly) thanks, I want to acknowledge the support of so many people (in no particular order!). I would like to thanks my friends and colleagues at MSI. Antoine (who spent so much time reviewing this thesis. François (let's go back to Ha Giang!), Benoit, Patrick (it seems that the Vietnam had one similar consequence on the three of us) and Guillaume with whom we had motivating talks and great times exploring Vietnam. Minh Thu (don't forget to succeed in your new objective!) and Doanh (remember the Senegalese adventure?) whom welcomed in their homes. I am also grateful to my housemates (Cela, Georg in particular) for the nice adventures we had. I do not forget Nelly who gave me so much support in addition! I must also mention (more than a mention would be better) Caroline (don't settle!), Ronan, Baptiste, Marielle and Anna. Finally, Nam who is much much more than a friend...

I did not forget my many friends in France who had the courtesy to welcome me any time I came back to Paris: Sylvain (flatmate forever!), Phillippe (breton forever!) & Ran, Mathieu ("l'homme au chapeau"), Olivier & Ludmilla, Geoffroy & Pauline, Zach & Ido, Ronan & Hanne (skiing was "bref mais intense"), Julien & Béatrice (both almost "there").

Much south of Paris, there was Montpellier where I was welcomed in CIRAD. I would like to thanks the AGIRs member: Raphael, Bruno, Carlène, Vladimir & Eve but also Colette who sheltered me (I felt like a wandering phd candidate sometimes). My special thanks goes to Stéphanie who bore the inter-disciplinary collaboration with a freshman!

Many more friends have been forgotten in this text but not in my heart (though do not hesitate to remind me, haha...)!

Last but not least, the family which is an important variable in a PhD equation. I thank my father and brother for their taciturn (family personality feature) but strong support... especially nowadays: distance is nothing. I also thank Laurence and her family which I progressively discovered at each new come back.

I dedicate this thesis to my mother who gave me much more than life.

Contents

List of Figures

Introduction

The use of models in scientific domains has a long record in history. Initially, models have been mainly designed as a support tool for prediction and generalisation. The former objective assigns to the model the task of forecasting an upcoming state of a given system given its actual state, while the latter is used to underline the common characteristics between a set of systems in order to draw generic laws, or even theories, from them. Since a few decades, however, one can see in various domains that two new objectives are being assigned to models: those of exploration and explanation. Whereas the task of prediction implied that the model be able to produce outcomes with a certain confidence level, the task of exploration is more concerned by the «how» and under which hypotheses or conditions these outcomes are obtained. Similarly, whereas the task of generalisation demands that the model supports formalising and generalising relevant common outcomes of a set of systems, the task of explanation is more interested in the «why» these outcomes are common -- and «why» they may be different.

This evolution from «classical» prediction and generalisation objectives towards «new» exploration and explanation objectives are partly the result of a change of point of view on the notion of «model». Greatly influenced by the «complex systems» paradigm, the models designed to pursue these new objectives do not consider the system studied as a whole but, instead, pay attention, with great details, to its constituents and their interactions. A methodological opposition may be used to characterise this evolution: whereas «classical» models were mainly driven by a simplicity-oriented approach (Occam's razor), the «new» models are much more details-oriented: they focus on explicit and detailed representations of the components of the system, which in turn provide great capabilities of flexibility both in terms of parameters and model structure. They are not «simple» anymore. And it is a fact, besides any discussion on the ontological status of such models and their place in scientific processes, that these characteristics are now more and more requested by their «final users» (researchers, decision-makers or stakeholders), who find in these models a way to evaluate their hypotheses, to conduct «as-if experiments», to design rich and informative scenarios, and finally to explore their outcomes by way of repeated simulations.

This evolution can be observed in virtually all domains but Epidemiology is the most interesting for two reasons. Even though it has a long modelling history, all the epidemiological models produced so far have predominantly been used for prediction and generalisation tasks, at least until very recently. Moreover, while this shift in modelling is recent, the benefits envisioned seem to be particularly important. Indeed, recent situations such as the re-expansion of diseases kept at bay since a long time, (i.e. chikungunya or malaria), but also the apparition of unknown diseases such as the SARS and, more recently, avian influenza, argue in favour of these new models. Avian influenza, in particular, remains endemic in South-East Asia, but the causes are poorly understood and this ignorance limits the possibility to build predictive models. As a matter of fact, it seems necessary, to explain or understand «how» and «why» this epidemic has not been eradicated yet, to take into consideration the interplays of very different dynamics (social structure, economical networks, ecological context, climatic conditions, biological dynamics, etc.). When predictive models are tentatively adapted to take some of these into account, it results, at best, in ad hoc solutions, and these models often become dependent on incomplete or qualitative data (for instance, the decision-making processes of social actors or the behaviours of birds), which prevent them to be used for any serious predicting purposes and do not really advance the exploratory objective. Therefore, in a few cases, the epidemiological community is slowly beginning to turn to models that are specifically designed to serve exploratory purposes, and where the model is built to allow understanding the links between different dynamics, generating and studying various hypotheses and measuring, with respect to these hypotheses, the impact of local or global policies in complex scenarios. Scattered tentatives have been conducted so far, but their results are far too specific to envision proposing a generic formalism or methods. Even more, these tentative models have revealed serious methodological issues that could not be addressed by existing modelling and simulation practices and tools.

The objective of this thesis is three-fold: (1) to provide a characterisation of these explorative models (hereafter called «world models»), (2) to report on the design and implementation of such a «world model»

applied to avian influenza in South-East Vietnam, (3) to draw from this experience a proposition of modelling method (KIMONO), developed from the outset to support the design of «world models».

The first chapter of this thesis presents a series of examples from various domains (road traffic, social segregation, soil dynamics and more extensively epidemiology) in order to illustrate the rapid expansion of exploratory models. These examples allow me to introduce the requirements and first solutions proposed to design and build these models, among them the now well-known «KIDS» principle. In Chapter 2, I propose a more formal definition and characterisation of these explorative models under the name «world models» and a general review of the tools and methods modellers can rely on to build these models. The third chapter introduces the application I have been working on during my thesis: a thorough investigation, by modelling and simulation, of the local mechanisms causing the persistence of avian influenza in North Vietnam. I also review some other models that are of specific interest for this application. The whole modelling process is detailed in Chapter 4 and serves as a practical introduction to Chapter 5, where I describe KIMONO, the method I propose to facilitate the design and implementation of world models.

KIMONO focuses on two main elements: a deep implication of the experts and a great attention to a detailed representation of the system components. As the objective of the proposed modelling process is to explore and allow the assessment of the experts hypotheses, it is done at every stage of the modelling activity, starting with the extensive descriptions of the experts knowledge in their own formalisms, which are then conceptually linked together before considering the implementation process. In practice, it results in an incremental design and implementation of a common model, following an arborescent organisation of the modelling process, that progressively integrates the knowledge available. This iterative process comes to an end only when the specialists consider that they have a sufficient insight of the system or when further investigations require field experiments.

This methodological proposal is supported by adapted implementation tools. More specifically, I propose to combine Agent-Based Modelling (ABM) and Geographical Information System (GIS). Whereas the first one has long demonstrated its versatility in almost any domain, the second allows for a detailed and complex representation of the spatial environment. I demonstrate that such a combination allows to represent a system and its components at any arbitrary level of description that may fit the experts requirements. Moreover, it allows to integrate both quantitive and qualitative knowledge that may be needed to investigate their hypotheses and their evolution throughout the modelling cycle.

KIMONO proposes a clear sequence of modelling steps and relies on a heavy use of the ODD model communication protocol. ODD has the advantage of being generic enough to be adapted to any domain but specific enough to agent-based modelling to make modellers and experts provide the necessary details in their model. In addition, its structure of progressive complexity allows them to communicate effectively from the outset on the description of the model. Eventually, since the computer implementation of a model can also be source of uncertainties, I propose to overcome most of them by using a high-level modelling language designed to give experts an access to the implementation and a way to collaborate with modellers on this aspect too.

Chapter I : Evolution of modelling

The modelling activity has become a standard tool for many applied researches **[Mar06c, Lin09, Cro10]**. Consequently, there are a very varied set of modelling practices depending on the considered field for instance. Nevertheless, we can see transversal modelling evolutions in term of objectives, approach and representation since a few decades. While "classic" models used to represent a (real) system from a global point of view, with a strong concern on "generalisation" to other systems, and were designed according to a simplicity-oriented philosophy, there is now a flow of recent works that do not satisfy anymore these principles, because of related evolutions in terms of objectives, approach and representation. This chapter is dedicated to the characterisation of these evolutions of the modelling activity based on examples taken in various domains.

The "classic" modelling approach is described and core notions like *reference system* or *model* are defined in the first section. The four general objectives of model I consider in this thesis: *Prediction*, *generalisation*, *exploration* or *explanation* are also definition. Then, three examples are used to illustrate the recent trends in modelling. Each of these examples start with a prediction or generalisation oriented model and ends with exploration or explanation oriented ones. The consequences of this model usage evolution, along its consequences in terms of representation and approach, is summarised. Finally, and despite the relative youth of exploration and explanation models usage in this domain, I present several recent examples of epidemiological models that, because of their anticipated usage, highlight a similar evolution.

1. Basic definitions

Before characterising the evolution of models, I present essential definitions of the modelling activity. In particular, I define the notion of model and its relation to a reference system according to a specific research question. I also define three more definitions; meta-model, modelling cycle model clade and contributors; related to the construction of model.

1.1. System and reference system

A model is created after a given reference system. Thus, I need to define what is a system first. A system can be considered in two ways: As a whole or as a set of elements, depending on the observer considering the system. **[Hea10]** provides a definition in accordance with the first point of view:"*A system is defined as something that transforms inputs into outputs*" whereas **[Ber68]** adopts the second point of view: "[systems are] *Complexes of elements standing in interaction*". In the second case, there are two ways to consider the *complex of elements*. On the one hand the, the *analytical approach* (introduced by Descartes) considers that the system's elements are clearly dissociable until the primary elements exhibit basic enough behaviour. On the other hand, the *systemic approach* (**[Ber68, Der75, Lem77]**) considers that the system's elements are specifically organised and strongly related. Thus the elements are hardly dissociable and the focus is set on these relationships and not solely on the elements anymore.

Finally, the complex system approach **[Wal92]** can be considered to be the synthesis of the analytical and the systemic approaches (more details are given in **[Gli07]**). A complex system is a large set of elements strongly interacting with each other and producing a macro-structure not directly derivable from the characterisation of the elements. The reference systems considered in this thesis are considered from this perspective.

How the system is considered, (through analytical, systemic or complex system approach) greatly influences the model and how it is designed. The modelling evolution I present follows this evolution of system consideration, from global point view that may even consider the system as a black box to the complex system approach which considers the system as an indivisible set of elements in interaction producing an emergent macro-structure.

Now that I have defined the system concept and introduced a few ways to consider it, I can define in a general manner what is a *reference system* **[Tre08]** (or *target systems* **[Gau09]**) in the theory of modelling context;

Definition 1.1 [Reference system]: a reference system is defined by an observer and is constituted of interdependent entities interacting with each other and with the outside in a limited way

1.2.Model

Numerous definitions of the term "*model*" exist in the abundant modelling literature. Minsky's definition is probably the most classic and generic definition: "*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*" **[Min65]**. If we replace "A" by "reference system", this definition highlights the purpose of a model (A*): to help someone (a modeller, a scientist, an expert, a decision-maker) answer questions raised by the study of a reference system. It also highlights the fact that a model is not something that ought to follow a pre-established structure: anything can be a model of any reference system as long as this ***representation*** makes sense to "B" in answering questions about this system.

A model is thus **a simplified *representation*** which is designed when it is not possible, or when it is not preferable, to investigate the reference system directly. The limited scope of the research question allow the simplification of the representation. Some elements existing in the reference system may be omitted in the model or considered as negligible if they do not seem to play a role in the question. Of course, an infinity of concurrent representations of the same reference system can coexist and be used as models by different researchers, as they will likely correspond to different questions. In this thesis, I will then use the term "model" in this sense:

Definition 1.2 [Model]: A model is an abstract representation of a reference system, defined to answer one (or several questions) about this system. This representation is a simplification of the reference system that omits elements considered as irrelevant with respect to the(se) question(s).

1.3.Meta-model

In the context of the modelling activity only one type of model are considered, abstraction of the reference system (in opposition to mock up for instance). Such models are expressed using formal languages in order to make them manipulable. Such manipulations are constrained by a ***meta-model***. According to Woody Pidcock (cited in **[Mar06c]**), a metal-model is:

Definition 1.3 [Meta-model]: *an explicit model of the constructs and rules needed to build specific models with a domain of interest*

In other words, a meta-model is a set of concepts and how they relate to each other. Meta-models can be defined at various levels of abstraction. Example of meta-models includes differential equation systems (DES), cellular automata (CA) and agent-based model (ABM). These meta-model actually covers a range of "*lower meta-models*". For instance, ordinary differential equation systems (ODE) or partial differential equation systems (PDE) are lower meta-models derived of DES while agent-group-role (AGR) **[Fer03]** or Capacity, Role, Interaction, Organisation (CRIO) **[Gau07]** are meta-models derived from ABM.

1.4.Modelling cycle

I have defined what is a model and how it is structured but the conception of the model remains to be defined. In order to ease the creation of a model numerous modelling cycles have been defined **[Shan98, Sarg05, Dro02, Que09]**. They are very diverse yet they all have in common the following steps (though some are implicit): 1) problem definition; 2) reference system definition; 3) creation of the model (or modelling step); 4) model analysis.

Definition 1.3 [Modelling cycle]: a modelling cycle is a sequence of steps facilitating the progressive creation of a model after a research question. The minimal sequence considers the definition of the research question, the representation of a reference system, obtaining results from this representation, evaluation and analysis of these results.

Depending on the complexity of the reference system and the objective assigned to the model, the "model creation" step can be much more detailed. In particular, the reference system can be expressed initially as a conceptual model, potentially informal, before the definition of the actual model.

For instance, in the agent-based models context **[Dro02]** proposes three step models: domain, design, operational models and the actual one called implemented. These models correspond respectively to the first informal model, a first formalisation, an adaptation the implementation constraints, a fully implementation compatible and the executable models.

1.5. Model clade

In the best-case scenario a single iteration of the modelling cycle is needed: The actual model is defined once and for all. Nonetheless, the "model creation" step is usually decomposed as an iterative and incremental process: After expressing an initial model, often called the **domain model**, the actual model is gradually created and formalised by considering different aspects of the reference system or it is tweaked to conform more closely to the reference system. This iterative construction is called a ***model line***. In addition, several research hypotheses and modelling alternatives can be evaluated before making a choice. Each of these evaluation results also in a different model line. For example, when modelling a population two model lines can be created in order to evaluate the population representation: As a whole or as a set of interacting individuals.

The set of model lines, which all represent the same reference system and aimed at answering the same question, are regrouped as a ***model clade*** (adapted from a definition in **[Tre08]**). Figure 1.1 illustrates a model clade constituted of 3 models lines: line 1 and 2 are diverging at the beginning of the modelling step while line 3 is created as a divergence from line 2.
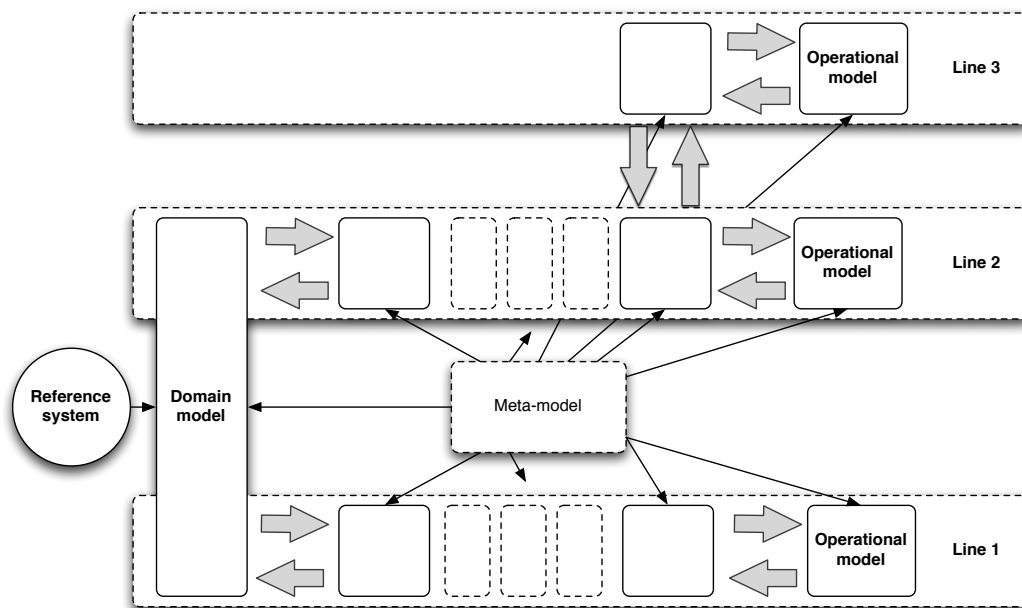
Fig. 1.1 A model clade with different model lines corresponding to one domain model (Adapted from **[Tre08]**)

1.6. Contributors

In order to study a natural phenomena, one can create a model after the reference system according to a modelling cycle. The model creation usually implicates different contributors. Two contributors roles are always differentiated: ***domain expert*** and ***modeller***. The domain expert is a knowledgeable person concerning the reference system. The role of the modeller is to propose a formal abstraction of the reference system based on the knowledge of the domain expert.

Several other important roles can be considered depending on the model objective and how it is designed. For instance **[Dro03]**, add the ***implementer*** which role is to transform the modeller's abstraction into an implemented model to allow its execution or simulation. **[Eti11]** add two roles: the ***facilitator*** and ***observer***. Their purpose is to facilitate the creation of the initial model by domain experts.

1.7. Model objectives

Any model is defined after a specific research question but it is possible to generalise these questions in terms of model objectives. In this thesis, I consider four model objectives: Prediction, generalisation, exploration and explanation.

Prediction

Models are the only way (other than imagination, that is) to make some assumptions about the future, and are heavily used for this task in almost every domains. In such cases, it is necessary to understand or hypothesise "how" the system evolves in order to be able to forecast its evolution, but the model is not intended to address this question. What is required is that it provides "valid" forecasts, not answers to "how" and "why" these forecasts are obtained. This kind of model, since their innate functioning need not be realistic, often function as a "black box" transforming inputs into outputs (Heath's definition of system in Section 1.1): The internal dynamic of the system does not resemble in any way the reference system's one.

Weather forecast is an applied science with many examples of models aimed at prediction. "*Numerical weather prediction*" (NWP) models are a type of weather forecasting models derived from atmospheric models **[Lyn08]**. They consider the atmosphere as a continuous fluid medium which allows to describe the atmospheric dynamic using fluid flow equations. A NWP model represents a reference system (e.g. a country) at its global level through a set of at least seven differential equations **[Lyn08]**: gas law (Boyle's law and Charle's Law), continuity equation, water continuity equation, equations of motion (Navier-Stokes' law) and thermodynamic equation. It is a very general model thus for each reference system the model is adapted through parametrisation based on previous situations.

Generalisation

Similarly to this prediction usage, the generalisation usages does not either consider the "how" and "why" the reference system evolves as a question to answer. In such case the model relies on previous information in order to be able to determine what are the characteristics shared by a set of reference systems. What is important in this case is that a set of systems have similar dynamics under similar circumstances. Thus, a model oriented towards generalisation is designed to generate output that are in accordance with the dynamics of different but similar reference systems.

Population dynamics provides more than a few models that can be used for generalisation. One of the first population dynamics model have been proposed by P.F. Verhulst **[Ver38]**. This model represents a population (as a whole) and relates its growth to a "*carrying capacity*", *K*. The carrying capacity is an aggregated parameter representing the capacity of the environment to sustain a maximal level of population. In practice, the model is expressed as a single ordinary differential equation (ODE): $dP/dt = rP (1 - P/K)$ where dP is the population growth and $r(1-P/K)$ corresponds to the *per capita* birth rate. It has been proposed to represent any human population. It has been successfully applied by **[Pea25]** to census data from the France, Sweden and the USA, demonstrating its generality for human population. Nevertheless, its generality is much wider than human population. For instance, in **[Cro94]** the authors use the logistic model to represent the growth of tumour cells. In practice, they define this growth with the following equation: $Nt+1 = rNt(1-Nt)$ where *K*, the carrying capacity is considered to equal 1. Many other examples can be found in the population dynamics **[Mur89a]** models such as the *Lotka-Volterra prey-predator* model which represents the evolution of the populations of preys and predators.

Exploration

Whereas prediction usage oriented models are defined in order to generate outputs with a certain confidence level, exploration ones are more concerned with uncovering in which conditions and under which hypothesis some reference system outcomes can occur. It is definitely not limited to "parameter space exploration" (i.e. varying the parameters of a model in order to know all the possible dynamics it can express) it is much more

general than this. It includes it but also considers, the exploration in terms of system structure, internal processes and relationships between them.

The MALCAM (Section 3.5.2) is a good example of an exploration model. This model intends to evaluate how a disease (Malaria), absent at the time of the study, could re-appear and how it would be propagated. To attain such objective, this model is expressed using the agent meta-model which allows to evaluate the interactions of the system's entities and assess possible emerging risk factors which would lead to re-emergence of the virus.

Explanation

Whereas the generalisation usage oriented models are defined to support the formalising common (system-level) characteristics of a set of systems, explanation ones focus on the internal processes of these system. The objective here is to define what drives the system dynamic not to define its general form.

The various models proposed by Crook (Section 2.1) are good examples of explanation oriented ones. The author considers several hypotheses (e.g. influence of the environment, individuals heterogeneity) in order to determine which one could explain the segregation dynamic observed in cities.

Prediction and generalisation objectives are related whereas exploration and explanation are related to each other. Indeed, the models oriented towards the two first objectives focus only on the generation of appropriate outputs whereas the models oriented towards the two latter ones focus on deriving knowledge from the model and its construction. In this chapter, I show that the design methods of a model defined after a question belonging to prediction or generalisation objectives are well characterised. In contrary, few model design methods have been proposed for exploration or explanation oriented models despite the growing interest for such models.

2. Modelling evolutions

Two elements are driving the evolution of models. First and foremost, new questions are continuously addressed to models. These new questions results in new requirements to which previous models cannot always address thus new modelling technics are proposed. Conversely, these new technics sometimes allow to address questions that have remained unanswered so far. This evolution is a transversal one as it is occurring in virtually all fields of applied science. More specifically there are three evolutions: objective, approach and representation. Models objectives shift from prediction or generalisation towards exploration and explanation of the phenomenon of the reference system. Models were often designed after a single point of view but more and more models are integrate knowledge from different fields thanks to the collaboration of different domain experts. Finally, the representation of the reference system is less and less considered at its global level and more at its constituents level. This section illustrates such general evolutions through three examples from very varied domains: segregation dynamics, soil dynamics and traffic system.

2.1. Objective shift : segregation models

The evolution of models is driven by the evolution of questions addressed to them which tends to belong more and more to exploration and explanations usages instead of generalisation and prediction ones. Such evolution can be illustrating using the segregation domain.

2.1.1. Generalisation oriented model: Schelling's

In the context of segregation studies, one of the first questions addressed to models is to propose a general characterisation of these dynamics. The model proposed by Schelling **[Sch71]** can be used for such purpose. This model represents located individuals (e.g. households) that prefer to reside in an area with a minimum number of individuals similar to them. In practice, the model contains two types of individuals (e.g. represented by different symbols) scattered over a grid space (initially it was a simple a draught-board). Individuals prefer to live among other individuals similar to them: If an individual is not satisfied, it will move to the nearest spot that satisfies its preference. For instance, an "O" individual is satisfied to live in a place where the eight neighbour cells of the draught-board contain at least three "O" individuals; if, it is not the case it will move to the closest spot with three red individuals.

With such a simplistic model, the author succeeded to provide a general segregation characterisation. Moderate preferences for living among similar individuals lead to marked global segregation. In Figure 1.2 two types of individuals ("O" and "#") are segregated in four distinct clusters. This situation emerged from a population of "O" and "#" individuals originally randomly scattered on the draught-board after computing individuals' movement based on their neighbourhood preferences.
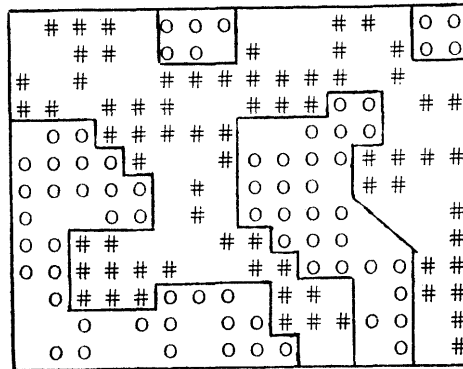
Fig.1.2 Schelling's original model **[Sch71]**. "O" and "#" types of individuals and hand drawn territories (or clusters).

This model proposes a general description of the segregation mechanisms. Such mechanisms are generic enough to be adapted to any city or even any area. Moreover, these mechanisms can be applied to represent the segregation of individuals of any species. The author postulates that segregation can be simply caused by individuals' mild preference for similar-individuals-neighbourhood.

2.1.2.Exploration and explanation oriented models: Crook's

The simplicity of the Schelling's model is a great feature but it is also its main defect as it cannot be used for other objectives than generalisation (or potentially prediction). Substantial modification of the model are needed if the questions addressed to the model move from generalisation to understanding and/or exploration of the segregation causing mechanisms. More precisely, integrating more complex individuals behaviours or detailed environment representation is difficult if not impossible in the Schelling's model as it is. For instance, if the question is "understanding the influence of the environment topology onto the segregation dynamics", the representation of the environment has to introduced and much more details accounted for.

Crooks **[Cro06, Cro10]** proposes a series of models to address exploratory and explanatory questions. Even though each of them investigates specific phenomena causing segregation, they all represent the same reference system: A city (London), and are expressed using the same meta-model: agent-based model[1]. These models constitute a clade which has for (common) research question: investigating the factors impacting the segregation dynamics.

An important factor, after the preferences, is undoubtedly the environment. Crooks introduces a much more detailed representation of the environment. The grid-based environment is left for a continuous one, composed by dots, lines and polygons. It allows to represent much more precisely the topology of the environment (i.e. London city presently). The individuals neighbourhood is not restricted to adjacent cells on a draught-board. A radius-based representation is used to model the individuals neighbourhood. This neighbourhood representation is constrained by the topology of the environment (blocked by a street or a building for instance).

Crook's model also introduces N different groups of individuals while Schelling's model is originally restricted to two types of individual. In addition, he allowed each group to have different neighbourhood preferences which results in different sizes of cluster. Figure 1.3 illustrates such situation with four populations (red, blue, green, white) located in a city of four districts (using a continuous representation of the environment). A mark segregation emerges from interactions of the individuals. The colour of a district is determined by the

---

1 Schelling's model can be considered as a located micro-simulation

colour of the majority of its individuals. Nevertheless some individuals of a different colour still reside in a given district.
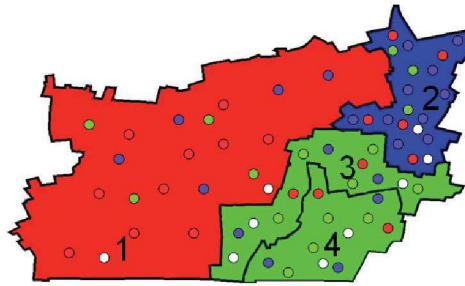
Fig1.3 A segregated city with four different populations **[Cro10]**.

Finally, Crooks integrates a "*population dynamic*" as individuals can be removed and introduced into the simulation. Such process presents a new source of dynamic and makes the cluster more unstable. In practice, it can represent ageing or immigration processes.

From a practical point of view, Crooks' model represents a small part of the city of London (1.5 sq. km.) thanks to the use of a GIS (Geographic Information System). At each time step, every individual has the possibility to move around its location if it is not satisfied by its neighbourhood (see **[Cro10]** for more details).

Crooks' model confirms Schelling's results: Even moderate preferences lead to marked clustering. The clustering speed depends on the preference strength: the stronger the preferences, the faster appears the segregation. Nevertheless, these models also explore different phenomena that may explain segregation. The author shows the impact of the movement speed: A high movement speed increases the clustering formation speed, stronger preference leads to more frequent moves.

Many other interesting results are also highlighted. Increasing the preference radius increases the size of clusters. To constrain the preference radius affects the size and the shape of clusters, especially with less contact points between clusters, as shown in Figure 1.4. In this figure two populations are represented. In The half left of the figure the geometry (or topology) is constraining the perception of the individuals while the right one show when it is not the case. In the latter situation (zoomed part of the right half of the figure) we can see that the influence of a group in a given area can influence the segregation in a close but disjoint area (two parallel residency areas or U shape presently). The spatial resolution of the environment representation can deeply affect the dynamic of the model. An apparently well-mixed population (at a low level of details) may change into marked clusters (at a high resolution).
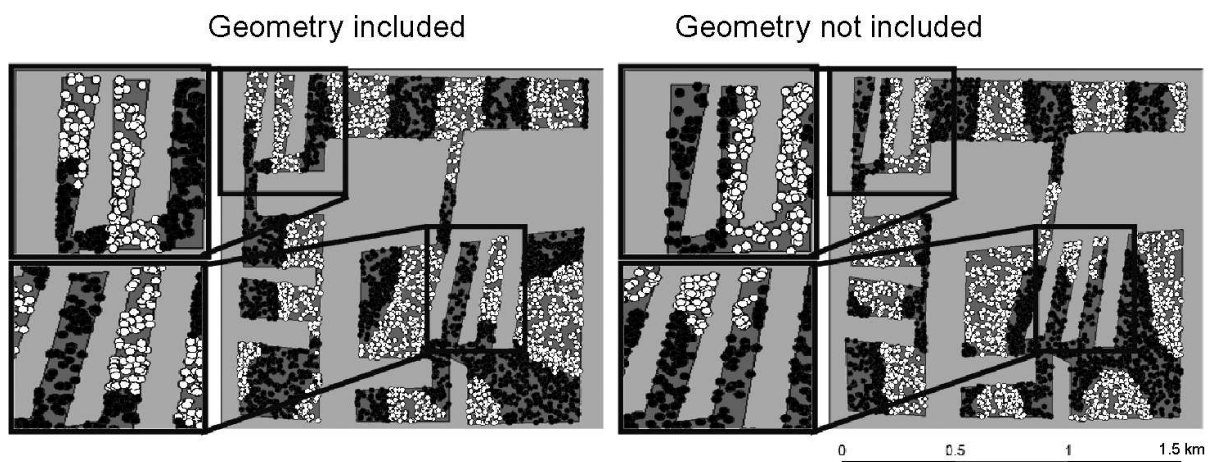
Fig.1.4 Comparison of clustering with and without taking into account the geometry of the environment **[Cro10]**

2.1.3.Discussion

The initial model proposed by Schelling is a general characterisation of the phenomena leading to segregation whatever are the area or the individuals' species. It is a very general model that may also be used for prediction with probably complementing it with some parameters. In contrary, Crooks model are much more "anchored in reality" through the consideration of specificities such as the real map of London (thanks to the use of GIS), heterogeneous individuals (different levels of social preference) and the possibility to represent N social groups instead of only 2. Such representation of the reference system are much more "realistic" and allows the author to explore various mechanisms that may generate segregation and ultimately provides an explanation by representing local phenomena which combinations result in the system's global dynamic.

Shift of the model objectives, and more practically the questions addressed to them, has two practical consequences: evolution of the modelling approach and representation types. Indeed, as introduced in the segregation model examples, exploration and explanation require to consider the reference system from different point of views (from different domain) whereas prediction and generalisation usages can be satisfied by a single field orientation. Finally, an even more practical evolution is the representation ones. Initially, systems were represented at their global level but more recent models operate at the individuals and possibly at different levels. I present these two resulting evolutions in the two following sections (2.2 and 2.3)

2.2.Approach shift: pedological models

In the case of prediction and generalisation oriented models, the reference system is generally considered from a single point of view and at the global level. When the objective is exploration or explanation the reference system is considered in more details. When the considered system constituents remains "conceptually homogeneous", it is possible to considered the reference system from a single point of view. Nevertheless, exploration and explanation objectives demand to consider the reference system in much more details than other objectives. It results in the need of considering the reference system from various point of views and integrate knowledge from different fields or more formally a inter-disciplinary approach of modelling is needed.

2.2.1.Representing earthworms actions on soil

A fit example of approach shift is given by the Sworm model **[Mar08, Bla09]**. The question at the origin of this work is the understanding of the earthworms impact on the soil structure. The hypothesis is that earthworms alter the soil structure by creating burrows (after ingesting organic matter) and create compact areas (when excreting compact cast or dejection). Despite the relatively easiness of earthworms representation one problem could not be solved by biologists or modellers: Having a realistic enough representation of the soil structure. Whereas the use of ABM is satisfactory to represent earthworms and their dynamics, a realistic enough representation of the soil (at the earthworms level) would be too cpu demanding to conduct any simulation.

Instead of developing a brand new representation of the environment, the authors of **[Mar08, Bla09]** collaborated with pedologist to adapt their representation of the soil structure. This model, called PSF (Pore-Solid-Fractal) **[Per02]**, represents the soil as a 3-d grid constituted of fractal cells.The resulting model is called APSF for Agent-PSF **[Mar08]**. Finally, a model coupling this soil representation and earthworms one has been developed **[Bla09]**. More details of these models are given as follows.

2.2.2.The Pore-Solid-Fractal model

Proposing a generic representation of the soil structure, condensed and preserving essential characteristics is not an easy task. The "*Pore-Solid-Fractal*" model (PSF) has been proposed in **[Nei89]** and enhanced in **[Per02]** to reach such objective. This model represents the soil as an organised discrete set of cells. Three types of cell exist: a cavity, a compact soil and a fractal cell constituted of heterogeneous sub-cells. The recursive pattern is repeated from level 1 to level *n*, as described in Figure 1.5. In practice, real soil does not have such a fractal organisation but this representation maintains essential soil characteristics **[Per02]**.
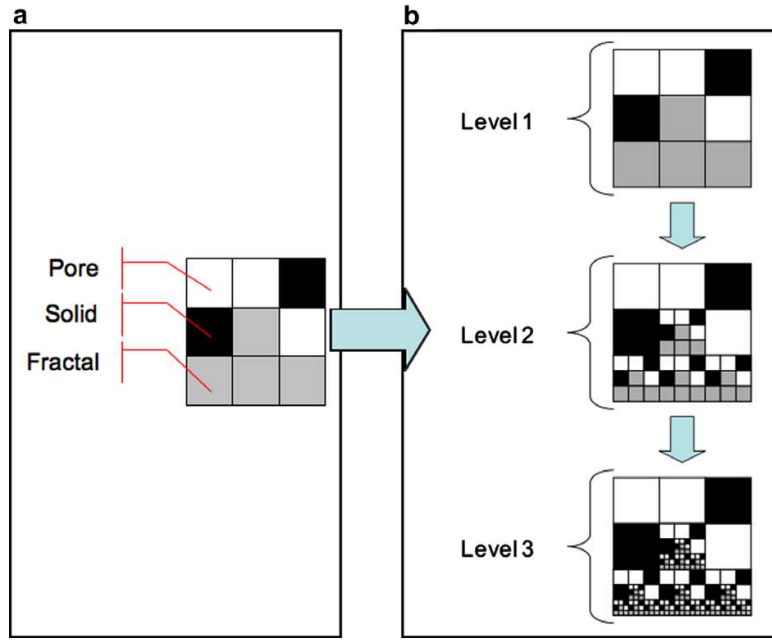
Fig.1.5 The PSF model as presented in **[Mar08]**.The a) part represents a given level constituted of pore, solid and fractals while the b) part shows the fractals decomposition

2.2.3.Sworm

The study of soil-worm interactions requires a detailed representation of the environment. The PSF model has been adapted for this purpose. A collaboration between different contributors was required. Three domain experts, a soil representation specialist, two biologists, and a modeller (acting as an implementer as well) have collaborated to adapt the soil fractal representation to interactions with individuals in a multi-agent system (MAS) **[Bla09]**.

- *APSF model*

This collaboration resulted firstly in the APSF model (Agent Pore Solid Fractal). The environment is still a set of organised discrete cells of different types (compact soil, cavity and fractal). However, instead of a repeated pattern, different "*canvases*" are used to simplify (from an implementation perspective) the environment representation. These canvases can be limited to represent the decomposition of a given level or can be recursive like fractals. Figure 1.6 presents an example of 2-d grid environment where three levels of decomposition are considered, two decompositions (C1 and C3) are recursive. The a) part of the diagram represents the recursive structure of the environment whereas the b) one presents the resulting environment representation.

In practice, the canvas-based decomposition process occurs when the need arises only. As long as a heterogeneous cell is not in interaction with another element of the model, its decomposition does not occur. With respect to this environment representation, much larger simulations can performed than with previous models.
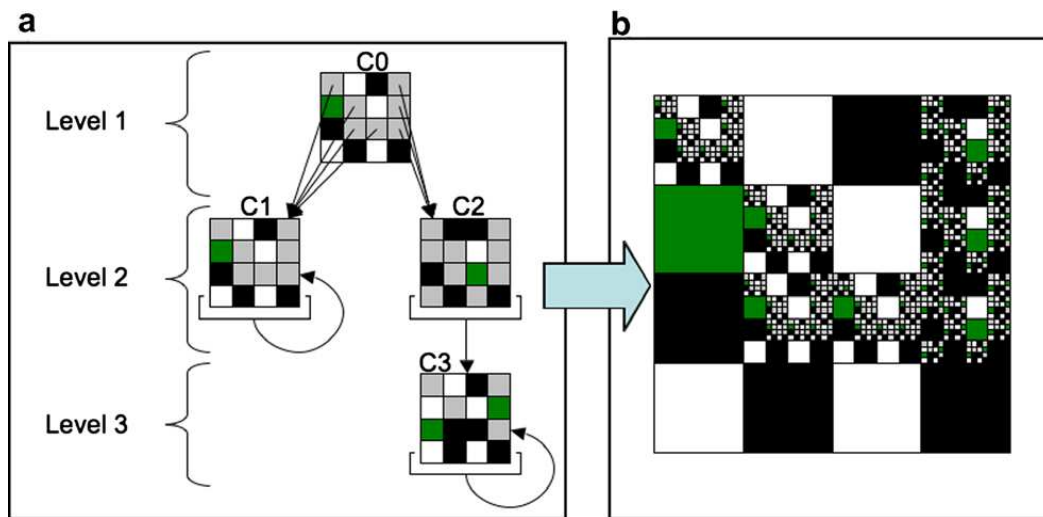
Fig.1.6 The APSF model as presented in **[Mar08]**.The a) part represents three decomposition levels of the environment b) part shows the general view of the environment

- ***The Sworm model***

In **[Mar08]** and **[Bla09]**, the authors coupled the APSF realistic representation of the soil to assess the impact of earthworms' actions on the soil structure.The earthworm behaviour consists mainly of moving towards organic matter, ingesting and assimilating it, and ejecting cast (dejection) in an empty space (pore).

Simulations were conducted using a 20x20x20 cm sample environment. Within this small cube, the authors verify the coherency of their model by tracking the evolution of organic matters in this sample. As described in Figure 1.7, the rate of organic matter decreases when earthworms are ingesting, then it abruptly increases when worms are ejecting casts. This cycle appears continuously, and in the long term the organic matter rate decreases as observed in the reference system **[Bla09]**.
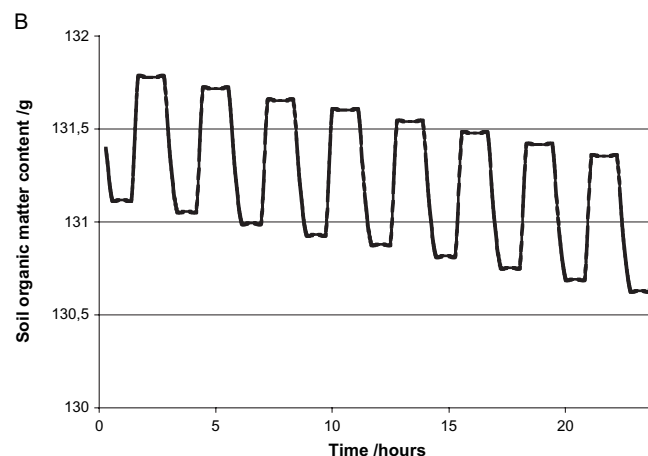
Fig.1.7 Organic matter level in the SWORM simulation **[Bla09]**

In another simulation **[Mar08]**, the authors tracked the impact of the earthworms action on the soil structure. It results in a specific structure constituted of "*clearly identified casts, areas of non-ingested soil, and large burrows*".

The model does not aim at studying the soil representation or the worms dynamic themselves. Instead, the model aims at evaluating the impact of earthworms action onto the soil structure. The obtained model is thus more complex than its elements.

2.2.4.Discussion

In this example, a static fractal representation of the environment, defined using an analytical paradigm, was adapted to a different application domain, defined using a computational paradigm. The PSF model, representing the soil structure thanks to the use of fractals, was adapted as APSF to be used as an environment within an agent-based model: Sworm. The PSF model is designed to study the soil structure whereas the Sworm models objective is to investigate the action of earthworms onto the solution structure.

The Sworm model can be considered as one of the simplest way to describe the emergence of complexity in modelling evolution. It shows that, to be modelled in an appropriated way, a reference system may need a complex representation coupling for instance two models representing two different reference systems.

The main interest of this example is that it illustrates the need, common to many exploration and explanation oriented models, to diversify the number of point-of-views on the reference system which results in an inter-disciplinary approach of the modelling activity. In order to design the Sworm model the collaboration of experts from two different domains were required. This collaboration was mediated by modellers whom proposed a way to adapt a soil representation to the meta-model selected by biologists to represent earthworms.

2.3. Representation shift: traffic models

The last modelling evolution I emphasis in this chapter is the representation one and is certainly the most practical one. The road traffic domain provides excellent illustration of such evolution. Indeed, from the early traffic models to the most recent ones: The representation moved from a global point of view expressed using differential equation to a detailed representation of the (heterogeneous) individuals.

Road traffic has been studied for a long time. Indeed, one the first published papers on this topic dates back to 1935 **[Gre35]** and the domain is still very active nowadays. Initially, the question addressed to model were related to the evaluation of the flow capacity of a road system while more recent includes safety improvements, optimisation of the flow, etc.

2.3.1. Initial model

The first question addressed to traffic model is probably to determine the maximum flow a road system can sustain and express this flow from the system characteristics. Greenshield **[Gre35]** conducted one of the earliest study on traffic flow and defines his work as "*a study of traffic capacity undertaken to determine the approximate hourly density on a roadway of given width at which congestion or the slowing of traffic begins, and the amounts by which traffic congestion*". The outcome of this study is a "*linear relationship between speed and density which translates into a parabolic relation between flow and density*" **[Leclercq02]**. The flow increases with density of vehicles until a certain density is reach then the flow decreases (representing the effect of traffic jams). This relation between a flow and a density of vehicle is called the *fundamental diagram* and is specific for each traffic system.

In order to effectively compute the flow of a traffic system, a first dynamic model was proposed incrementally by three authors **[Lig55, Ric56]** and called the LWR model (named after the authors: Lighthill, Whitham and Richards). This model re-use the fundamental diagram and computes the flow of a traffic system given its vehicle density. This model is based on the hypothesis that traffic flows are similar to fluid flows. Thus, it is possible to express the link between traffic flows and density using partial differential equations similar to the one used for fluid flows. The equations are as follows:

$$\begin{cases} \frac{\partial Q(x,t)}{\partial x} + \frac{\partial K(x,t)}{\partial t} = 0 \\ Q(x,t) = K(x,t)V(x,t) \end{cases}$$

In these equation: *Q(x,t)* is the traffic flow; *K(x,t)* is the density; *V(x,t)* is the speed. The maximum flows is obtained for a certain density of vehicles then it diminishes due to falling of the average speed of vehicles.

In order to obtain fitter models it would valuable to take into account vehicles characteristics such as their limited acceleration capability. The LWR model cannot integrate such parameters but "second-order" models have been proposed to address such issue **[Wan09]**. These models does not consider macroscopic velocity as a fixed variable but as a dynamic one.

According to **[Wan09]**, the traffic flow is intrinsically a stochastic system but even "second-order" models cannot represent such situation. The stochastic models **[Wan09]** provides flows of density that fits better

empirical data. Moreover, they allow easier representation of new situations dynamics such as off-ramps dynamics and more generally any location-based dynamics.

2.3.2. Modern models

LWR and its derivative models use fixed fundamental diagram (linking traffic density and flow) and can be used to study the global traffic system dynamic in order to forecast the needed development of the traffic system for instance. Nevertheless, new questions emerged as many phenomena were not understood yet, such as stop-and-go waves (vehicles stop for a short duration then start and so on), vehicles changing lane, vehicles of different speed needed to be investigated **[Kesting08]**. To do so it is necessary to represent explicitly various reference system's constituents. The LWR model is incapable to do so as it is expressed in ordinary differential equation (ODE) representing a road system at its global level.

A new representation paradigm emerged to obtain more malleable models. The global representation of the reference system shifts towards a representation of its elements at various levels which allows the introduction of *heterogeneity* in the model. These new models are called *microscopic models* by opposition to *macroscopic models* which represent the system as a whole. Actually, it is possible to differentiate four levels of representation of the system: *macroscopic* (or system level), *mesoscopic* (cluster of vehicles level), *microscopic* (vehicle level), *nanoscopic* (driver interacting with the vehicle). These different levels of representation are imposed by the questions addressed to the models. I already presented the LWR model which is a macroscopic one and the three other levels are presented in the following sub-sections starting with microscopic models.

• *Microscopic models*

The emergence of stop-and-go waves cannot be explained by LWR models and its derivatives because it is a local phenomena whereas these models consider the system globally. Microscopic models are capable to represent such local phenomenon and others in a generative manner. For instance, the *NaSch* model **[Nag92]** proved to be capable of generating stop-and-go waves and how they propagate.

The *NaSch* model is based on the "car-following model" proposed in **[Sch02]** which proposes that vehicles behave as follows: "*going at maximum legal speed when possible, decelerating progressively when getting closer to other vehicles and accelerating progressively when possible*". In the NaSch model, vehicles are explicitly represented and placed on a cellular automaton. Figure 1.8 presents a 1-d traffic cellular automaton: Each number represents a vehicle (speed and location) which are moving from left to right, three iterations of the models are represented.

Iteration *i*	3				2				3			1	
Iteration *i* + 1	2				4			3			2		
Iteration *i* + 2				3			3			3			2

Fig.1.8 A one dimension traffic cellular automaton

According to the NaSch model, vehicle behaviours are accounted for and the dynamic emerges from their interactions, while according to the initial LWR model, the dynamic is defined and suitable input parameters must be identified.

The NaSch model is able to reproduce new dynamics: stop-and-go waves, free-flow, traffic jam as illustrated in Figure 1.9 and synchronised flow **[Sch05]**. This figure represents a traffic jam in an early model **[Nag92]**. It is similar to the previous one as it represents a 1-d traffic cellular automata where the vertical series of lines shows the system's evolution. The traffic jams occur where vehicles are stopped (0 velocity). The movement of the traffic jam, in the opposite direction of vehicles, can also be observed. This phenomenon is visible at the global scale of the system but is emerging from local interactions (from the "car-following" model) between vehicles.

space (road)

time

```
.........4........5......5......5.....00.1.......4.........5
..4......5.......5......4....5.....2...00..2............5...
....4.......5.......4.....4.....3...0.01...2...........5...
.4....5.......4......4.....5......01.1.2....3............
...5....5......4......5......0.1.0.2..3....4..............
.4.......5......4......5.......1..00...3...3.......5....
.....4....4....5......4......3....000........2..3.........
.............4......5......4......0.001..........3...3....
......4.......4......4......5....4..0.00.2......4....4...
.4.........4......4.......5...2...1.01..2...........3....
....4......4.......4......5......2.1.01.2..3..........4
......5....................5......0.1.0.2..2....3........
.......5......5......4......5......0.00..1..2...3........
.4........4......4......3..1..00...2..3....3.......3...
...5......4......4...........2..0.00......2..3...3...3
.5....4...4......5...........01.00......3...4...4....
.5....4....4......5.......5.....0.000.........3....4...
....4....5....5......5......4...1.001..........3....4
4.....5....5......4......1.000.1..............4......
...5.......4....5......5......0000...1...............5
.....4....5....4......4...0000...2...................
......5......5......4....6....0.0001.....2...........
.......4.....4....4....1..0.000.2...2................
.5........5......4......2...0.1.000...3...3.........
.........5......4......5.....1.0..0001.....4...4.....
..5......5......5.......2..01..001.2.......4....5....
....5.....5....4....5.......1.0.0.01.1..3.........5....4
...5.....5....5......5.......00.0.0.1.1...3..........4...5
...........4....5......5......01.1.0..0.1....3..........4
...........4......5......2..1.0.00..1..2.......4.........5
..........5......4....2..01.01..2..3.........6........
5.........4......4....2..01.00.2...3...3.........4...
.....4..........4........5.....00.000...3...3....4............5...
```

Fig.1.9 A traffic jam in the **[Nag92]** cellular-automaton model.

This model can be considered as an explanatory one as it allows to evaluate an hypothesis on stop-and-go behaviour and traffic jam generation. It can be seen also as exploratory as it is possible to tweak the parameters values and behaviours in order to explore resulting dynamics at the system level.

Indeed, this model has been widely extended to design intersections, to regulate traffic lights, to study the impact of variable message signs, of ramp metering or any other form of traffic control (e.g. **[Mah01, Liu99]**). The change of representation, from global to individual, allowed scientists to investigate the cause of traffic phenomena.

- *Mesoscopic models*

Macroscopic models allows to study traffic systems based on their density of vehicles while microscopic models represent mechanisms generating local congestions problems. Studying the within-day and day-to-day dynamic of a traffic system and evaluating new policies are very different issues. Macroscopic cannot investigate the processes generating such dynamics while microscopic ones represent many details (local interactions) not directly contributing to these dynamics. Consequently, mesoscopic models, considering certain details of the traffic system but less details than microscopic ones, have been proposed. This intermediate solution considers vehicle clusters moving altogether on a road section **[Ben85]**.

Based on this idea, many models have been proposed and one of them is of specific interest: the *Metropolis model* **[Dep02]**. In this model, the road system is represented as a graph where edges represent roads. When a vehicle enters an edge, its travel time is computed according to the edge vehicle flow. This representation may be viewed as a cellular-automata where each link is a unique cell (of the matrix). Although individual vehicles are considered, they are all homogeneous and grouped by edge. In addition, no interaction can occur on an edge, which can thus be considered merely as a *FIFO* (First In, First Out) stack.

This model also aims at evaluating driver's information as a mean of regulating the traffic system. It includes the representation of past congestion knowledge and also integrates representation of public information as "A*dvanced Travel Information System" (ATIS)*. ATIS are part of an "A*dvanced Travel Management System (ATMS)*" which monitor (using cameras, speed censors, etc.) and improve flows (using information system such as ATIS) within a traffic system. The authors successfully reproduced the dynamic of a large traffic network (12,000 sq. km) of a real province road system, in Île-de-France, with sound departure rates and congestion dynamics. They also argue that this model is ready to be used on other traffic systems of similar size. Moreover, they claim that the global analysis of the simulated system is now relevant and can lead to planning analysis (to improve flow in traffic systems). This should allow decision-makers to evaluate new traffic policies targeting the decrease of traffic jam. Consequently, I can say that this model can be used with an exploratory objective in mind and thus illustrating the shift of model objectives in addition to the representation shift illustration.

In order to represent dynamics that could not be represented by macroscopic neither microscopic models, the *Metropolis* model (and more generally mesoscopic models) used an intermediate representation level. It introduced the vehicle clusters which is derived from the reference system but has no physical existence.

- ***Nanoscopic models***

Being able to forecast traffic flow and investigate the cause of traffic jams, stop-and-go waves, etc., allows to enhance traffic safety. Nevertheless, another element impacted traffic safety and the system in general is left out: The driver and his/her behaviour. Unfortunately, even microscopic models are not well suited as they consider homogeneous drivers behaviour and do not integrate sufficient details influencing it. Consequently, models differentiating the vehicle, its driver and the environment **[Ni03]** have been proposed, they are usually referred to as *nanoscopic models*.

HUTSim **[Kos06]** is a large project which intends to couple several representation levels to obtain a realistic simulator which focuses on the study at the driver behaviour and enhancing the traffic safety.. This project covers a traffic simulator applied to various environment (e.g. inside or outside a city). The simulator is based on a cellular automata at the mesoscopic level where drivers' behaviours are usually homogeneous and are derived from safety regulations. However, people do not always drive safely in practice. Thus, a better representation of drivers' behaviours is required. The authors propose a model based on the knowledge of human perception and cognition in order to represent the heterogeneity of driving behaviours. This more realistic model can be used to envision interactions between different categories of road users (especially pedestrian). This model pushes forward the individualisation process by differentiating vehicles and drivers. Moreover, this model explicitly consider several levels of representation and their interactions. Instead of forcing traffic safety issues by considering only different types of vehicle behaviours, this model generates them by representing explicitly different elements of the reference system: the driver, its vehicle and the environment and their interactions.

2.3.3. Discussion

The initial LWR model is designed to determine the flow of a road section from its vehicle density while more recent models intend to understand traffic system phenomena (e.g. traffic jam and stop-and-go waves) and/or allow evaluation of the possible regulations (e.g. ATIS). This evolution in terms of question addressed to the model leads to a double evolution of representations: scales and dynamic. On the one hand, the model scale of representation moved from a system level (e.g. LWR model) to an individual level (e.g. HutSim). On the other hand, the model dynamic was initially forced onto it (e.g. using LWR fundamental diagram) while more recent models generate the overall dynamic through their elements' interactions (e.g. HutSim). The evolution denotes first a shift towards representation of "more local" issues. Moreover, this evolution goes much further as the model dynamic is "emerging" from lower levels. It has the two advantages of allowing investigation of the causes of the overall system's dynamic and evaluating new policies to modify it.

2.4. Conclusion

As modelling is becoming more and more popular and mastered by scientists, the urge to stick to (conceptually) simplistic models is becoming less and less relevant. Thees evolutions in terms of objective, approach and representations are illustrated by the three series of model examples of this section. Each series demonstrated a specific evolution. Nevertheless, the three evolutions are linked to each other and each example exhibits the three evolutions to different extents. For instance, the segregation models by Crooks **[Cro06, Cro08, Cro10]** illustrate the objectives shift but it also exhibits an approach shift as they integrate not only the sociologist point of view but also the urbanist one as the city topology is also part of the model. Similarly, modern traffic models illustrate the representation evolution but also the objective one as "modern" models are addressing explanation questions.

In addition, I can draw out several transversal tendencies opposing prediction or generalisation to exploration or explanation models. In particular, modelling practices for prediction and generalisation oriented models are fairly opposed to exploration and explanation oriented ones. The model itself is not of much interest for the former ones. Indeed, as long as its output is considered valid (fitting the observation of the reference system), the question answered. Thus the validation of the model in terms of fitting outputs is of outmost concern for such models. In contrary, exploration and explanation oriented models does a great deal of the

model itself and how it is constructed. Emergent characteristics of the model are used to gain knowledge over the reference system whereas output fitting of these models are of lesser concern. Finally, prediction and generalisation oriented models result of domain specific traditions. The representation of the system is considered adequate based on the domain habits. Consequently, the representation paradigm is **"imposed"** before the system is even considered.

3. A similar evolution in Epidemiology

The previous section demonstrates the three modelling evolutions: Objective, approach and representation, using model examples from three different domains in order to emphasis their generality. These evolutions are also occurring in Epidemiology and are of special interest due to the new challenges this field is facing. New diseases such as the Avian Influenza are defying understanding of the epidemiologists whereas well-known diseases such as Malaria, are expanding after specialists considered them under control.

The present section is organised as follows. First, I introduce the origin of Epidemiology, the discovery of the relationship of the environment and outbreaks of diseases in Section 3.1. Then, the initial "compartmental model" by Kermack and McKendrick **[Ker27]** and its extensions are presented in Section 3.2. Then more recent models are presented in order to show the three modelling evolutions in the epidemiological context in Section 3.3 to 3.5.

3.1. Epidemiology history

Epidemiology is the science that studies the appearance, the propagation and the evolution of diseases. Contemporary epidemiology can be traced back to a few historical events. The first **formal indicator** of a disease appeared later in 1662, formulated by *J. Graunt*: the morbidity rate **[Gra62]**. The first **link** between a disease and its cause has been found by *J. Lind* in 1753. He proposed a remedy to Scurvy: simply eating citrus fruits because they probably contain an element to prevent the disease. For a long time, it is well known that scurvy is due to a lack of vitamin C. which can be found in important quantity in citrus fruits. The actual birth date of modern epidemiology is considered to be in 1854 when *Dr. J. Snow* demonstrated that a Cholera epidemic was linked to a public water pump during the "Soho cholera epidemic" **[Vin03]**. Snow noticed a significant higher death rate around this water pump and decided to treat it using chlorine and to remove the handle of the pump. This action did not only stop the cholera outbreak but also showed that the prominent theory of miasma was incorrect. Indeed, he showed that the origin of the outbreak was the water pump and not some "noxious air" as usually believed by proponents of the "miasma theory". He clearly demonstrated that a specific **factor** was the source of the disease.

Modelling in epidemiology or mathematical epidemiology appeared in late 19th - early 20th century with scientists like *Ross*. Ross is famous for his study on Malaria **[Ros16]** in which he gave a detailed description of the Malaria cycle, which can be briefly summarised as follows. An infected *anopheles* mosquito can transmit a parasite, *Plasmodium falciparum* (being the most common one), to a human who will declare the Malaria. A susceptible (not infected yet) mosquito can get the parasite from an infected human and continue the cycle.

3.2. Classic models

The initial model used to represent the propagation of a communicable disease in a population has been proposed by Kermack and McKendrick **[Ker27]** who also collaborated with Ross.

3.2.1. Compartmental model

One of the initial objectives of epidemiology is to determine the outcome of an epidemics: "*How many individuals will be infected when the epidemic ends?*" Kermack and McKendrick **[Ker27]** proposed the compartmental model to address such question. This model relies on a general principle: decomposing the population in three compartments according to their infection status:

• Susceptible (**S**): Individuals that are naive to the disease (and with no immunity)

• Infected (**I**): Individuals that are currently infecting and expressing the disease

- Recovered (**R**): Individuals that have been infected but have no more expression and cannot be infected anymore (immunity or death)

From these three compartments, the dynamic of the model is formulated by *Ordinary Differential Equations* (ODE) based on two parameters: β the transmission rate and ν the recovery rate. The original compartmental model is shown in the following equation system:

$$\begin{cases} \dfrac{dS}{dt} & = -\beta SI \\[2em] \dfrac{dI}{dt} & = \beta SI - \nu I \\[2em] \dfrac{dR}{dt} & = \nu I \end{cases}$$

Despite its simplicity, this model admits no generic analytic solution. Nevertheless, this system has an important advantage: it is possible to compute the R_0 ratio that defines whether the epidemic will go on or fade out. This is called the "*basic reproductive ratio*" or the "*basic reproductive number*". This ratio is defined after 2 parameters β (effective contact rate) and ν (removal rate) and computed as: $R_0 = \beta\ /\ \nu$ **[Jon07]**. If $R_0 < 1$ the epidemic is going to fade out while $R_0 > 1$ means that the epidemic is going to spread all over the population. Evidently, it is not so simple many factors, that cannot be evaluated simply, are determining the value of the two aggregated parameters of the model, β and ν.

The main advantage of this model is its simplicity, it is difficult to envision a simpler model to represent the propagation of a disease in a population. It is also its main defect as it forbids the model to take into account many specificities of the reference system.

3.2.2.Extensions

From the original compartmental model of a disease propagation, many improvements are required to represent more precisely the evolution of different diseases. The initial compartmental model considers a communicable disease where once an individual interact with a disease, it becomes instantaneously and irremediably infected (and infectious at the same time). After some time, this individual becomes also irremediably recovered. This is correct for a limited set of diseases while many others have more infectious stages. For instance, the Measles disease has an asymptomatic stage when the individual is infectious but does not show any sign of infection. Many extensions to the original compartmental model have been proposed to take this situation into account and many others.

- ***SIRS and SIS extensions***

The first extension considers that recovered individuals may not be recovered for ever. There are actually three possible states: individuals can clean out the infectious agent (virus, bacteria, etc.), individuals may develop immunity to it, and last individuals may die from the infection.The original SIR model considers life-long immunity and death within the same compartment because both types of individual are not interacting anymore with the other two compartments.

Some disease are known to induce temporary immunity, like Influenza[2] for instance **[Kyr05]**. In order to handle this issue, the SIRS model has been proposed. The original SIR model's equations have been simply adapted to allow a loss of immunity according to a parameter *f* representing the speed of immunity loss. The updated equations are :

---

2 Immunity for a given strain or "type" of influenza

$$
\begin{cases}
\dfrac{dS}{dt} & = fR - \beta SI \\[2em]
\dfrac{dI}{dt} & = \beta SI - \nu I \\[2em]
\dfrac{dR}{dt} & = \nu I - fR
\end{cases}
$$

The SIS model has been proposed to represent epidemic of a disease without immunity and no death. It does not have a "recovered" compartment, infected individuals just become susceptible again after the infection. This model has been used for bacterial diseases, such as Meningitis or Malaria **[Het76]**.

- **SEIR extension**

The SEIR model is another example of extension, which is suitable in the case of many childhood diseases like Measles or Mumps. These diseases are characterised by an asymptomatic period **[Sim06]**: the individual has the virus (or bacteria) in its organism and is generally contagious (depending on the disease) but expresses no symptom. A new compartment of population, called "Exposed" (**E**), is added in the SEIR model. The dynamic of this new compartment is defined by the update of the SIR model as follows:

$$
\begin{cases}
\dfrac{dS}{dt} & = -\beta SI \\[2em]
\dfrac{dS}{dt} & = \beta SI - \theta E \\[2em]
\dfrac{dI}{dt} & = \theta E - \nu I \\[2em]
\dfrac{dR}{dt} & = \nu I
\end{cases}
$$

In order to account different factors such as immunity (maternally derived, cross-immunity[3], etc), asymptomatic carrier, co-infection, etc. it is possible to add one or several compartments to represent more closely the disease(s) **[Ack05]**.

- **Vaccination and maternally induced immunity**

Vaccination can be studied by means of a new population compartment, called "Vaccinated", as done in the previously described extensions. Most of the time, models considering vaccination also consider population dynamics thanks to the addition of a parameter "μ" called the renewal rate. The equations can be updated as follows (from **[Bra08]**):

---

3 Immunity to one disease which is also effective against a similar (but different) disease

$$
\begin{cases}
\dfrac{dS}{dt} & = \mu N - \beta S I - \mu S \\[2.2ex]
\dfrac{dI}{dt} & = \beta S I - \nu I - \mu I \\[2.2ex]
\dfrac{dR}{dt} & = \nu I - \mu R \\[2.2ex]
\dfrac{dV}{dt} & = \alpha N P - \mu R
\end{cases}
$$

Another common extension of the original model has been proposed to represent maternally induced immunity. For instance, newborns can be immune against disease like Measles for some time **[Bjo02]**. The previous models can be adapted to handle this case by simply considering that 100% of the new individuals are immune (or vaccinated).

- *Combinations*

Eventually it is possible, if it represents an effective disease, to combine or extend as much as one requires. For instance, a detailed infection model of Avian Influenza is the SEIIR used in **[Ren07]** where the infectious state is decomposed in two different states due to the different levels of mortality and virus excretion.

3.2.3.Limits and conclusion of compartmental models

Compartment-based models can consider many factors affecting the evolution of an epidemic. Different factors can be studied like the effect of vaccination or various kinds of immunity. However, it is difficult to integrate these factors altogether in a single model. Indeed, it would lead to intractable equation systems with numerous parameters like in **[Bow10a]** which presents a model with n "exposed" states to study the tuberculosis dynamic in a generic way, as described in Figure 1.10.
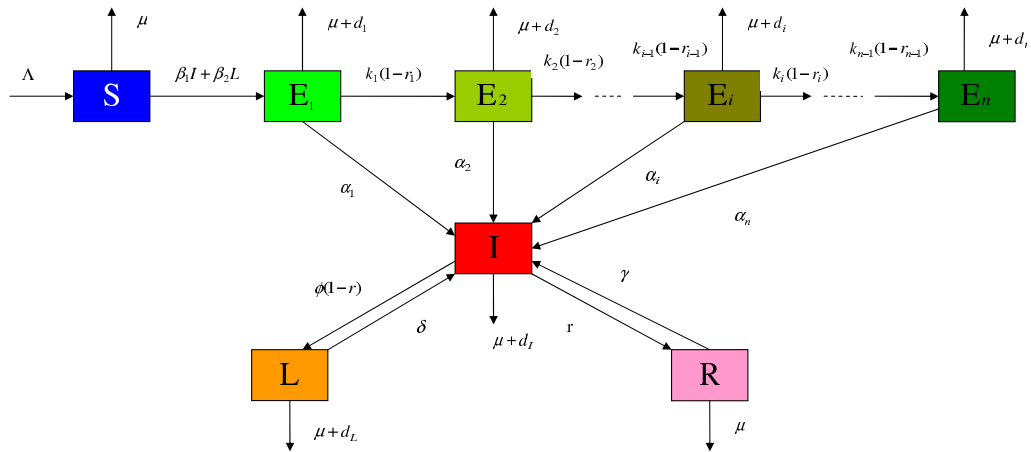
Fig.1.10 A tuberculosis compartmental model with two differential infectivity and n latent stages **[Bow10a]**

The same authors also recently proposed an almost intractable model of the malaria transmission summarised **[Bow10b]** in Figure 1.11. This model considers numerous infectious status: an infective one, four exposed, four infected, etc. It also presents all the possible transitions between them and the associated probabilities.

Fig.1.11 A model of hepatitis and malaria co-infection **[Bow10b]**

Even though the extensions of the original SIR model complexify the system representation, the conceptual complexity remains equivalent. The addition of compartments is imposed by the reference system, for instance Measles has a significant exposed phase. Despite their apparent complexity, the Tuberculosis model **[Bow10a]** and the Malaria co-models **[Bow10b]** are not more conceptually complex than the original SIR model.

• **Conclusion**

The compartmental model proposed by Kermack and McKendrick **[Ker27]** and its extensions are similar to the initial models of the general examples, presented in Section 2, for the following reasons. Their objectives are prediction: "What proportion of the population will get the infection?"; or generalisation: "All the diseases causing a given status set can be represented with a single model (adapted through parametrisation only)". The reference system is considered only at the global level: The representation does not consider the system constituents and the dynamic is expressed globally. Moreover, these models do not explicitly consider any elements that may impact indirectly the system's dynamic, models are parameterised with global probabilities only..

Compartmental models are probably the cornerstone of epidemiological modelling. To such an extent that more than often the compartment-based representation of the system is selected even before the reference system is clearly assessed and thus the representation is "imposed" onto the system. This "modelling habit" of relying on a classic representation of the reference system can result in poorly adapted and intractable models (e.g. **[Bow10b]**).

3.3.Shift of representation

According to **[Koo04]**, compartmental models are based on seven hypotheses but two general ones, of particular importance, can be expressed as: The considered population is homogeneous (i.e. individuals are identical) and "well-mixed" (i.e. the probability of contact between two individuals is the same for all of them). It is not always possible to consider such hypotheses relevant and, thus, the compartmental models are not satisfactory anymore. They actually lack the representational power to be able to consider heterogeneous population or its structure (i.e. not well-mixed of the population). Consequently, new representation paradigm or meta-models are needed: micro-simulation can represent heterogeneous population while cellular-automaton can take into account the (spatial) structure of the population. Finally, MS and CA have some limitations that can be addressed by the use of the agent-based models.

3.3.1.*Individuals representation: micro-simulation*

Considering diversity within a population by using analytical characteristics (e.g. mean or variance) would be a limited approach as it simplified representation of the individuals heterogeneity. A direct way to represent this population heterogeneity is the representation of the individuals themselves. One of the early paradigm used in this purpose is the *micro-simulation*. The model itself is formulated as a matrix where each row (or column) represents the characteristics of an individual. The dynamic of the whole system is expressed as global rules which are applied to each row (or column) of the matrix. Table .1.1 presents an example of micro-simulation matrix of individuals: Each row represents an individual and each column an attribute. An example of global rule to apply on this matrix could be: "*Each infected individual potentially infect a susceptible one at each time step. The probability of effective infection depends on the susceptible individual characteristics*".

	Age	Sex	Weight	Infectious status
Ind *1*	18	Female	60	Susceptible
Ind *2*	42	Female	56	Infected
...	...	...	…	...
Ind *n*	29	Male	71	Recovered

Tab.1.1 A micro-simulation matrix

The integration of the population heterogeneity in terms of sex, age, natural resistance to disease or cross-immunity overcomes the limits imposed by a limited number of compartments. For example, a model studying the "*sleeping sickness*" has been proposed in **[Art01]**. In this case the recourse to micro-simulation model was required because the population size was too small too consider it homogeneous: The difference of a few individuals changing dramatically the dynamic of the system. The model enables the authors to show that a slight difference in the number of the Tsetse flies population (the vector of this disease) can lead to very different situations that cannot be handle with compartment-based models.

Micro-simulation models are fitted to represent heterogeneity of a population but are not adapted to represent the (spatial) environment.

3.3.2.*Environment representation: cellular automata*

It is not always possible to consider the population as "well-mixed" which means that its structure should be accounted for. The population's structure is largely dependent on the spatial environment thus its explicit representation is needed. This is not possible with compartmental models, thus another meta-model has to be used.

Cellular automata (CA, Section 2.3.2) represents explicitly the environment and allow to investigate its influence on the system's dynamic. By decomposing the environment into inter-connected places this representation allows to consider the heterogeneity of the environment and its spatial effects. In an

epidemiological model, each cell represents a located population[4]. Each population (of the cells) are represented as a compartmental model to which is added equations to represent interactions with population situated in the neighbouring cells. These interactions usually represent migration of populations. Thus, the impact of the environment is accounted for as shown in **[Fu03]**.

3.3.3.Individual and environment representation: agent-based models

• *Minimal model*

An effective way to couple individuals and environment representations is the agent-based model. For example, the model depicted in **[Yor79]** has been extended by Willensky and implemented in the NetLogo platform **[Wil98]**. In this model, agents (representing the individuals) are located in an homogeneous environment represented by a grid. This model remains quite theoretical as it represents a generic communicable disease which is transmitted through contact (being on the same cell). Despite its simplicity, this model is interesting as it allows experts to "play" with parameters and have a direct feedback of the simulation. Parameters not related to the disease, but influencing it, can also be included like the individuals' movement speed, the density of population, etc. An enhanced model has been proposed by **[Dun05]**. The reference system is still a theoretical system where individuals are place in environment with home and work location and they go back and forth. This model is interesting because it details the core representations shared by most of the epidemiological models, subsequent agent-based models add numerous details but remain conceptually equivalent.

These models consider the spatial environment as a medium where individuals are moving but the emphasis is on the social environments (i.e. the individuals contact network). Thus the spatial environment is relegated to be considered as a second class constituent of the real system. Recently, new models are being proposed to address situations where the natural environment is much more important (than the social for instance).

• *Improved model*

Muller *et al.* **[Mul04]** studied the dynamic of Human African Trypanosomiasis (HAT), also known as sleeping sickness, at the village scale with data from Southern Cameroon. This disease is different from the ones previously presented. Indeed, its diffusion does not rely on inter-individual contact but is vector-borne (transmitted through individuals of a different species). The vector is the tsetse fly and hosts are humans and, at a lesser extent, pigs. The modelling choices are quite different from the choices made in previously presented models. Here, the spatial environment is decomposed in "agentified" places. These agent-locations are linked together, constituting a simple network of places. The other originality of this model is that it represents an heterogeneous environment (pond, village, field) where the densities of hosts (humans and animals) and vectors (tsetse flies) vary greatly whereas CA based model considers homogeneous environment and previously presented ABM used the environment's topology only (no environmental dynamic) . The spatial behaviours of both, hosts and vectors, are explicitly represented, thus their influence on the diffusion of the disease can be studied. The main result of this study is that a sustained low prevalence can occur for long period of time (13 years simulated) which is impossible to simulate using compartment based models. The main drawbacks of this work are the limitation of the environment representation to a set of interconnected locations and some simplifications: Populations remain constant (except animal hosts population which has a specific dynamics, e.g. with a mass pig slaughter during festivities). Finally, the authors note that the environment representation is not extensive enough. For example, animal's environmental attractors exist (e.g. ponds, spatial discontinuities or ecotones), at least for pigs, and are not represented in the model and the human dynamics is very limited.

Whereas micro-simulation and cellular-automata allows one hypothesis relaxation each, respectively population's homogeneity and "well mixed" population, the agent-based models allow to be free of both.They can represent the spatial structure of the population and its heterogeneity, by the explicit representation of each individual and locating them within an environment representation.

3.3.4.Conclusion

---

4   It is also possible to consider a set of individuals in each cell, but that is beyond the scope of this subsection.

The representation evolution of the epidemiological models started with models (i.e. compartment-based) considering reference systems at their global level with little to no interest of the internal mechanisms leading to the global dynamic. On the contrary, many recent models takes a great deal to consider such mechanisms in order to assess new situations. For instance, **[Art01]** can represent the sleeping sickness propagation according to a finer granularity, **[Fu03]** can study spatial patterns of an epidemic and **[Dun05]** can investigate the impacts of commuting between home and work on a disease propagation.

Similarly to the general examples presented in Section 2, the representation evolution of the epidemiological models is required by the evolution of the models objective. Compartmental models are targeted at prediction (outcome of an epidemic) and generalisation (general characterisation of the epidemic) but cannot allow exploration and explanation of the mechanisms governing an epidemic due to their limited representational power. Consequently, new models, based on different meta-models, have had to be defined in order to have a fitter representation of the reference system. The evolution is also perceptible in other modelling facets, like the modelling approach as described in the next section.

3.4. Shift in modelling approach: EpiSIM and Episims

The previous section highlights the shift of representation from global ones to explicit representations of the reference system constituents by presenting CA, MS and ABM models. In the present section, I illustrate the shift of modelling approach with mostly agent-based models though limited approach shift occurred with two notable extensions of the compartmental model.

The initial model by Kermack and McKendrick **[Ker32a]** and its derivatives always consider epidemiological systems at their global level (i.e. population level) and consider solely the epidemic and no factor affecting it, except through the very aggregated parameter basic reproductive ratio "R_0" (Section 3.2.1). Such approach has been recurrent in Epidemiology as the variety of compartmental-based model shows (Section 3.2). Nevertheless, the modelling approach evolved and more and more factors are considered. The models presented in this section consider the reference system containing several interlinked dynamics.

3.4.1. Adding new dynamics

- **Population dynamics**

Even if the original Kermack & McKendrick model **[Ker32a]** has many offspring models considering solely the epidemiology, the original authors also saw that some dynamics not directly related to the epidemiology impact greatly epidemics both quantitively and qualitatively. The dynamic they consider, which is probably the most important, is the population dynamic (i.e. renewal of the population). To take it into account, the authors amended their initial model by adding a parameter "μ" which represents the renewal rate. Differentiating birth and death rates does not deeply alter the expression of the model.. The resulting model is the following (from **[Ker32a, Ker32b]**):

$$
\begin{cases}
\dfrac{dS}{dt} & = \mu N - \beta SI - \mu S \\[2ex]
\dfrac{dI}{dt} & = \beta SI - \nu I - \mu I \\[2ex]
\dfrac{dR}{dt} & = \nu I - \mu R
\end{cases}
$$

The expression of the R_0 ratio must be updated as: $R_0 = \beta / (\mu+\nu)$. According to the value of this R_0 ratio, two outcomes are still possible. If $R_0 < 1$, the epidemic is going to fade out and the situation is called a Disease Free Equilibrium, while if $R_0 > 1$ the epidemic is going to spread through the population. However, the epidemic cannot reach 100% of the population due to the renewal dynamic. This situation causes an endemic and is called an Endemic Equilibrium.

- ***Environmental transmission***

Another amendment of the initial compartmental model has been proposed much more recently in **[Roc09]**. This model has been defined to demonstrate the influence of the environment on the Avian Influenza propagation among wild birds. The addition of an environment compartment allows the authors to question its interaction with the host population, and thus to represent new dynamics. The equations are updated as follows where b_m=*birth rate*, d_m=*death rate* (from **[Roc09]**). The transmission rate, λ, is actually computed in two ways: As a contact rate and a mix between contact rate and environmental infection rate based on the virus level in compartment *B*.

$$
\begin{cases}
\dfrac{dS}{dt} & = b_m N - (\lambda + d_m)S + \varepsilon R \\[2ex]
\dfrac{dI}{dt} & = \lambda S - (d_m + \sigma)I \\[2ex]
\dfrac{dR}{dt} & = \sigma I - (d_m + \varepsilon)R \\[2ex]
\dfrac{dB}{dt} & = \gamma I - \pi^* B
\end{cases}
$$

Even though, these amendments allow to take into account other dynamics that influence the epidemiological one, models objective do not usually leave prediction and/or generalisation totally. Indeed, one of the question addressed to the model described in **[Roc09]** is to fit better the observed evolution of the Avian influenza Epidemic in wild birds thanks to the consideration of the environment transmission.

Adding the representation of the population dynamic or the environmental one do not alter much the expression of the model. In contrary, it denotes a change of modelling approach. Instead of increasing the fitting of the model by having a more detailed representation of the population infectious statuses by adding a new compartment, this model adds a compartment to represent dynamics that were not accounted for in previous models.

3.4.2. Considering a multi-scale environment: EpiSIM

The EpiSIM model proposed by Simoes **[Sim06]** is a good example of how different concerns can be integrated in a model. EpiSIM aims at investigating a Mumps epidemics which occurred between 1993-1996 all over Portugal. This model is constituted of two sub-models respectively dedicated to movement and infection. The infection sub-model is based on a classic SEIR compartments which is expressive and general enough. The movement sub-model is based on a combination of GIS and ABM. Based on real data, the movement of individuals is represented at 4 levels through different social networks (within a block, inside a city, between neighbouring cities and inter-provinces). The nodes of these networks are then located thanks to the GIS as shown in Figure 1.12 where Lisbon appears to be the infection network centre as in the actual epidemic.This model is interesting because it adapts the compartmental model to the agent meta-model and couple it with a GIS integrating real geo-reference data.

Fig.1.12 Example of infection network

Even though this model highlights the approach shift through the consideration of the spatial and social environment influences, it also illustrate the objective shift: Instead of predicting the outcome of an epidemic it re-enacts it in order to understand how it unfolded.

3.4.3. Coupling social contacts and spatial environment : EpiSims

Another set of model examples illustrates this approach evolution. The EpiSIM model was implemented from scratch and the resulting tool focuses on one specific application: Measles epidemic in Portugal at a country level. In contrary, examples presented in the present section are all based on generic platforms which has the advantage of being re-usable. **[Eub04]** proposed an epidemiology oriented Multi-Agent Based Simulation (MABS) platform called *EpiSims*. This tool mainly focuses on the transportation network within an urban area and on the emergence of individuals' contact network. Recently, it has been improved to obtain a more scalable tool called EpiSimdemics **[Bar08]**.

The EpiSims agent-based simulation platform has been used at various scales (from city to state level). For example, **[Str07]** presents a study of the diffusion of Influenza within the Southern California. One question addressed to this model were to link the severity of the epidemic to the characteristics of the system's constituents. Their model integrates various representations: each individual, its household, its work, its leisure places, its schedule and the transportation networks as well. Figure 1.13 shows the three typical schedules (work, study and leisure) of the represented individuals (worker and student). Each line represent a typical day schedule and each node of the graph represents a location or activity where infectious contacts can occur. This schedule graph and a GIS are associated to represent the environment: realistic movements emerge as shown in Figure 1.14. In this figure, the movement routines (home-work, home-shopping, etc.) of an individual are located in the spatial environment thanks to the GIS.
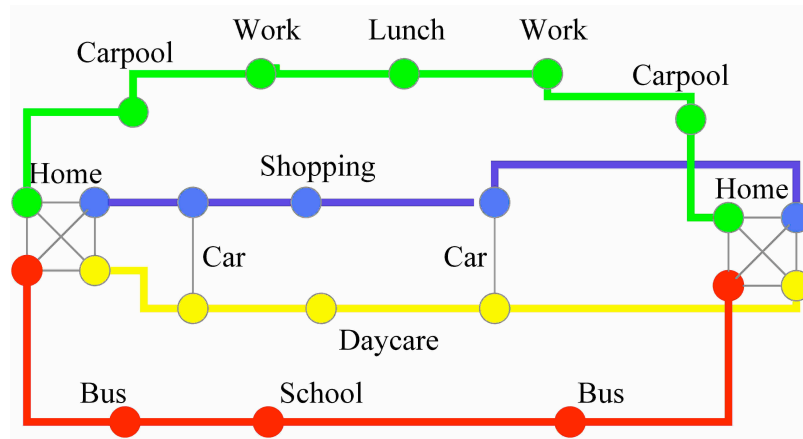
Fig.1.13 Individuals schedules and locations/activities **[Str07]**
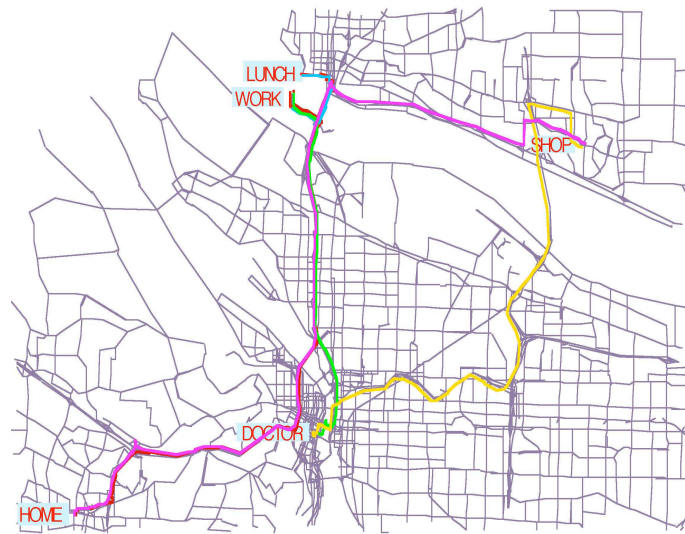
Fig.1.14 Generated movements of an individual in the city **[Str07]**

This study established that the best predictor of the local severity of an epidemic is the household size of a community while at the opposite, the population density is surprisingly a poor predictor. Linking the epidemic's severity to characteristics of a constituent improve the understanding of the system and thus this model can be considered as a explanation oriented one.

Before this state-level model, *EpiSims* has been used at the city level for the simulation of a bio-terrorist attack **[Eub04**, **Bar05]**. This model simulates the diffusion of Smallpox in Portland (Oregon, USA) and investigates the efficiency of different kinds of mitigation measures. In particular, the authors investigate the contact network structure which emerges from the combination of the typical human schedule, the transportation system, the environment topology, etc. The authors observe that this network has a very specific characteristic. Indeed, the contact network is an *expander graph* **[Sar04]**, a specific class of scale-free networks. While scale-free networks can be shutdown by cutting a few hub vertices (i.e. individuals), it is not possible in expander networks. Indeed, they established that "*everyone but the most devoted recluse is effectively a 'small hub', thus classical 'hub individuals' are not so crucial anymore*" **[Bar04a]**. Consequently, they showed that "*a policy of targeting only hub individuals would probably do little to slow the spread of a disease through the city*". It is a very interesting result: the network is emerging from local interactions of the individuals (i.e. contacts). They discovered a characteristic of the real system using the model without a priori information of this characteristic. However, field validation has still to be conducted. This model can be considered as a exploration oriented one as different mitigation measures against a bio-terrorists attacks have been evaluated instead of a "mere" prediction of such an event.
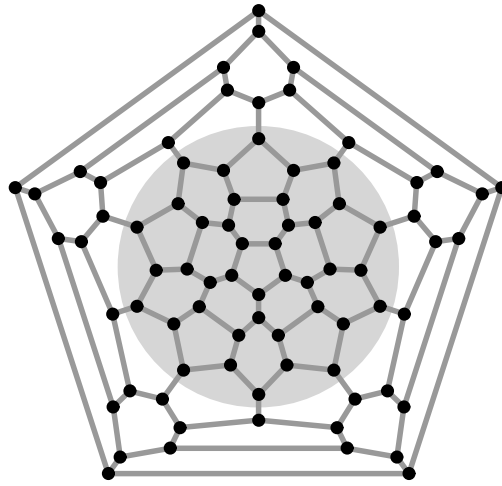
Fig.1.15 A simple expander graph (sparse graph with strong connectivity) **[Sar04]**

Even the initial Kermack and McKendrick model **[Ker27]** can be amended when a shift of approach is occurring as they were able to take into account several dynamics at a time (e.g. disease direct transmission and population dynamic or environment transmission). Nonetheless, this approach evolution is very limited as the representational power of compartmental model is also (Section 3.2). Moreover, these models are still designed in order to address prediction and/or generalisation objectives.Nevertheless, the other examples **[Eub04**, **Bar05 Sim06, Str07]** constitute the real approach evolution. Indeed, they are able to represent much more various dynamics, with a great deal of details and at various representation levels.

In the previous section I presented the epidemiological models evolution in terms of representation while the current one depicts the approach evolution. In these two sections, the emphasis is not set on the objective evolution, though it is mentioned due to the intertwining relations, it is the focus on the following section.

3.5. Objective shift: MALCAM and Simpest

This section presents an evolution in the objective assigned to models in Epidemiology. The "classic" approach only considers prediction and generalisation as objectives while the new approach targets exploration and explanation.

3.5.1. Exploration usage: the MALCAM model

MALCAM **[Lin09]** is an agent-based model focusing on the assessment of the risk of re-emergence of Malaria in Camarges (Southern France). This disease is endemic in many African countries and its endemic area is moving North towards countries up to France. The transmission cycle of this disease includes mosquitoes as vectors and humans as hosts. The transmission occurs in areas where both types of individual meet. To assess such situations, considering the spatial location of individuals is essential in addition to the epidemiological mechanisms. Thus, the explicit representation of the environment is required. Indeed, this risk is tightly related to the land-use (e.g. rice-fields are perfect breeding sites) and to the human presence. In this model, the environment is a lattice of regular cells. These cells are responsible for the environmental dynamics and the density of mosquitoes in a similar way as cellular automata initialised with field data stored in a GIS. Humans and animals (horses and bulls) of the system are represented by agents. The main originality of this model relies on the combination of numerous data sources: land-use covers, social investigations, biological and epidemiological data. In addition, in contrary to other models, this model considers seasonalities of the different entities: rice-fields can be flooded or dried, the touristic season, mosquitoes peaks of reproduction.

The MALCAM model investigates the risk associated with individual kinds: It highlights that more risky areas correspond to human-dense ones close to rice-fields (i.e. ecotone between human residency area and rice-fields). Nevertheless, the humans more at risk are hunters as their activities coincide more with mosquitoes activities despite their preventive measures (i.e. use of mosquito repellent and adapted clothes).

41

This model presents a clear situation of what would happen if the vector, mosquitoes, is moving to the north, due to climate change. In addition, it demonstrates where the disease is likely to emerge: At the ecotone of human habitat and rice-fields (i.e. breeding sites for mosquitoes). This model is clearly an exploration oriented one but also illustrate the need for new representation and new modelling approach.

3.5.2. Explanation usage: SimPest

The *SimPest* **[Lap09]** model is dedicated to the study of the Plague in Madagascar. Plague is endemic in Madagascar since 1898 and is mainly transmitted to humans from rats, which act as a virus reservoir. The question addressed to this model is evaluating the impact of rats behaviours onto the disease dynamics in the human population. Rats are important in this context for two main reasons. Rats as humans can be hosts to the disease. In other words, even if the disease is absent from the human population, it can re-emerges later from the rat population. Rats are also important since they carry the actual vector of the disease: Fleas (which are considered motionless if not attached to a rat). Thus, their spatial patterns are pertinent to understand the disease dynamics in human population.

In practice, *Simpest* **[Lap09]** is similar to MALCAM **[Lin09]** as it exhibits an explicit representation of the environment, represent the different movement behaviour of individuals and considers seasonalities too. Nevertheless, the spatial resolution is quite different: Cells' side are around 1.6 meters long instead of 30 meters. Two implementations have been performed: a schematic one and a realistic one using field data from 2 villages, stored in a GIS. The schematic environment model shows the possibility to have a local (village scale) cyclic endemic with very clustered transmission of infected fleas. Figure 1.16 illustrate the schematic representation of a village: A village is surrounded by cultivated land and then natural land, rats and non attached fleas are located in these areas (their colours represent their infectious status).
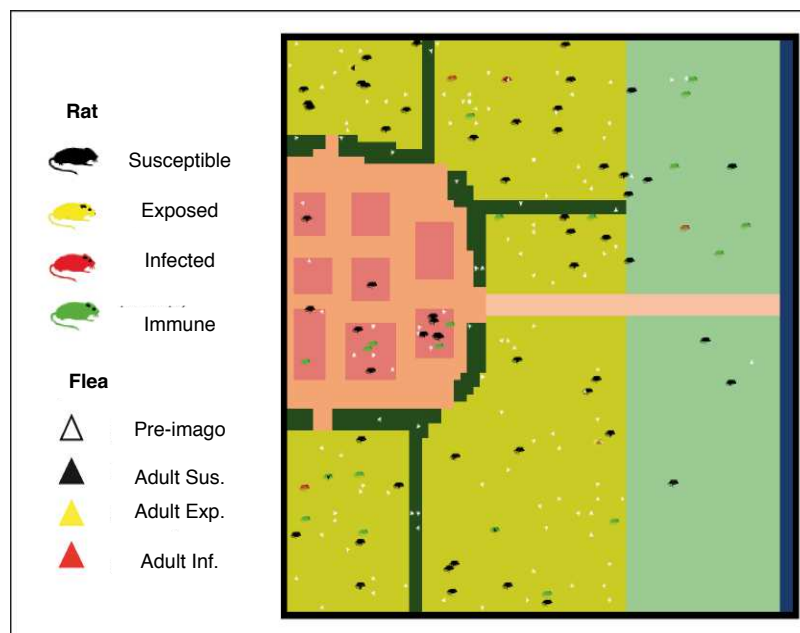
Fig.1.16 Simpest schematic model visualisation **[Lap09]**

In one of the modelled villages (*Amboasary*), simulations never succeed to exhibit an endemic even if it the case in the reference system; the authors assume that some field data were largely underestimated. In the other village (*Ankadimafana*), the endemic is observed throughout the simulation (400 simulated days); this experiment showed a link between the size of the rat population and the prevalence of the disease in the human population.

This model explicitly represents the environment as an euclidean space using land-cover data, it considers the seasonalities and focuses on a village environment. It is intended to be used for "virtual experiments".

Considering the initial epidemiological models and the two models of this section, clearly show the evolution of models in terms of objectives. Whereas compartmental models can be used to address questions related to prediction and/or generalisation, MALCAM and Simpest demonstrate their capabilities to address questions related to exploration and explanation respectively. Evidently, the representation these two models use is much more detailed that the compartmental models one, focusing on system's constituents or the global system itself respectively. Finally, the modelling approach used for these two models differs greatly from compartmental models. Indeed, MALCAM and Simpest integrates various dynamics and different point of views on the system.

3.6.Conclusion

This series of examples shows that the epidemiological models are following evolutions equivalent to the ones illustrating in the general examples of Section 2. Compartmental models are largely predominant in epidemiology which certainly impedes the evolutions of the modelling activity in this field. Nonetheless, the need for "modern" models addressing questions related to exploration and explanation is growing stronger due to the emergence of new epidemiological threats (e.g. new diseases, evolving ones and also bio-terrorism).

The initial model might be improved to address questions related to exploration and explanation though they showed very limited success as presented in Section 3.4.1. On the contrary, brand new types of model as illustrated in Section 3.3, 3.4 and 3.5 were needed to allow better understanding of existing situations (e.g. plague in Madagascar **[Lap09]**) or explore the consequences of situation yet to come (e.g. introduction of Malaria in South France **[Lin09]**). In addition to this shift in objectives this section illustrates the shift in approach: Various dynamics and system's constituents are now assessed to be integrated in the model, and a shift in representation, both individuals and environment are more and more frequently represented at different levels.

4. Synthesis

4.1.Summary

In this first chapter, I presented the evolution of models along a main axis: from prediction and generalisation objectives to exploration and explanation ones. This evolution has two direct consequences: A shift of approach and a shift of representation. I support each of these three evolutions in three different domains: segregation, pedological and traffic models illustrate respectively objective, approach and representation evolutions. In Section 2.1, the initial segregation model **[Sche71]** is a generalisation of an observed phenomenon whereas recent models **[Cro06, Cro08, Cro10,]** allow exploration and explanation of such phenomenon illustrating the objective evolution. The shift of approach is illustrated in Section 2.2: the authors of **[Mar08,Bla09]** collaborated with pedologists to adapt their soil representation to their concern, instead of developing a brand new representation specific to their needs and domain. The last evolution, the representation one, is detailed in Section 2.3. Early traffic model (especially LWR) represents the system at its global level allowing no adaption whereas numerous more recent models consider it at various level and especially the individual one. Even though three different examples are used to present the three evolutions, it is clear that these evolutions are entangled altogether. Indeed, the evolution of objective, or more practically evolution of question addressed to the models, requires new approach and representation and, conversely, any improvement of these two modelling aspects can address of unanswered questions.

These evolutions are occurring in many applied science fields but is certainly of specific importance in Epidemiology. Indeed, many recent epidemiological crisis, such as the Chikungunya epidemics or the Avian Influenza one, are challenging epidemiologists and their models. Whereas prediction and generalisation model have been largely predominant in the field, these new events require models oriented towards understanding and exploring probable epidemic evolutions. Similarly to the first series of examples illustrating the modelling evolutions, I present the three evolutions separately even though they are also entangled in epidemiological models. The initial compartmental model **[Ker27]** and its extension are emblematic of the starting point of these three evolutions (prediction and generalisation oriented, focusing on epidemiological parameters solely and considering the system at its global level). I presented several models highlighting the evolutions that occurred in epidemiology. In practice, the representation evolution is illustrated in Section 3.3 with CA, MS and agent-based models replacing the compartment-based models to

allow much more varied and adaptable representations. These meta-models allow the explicit representation of the reference system constituents. Similarly, Section 3.4 highlights the approach evolution with the EpiSIM **[Sim06]** and Episims **[Eub05,Str07,Bar08]** models. These models integrate large knowledge from different domains in order to investigate epidemiological phenomena.

We can see that these modelling evolutions lead to the differentiation of two model types based on their objectives. Whereas prediction and generalisation oriented models are satisfied by considering the system at its global level (and even possibly as a black box), the exploration and explanation ones require the thorough consideration of many system's components. Such opposition can be linked to the opposition that Edmonds formulates in **[Edm04]**. In this article, he proposes an opposition or a response, called the "*KIDS approach*", against the "*simplicity principle*", advocated and named "*KISS* " by Axelrod **[Axe97]**.

4.2. KISS and KIDS

4.2.1. KISS

In **[Axe97]**, the author advocates for simplicity-oriented modelling in general, though he focuses on the agent-based models, and states that "*while the topic being investigated may be complicated, the assumptions underlying (an) agent-based model should be simple*", following in that stance the KISS («Keep It Simple, Stupid») principles initially coined by Kelly Johnson **[Wik01]**.

The KISS principle is no more than a rephrasing and application of the Occam's razor (or "law of parsimony") principle to the modelling activity. This principle states that "*the simplest explanation is the best*" **[Cho08]**. In other words, when two sets of hypotheses can explain the same phenomenon, the smallest one should be chosen. Applied to the model design, this principle states a more complex model (e.g. adding a hypothesis) is relevant only when the result model is more informative than the initial model.

Prediction and generalisation oriented models are in accordance with the KISS principle as these models can be easily evaluated. Indeed, the predictive power of a model corresponds to its fitting to existing data. Similarly, a generalisation oriented model can be evaluated through the number of different reference systems it can represent (in addition to its fitting to these different reference systems). Thus, the addition of a new concept or hypothesis in the model can be evaluated by comparing the model's fitting with and without it.

Unfortunately, KISS is not fitted when we have no direct and objective evaluation of the model. This is the case for exploratory and explanatory models. Exploratory models are conceived to investigate the dynamic of a system before the conditions applied to the model have been actually observed on the reference system thus making it impossible to evaluate, at least as previously. Explanatory models are also difficult to evaluate for two reasons. First, an explanatory model cannot be evaluated in terms of fitting, the explanation it proposes requires further field survey to be validated. Second, the complete explanation of an observed phenomena can be an emergent one requiring numerous details to be represented before it could be generated within the model and thus the modelling adapted Occam's razor cannot be used.

Axelrod's paper **[Axe97]** was published while the design methods for agent-based model were still immature. Thus the simplicity orientation was relevant in order to keep a clear track of all interactions between agents in ABMs.

4.2.2. KIDS

The design methods for agent-based models have greatly evolved since the publication of Axelrod's paper **[Axe97]**. Based on this observation, Edmonds **[Edm04]** tackles the KISS modelling principle and proposes a set of principles called KIDS (*Keep It Descriptive, Stupid*) which encourages modellers to consider the modelling process in an upward way. The representation is not defined at a global level of description nor is "imposed" onto the system. He advocates to build the model by representing explicitly the system constituents and let the dynamics results from the interactions of the represented constituents. Moreover, he admonishes to assess the integration of representations of as many as possible systems constituents.

In order to envision representation of as many as possible system's constituents, the modeller, must not rely on modelling habits of the domain expert's field. For instance, probably due to the importance given to compartmental models, epidemiologists almost always consider the system at the population level and

overlook most of the within-population dynamics and its heterogeneity. In contrary, different levels of representation should be assessed and data or knowledge used to design the model should not be limited to the one domain in which occurs the question. Logically, it complexifies the modelling process, instead of a single or few model lines it leads to large model clades. Moreover, it may lead to several clades as different domain models can be envisioned when contradictory knowledge are expressed due to the multiplicity of point of views.

KIDS has been proposed as a response to the failure of the KISS principles to deal with these "new models". It highlights the need to consider the complexity of the reference system. It stipulates that this complexity might be reflected in the model but has, for sure. This complexity should not be avoided: simplicity should not be pursued for its own sake, as illustrated in Figure 1.17. In this figure, we can see that the complexity of a KISS oriented model only grows while the complexity of a KIDS oriented one can grows or shrinks during the modelling process. It clearly illustrates which quality each approach favours: representation simplicity for KISS and representation details for KIDS.
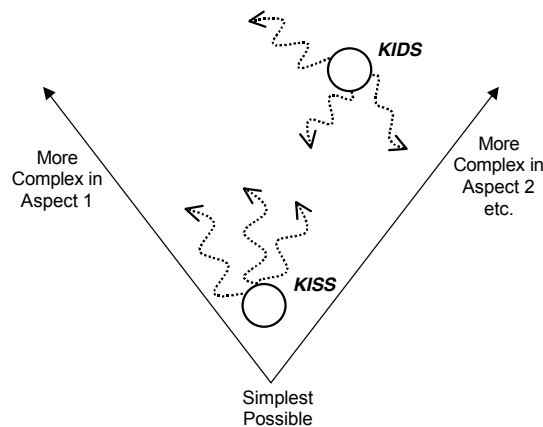
Fig.1.17 Illustration of KISS and KIDS **[Edm04]**

KISS principle and KIDS ones provide advices on how to design a model from a reference system. Unfortunately, these advices are very general, not to say vague, and can be summarise as: Simplicity orientation against descriptive representation. Easing the design of exploratory and explanatory models cannot rely on the KIDS approach, a complete modelling method should be proposed.

Numerous modelling methods have been proposed **[Sar05, Fis95, Zei89]**. They are usually well-suited for prediction and generalisation but their adequacy for exploratory and explanatory models is unknown. In order to be able to assess their adequacy, a clearer characterisation of the modern models introduced in this chapter is needed. In the next chapter, I give a characterisation of such models which I call ***world models***. Defining these models allows me to express their design requirements. From these requirements, I evaluate existing modelling methods and tools to support their design.

Chapter II : World models

Nowadays, modelling has become a common tool in many applied research domains (e.g. Sociology, Ecology, Economy or Physics). The research context and the question to investigate determine the usage, the representation and the modelling approach that should be considered. Chapter 1 has shown the recent evolution of these considerations, which have led to the emergence of a new type of model. None of the legacy modelling approaches, KISS and KIDS, can easily handle. I propose in this chapter to call these models ***world models*** and to characterise them in more details.

Even though a sketch of world models can be easily derived from the examples of Chapter 1, it is necessary to define them more formally and list explicitly their characteristics. After such characterisation, their design is the following logical concern. Thus, I review two main types of methods, from a very general one to a very specific one: Software engineering and "complex system modelling". None of them really addresses the issues raised by the design of world models but provides several relevant practices. Finally, tools that can be used to design and implement world models are introduced. Such tools and methodological practices constitute the basis on which I ground my proposal of a modelling method: **KIMONO** (Keep It Modular, Open-ended, and Neatly Organised), which focus the creation of world models.

1. World models

In Chapter 1, I conclude that prediction and generalisation targeted models are well-known and adapted modelling methods exist. Moreover, the KISS principle **[Axe97]** provides a general orientation to design such models. In contrary, exploration and explanation targeted models are poorly characterised and no adapted (and general) modelling methods have been proposed so far. Only the KIDS approach, proposed by Edmonds **[Edm04]**, provides a general orientation to design such models. Unfortunately, it does not provide any practical advice on how to design exploration and explanation oriented models and very few characterisations of these models . However, a clear characterisation of these models is essential before addressing modelling method issues.

In this section, I propose a characterisation of these models named: **World model** which is first introduced by considering the exploration and explanation models from Chapter 1. I come back on these model objectives evolution and the co-evolutions of questions and modelling techniques that lead to the need characterisation of the world models. Then, I review the issues to design world models: Selecting a meta-model and a modelling method. Chapter 1 model examples hint towards the use of the agent meta-model but provide scarce information concerning the method that was used to create them.

1.1. Introduction

1.1.1. Towards explorative use of models

Introductory examples described in Chapter 1 can be used to derive general characteristics of these models. Despite their differences in terms of application and formalism, these examples have a common trait: they investigate questions using a rich representation of the reference system, a wide consideration of its constituents and adopting a new point of view on the whole modelling process: The model and its design process become of interests, not only its outputs.

For instance, the traffic flow models show that in order to investigate the emergence and the dynamics of various phenomena such as start-stop waves or synchronised flows, the representation of individuals, with their behaviours and internal states, was essential. The segregation model highlights the profits of integrating modern tools like GIS, which upgrades the representational power of the model and allows the study of new dynamics. Finally, the SWORM model shows that it is relevant to adapt a representation (APSF) proposed in a given domain (soil structure) to other related domains (e.g. organic matter dynamic in the soil) within a different formalism (ABM instead of equations).

1.1.2. Co-evolution of questions and models

The evolution of modelling techniques and research questions to investigate has not been straightforward. The emergence of new research questions and the development of new modelling techniques are clearly distinct processes, their evolution is actually tightly related. For instance, in Epidemiology, populations were initially considered as homogeneous and compartmental models have been adopted. Then, the knowledge about disease dynamics were gradually integrated and thus models based on micro-simulation and cellular-automata have been conceived. In the case of the Sworm model, two existing models were coupled to address an existing question more effectively. Sworm adapts the soil representation of the analytical model, PSF, as a computational model, called APSF, to investigate earthworms actions on the soil structure whereas the original PSF model was developed to study the (static) soil structure. Finally, tools, such as GIS, enable to address existing questions unanswered so far. For instance, MALCAM and Simpest studies could not have been done without a very detailed representation of the environment which is now made possible through the recourse of GIS.

The following section bases the review of world models' general characteristic on the examples used to portray this evolution. Among these examples, the epidemiological ones are very representative of the changes emerging in modelling approaches. This domain has a wide corpus of knowledge and a priori modelling traditions.

1.2.Characterisation of the world models

World models are a new step in the co-evolution of research questions and modelling techniques. In this section, characteristics of world models are first extracted from examples, leading to a clear definition of world models.

1.2.1.Extracting general characteristics from the examples

- *General examples*

The segregations models illustrate the evolution in terms of objectives. This evolution leads to a detailed representation of the environment and to allow the assessment of knowledge and hypotheses that have not been questioned before (e.g. impact of the topology, of the number of social groups or of heterogeneous group preferences).

The evolution of modelling approaches is illustrated with the Sworm model which combines different modelling technics and knowledge from distinct domains. This model also highlights the necessity of inter-disciplinary collaboration between domain experts and the pressing needs for dedicated tools **[Gau09]**. Finally, it also stresses out the importance of the environment representation and its potential complexity (multi-scale for instance).

The traffic flow models illustrate the representation evolution: From a global description of the reference system to consideration of its constituents and their interaction. In addition, it highlights that: (1) recent models aim at exploring and explaining; (2) knowledge from different fields must be integrated to design such models (e.g. road regulations, sociology and psychology of drivers or physics). The global dynamics of the observed reference system emerges from the interactions between the different constituents of the model. Emerging dynamics are at the heart of the modelling process of complex systems: by representing extensively the constituents of the system, *emergence* is likely to happen.

- *Epidemiological examples*

In Epidemiology, the objectives evolution is introduced with the MALCAM **[Lin09]** and Simpest models **[Lap09]**. The former is dedicated to the exploration of the possible re-emergence of Malaria in South France. The latter investigate the recurrent plague epidemics in Madagascar in order to assess their unknown causes. To attain such objectives both of these models integrate numerous field-data: environmental, sociological and epidemiological. In particular, the environmental data are used to generate a detailed environment representation.

The modelling approach evolution in Epidemiology is introduced by the EpiSIM **[Sim06]** and EpiSims models **[Eub04, Bar08,Str07]**. These models integrates knowledge from different domains: Sociology, Epidemiology, etc. They also account for the influence of the spatialisation.

Finally, the representation evolution is illustrated by various models **[Art01, Fu03, Dun05, Mul04]** grouped by the meta-model they use: CA, MS or ABM. Instead of representing the global system, its constituents (individuals and/or environment) are explicitly represented and interact with each other to generate the global dynamic.

1.2.2. Definition of the world models

From the characteristics of the Chapter 1 model examples summarised in the previous section I can formulate the definition of a world model. The aim of world models is the understanding of a reference system dynamics by domain experts and/or exploration of its future state. In these purpose, world models explicitly represent the constituents of the reference system in a descriptive way. More precisely, an homomorphism between the reference system, at a given level, and the model should exists: Each (relevant) reference system constituents is represent by an element of the model. The environment must be considered as any other constituents of the reference system. The reference system dynamics emerge from the interactions of various elements at different scales. In order to allow investigation of the emergent phenomena the whole model should remain accessible and manipulable by domain experts and capable of integrating information from different and possibly numerous domains. Before giving a definition of a world model I can summarise their characterisation in four elements:

- •(W1) Reference system constituents are explicitly represented

- •(W2) Reference system constituents are described in terms of internal state and dynamic

- •(W3) Individuals and environment are given equal concerns

- •(W4) Domain experts can manipulate the model directly

Definition 2.1 [World model]: A world model is a model clade allowing domain experts to address an exploratory or explanatory question. It is used as a support tool for domain experts to evaluate their hypotheses and orient their (field) research. Concretely, a world model represents explicitly the reference system constituents, both individuals and environment.

1.3. How to design world models?

Proposing a characterisation of world models is the first step to facilitate their design. The following steps include the definition of a **meta-model**, a **modelling cycle** and **contributor roles**. In addition, selecting or proposing **support tools** would also facilitate such process. The world model precursors presented in Chapter 1 propose different solutions. Nevertheless, modelling papers are more or less concerned by these aspects but only the meta-model is always mentioned.

1.3.1. Meta-model

The meta-model used by "world-model precursors" described in Chapter 1, are limited to cellular-automata **[Nag92, Fu03]**, micro-simulation **[Sch71, Art01]** and agent-based model **[Kos96, Wil98, Ni03, Sim06, Nor06, Lin09, Lap09, Cro10]**. Such meta-models allow the representation of the reference system at the constituent level (W1). CA and MS focus on the representation of either the (spatial) environment or the individuals thus fails to address (W3). In contrary, both individual and environment are representable with equal concern in an **agent-based model**. Moreover, the agent paradigm allow to represent both a reference system constituent's internal state and dynamic (W2). In contrary, CA and MS models are governed by global rules. Finally, agent-based model are easily manipulable by domain experts (W4) thanks to the direct link between a system constituent and its agent representation. CA and MS are also accessible meta-models but do no provide a direct link between the system's constituents and model's elements (their formalism are more constrained). In summary, the agent meta-model is fitted to be used to designed world models.

The selection of the agent meta-model constrains the choice of modelling cycle, contributor roles and support tools. In particular, ABM are computational models which means they have to be implemented and executed. Then, an ABM consists in a conceptual model, static representation of the system, and in an implemented model that can be executed. Implementation of an ABM usually requires specific programming skills which domain experts do not have in general. Nevertheless, (W4) implies that experts are

also part of the implementation process thus specific support tools are needed to allow their access to the implementation.

1.3.2.Modelling method

The agent meta-model has been selected as it allow expression of models fitting world model characteristics but it has also been hinted by the review of Chapter 1 precursor models. Nevertheless, these examples and their original publication do not provide any information on the modelling cycle used or the contributor roles.

The selection of the agent meta-model as the most adapted one to the expression of world models have been suggested. Consequently, I have to review specifically proposals about modelling cycle and contributor roles. Nevertheless, definition of these two elements do not constitute a complete modelling method yet. Indeed, a modelling method for agent-based model is constituted of:

(G1) The definition of the roles of the various model contributors,

(G2) A modelling cycle,

(G3) Modelling guidelines,

(G4) Support tools,

Most of the modelling methods focus exclusively on (G1-2) and leave the choice of (G3-4) up to the modeller. Consequently, my review of modelling methods, presented in Section 2, focuses on the two first elements while (G3-4) are considered in Section 3.

Chapter 1 examples are slightly more informative in terms of used support tools than concerning modelling cycle and contributor roles. For instance, there are two (main) options for the implementation of a model: a **general purpose programming language** as C++ used by EpiSIM **[Sim06]** or a **dedicated modelling language** executed thanks to a **multi-agent based simulation (MABS) platform** as Netlogo **[Wil99]** by Simpest **[Lap09]**. In contrary, both of these publication rely on **UML [Ben05]** for the expression of the conceptual models. A third implementation option has been used in **[Cro10]**, the implemented model is expressed in a general purpose programming language (Java) but profits from the resources provided by an MABS platform (RepastJ **[Nor06]**). Another interest of this work is the use of a **GIS** to allow a representation of the environment as detailed as the agent ones. Indeed, many agent-based models (and simulation platforms) lack a detailed representation of the environment. Indeed, the common interpretation of the agent meta-model considers that the environment is of much less concern (or complexity) than the individuals. Finally, the documentation of the model is an important part of the modelling method as it eases the modelling process and facilitate the involvement of domain experts. Unfortunately, it is not strictly considered in any of the presented examples. Nevertheless, the MALCAM model **[Lin09]** is presented using a communication protocol called **O.D.D** which may be used as a model documentation **[Grim09]**. Evaluating these support tools is possible only after the selection of a modelling method. Consequently, the support tools review is presented in Section 3.

1.4.Conclusion

The evolution of questions addressed to models lead to a more general evolution of model objectives. A consequence of such evolution is a need for new models which are dedicated to exploration and explanation. I defined a set of characteristics that a model has to exhibit in order to target such objectives and called these models: World models. Namely, they represent the reference system constituents explicitly, in terms of internal states and dynamics, give equal concern to representation of individual and environment and, are manipulable by domain experts.

Facilitating the design of world models is the subsequent need. The precursor model examples described in Chapter 1 provide only ad hoc and partial solutions. Indeed, the tools (and methods) used are generally application domain dependent. Moreover, publications usually do not mention which modelling method was used. In order to propose a general solution, I listed the required elements easing the design of world models: Meta-model, contributor roles, modelling cycle, modelling guidelines and supporting tools. An adapted meta-model has allow a model to fit the world model characteristics (W1-4). Precursor models from the example of

Chapter 1 hint that the agent meta-model is a possible solution and, indeed, it is as it addresses all these requirements.

Selecting the agent meta-model has been made clear by the description of Chapter 1 examples. Selecting a method is much less clear, for the least. Indeed, many modelling papers present their results in details, their models to a limited extend but their modelling method is rarely accounted for. For instance, of the examples of Chapter 1; only papers about Sworm **[Mar08, Bla09]**, MALCAM **[Lin09]** and SimPest **[Lap09]** provide some insights on their method. Consequently, I review existing modelling methods that may fulfil world model modelling method requirements in the following section. A similar issue occurs with the supporting tools: GIS, some MABS platforms and the O.D.D protocol are mentioned but no consensus exist thus Section 3 is dedicated to the selection of the supporting tools.

2. Methods to design world models

The first section of this chapter characterises the world models, it also defines the agent meta-model as a suitable meta-model with keeping reference to the examples of Chapter 1. It also expresses the need of a modelling method but the mentioned examples propose no solution. Consequently, the current section is dedicated to the evaluation of "general" modelling methods against the world models methodological requirements which are expressed first (Sub-section 2.1). Then, the method review is separated in two sections based on their specificities: Section 2.2 reviews focuses on engineering methods whereas Section 2.3 is dedicated to "complex system oriented" methods..

Providing a "complete" modelling method constituted of (G1-4) is not always mandatory. Nevertheless, the modelling cycle (G2) is certainly the most important element of all. In Section 2.2 and 2.3, most of the presented methods focus on this element, though agent-based modelling and simulation (ABMS) methods (Section 2.3) also consider the roles definition (G1). Nevertheless, the two other elements (G3-4) are not consider the slightest. Thus, these elements are reviewed in Section 3 in order to envision a complete methods.

2.1.Method requirements

The requirements in terms of modelling method can be derived from the world model characteristics (W1-4).. In practice, due to the exploratory or explanatory objective, the design of a world model is likely to face lack of data and multiplicity of possible representation and hypotheses.Testing hypotheses and representation alternatives are common though they do not necessarily require to be based on a "complete" model and thus can be tested before the completion of the model. In order, to ease the design of the model and facilitate the integration of different data, the model should remain accessible to all model contributors. Finally, a world model cannot be validated in a classical way (e.g. comparing its output to the reference system observation given inputs) thus a way to determine if the model is informative has to be defined. Consequently, the requirements the method has to address are:

(R1) Place domain experts at the centre of the modelling process

(R2) Provide tools to ease the incremental and modular design of the model

(R3) Capability to cope with temporary lack of data

(R4) Provide a model evolution assessment mechanism/practice

The design of a world model is likely to require numerous sources of data (Section 1.2.2) in order to obtain a model descriptive enough to allow domain experts to evaluate their hypotheses. Nevertheless, the model should remain accessible to these experts.Consequently, the method must provide tools to cope with different data formalisms and possible different meta-models. Thanks to the integrative capability of the agent meta-model, it is possible to translate other formalisms into it which is the favoured solution over model coupling. This requirement also express that the understanding and the manipulation of the model by domain experts has to be eased by the method. A direct consequence is the need for extensive description of the model all along the modelling process.

Incremental design of the model allows to progressively integrate data and knowledge of the domain experts in the model. The modularity allows also to conduct evaluation of modelling and hypotheses alternatives

evaluation. In practice, the branch organisation of a model clade has to be facilitated by the method. In the end, such tools also help the modelling process to cope with lack of data which is a common issue for world models as all the hypotheses to evaluate cannot be expressed at first.

A world model is not simplicity oriented thus the KISS principle (Chapter 1) cannot be used to assess the correct evolution of the model during its creation process. Although world models are in accordance with KIDS principles (Chapter 1) but they do not provide any method to evaluate the model evolution. Consequently, it is the duty of the modelling method to propose workaround to evaluate the model ongoing evolution.

Validation of a model and its result is an important task of the modelling process, nevertheless it is rarely conducted completely because of its difficulty (time constraint, lack of data, parameters space to important to explore, etc.) **[Hea10]**. It is, actually, an "*ongoing issue*" according to **[Mos08]**. Nevertheless, some workarounds have been proposed. For instance, one called "weak validation" **[Axe97]** which focuses on increasing the domain experts confidence in the operational model. In practice, **[Axe97]** proposes to replicate the model : "*Replication can also be useful for testing the robustness of inferences from models*".

2.2.Software engineering methods

This section reviews several conception methods from software development in general (Section 2.2.1) and agent-oriented software development (AOSE) methods (Section 2.2.2). Even though the first methods have not been designed for the development of agent-based models, they introduce several major concepts and useful practices. In contrary, AOSE have been defined specifically for the development of agent-oriented software and thus are more likely to fit my requirements.

It is important to review recent advances in software engineering processes for several reasons. Designing computational models, or at least agent-based models, is clearly a sub-category of software engineering process. Modern methods share some concerns with the requirements I expressed in Section 1.3.2, wide yet not-so-formal requirements, need of adaptation and involvement of the "client" (or the thematician in my concern). In addition, the agent paradigm is often considered to be an extension or an improvement of the object one **[Sho93]** and many existing agent-based model design methods are based object-oriented ones **[Woo01]**.

2.2.1.Classic methods

• *Waterfall model*

The waterfall model has been conceived during the 50's, became popular in the 70's and is still in use nowadays **[Cer05]**. It is a series of phases that have to be completed sequentially. These steps are: (1) Requirements, (2) Design, (3) Implementation, (4) Verification, (5) Maintenance.

It is a very rigid structure as each step requires the completion of the previous to be started.. Even though there are feedback loops between each contiguous stage, the risk management is poor (if an error occurred in an early stage the consequences will be large). It also totally leaves the client out of the process except for the requirements stages. Unfortunately, most of the clients are not able to express completely their needs at first. This development model does not fulfil any specific requirements of designing world models but it has the interest to put the emphasis on two very general desired qualities: A clear and complete design structure and a detailed documentation which ease the involvement of domain experts (R1).

• *Spiral model*

The rigidity of the waterfall limits it to standard and not-so-large projects; for larger and more complicated project, the spiral model have been proposed **[Boe86]**. It focuses on prototyping and extensive initial In particular, requirements are expressed through interviews with all external and internal users of the system.

As shown in Figure 2.1, each iteration of the development cycle is divided in four phases: (1) objective definition, (2) risks assessments and resolution, (3) development (code and tests), (4) planning for the next iteration. Every iteration results in a prototype and each of them addresses more and more of the initial requirements.

These software development process has several advantages. Client are much more implicated in the development process: expressing the initial requirements through interviews
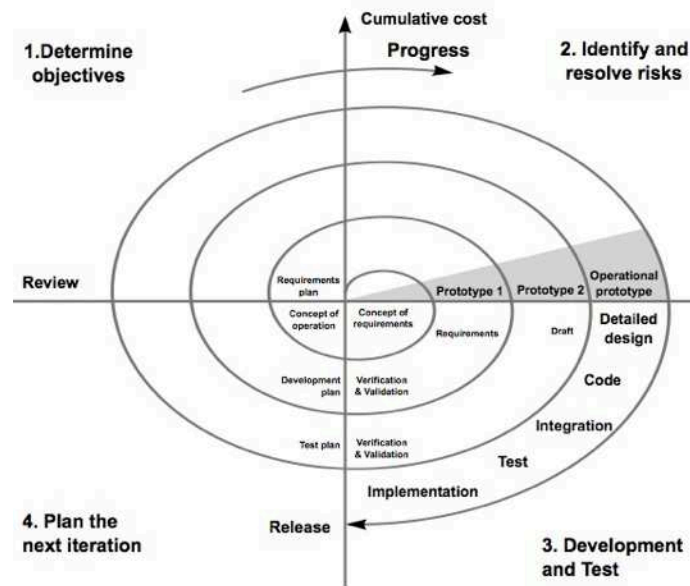
Fig.2.1 The spiral model

Even though this model addresses the issue of uncertainty or incompleteness of the initial requirements thanks to the involvement of the client and the incremental development cycle, it is not possible to use as it is. A world model is designed when all the hypotheses (and modelling issues) cannot be easily evaluated according to classic methods. Moreover, such hypotheses are not even expressible at first due to lack of data (R3). In addition, the model clade structure required to design a world model is not supported by this spiral cycle.

- ***Other classic models***

Many other software development cycles exist such as the V-cycle (links each development phase with a testing one), iterative-and-incremental cycle (programming and design of subset are conducted iteratively) and Rapid Application Development (RAD, where planning is interleaved with programming) **[Mcc96]**. All of them are more or less related to the ones presented previously and propose marginal improvements (e.g. emphasis on tests, focus on prototyping, etc) with respects to the methodological requirements.Finally, too little concern is given to the client (respectively the domain expert in my context) which is a strong requirement for the desired method as expressed in (R1).

- ***Agile methods***

Even though Agile methods are not intensively used in the agent-oriented modelling community, they have many sound and applicable ideas to integrate in a modelling method **[Gau11,Cer05]** for world models. As shown in Figure 2.2 agile development is defined by a deeply incremental development process. Each iteration focuses on a specific functionality (or a limited set) to implement and end by a deliverable that can be evaluated by the client. In an agile development, clients are generally largely involved and every users of the software are accounted for, not only the buyer or the client manager.
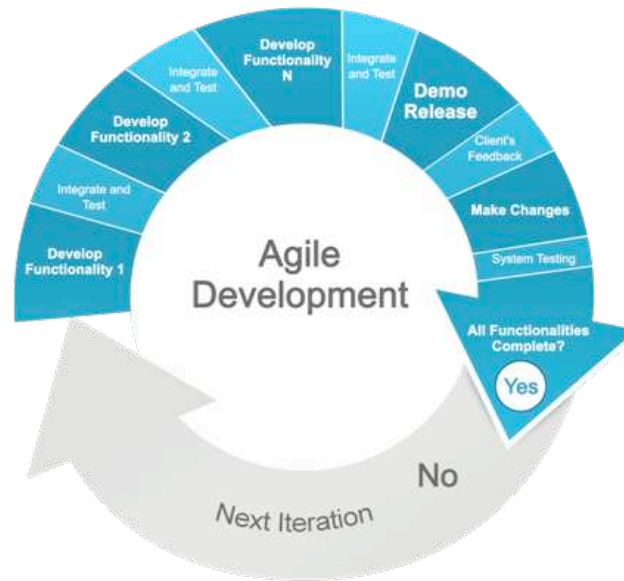
Fig. 2.2 The general agile development cycle**[Wiki08]**

Even though roots of the Agile methods date back to **[Edm74]**, a much more comprehensive and short definition has been given as the *Agile manifesto* **[Bec01]** which is as follow entirely:

> *We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:*
>
> - *Individuals and interactions over processes and tools*
> - *Working software over comprehensive documentation*
> - *Customer collaboration over contract negotiation*
> - *Responding to change over following a plan*
>
> *That is, while there is value in the items on the right, we value the items on the left more.*

All these points can be matched to the methodological requirements except the lesser consideration given to the documentation (which is important when having an incremental design of the model (R5)). Specifically, the focus on deliverable is related to the incremental design of the model. A deliverable focus on one (or a limited set of) functionality to add to the existing software whereas the evaluation of hypothesis or a modelling choice can be the focus of one modelling cycle iteration. I come back in Section 2.4 on the specific elements of the agile methods that can be integrated in a method to design world models.

2.2.2.Agent-Oriented Software Engineering (AOSE) methods

The previous methods have been designed to ease the development of a software in general. Although, they have been conceived with functional or object-oriented programming language in mind. In the current section, I present two well-known AOSE methods: GAIA **[Woo00]** and TROPOS **[Bre02, Gio04]** though many similar methods exist (Prometheus **[Pad01]**, Ingenias **[Pav05]**, Adelfe **[Ber03]**). They have the advantage to be specifically tailored for the design of agent-based model while remaining general in terms of application domain. Unfortunately, they totally overlook the implementation issue.

- *GAIA*

Gaia **[Woo00]** is a general (not domain oriented) method devoted to design and analyse multi-agent system based on more traditional software engineering methods. As illustrated in figure 2.3 The method proceed in three main steps: requirements statement, analysis and design which are implicitly followed by a "*direct implementation*". The elicitation of the requirements is accounted for but not covered by the method though it is a crucial step, especially for world models.
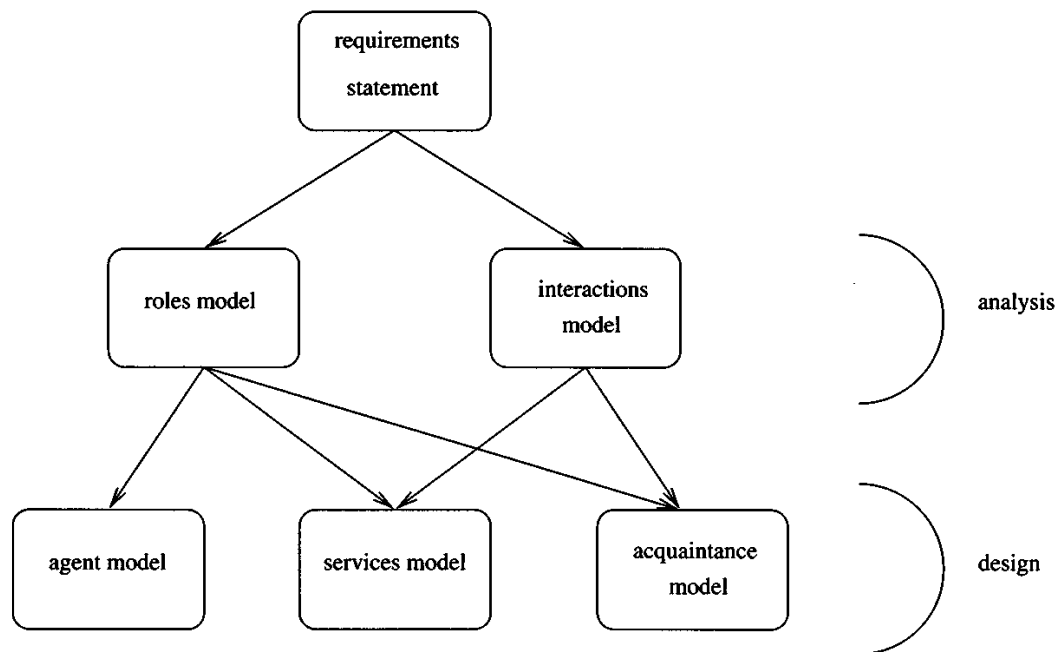
Fig. 2.3 Relationships between Gaia's models **[Woo00]**

The structure of the system is defined during the analysis step: Roles and interactions of the entities are listed. A role is the description of an entity's functions and is express by four elements: responsibilities (the tasks of the entity), permissions (allowed resources and interactions with the system), activities (internal actions or process) and protocols (allowed interactions).The interaction models define which roles interact together and through which protocol. A protocol consists of: Purpose (brief textual description), initiator (role responsible for starting the interaction), responder (role with which the initiator interacts with), inputs (information used by the initiator), outputs (information supplied by the protocol), processing (process that the initiator performs during the interaction).

The design step is decomposed in three models: agent model, services model, acquaintance model. The first model is where the agent are defined along their multiplicity in the system. An agent is coherent collection of roles (*e.g. a secretary will have the role to answer phone and manage a c.e.o calendar but not to decide when the board of director meeting should occur*). The relationships between agents, based on the roles they perform and the interactions model, are not represented in this model as it is the purpose of the acquaintances model. The services model list all the services of the agents. A service is coherent block of activity of an agent and is defined by: Inputs, outputs, pre-conditions and post-conditions. Each role has at least one service.

This method facilitates the development of an agent-based system and prepares its implementation. The incremental model refinement cycle partially addresses the requirement (R2). More precisely, this method presents the progressive refinement from a very abstract model to a model "*at a sufficiently low level of abstraction that they can be easily implemented*". Unfortunately, it does not provide any implementation advice or practice which is one of the four aspects of an ABM modelling method (cf Section 1.3.2). Even though GAIA structure clearly eases the expression of the requirements, it is a rigid structure: The modelling process does not account for the consequences of any change (R3). Finally, it provides a analysis step of the produced conceptual model but no evaluation of the model (R4).

- *Tropos*

Tropos **[Bre02, Gio04]** is another generic AOSE method. It puts the emphasis on the early phases of requirements analysis . This idea is reminded in the two main features claimed by TROPOS: (1) "*the notion of agent and related mentalistic notions are used in all software development phases, from the early requirements analysis down to the actual implementation*". (2) "*the method emphasizes early requirements analysis, the phase that precedes the prescriptive requirements specification*".

In order to attain such objective and facilitate the modelling process, TROPOS supports four steps of the (software) development process: (1) *early requirements analysis*, (2) *late requirements*, (3) *architectural design*, (4) *detailed design*.

(1) Early requirements analysis' objective is to list the "stakeholders of the system-to-be", their intentions and the inter-relationship between stakeholders to achieve their intentions. Stakeholders are represented as social actors which have goals and who provides resources and can perform tasks enabling fulfilment of the goals. The The *i** language **[Yu95]** is used to represent the dependencies between goals. A network of actor's relationship is then constructed from these goal dependencies. This phase leads to the definition of functional and non-functional requirements.

(2) Late requirements focuses on representing the system-to-be, its functions, its qualities and its environment. The system is considered as actor with goals and dependencies with other actors.

(3) Architectural design's objective is to design the system architecture in terms of subsystems (actors) and connection through and control flows. It proceeds in three sub-steps. First, the overall architecture is defined using an extended actor diagram (actors can be decomposed during this step). Second, actors required "capabilities" (action the actor can perform) are listed using the extended actor diagram. Third, this step defines a set of agents and assign them capabilities they have to provide (to actors).

(4) Detailed designs helps to define the full specifications of the agents (goals, beliefs, capabilities and communication). Three UML-based **[Ben05]** diagrams are proposed to fulfil this objective:

> (1) C*apability diagram* is based on the UML activity diagram.
>
> (2) P*lan diagram* details each node of a capability diagram.
>
> (3) *Agent interaction diagram* is based on the AUML **[Bau01]** sequence diagram to represent asynchronous communication.

This method provides a clear modelling cycle and tools (i.e UML) to facilitate such process which partially address requirement (R2). The concern given to list all the system-to-be stakeholders can be seen as a first step to integrate domain experts in the modelling process but is too limited to be considered to address requirement (R1) even partially. Defining entities after their goals and their interactions through interaction protocol is a major defect as it specialises the method for representing reference system where such goals and interactions can be formally expressed (for instance human societies).

- *Conclusion*

In summary, none of these methods do not fit for the methodological requirements. At best they acknowledge the existence of system stakeholders whereas the first method requirement (R1) expresses the need for domain experts to be the contributors at the centre of the modelling process. These methods provide a clear modelling cycle but only *Agile* methods with their prototype orientation partially enforce the incremental and modular model design requirement (R2). All the methods consider that all the required information are available at the beginning of the development process which is in contradiction with (R3). Few of these methods propose a way to evaluate the advancement of the model (deliverables of the spiral model and agile methods, for instance) (R4). Nevertheless, they provide some development or modelling advices and list several good practices, especially the agile methods.

2.3.Complex system oriented modelling methods

The GAIA method, presented in the previous section, introduces the incremental design of the model through a sequence of more and more detailed models. We see in the current section that this characteristic is shared by all agent-based modelling and simulation methods.

As the agent meta-model the world model requirements, the agent-based oriented methods should prove to be adapted even though they were formulated without the exploratory or explanatory objectives in mind. Nevertheless, they should at least provide a general frame for the modelling process.

2.3.1.Introduction: Fishwick and Gilbert & Troitzsch

Fishwick **[Fis95]** formalised computer simulation development (which includes ABM simulation) as an incremental process of three iterative steps: *model design*, *model execution* and *execution analysis*. Domain experts' knowledge and data from the reference system are gathered in order to create the model with the support of a specific formal semantics. This model is then transformed into algorithms executed on a computer and producing outputs. These outputs are analysed and confronted with observation from the reference system.

As pointed out in **[Gil99]**, model implementation is not a trivial work and thus the authors proposed to add a new step after the model design one: *model building*. Unfortunately, according to the authors the problem is just pushed forward as there is still no effective procedure to translate the model into an algorithm.

These two methods inspired many methods. For instance, the method presented in **[Aum07]** which is targeted at complex system representation and study through simulation models. Another example is the method proposed in **[Dro02]** which proposes a modelling method for agent-based models. These two methods are presented in the two following Section (2.3.2 and 2.3.3).

2.3.2. Auman's method

In **[Aum07]**, the author claims that the proposed method can be used in any kind of ecological understanding objective. This method focus on the definition of a "*focal system*", the core of the reference system which needs investigation, and the "*focal -1*" and "*focal +1*" systems to which its related. To link them, the author rely on the DEVS formalism **[Zei89, Zei00]** (see Section 3.3.2).

This method rely on a modelling cycle decomposed in four steps: (1) *specification of the modelling objectives and knowledge synthesis*, (2) *model design*, (3) *model implementation* and (4) *model assessment*; which are decomposed as in Figure 2.4 and explained as follows.
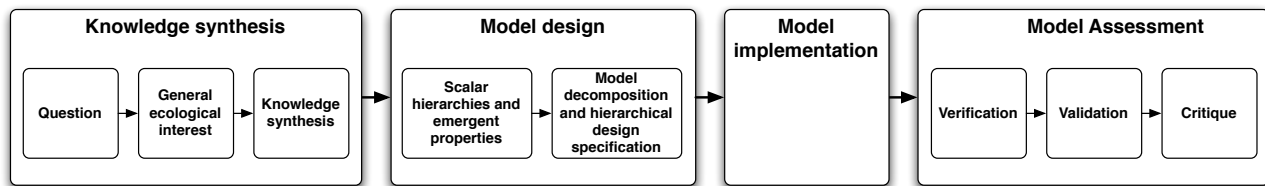
Fig.2.4 Aumann's modelling cycle

- **Modelling cycle**

In this method, specification of the models requires to define three elements. First, the question to be addressed through modelling has to be clearly defined, for instance: The reference system and the hypothesis to verify. Second, the general ecological interest of answering such question has to be assessed: Is there any more general ecological understanding interests. Third, how and why the model will allow to address the initial question. This modelling step, called *knowledge synthesis*, as for single objective to ensure that enough information are accessible to effectively design and implement the model.

This step acknowledges the importance of defining early and extensively the question addressed to the model and the elements on which rely the rest of the modelling process. A world model relies on numerous data sources but some data lack can appear during the modelling process, thus it is important to assess them as early as possible. This step by itself only establishes whether or not lack of data are present but does not resolve the issue of (R3).

The second step, *model design* has two main sub-steps: (a) *scalar hierarchies and emergent properties definition*, (b) *model decomposition and hierarchical design specification*. Sub-step (a) defines which entities of the system are considered and how they are arranged as in a containment structure (*e.g.* an estuary contains clams and crabs which contains their internal organs **[Aum07]**). Within this hierarchy (expected) emergent properties at the different level have to be defined. According to the author an "*emergent property is something that is influenced by the properties and relations characterizing the lower levels, but is not strictly predictable from, nor reducible to, these lower-level entities*". Sub-step (b) focuses on assigning the knowledge to the representation of the different systems (focal, focal-1, focal+1). The method states first the primary system representation has to be describe in terms of its

56

"*main phenomenological characteristics*" and the related experimental frames[5] **[Zei00, Uhr09]** have to be expressed. Then, similar work has to be conducted on the focal-1 and focal+1 system representations. Finally, processes and relationships linking the different level have to be defined and their *level* of dependency on lower level processes. The (adapted) scalar hierarchy example by **[Aum07]** is presented in Figure 2.5. This example focus on crabs, focal system, located in an estuary, focal+1. Crabs also interact with clams and the environment which are at the focal level.
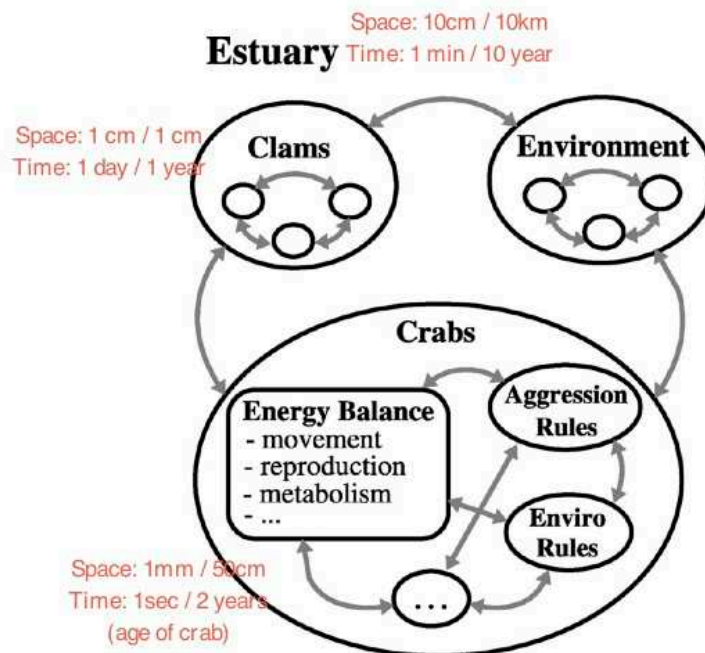
Fig.2.5 The three focal system of the **[Aum07]** example (adapted)

The *model design* step is certainly adapted to the modelling of natural reference systems when the knowledge is available from the start. Unfortunately, in a world model context, this is not the case and the design of the model has to be adapted to the progressive increase of knowledge. Moreover, this step cannot deal with the possible lack of data even if such data were envisioned. Finally, the definition of focal, focal-1 and focal+1 is certainly a good idea because it structures the model: The focal system is the centre of interest and observation, the representation of focal-1 system allows the representation of emergent phenomena and the focal+1 system enables to take into account the influence of the wider "environment" onto the focal system. Such idea may be integrated into a modelling method for world models. Nevertheless, it cannot be adopted as it is because of the evolution of domain experts' hypotheses (and available data): It can be decided during this design step to aggregate or desegregate some entities.

The third step, *model implementation*, is the process to translate the conceptual model obtained in the previous step into in an executable one. The author reminds it implies the creation of another morphism linking the conceptual model to the implementation. The author clearly acknowledges the importance of the implementation step. In particular, he reminds that the creation of another morphism is yet another source of possible discrepancies. He even provides in the next step an evaluation mechanism for this issue. Nevertheless, the implementation step is also considered to occur once and for all whereas a world model is very likely to know numerous implementation due to the evolution of the model and due also to the evaluation of different hypotheses all along the modelling process.

The final step, *model assessment* proceeds in three steps: (a) *verification*, (b) *validation* and (c) *critique*. Verification consists in demonstrating that the implementation is in accordance with the conceptual through checking each implementation entity against its conceptual counterpart. Concerning validation, the author proposes two way to ensure validity of the model: "replicative validity" and "structural validity". The former one is defined as given the experimental frames, all the possible experiments on the model are in accordance with

---

data from the reference system). The latter one implies that the model is "mimicking" the reference step-by-step and component-by-component. Critique of the model is a higher level process, its aim is to detect inadequacies in all the step of the modelling process. For instance, even if the model is verified and validated, which is difficult as it is an open-ended process **[Aum07]**, the early specification of the model may be inadequate making the whole process irrelevant.

- *Discussion*

The verification step may be adopted as-it-is for a world model as it simply ensures the adequacy between the implemented model (*i.e.* implementation) and the conceptual model but this is not the case for validity. It partially addresses (R4) and also increases domain expert confidence in the model which derives from (R1). Replicative validity is based on comparison of the model's output against the system's observation given. Due to the exploratory or explanatory objectives of my context, it is not required (not even desirable) to attain replicative validity. The structural validity might be more fitted as it requires each model component to "mimic" the reference system component which can be linked to the (W1) world model characteristic. Nevertheless, this validity is not either required. Finally, the critique demands that a overall assessment of the model and the modelling process. It is a important consideration but it is also very liable to subjectivity (in accordance with the author) and the author do not provide any advice except the "*continuous application of just and effective criticism across all stages of the modeling process.*" Such principle is to be kept in mind for a world model design method as it is linked to (R1).

The proposed modelling cycle seems to be generic enough to be adapted to the world model methodological requirements. Nevertheless, this method does not specifically address the method requirements (R4-7) and important adaptation should be made. Moreover, this method consists mostly of a modelling cycle (G2) but does not propose information and tools for (G1,3,4)

2.3.3. Drogoul et al's

In **[Dro02]**, the authors start by reviewing **[Fis95]** and **[Gil99]** and claim that a method should more clearly focus on defining what information is needed to build a model and whom provides them. They first define the roles of the "stakeholders of a model" (i.e. model contributors in my context). The *thematician* (i.e. domain experts) who provides both the question and the knowledge about the reference system. The *computer scientist* whose role is to implement the model as a program and the *modeller* whose role is to link the domain model to the design model which is used by computer scientists.

In this method the model is created iteratively through four models: (1) *domain*, (2) *design*, (3) *operational model* and (4) *implemented model*. The domain model defines the objectives of the model, the reference system (its constituents and their relationships) and can be informal. It is build after the domain expert's micro knowledge. According to the authors: "*micro-knowledge is 'local' knowledge about the 'individuals' without which the target system would not exist; it is composed of both observations (behaviors, etc.) and assumptions.*".The design model allows the modeller to elaborate a formal model from the previous one using the agent meta-model. The operational model is expressed in "implementable" formalism (such as UML) by the computer-scientist. The implementation constraints are accounted for in this model. Finally, the implemented model transformed the static models into an executable one expressed in a specific modelling language or in a general purpose programming one.

Figure 2.6 illustrates the creation of the model as previously presented, but also how it is verified and then used. It shows that the domain model description of the real system is based on micro-knowledge (of the thematicians) which is formalised into a design model. This model is then translated into an operational model much easier/directly implementable. It shows also that there are two feedbacks between this 3 models, whenever a incoherence or inexactitude is discovered in a given model, it might force to go back modifying the previous. When the operational model, the last one, is finished, the construction or implementation of the model is "simply" conducted. Then, the implementation is tested (verification) and the model is validated through simulation (internal and external validation). Finally, simulations can be conducted to generate results which are then analysed in comparison to macro-K.
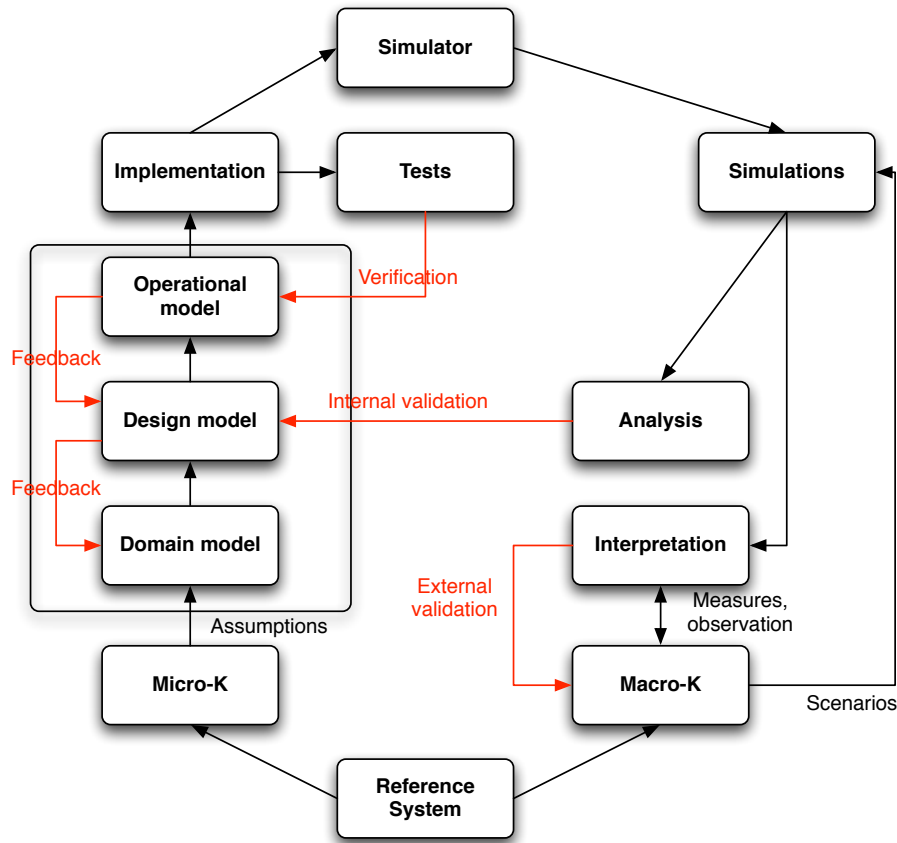
Fig.2.6 The modelling cycle proposed in **[Dro02]**

- *Discussion*

This method provides a general modelling cycle (G2) and defines the roles of the different model contributors (G1), acknowledging both the thematician (i.e. domain expert) and the modeller roles, and guide them with high level modelling steps. Nevertheless, the generality orientation of this method has for consequence that it remains at a high conceptual level. More practical advices could be profitable and no tools (conceptual nor technical) are proposed (G3-4).

Domain expert's micro-knowledge is used to define the model whereas the macro-knowledge is used to validated the model as it "*provide scenarios and measures for the simulation*". Both of these definitions acknowledge the importance of the domain experts or at least their knowledge. Nevertheless, some changes have to be made in order to place the domain expert at the centre of the modelling process (R1). The decomposition of the design cycle in several different model types eases the incremental design of the model (R2) while the feedback loops are a first way to cope with lack of data (R2). Finally, the authors define three types of model assessments but the model design evolution assessment is not considered (R4).

Despite classifying the different possible objectives of a thematician (predictive, speculative or theoretical), the authors does not take into account their effects on the modelling process. The thematician's objective directly defines the model's one and thus impact the design process to use. For this thesis I envision to present a design process adapted to world models, which corresponds to thematician's speculative and theoretical objective in Drogoul *et al*'s. terminology. Nevertheless, this method may be adapted and specialised to enable the design of world model as it addresses many of their methodological requirements.

2.3.4.Elicitation oriented methods

- *ComMod*

The previously presented methods focus on the design of model after the necessary knowledge has been made available. Unfortunately, one of the most important issue a world model modelling method has to face is the incompleteness of this knowledge and its scattering among different domains. Consequently, a step of

data collection or gathering should be proposed before the design cycle. This issue is at the origin of the "companion modelling" methods. Companion modelling means that the elicitation of the model is conducted in collaboration between reference system stakeholders and model contributors.

In this section, I review two "companion modelling" methods: ComMod and ARDI. The elicitation process is conducted in collaboration between stakeholders and a facilitating person (ARDI) or apparatus (ComMod).

ComMod **[Bou05]**, is a modelling method in which researchers (corresponding to more or less model contributors), policy-makers and stakeholders of the reference system collaborates to design and use the model. It has been developed with a very specific objective in mind: understanding and strengthening the collective decision-making process in a natural resource management. It proposes an iterative modelling cycle with a multi-agent system at its facilitating apparatus. The three steps of an iteration are **[Bou05]**:

1. *Field investigations and a literature search supply information and help to generate explicit hypotheses for modeling by raising a set of initial key questions to be examined by using the model.*

2. *Modeling, that is, the conversion of existing knowledge into a formal tool to be used as a simulator.*

3. *Simulations, conducted according to an experimental protocol, to challenge the former understanding of the system and to identify new key questions for new focused investigations in the field.*

Finally, simulations or RPG (role playing game), conducted on the CORMAS simulation platform, are used to enhance the cooperation of the different stakeholders of the reference system.

ComMod is interesting for my concern due to the strong involvement of stakeholders of the reference system (similar to R4) and for its incremental procedure of model design (R2). Nevertheless, it is a very oriented method towards enhancement of cooperation within the reference system. It has no concern of lack of data (R3) nor assessing the model design evolution (R4).

- **ARDI**

ARDI **[Eti11]** is a method which focuses on creating a conceptual model through collaboration of the different stakeholders of a real system. Similarly to ComMod, this method is oriented towards natural resource management. This method proceeds in four steps concretised by workshops where all the stakeholders of the reference system collaborate with a *facilitator* and an *observer*. The facilitator leads the process by guaranteeing clarity an general agreement (of the model under construction) from all stakeholders. The observer keeps track of stakeholders interaction and how the model is built. The four steps of the collaboration are as follows. (1) *Actors* of the system are listed, they have to be direct or indirect (influencing the system but not part of it, i.e. regulators) stakeholders of the system. (2) *Resources* of the system are listed and characterised. (3) *Dynamics* of the actors of the system are defined (human and natural dynamics are clearly distinguished). (4) *Interactions* between users and resources are defined using the answers of the three previous steps.

All these steps are achieved through a series of workshop close in time (authors advocates for less than a month) where the central form of the model is a diagram, Figure 2.7 presents such diagram at its final stage **[Eti11]** from an application on water management in a "Crocodile river". In this diagram, the white boxes represent stakeholders of the system and green boxes the resources. Boldfaced letter codes indicate processes, and blue text indicates actions. Words or concepts written in red denotes addition concepts made after their dedicated steps (e.g. a resource is added during the dynamics step).
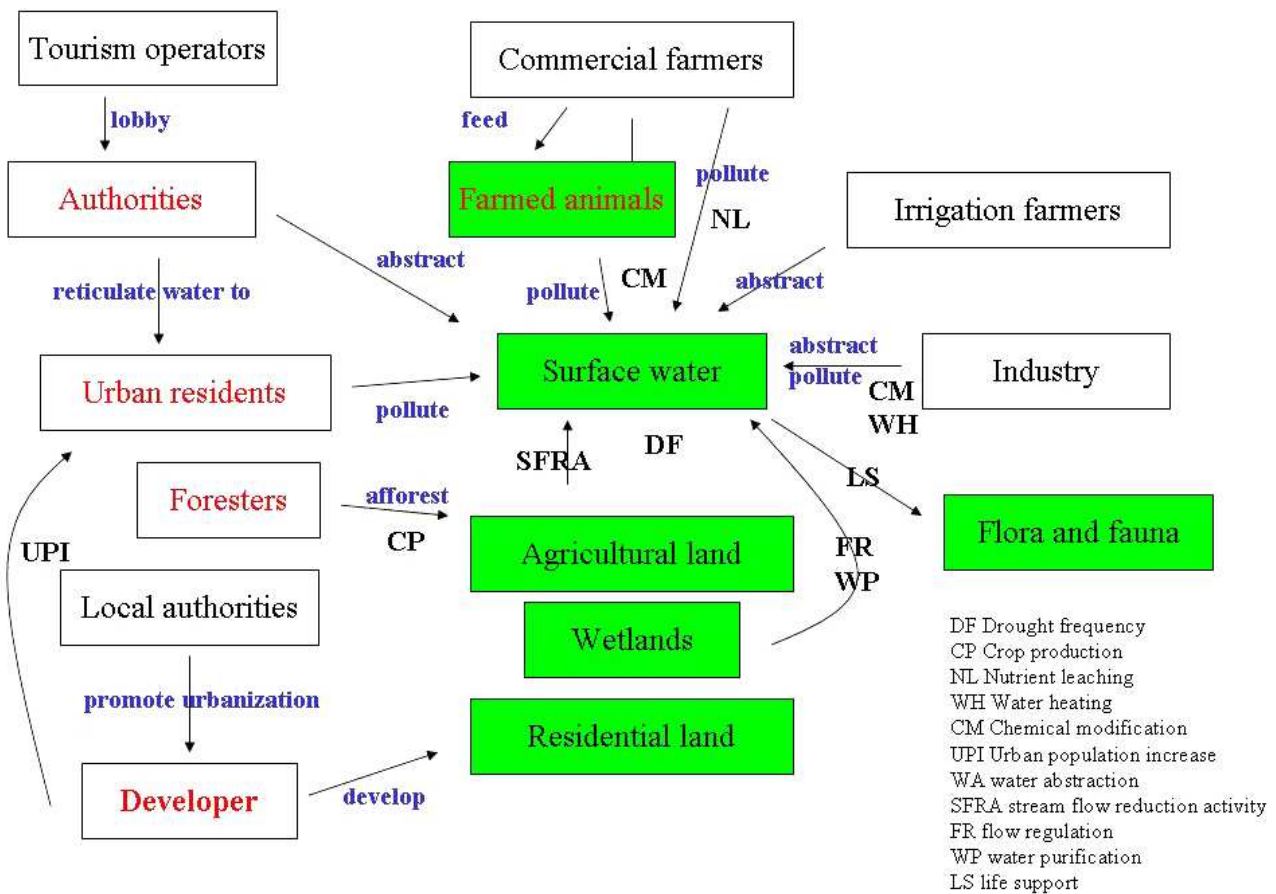
Fig2.7 ARDI generated model of the water management in the "Crocodile river" **[Eti11]**

This method address to some extent all the general method requirements (G1-4): It provides a list of roles (G1), a coherent modelling cycle (G2), guidelines (G3) and documentation tools (G4). Nevertheless, the list of roles is not coherent with the world model approach which considers domain expert as the source of knowledge, not the reference system stakeholders. The modelling cycle, guidelines and documentation tools are limited to the design of a conceptual model only (*e.g* no implementation).

Similarly to ComMod, this method have great concern of the involvement of system's stakeholders (similar to R4), provide a incremental sequence of model design based on shared representation (partly addressing R5) but is not concern by lack of data (R3) nor assessment of the model evolution (R4).

The way these methods propose to elicit a model through collaboration of the reference system stakeholders and manage the collaborative process may be re-used to design world models. Indeed, they propose a chronology where each system stakeholders of the system can amend the shared representation of the system. The facilitator ensuring the convergence of the process whereas the observer allow trackback of the model iterative construction. Finally, they propose a limited yet adapted set of conceptual diagrams.

2.4.Synthesis

2.4.1.Summary

In Section 1, I selected the agent meta-model to express world model as it allows creation of models with the fitted characteristics (W1-4). In this section I review software development and complex system oriented modelling methods. More precisely, the presented methods propose mainly modelling cycles and only **[Dro02, Bou05 Eti11]** propose a definition of the modelling contributors roles.

General software development methods (e.g. waterfall model, spiral, etc.) are not really fitted to design world model and cannot be adapted. Nevertheless, the Agile methods propose a few good practices than can be re-used. The client is much more implicated in the development than before and is included in most development steps. These methods also focus on small but frequent incrementation. It has the advantage to

clearly organise the development on specific functionalities and favour the recourse to frequent testing using "*unit test*". Finally, the pair programming **[Wil00]** which ensures code quality by coupling two programmers (an driver and an observer) may be adapted to increase domain experts confidence in the implemented model by reviewing its code with the support of a computer-scientist.

AOSE are specifically dedicated to the development of agent-based software and thus are more fitted to design world models as they rely on the agent meta-model. These methods introduce incremental refinement of model, for instance: (1) early requirements analysis, (2) late requirements, (3) architectural design, (4) detailed design as in **[Gio04]**. They also state that UML is satisfactory to express conceptual agent-based models. Unfortunately, they barely address any methodological requirements and their adaptation is not advisable.

The presented agent-based modelling and simulation methods **[Aum07, Dro02]** are definitively the most fitted methods of the lot as they are dedicated to design an agent-based model of natural (and human) reference system. Auman's method **[Aum07]** proposes a general modelling cycle that may be adapted and some good practices to adopt but still does not address most of the methodological requirements (R4-7). In particular, this method does not provide a definition of the roles of the various model contributors.

In contrary, Drogoul et al's method **[Dro02]** provides a definition of the roles of model contributors acknowledging the importance of domain experts importance. Its modelling cycle is also more fitted to the world model approach as it includes feedback loops. Nevertheless, the modelling cycle gives limited concerns of the creation of the initial domain model and the rollback during a modelling process are limited. The first issue is the specific interest of the two last presented methods: ComMod **[Bou05]** and ARDI **[Eti11]**. These methods provides a modelling cycle focused on the elicitation of the domain model, some conceptual tools and good practices.

2.4.2.Introduction to the proposed roles and modelling cycle

As this modelling review highlights, there is no existing modelling method sufficiently adapted to the world model approach. Consequently, I propose a new modelling method called KIMONO (*Keep It Modular, Open-ended, and Neatly Organised*) which addressed all the methodological requirements, eventually, and streamline the design of world model.

The contributor roles in KIMONO are similar to **[Dro02]**: domain expert, modeller, computer scientist. The role of domain expert is to provide knowledge about the reference systems and leads the modelling process. Modeller and computer scientist are only support roles. The first one provides support for design regarding conceptual models while the latter's area of expertise is the implemented model (development and use).

[Dro02] proposes a detailed yet generic enough modelling cycle. Nevertheless, this cycle do not address the lack of data issue nor the multiple sources of knowledge. Consequently, model elicitations steps as in **[Eti11]** are added while the whole domain, design and operational model is considered as a loop. Moreover, the data collection process is considered to be conducted in parallel of the the modelling one.

Figure 2.8 illustrates the general structure of the KIMONO modelling cycle. The elicitation of the domain model is preceded by an informal description of the reference system.It allows domain experts to progressively refine their description of the reference system and list all the needed knowledge from different domains. During these steps, domain experts progressively organise and formalise their micro-knowledge of the system. The definition of the domain model can be conducted similarly to the ARDI's elicitation of the conceptual model, more precisely: Domain experts create it collaboratively with support of the modeller. Afterward, a general data assessment is made in order to determine which data are available or missing (or potentially uncollectible). This assessment allows to parallelise the modelling process and the data collection one.

The second part of the modelling cycle is similar to the modelling cycle of **[Dro02]**, though it is an incremental step in my method and is interleaved with a data assessment step. Each iteration of this step focuses on the modelling of a limited subset of the reference system. It allows model contributors to focus on a specific aspect of the reference system, to evaluate a given hypothesis or a modelling alternative one at a time.
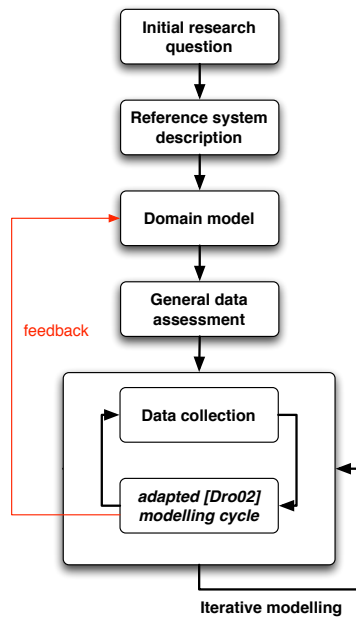
Fig.2.8 KIMONO general modelling cycle

Even if most of the modelling methods limit themselves to the definitions of contributor roles and modelling cycle, they do not constitute a complete method (G1-4). Tools and processes for the documentation should be provided. Moreover, simulation and agent-based model methods has to come with modelling guidelines. Such concerns are addressed in the following section.

3. Tools to design and implement world models

The selection of the agent meta-model to express world model has for consequence that two models have to be defined: a conceptual one and an implemented one. Actually, the modelling cycle of **[Dro02]**, on which is based the one proposed in Chapter 5, defines three conceptual models: domain, design and operational models. I present in Section 3.1 what are the fitted formalisms to express them. Actually, simply expressing the model is not enough because of requirements (R4-5). For this reason, I introduce (in Section 3.2) a model communication protocol, O.D.D. **[Grim10]**, that can be used for model and modelling process documentation.

An agent-based implemented model can be expressed in a general purpose programming language from "scratch" or within a simulation platform which facilitates the implementation by providing many primitives and possibly a dedicated modelling languages (see Section 1.3.2). The former does not fit into a world model design method as (R1) impose that the domain expert is capable to intervene in any step of the modelling cycles. Consequently, Section 3.3 is dedicated to the review of modelling platforms. The selected modelling platform is, then, describe with more details in Section 3.4.

3.1. Expressing conceptual models

3.1.1. Introduction

The expression of the conceptual agent-based models usually relies on UML (Tropos for instance, Section 2.2.2). This is justified by the fact that ABMs can be defined as an improvement of the object paradigm and UML is adapted for conceptual description of an object-based software. In contrary, the model and the modelling process documentation is scarcely addressed in the literature though a specific "communication protocol" called O.D.D **[Grim06, Grim10]** fulfils the documentation need.

The **[Dro02]** method on which is based my method relies on three conceptual models: Domain, design and operational models. According to (R1) the main requirement for the formalism to express these conceptual models is its accessibility to domain experts. Thus, the formalism must be accessible to the expert in addition to be adapted to the agent meta-model.

3.1.2. Domain model

The domain model is the first extensive but generally informal description of the reference system. It has to be understood by both domain experts and modellers. The domain expert is responsible to create it thus it is likely that this model is express in her/his specific formalism. Nevertheless, the final domain model should be written in a formalism accessible to the modellers in order to allow them to create the design model. The ARDI method **[Eti11]** proposes four ontologies of progressive complexity: Actors, resources, dynamics and interactions (see Section 2.3.4).

The observer role who documents the collaborative elicitation of the different ontologies, introduced by this method points out the need for a documentation of this first model. I present in Section 3.2 the O.D.D. protocol which can be adapted for such purposes and can be used transversally for all model iterations. In addition, if the domain expert require more details, a model (or a part of the model) description can be formulated in the domain formalism and integrated in the documentation. Such document would provide details whenever an incertitude is occurring in the later stages of the elicitation of subsequent models.

3.1.3. Design model

According to **[Dro02]**, the design model is the first formal model and is the duty of the modeller. According to requirement (R1), the domain expert also intervenes in the conception of design model though the modeller support is stronger than for the domain model. In order to obtain a formal yet accessible model, it can be expressed in the unified modelling language (UML) **[Ben05]**. More precisely the class diagram is generally sufficient. This diagram represents the structure of the whole model: all the entities and their relationships. Figure 2.9 provides an example of an UML class diagram from **[Lap09]**. There are two main types of agents: *Rat* and *Flea*, interacting with each other. The *Turtle* and *Patch* are class defined by the Netlogo platform and represent generic agent and environmental cell, providing basic primitives.
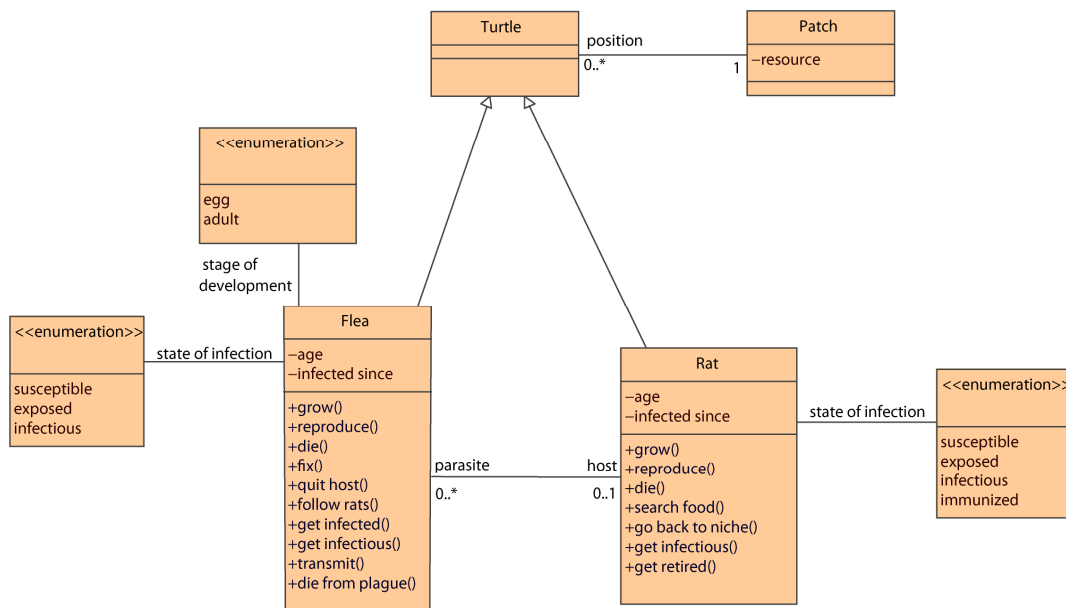
Fig.2.9 Class diagram of the SimPest model **[lap09]**

3.1.4. Operational model

The operational model objective is to prepare the implementation of the model by taking into account the constraints of the implementation formalism. It allows the domain expert (and the modeller) to provide a feedback to the implementer concerning the implementation.

Similarly to the design model, the operational one can be expressed in UML. Nevertheless, it should be more detailed with respect to implementation constraints. Thus, the UML class diagram should be complemented with the object, the activity and the state transition diagrams. The object diagram can be used to represent specific interactions (in addition to the class diagram). The activity diagram can be used to describe precisely the schedule of action and interactions of the implemented elements. The state diagram provides a complete

view of the "lifecycle" of an agent. An illustration of this diagram, from Simpest **[Lap09]**, is presented in Figure 2.10: It represents the life evolution of flea start as an egg which grows for a given period of time. Then it becomes a susceptible adult that can be exposed and infected through its host. Finally, it dies of age or of the disease. In addition, the adult flea can attach itself to an host (rat) if they have the same location and becomes free only when the host is dead.
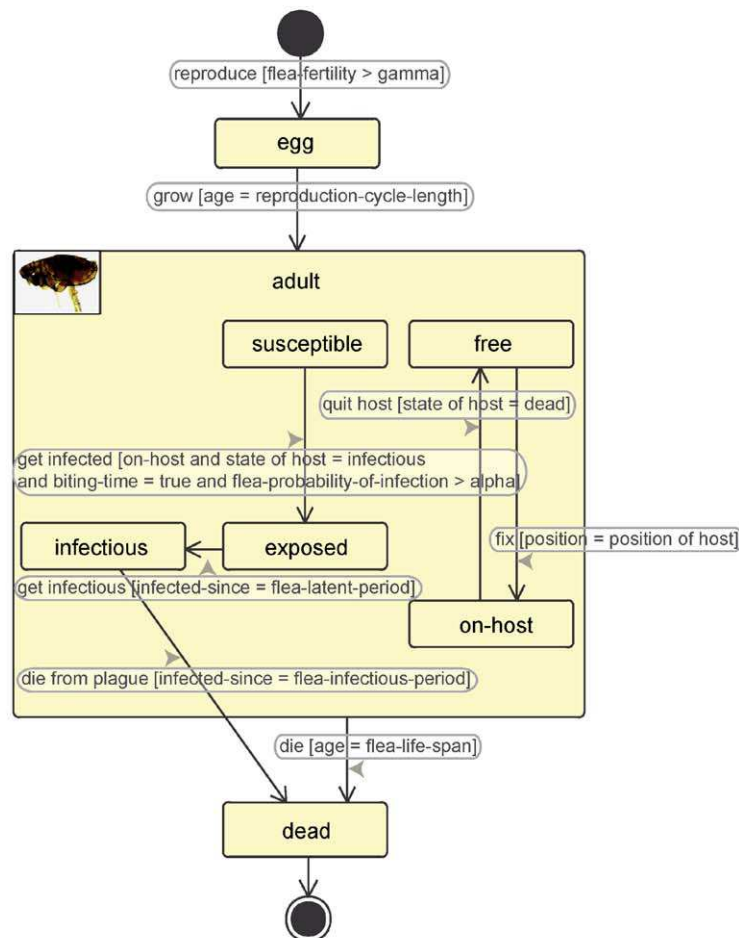
Fig.2.10 State-transition diagram of a flea (SimPest model **[Lap09]**)

3.2. Documentation

According to definition 2.1, a world model is the result of a succession of models organised as a clade. This multiplicity of intermediate models is a consequence of the exploratory or explanatory objective. In order to facilitate the modelling process and especially the work of the domain experts (R1), it is necessary to provide a documentation of these models and the overall modelling process. A basic tree graph can be used to keep track of the various models and their chronological relations.. In addition, a description of each model evolution has to be made available in order to justify the modelling and hypotheses choices and more generally the global model evolution. It is possible to use the O.D.D. **[Grim06, Grim10]** protocol to do so.

3.2.1. O.D.D. description

O.D.D. is a textual template designed "*for describing individual-based and agent-based models*" **[Grim06]**, ease its reading through standardisation and thus facilitate analyse, comparison and replication of the model (to some extent). It has been proposed and applied firstly in the ecological modelling community where it is largely used but new domains such as Sociology starts to use it **[Pol08]**. This textual template is organised in three main section: Overview, Design Concepts and Details. These sections and their sub-section are listed in Table 2.1: the light grey sub-sections have been added or modified in the updated version of the template **[Grim10]**.

Categories	Elements
Overview	Purpose
	Entities, state variables, and scales
	Process overview and scheduling
Design concepts	Basic principles
	Emergence
	Adaptation
	Objectives
	Learning
	Prediction
	Sensing
	Interaction
	Stochasticity
	Collectives
	Observation
Details	Initialisation
	Input data
	Sub-models

Tab.2.1 The O.D.D. protocol structure **[Grim10]**

The section *Overview* is decomposed of three subsections: (1) *Purpose*; (2) *Entities, state variables and scales*; and (3) *Process overview and scheduling*. It is a detailed introduction to the model, which should give the reader enough information on the research question and the reference system of the model. A program skeleton with the main entities of the model should be derivable from the information contained in this section which correspond to an extended textual description of the domain model.

The section *Design concept*s consists of eleven subsections: Basic principles Emergence, Adaptation, Objectives , Learning, Prediction, Sensing, Interaction, Stochasticity, Collectives, Observation. This section allows the modeller "to clearly communicate important design aspects of the models" **[Rai10]**. Completing all the eleven sub-sections are not mandatory as usually one or several of them would be irrelevant depending on the model. Counterintuitively and against OOP/usual software engineering practices these subsections scatter the description of the individuals' behaviour (and the environment). Nevertheless, the comprehension remains and the processes where different individuals interacts inextricably (and emergent phenomena) are, actually, even better described this way.

- Basic principles: presents the theory and hypotheses on which is based the design of the model. It also includes the objective of the model (prediction, exploration, explanation)

- Emergence: results that are unpredictable (to a limited extent) and more predictable observed phenomena given the model design

- Adaptation: describes the adaptive behaviours of the agent wether it be driven by fitness improvement or reproduction of observed phenomena

- Objectives: if adaptation is present what criteria drives it and how agents proceed to adapt

- Learning: does the adaptive traits evolve over the simulated time

- Prediction: presents if any planning or evolution of the surrounding is conducted by the agents

- Sensing: describes which environmental and other agents' variable are accessible to the agent

- Interaction: presents the agents (and environment) interactions and how they affect them

- Stochasticity: describes the random process occurring in the model, what they should represent and how it is implemented

- Collectives: describe imposed or emerging group of individuals and their potential behaviour, internal state and how they affect the member and other agents

- Observation: present the outputs of the model and their parameters (frequency, accuracy, etc.)

Details presents the model in details through three subsections: Initialisation, Input, Sub-models. This section may be extensive so it can be included in an online appendix but has to be thoroughly completed, especially in the case one wants to allow for replication. Initialisation presents the initial state of the system at the beginning of a simulation while input is about the data used to initialise the simulation and the time series used during the simulation. The sub-models section makes the protocol recursive as each sub-model can be considered as an independent O.D.D description of another self-contained model.

3.2.2.Conclusion

O.D.D. includes the description of fundamental and specific characteristics of complex system models, which include world models, such as scales, environment, accounting of both individual and collectives behaviours, etc. **[Amo10b]**. This is a major advantage over non-dedicated software engineering tools, like, for instance, UML. Moreover, O.D.D. seems appropriate to document models of virtually any domains even though it has been conceived within the Ecology field **[Pol04, Grim06]**.

The representation of both environments and entities include qualitative aspects that would be difficult to mention in traditional software engineering languages (for example: quality of data, lack of data, etc.). Consequently, O.D.D. is a good trade-off between domain issues (data, measure, initial application problematic, etc.) and the modelling activity (with entities, hypotheses, simplification, experimentations).

Nevertheless, this protocol has several flaws as I describe in **[Amo10b]**. For instance, O.D.D. has no section dedicated to the evaluation of used data and the parameterisation of the model is awkwardly relegated to the sub-models section which lacks a way to link together the sub-models. Hopefully, some of these concerns have already been addressed and proposal of improvements are welcomed **[Grim10]**.

Some major defects remain, in particular O.D.D. is inherently ambiguous **[Pol08]** as it is a textual template. Numerous examples have been proposed to clarify its concepts (see annex of **[Grim06]**). Examples have never replaced specifications, and the lack of a formalism to express the model (rather than a list of items) is clearly a drawback. The use of UML to express conceptual models address such issue. In addition, O.D.D. does not considered sufficiently the implementation issues. Even the combination of this protocol and the previously presented methodological concepts (Section 2) does not allow the involvement of domain experts in the implementation phase. Such requirements is then reported onto the agent-based simulation platform.

3.3.Implementation tools

In Section 1.3.1, I present the agent meta-model as the most fitted one for expression of the world models. In Section 2, the design of conceptual agent-based model is addressed while their expression and documentation is assessed in Section 3.2. Nevertheless, the implementation of these models have not been considered yet. Consequently, the present section addresses such issue by reviewing agent-based simulation platforms. Section 3.3.1 presents the requirements such platforms have to address. The review is conducted in Section 3.3.2 and the selected platform is extensively presented in Section 3.3.3.

Numerous agent-based simulation platforms have been proposed since a few decades, Figure 2.11 presents the lineage of important ones. I may classify the lineage in three components: all-around platforms (Swarm, Repast, Mason and GAMA), education oriented (Starlogo, its derivatives and Netlogo), and collaboration oriented (Cormas and Mimosa). This figure also illustrate the implementation formalism of these platforms: general purpose programming language (orange), dedicated modelling language (yellow) and graphical modelling (green).
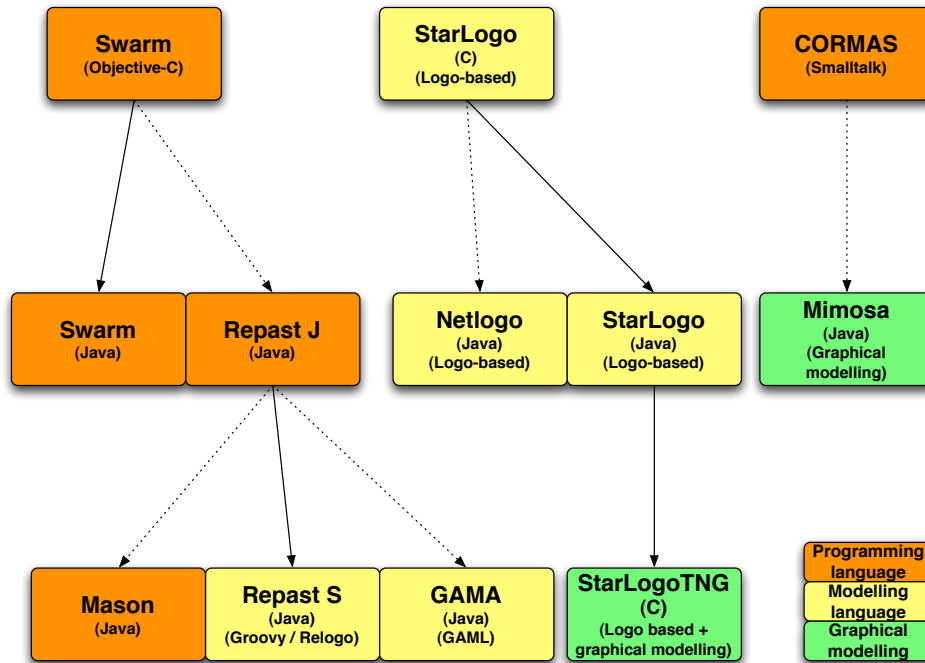
Fig.2.11 Genealogy of the considered ABMS platforms (adapted from [Cas06])

Before, reviewing these platforms I must derive the implementation requirements from the world model characteristics (W1-4) and methodological requirements (R4-7).

All the general purpose ABM simulation platforms allow to represent both the individuals and the environment of the model although *world models* require these descriptions to be thorough. The same level of representational complexity is demanded for the environment and the individuals.

3.3.1. Requirements

An adequate simulation platform has to support the method in addressing the methodological requirements (R4-7) and allow implementation of the conceptual model to fit in the world model characteristics (W1-4). In practice, this leads to the definition of several technical requirements as follows.

(P1) The agent meta-model allows to design conceptual models fitting the world model characteristics (W1-4) in particular. The selected simulation platform has to allow the same characteristics to the implemented (world) model. Agent-based model simulation platforms always allow explicit representation of the reference system constituents (W1) in terms of internal state and dynamics (W2) but often overlook the **environment**, limiting the complexity of its representation (W3). The use of a GIS seems to be the best option to date, thus, the platform must integrate GIS functionalities or allows seamless access to a GIS.

(P2) The simulation platform has to facilitate the **manipulation** of the implemented by domain experts (W4). In practice, the simulation platform has to come with an accessible modelling language. Graphical modelling tools are too limited [Gau11] to be considered. In contrary general purpose programming language have wide capabilities but forbid the involvement of the domain expert.

(P3) The selected simulation platform must be a **generic** one (not dedicated to one thematic) because phenomena from distant thematics are likely to be included in the model (for instance, sociology of poultry trading and virus pathology).

(P4) Having a platform offering a large of **primitives set** is not only a way to speed up the development of the operational model but also it allows early evaluations of the conceptual model by developing prototype implementations. In addition, some mechanisms to handle, use and analyse the simulation should be provided (visualisation of the simulation, graphical and file outputs). Nevertheless, the built-in primitives of the platform are unlikely to suffice so **extensions** should be allowed.

(P5) Representing the reference system constituents and allowing their disaggregation if required by domain experts can lead to **intensive computational requirements**. In practice, the selected simulation should be able to cope with large number of agents (in comparison to other simulation platforms).

It is very difficult for a platform to address all these requirements. Actually, most of the platforms are oriented whether on accessibility (often education oriented) whether on generality of the representation and execution efficiency. Accessible platforms the ones that do not require an extensive knowledge of programming, Netlogo being the most well-known one. I call "powerful platforms" those that focus on providing the most opened perspective in terms of representation capabilities (e.g. Repast) and which also usually focus on performance (Mason in particular) and/or "*extensibility*" of technical tool (GIS, scheduler, primitives). Unfortunately, such objectives usually result in an important complexity of the platform and limit its use and access to computer scientists only. In this case an ersatz of solution would be to use different platforms: a first one for the design of the final model in collaboration with the thematicians and then the final model would be implemented on the "powerful" one. Nevertheless, the situation is evolving and platforms, such as RepastS and GAMA, are trying to remain accessible while providing the widest representation capabilities possible.

In the following section I review major simulation platforms by presenting their advantages and defects according to these requirements before presenting in much more details the platform I used, GAMA.

3.3.2. Review of the ABMS simulation platforms

Numerous ABM simulation platforms have been proposed **[Rai06]**. Moreover, numerous reviews have been published that focus wether on the application domain **[Bery08]** or technical specificities **[All09]** for instance. The present review is defined after the requirements to implement world models (P1-5) thus the platform have to be generic and no specific technical advantage are reviewed.

- *Swarm*

Swarm **[Min96]** is a well-established simulation platform. It is coded in Objective-C and the model have to be implemented in this language as well which forbids the platform to address requirement (P2). The environment in Swarm is considered similarly to an agent, which is in accordance with (P1), but no built-in primitives are offered (for both the agents and the environment) (P4). Concerning the environment, a library called Kenge **[Box02]** allows to load layers of geographic vector data and some ad hoc access to geographical data has been developed for specific models (e.g. **[Hak01]**) (P1). Unfortunately, they do not provide any spatial primitives neither the possibility to store the resulted environments.

- *Cormas*

CORMAS **[Bou98]** is a platform initially dedicated to natural resources management (Section 2.3.4) issues (P3). It is developed in Smalltalk with the VisualWorks IDE (integrated development environment). This programming language is also used as the model implementation language which hinders the domain expert access to the implemented model despite the presence of a graphical user interface (GUI) easing the implementation process (P2). In addition, the representation of the environment is not as powerful as one would need for a world model. In practice, there are 2 environmental modes: a vector and a raster mode. Both modes only provide basic services such as localisation, neighbourhood, basic movement and perception.

A GIS support has been added recently **[Urb08]** but it is only a loose link **[Cas06]** (a data connection only) between the simulation platform and the ArcGIS software (P1).

Even though Cormas is still used and improved, the development of a new platform have been initiated. This platform, called Mimosa **[Mul05]**, goes further as it guides the modelling process and facilitates the translation of an early conceptual model to the implementation which would be very interesting to address requirements (P2) and the methodological one (R1). Unfortunately, the platform is still under development and support of the implementation phase is almost inexistent with released version at the present day.

- *Netlogo*

Netlogo **[Wil99]** is also a well-established simulation platform, largely used for educational purpose but also for research. Its strong advantage is the ease of access of the modelling language despite its object-oriented lineage (P2). The environment representations is originally grid-like, made of "patches", but the latest version of Netlogo introduced a GIS extension. Access to GIS is supported but very limited geometrical operation are available (P1). Unfortunately, Netlogo is not open source and other services that may be required by a model can not be added easily to the platform (P4).

Netlogo is not the only Logo language based simulation platform. I can mention StarLogo **[Res96]** and its derivative Starlogo TNG **[Klo05]** which are even more education-oriented (*i.e.* more accessible). StarLogo TNG has the advantage to propose a graphical user interface which allows graphical programming to a limited extent **[Gau11]**.

- *RepastJ and Simphony*

RepastJ **[Nor06]** is a modelling toolkit inspired by Swarm. It includes only basic services. The representation of the environment is whether based on a grid or a continuous environment and a link to a GIS library (OpenMap) offers basic GIS services: Importing/exporting of shapefiles and raster data, geometrical operations, access to data attributes, etc (P1). More complex geographical and most of agent-related primitives have to be coded by the modeller as very few built-in primitives are present (P4). There is no modelling language as implementation of the model has to be conducted with the language in which the platform is coded: Java thus this platform does not address (P2).

RepastJ has been replaced recently by Repast Simphony **[Nor07]**. This version greatly improves many technical aspects of RepastJ. For instance, a new and more powerful GIS library, Geotools, is used and a new modelling language, "Relogo" has been added. Nevertheless, there are still some technical lacks (e.g. no synchronisation between several environment). In addition, the modelling language is based on a general purpose script language called Groovy[6] which is not particularly accessible to domain experts. Thus, this improved platform does not yet address (P1) nor (P2).

Mason **[Luk05]** is another interesting platform which can be considered similar to Repast as it is a framework and not exactly a platform. This platform is performance oriented (P5) and thus the accessibility was not considered as a strong requirements and the built-in primitives are at best limited (P1, 2, 4).

- *VLE*

VLE **[Que09]**, or Virtual Laboratory Experiment, is a framework oriented towards multi-modelling with support of cellular automata, Petri net **[Mur89b]**, ODE, etc., but extensible to other formalisms, ABM for instance, through its API. One of the founding idea of VLE is to give an implementation of the DEVS formalism **[Zei00]**. This formalism allows to couple formally any type of discrete-time model and even continuous-time ones can be approximated. The Figure 2.12 shows on the example of the DEVS formalism: an agent coupled to a set of four cells organised as a cellular automata. Each sub-models or entities of the global model are linked to each other by input/output connections.

---

6 http://groovy.codehaus.org/
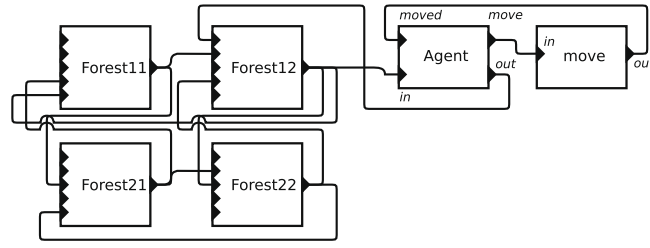
Fig.2.12 A DEVS multi-model schema example**[Que09]**

An advantage of the VLE simulator is its openness to other softwares. For example, it is possible to interface VLE with R, and to other modelling formalisms as long as available resources permit it. Interfacing with domain experts' software would ease their analyse of the simulations which is part of (P2). Unfortunately, no agent-based model is offered and thus existing agent-based models have been implemented using C++ which is not well-known to be "non-programmer friendly" (P2). In addition, there is no support of GIS (P1).

It would be of great interest to consider the DEVS model coupling approach on which this platform relies. Nevertheless, it is out of the scope of this thesis and I selected the agent meta-model as single one to use thanks to its integrative characteristic.

- ***GAMA***

GAMA **[Amo09,Tai10a]** is an open source ABMS platform programmed in Java which is developed by the MSI research team since 2007. It re-uses some of the libraries provided by Repast and rely on Eclipse libraries for the GUI. It is a non-specific domain simulation platform with a GIS support and offers a modelling language called GAML.

The environment representation can be discrete or continuous thanks to the GIS support (P1). The model implementation is made accessible to domain expert thanks to the modelling language GAML, *GAMA Modelling Language*, (P2). This platform has been conceived to allow the implementation of model where the representation of the environment is important which, in my context at least, do no limit its genericity (P3). The modelling language provides a large set of primitives: Movement (on discrete and continuous environments), interaction, behaviour management, etc (P4). Practical test showed that the platform can sustain hundreds of thousands of agent when executed on a desktop computer (P5). Finally, multi-scale capabilities of the platforms are under development **[Tai10a]** which would ease structuring of the model and may ease the domain access(P2)

All the simulation platforms address most of the requirements (to different extends) allowing the implementation of world models. Nevertheless, only GAMA addresses the whole set of requirements as shown in Table 2.2. Consequently, this platform is selected for the implementation of these models. A detailed presentation of this platform and its modelling language, GAML, is conducted in the next section.

Requirements / Platforms	P1 (model fit W1-4)	P2 (manipulable)	P3 (generic)	P4 (primitives)	P5 (CPU efficient)
Swarm	Yes	No	Yes	No	
Cormas	No (env.)	No	No	Yes	
Netlogo	No (env.)	Yes	Yes	No	No
Repast	Yes	No	Yes	No	Yes
VLE	Yes	No	Yes	No	Yes
GAMA	Yes	Yes	Yes	Yes	Yes

Tab.2.2 Requirements (P1-5) answered by the different ABMS platforms

3.3.3.GAMA and GAML

GAMA is a discrete-time, pseudo-parallel agent-based model simulator. It provides a ready-to-use agent-based modelling framework to domain experts thanks to the use of an accessible modelling language, GAML which is presented in details subsequently.

- **Introduction to GAMA**

It provides the user (domain expert or modeller) the possibility to use complex geo-referenced data as an environment or, even, to create agents directly from these data thanks to an advanced GIS library. Simulation can run with numerous agents (up to a million depending on the agents' complexity). Automated controlled run (i.e. batch) of simulations allows exploration of the model's parameters space through repeated simulations. The platform user interface is shown in Figure 2.13 with an example representing a simple "prey-predator model". The interface has four panels from left-to-right then top-to bottom: the parameter panel, 2 monitors (to track in real time values of global attributes), a chart (representing two population levels) and finally the visualisation of the simulation. The visualisation represents prey and predator individuals and the environment (grid-based representation). Prey individuals are represented by blue agents. Each cell of the environment represents a grass parcel that can be grazed by the prey agents. Predator, individuals are represented by red agents and "hunt" the preys. Depending on the parameters, no agents may survive, prey only or co-existence of both type of agents.
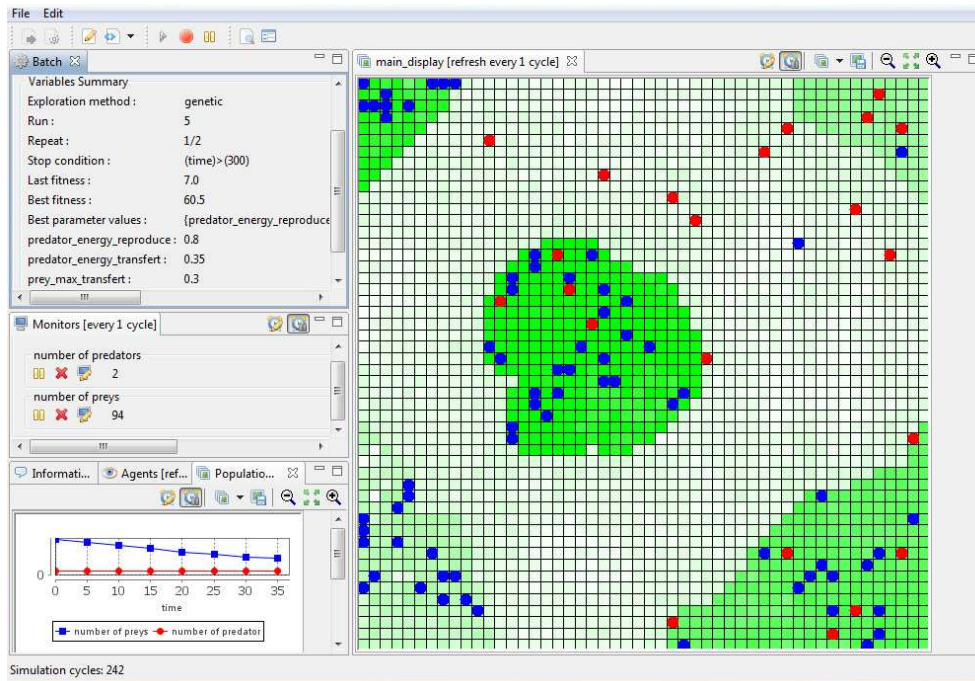
Fig.2.13 GAMA platform interface (with a prey-predator model example)
[Tai10b]

GAMA (version 1.3) is organised as presented in Figure 2.14. It consists mainly of a kernel, a meta-model and different controllers (of the simulation) linked to several tools: Eclipse for the user interface, Repast and GeoTools for GIS primitives. The kernel executes the simulation which can be controlled in different ways: Batch (e.g. parameter exploration), remotely (e.g. for collaboration) or interactively (e.g. for training the system from experts [Chu08]). The meta-model classes ensure that the implemented model is properly instantiated as a simulation. In particular, skills classes provide ready-to-use primitives (movement for instance) to the agents whereas the decisions classes provide control architectures: Finite state machine, reflex or task-based [Tai10a].
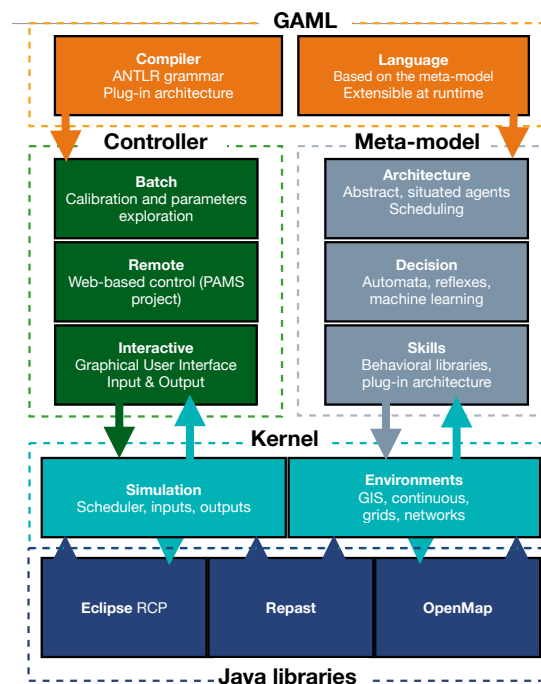
Fig.2.14 GAMA platform structure [Gam11]

- ***Entity and environment representations***

The agent meta-model used in the GAMA platform is presented in Figure 2.15. The three central elements are *agent*, *place* and *world*. *Agents* correspond to conceptual agents of conceptual models and ultimately to entities of the reference system. Similarly, *places* correspond to areas of the reference system environment, a field or a building for instance. *World* is the simulation manager, it holds parameters (input and time series), global variables and dynamics are defined. It is also within it that outputs (visualisation, graphics, and numerical outputs) are managed. *Place* and *world* are specialisations of agent in order to allow them to have similar characteristics in terms of dynamic and internal state.
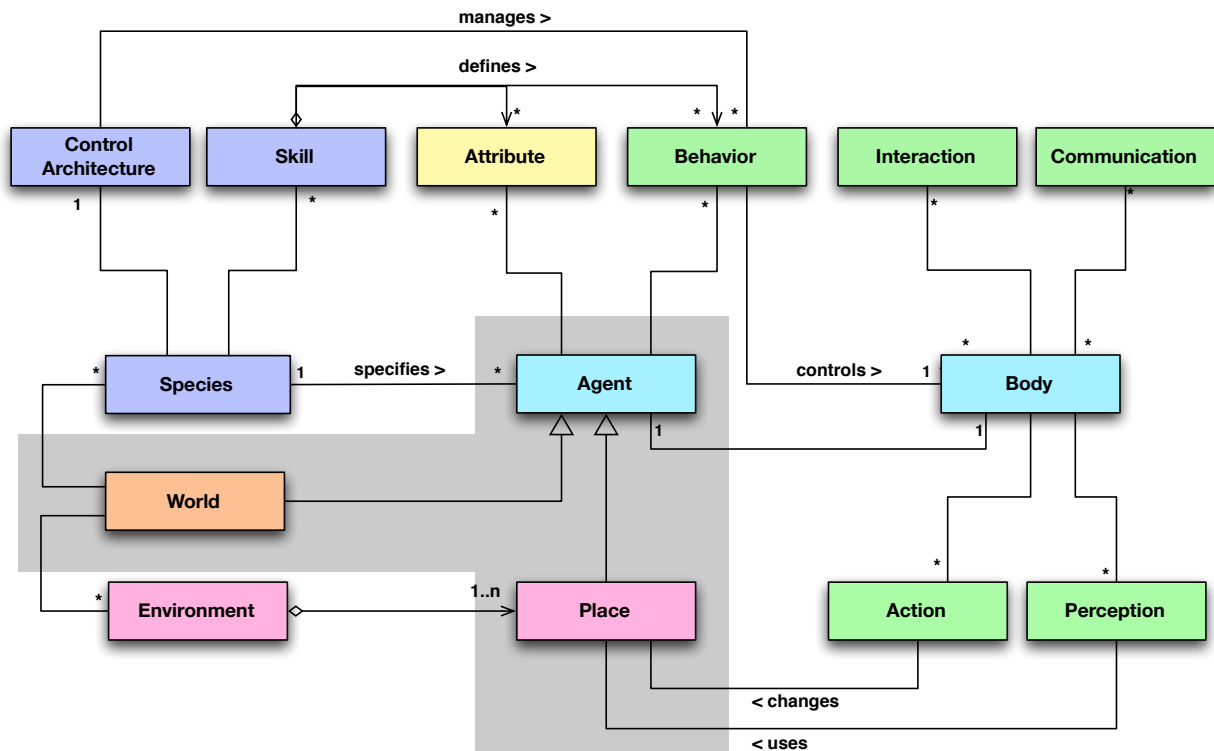
Fig.2.15 GAMA 1.3 meta-model **[Tai10b]**

The *Agent* class is responsible for holding the internal state of the individuals and their decision process. It is coupled to the *Body* class which enacts all the actions and interactions decided in the *Agent* class. The *Body* allows interaction and communication with other individuals and also interaction (e.g. localisation) and perception of the environment.

Similarly to object oriented programming, the concept of class is introduced in the meta-model by the *Species* concept. It represents generic generalisation of entities present in the reference system. For instance, in a traffic model, a species vehicle would correspond to a type of vehicle. It is where the behaviour and internal sate of the representation of the individuals is coded. It simplifies the implementation as it provides ready to use control architecture, or decision making architecture, such as reflex-based and task-motivation-based. It is also in this classes that the available *Skills* will be defined. In GAMA, a *Skill* is a set of defined primitives than can re-used in any models, a few built-in examples are: moving toward a target, computing a shortest path, establishing a communication using the FIPA protocol.

The idea of GAMA concerning the management of environment is to locate the agent in a continuous environment which is used as a reference **[Amo10a]**. It is possible to use a classic grid of cells and use it straightforwardly but the continuous environment remains available. In this continuous environment, all of the agents are provided with a geometric representation. The geometry of an agent can be simple (a point, a polyline or a polygon) or complex (composed of several sub-geometries). It is possible to use geo-reference

data to generate agents and places. In a segregation model, for instance, the environment can be generated from GIS data but also households agents can be initially located as in these data.

All of the computations related to spatial relationships are made onto the reference environment thanks to many built-in primitives based on the GIS library Geotools (previous versions relied on OpenMap, a more limited library). Thus, the distance between two agents, the overlapping of geometries, etc., are directly computed from it. In addition, the role of the reference environment is to adapt all situated actions of agents (movement, action, interaction) to all registered environments. Thus, for example, when an agent moves, the reference environment manages itself the movement of the agent on the potential grids. This operation is totally transparent for the modeller, who does not have to manage the synchronisation of the different environments. GAMA offer numerous other spatial features **[Amo10,Tail11]** such as: geometry transformations, distance between two geometries, shortest path, etc.

- ***GAML, a domain expert oriented modelling language***

In a world model context, the domain expert is at the centre of the modelling process (W4, R4, D2, P2). Tools and methods to attain such a requirement have been proposed for domain, design and operational models (Section 3.1) but the implemented model accessibility remains an issue. In such context, the main objective of a modelling language is to address this requirement (P2): making the implementation accessible and manipulable by domain experts.

The GAMA modelling language, GAML, is a generic XML-based modelling language created during the development of GAMA. It was inspired by EMF **[Dro92]** and the Netlogo platform's language. It was designed to allow fast and easy learning from the thematicians.

The structure of model in GAML is made of four sections:

(1) Global: global variables and simulator related command (parameters, temporal series, controller type, etc)

(2) Environment: different environment and places, their dynamics and internal-states

(3) Entities: agents with their dynamics and internal-states

(4) Output: The output is where the visualisation, charts, numerical output and interactions is described.

Figure 2.16 shows a basic example of a species definition. It defines a species called "animal" and which is capable of movement (″moving″ *skill*). Its internal state is defined by a "colour" and a "size". Its only behaviour is to "wander" according to two parameters the "amplitude" (how much it changes its direction at most at each time step) and the "speed".

```
<species name="animal" skills="moving">
     <rgb name="color" init="rgb 'blue'" const="true"/>
     <var type="float" name="size" value="(rnd 100) / 50" min="1"/>
     <reflex>
          <do action="wander">
               <arg name="amplitude" value="120"/>
               <arg name="speed" value="(rnd 200) / 50"/>
          </do>
          </reflex>
</species>
```

Fig.2.16 Example of a GAML description of an "animal" agent

In order to ease further the direct manipulation of GAML by domain experts, the future (for the present day) version of GAMA (1.4) will introduce an improved language which is developed using the Xtext framework.

This framework simplifies the development of a DSL, its grammar, a dedicated editor and an analyser. The new GAML version will still be similar to the original but simpler by getting closer to a classical programming language. Figure 2.17 shows the previous example in the new language.

```
species animal skills:moving {
      var  color type:rgb value:blue  const:true
      var  size type:float value:((rnd 100) / 50)  min:1
      reflex {
            do action : wander }
                  arg amplitude value:120
                  arg speed value:((rnd 200) / 50)
                  }}}
```

Fig2.17 Previous example expressed in new GAML

3.4.Summary of the selected support tools

The contributor roles and modelling cycle defined in Section 2 are complemented with tools described in this section in order to be able to propose a complete world models design method. These tools are UML, the O.D.D. communication protocol and the GAMA simulation platform.

Domain models can be expressed with simple ontologies (as in ARDI, Section 2.3.4) and if required by domain experts in their own formalisms. In contrary, design and operational models require more technical details and formality. To do so the UML diagrams (class, object, activity, state transition) should be used with the support of O.D.D. for explanation. The second purpose of the O.D.D. documents is to keep tracks of the modelling process and model iterations.

All the requirements to implement world models are addressed by GAMA. It is a generic platform and provides a large set of built-in functionalities. It allows detailed representation of the individuals and of the environment, supports large number of agents and have an accessible modelling language.

4. Conclusion

In this chapter I start by introducing the concept of world models. They come from the recent evolution of model uses from a predictive orientation to a more and more exploratory one. Unfortunately, they are not specifically considered nor defined in the KIDS approach and even less in the KISS one. Thus it is required to define them and propose a way to conceive and implement them. I review the introductory examples from chapter 1 (Sections 2 and 3) in order to extract their characteristics. Their main characteristics are the explorative or explanatory objective, their iterative construction which is driven by the domain expert as defined in Section 1. I also reviewed these examples to list all the tools that were used to construct their models. I selected the tools that can be re-used in a more general context to design world models but no general modelling method has been proposed.

The second section of this chapter reviews existing conception methods that may be used to design and implement world models. I use software development methods (Section 2.2.1) to introduce AOSE methods (Section 2.2.2) and because the recent "Agile methods" propose interesting elements in terms of adaptability and interaction with the client/domain expert. Section 2.3 focuses on complex system oriented methods which are more fitted to the design of world models. Still, these methods do not address all the world models methodological requirements. Thus, I introduce the contributor roles and the modelling cycle of the modelling method proposed in this thesis which is described in details in Chapter 5.

Section 3 presents the tools I selected to conceive and implement world models in order to propose a complete modelling methods in complements of the contributor roles and modelling cycle previously introduced. These tools are the UML to express conceptual models, the O.D.D communication protocol to explain the models and document the modelling process and the GAMA along its modelling language

GAML to express implemented model. The offer of ABM simulation platforms is vast but GAMA is certainly the most fitted for world models thanks to its advanced environment representation and its ease of access through its modelling language.

As one can see in this chapter, many existing tools address partially, at best, the requirements to design and implement world models. Some tools are generic enough to be adapted and linked together but many are too specifics to be of interest. All these tools were assessed and used during the application modelling process. In Chapter 5, I propose a more formal organisation in order to propose a complete and general modelling method for world model.

Chapter III : Application context and objectives

Epidemiology is the science that aims at characterising the spread of diseases within human or animal populations, forecasting their outcome and explaining epidemiological dynamic. In these purposes, models became standard tools since a long time, but they do not really evolve in time to meet the emergent needs of epidemiologists. Indeed, most epidemiological models are designed according to classic principles, like the compartmental models for instance. However, epidemiologists need to evaluate and integrate an increasing number of parameters related to varied facets of reference systems: Related to individuals, to the environment or to the population. Epidemiology would thus profit a lot in the design of *world models*.

The avian influenza epidemic is of special interest: Despite the good characterisation of its spreading at a country scale, spatial patterns at a province scale are still unknown. Thus, forecasting and preventing an outbreak at this level is barely possible. *Gripavi* is a project on avian influenza conducted by the CIRAD[7] institute in different countries. It studies the local spreading and persistence of this disease. Dr. Stéphanie Desvaux supervised the vietnamese part of this project, which started in 2007, soon after important avian influenza outbreaks that occurred in Vietnam between 2003 and 2005. I collaborated with Dr. Desvaux and veterinarians from the NIVR[8] institute to design a world model allowing investigation of their research questions on the avian influenza situation in Vietnam. Especially, their research focuses on the interplays between several dynamics: Poultry production, natural and cultivated environment, and human dynamics. Moreover, the environment is considered to play a key role in the disease persistence and propagation.

This chapter introduces the applicative context of my thesis. The general context in Epidemiology is presented first, focusing on the avian influenza. The collaboration context and the different model contributors are also introduced. Then, existing models about avian influenza are reviewed from the most general models (largest scale) to the most specific ones (smaller scale). This panorama of models provides insights and data on avian influenza which are re-used in the proposed model.

1. Avian influenza: history, context and research project

Avian influenza is only one type of influenza viruses that has been afflicting the human population since a very long time. Most of them only cause a mild "flu" but sometimes a virus mutates or is combined with another influenza virus infecting a different species which can generate a very lethal virus and possibly a large epidemic. For instance, four influenza pandemics occurred in the last hundred years. These pandemics had dramatic consequences and the current avian influenza virus may evolve and produce a equivalent pandemic.

1.1. Introduction to influenza

1.1.1. Influenza viruses

Influenza is caused by numerous viruses. They can be regrouped in three main classes (or *genera*) of RNA viruses (which use ribonucleic acid as genetic material): Influenza A, B and C **[Web92,Kam06]**. Influenza B and C viruses are known to infect humans as well as seals or pigs **[Kam06]**. Influenza A viruses can infect many different species of birds and some mammals including humans.

Influenza A viruses are differentiated according to their serotype which depends on the type of two surface proteins: *Hemagglutinin* and *Neuraminidase*. Figure 3.1 illustrates the general structure of an influenza virus. 16

---

[7] Centre de coopération Internationale en Recherche Agronomique pour le Développement (http://www.cirad.fr/)

[8] National Institute for Veterinary Research

hemagglutinin (H1 to H16) and 9 *neuraminidase* (N1 to N9) exist. These subtypes of virus, HxNy, are then classified in strains that appear almost as frequently as these viruses mutate. Finally, these strains are differentiated according to their pathogenecity: Lethal strains are called *HP (*Highly Pathogenic) while less dangerous strains are called *LP* (Low Pathogenic). According to this classification, the well-known *Avian Influenza* virus (AI) is often denominated as: *HPAI   H5N1* or *A/H5N1* or *HP H5N1*. In this thesis, *H5N1* denotes the virus whereas *HPAI* denotes the disease.
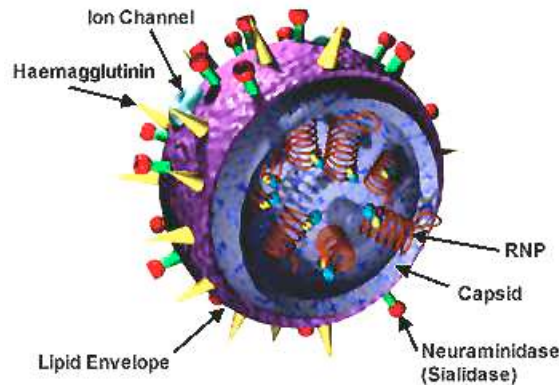
Fig.3.1 Influenza virus structure **[Wik02]**

1.1.2. Influenza pandemics of the 20th century

Nowadays, influenza or *flu* is usually considered as a winter annoyance. From a historical point of view, influenza is one of the disease that kills the most humans through numerous pandemics. The oldest known influenza pandemic dates back to 1580, and reoccurs several times a century since then: 1729, 1781, 1830, 1898, 1918, 1958 and 2009 **[Pot01]**.

Spanish flu

At the end of the first world war, an epidemic started to spread through America and quickly became the most important pandemic in human history: the *spanish flu* **[Pat91,Pot01,Bar04b]**. Indeed, around 500 millions people, about a third of the world population, contracted the sickness. Recent studies claim that up to 100 millions people died during this pandemic, between 2,5% and 5% of the world population. It was especially unusual because the mortality rate of other pandemics is below 0.1% **[Joh02,Tau06]**.

The "biological origin" of the virus is still not known for sure. One theory suggests that the virus "jumped" directly from birds (poultry from industrial farms), to humans **[Bel05]**. Similarly, the "geographical origin" remains unsure but a consensus emerges, pointing out Fort Reiley (Kansas, USA) **[Cro89,Phi04,Bar04b]**. Another study proposes the most probable spatial trajectory of the pandemic **[Pot01]**, as depicted in Figure 3.2. The first reported outbreak was on a boat travelling from England to Sierra Leone. Then, the epidemic has been located in Brest (West France) and then unleashed onto Europe and North America (through another boat travel). Wherever the exact origin of this pandemic is, the disease spreading is well-known but remains very complex and world-scale.
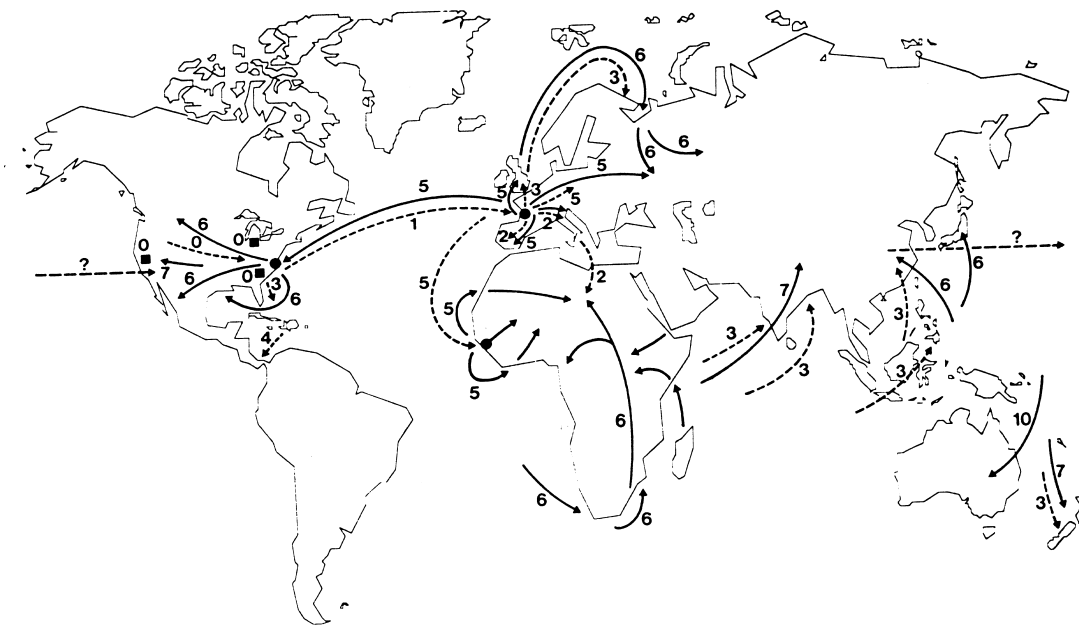
Fig.3.2 Spanish flu pandemic spreading **[Pot01]**.

Several other pandemics occurred during the 20th century but of lesser consequences. These pandemics were caused by viruses derived from the *spanish flu* virus (an H1N1 virus **[Tau06]**) but with less dramatic impact.

Subsequent influenza pandemics

In 1957, a pandemic spread from the Yunnan province (China) all over the world during 6 months. This pandemic virus was a recombination of the previous H1N1 virus and a H2N2 virus that roamed in poultry farms **[Por01,Bel05]**. This pandemic, called later the "*Asian Pandemic*", infected about 25-30% of the world population and caused around 1 million death (almost "nothing" compared to the spanish flu).

Another pandemic spread from Hong-Kong in 1968. The virus that caused this pandemic, which started in 1968, was a recombination of the *Asian pandemic* virus and a H3Nx virus (originally found in poultry farms) and caused around 1 million deaths **[Bel05]**.

The last influenza pandemic is probably the *swine flu* pandemic which started in 2009 in Mexico. This "flu" was caused by a H1N1 virus that is a recombination of three viruses usually found in human, swine and poultry populations **[Tri09]**. The use of the term "pandemic" can be questioned since the impact of this pandemic is definitively below others: 14,000 people died of the disease **[Ecd10]** whereas the smallest pandemics presented in this section caused at least 1 million death each.

1.1.3. Conclusion

Even if influenza viruses are often considered as a winter annoyance, they caused some of the worst human pandemics in recent history. All "pandemic viruses" have varied effects in terms of mortality rate, but may have apocalyptic consequences as the *Spanish flu* had.

Unfortunately, influenza cannot be eradicated as it happened with the *Smallpox* **[Hen77]**. Indeed, it would be much more difficult with Influenza A viruses because wild waterfowl act as a virus reservoir. Moreover, they are more resistant to the disease (lower mortality rate) and their eradication is out of the picture. Thus, only one possibility remains: disease prevention and control **[Web98]**. However, it is especially challenging due to the "considerable antigenic variation" of influenza A viruses. In other words, these viruses evolve quickly (new strains of viruses appear every year and minor mutations all the time) and thus limit the effect of immunity mechanism **[Web92]**.

1.2. H5N1 virus

All the viruses causing influenza pandemics since 1918 share a common origin: poultry influenza viruses. Such a disease that is transmitted from animals to humans is called an *enzoonose*. The evolution cycle of such viruses is illustrated in Figure 3.3. A first possibility of evolution is the *recombination* of a human strain and an avian one. The recombined virus can pass from birds to humans, and then more easily from humans to humans. Another possibility of evolution is the *mutation* of virus strains, passing by different types of poultry. In this case, the immunity mechanism of humans is much more limited because humans have not been confronted with a similar virus.
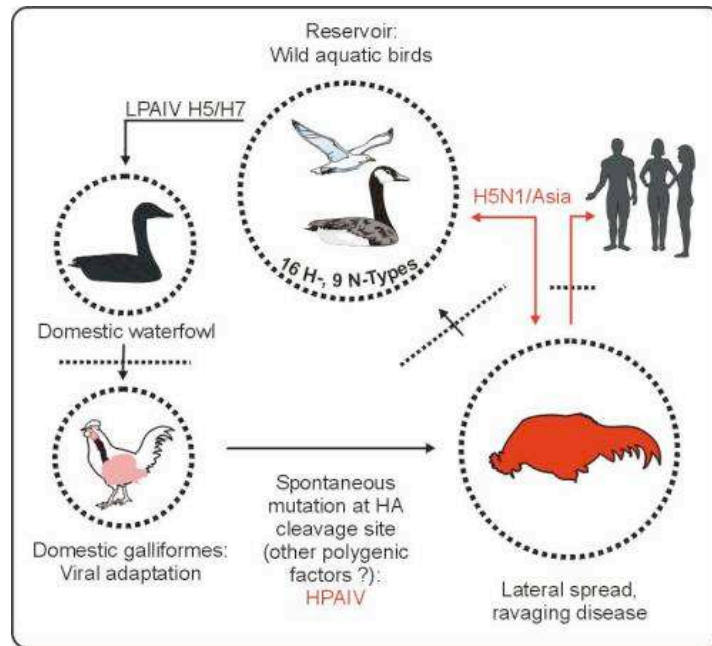
Fig.3.3 Circulation of influenza **[Har11]**.

1.2.1. H5N1 epidemic spreading

In 1996, the H5N1 virus causing HPAI has been isolated from a poultry of a farm located in the Guangdong province in China **[Who10, Oie11a]**. The following year, the virus has been isolated in both living animal markets and farms of Hong-Kong. It was the first epizootic (animal epidemic) event and the first attested case of transmission to humans (enzoonose).

This first HPAI epidemic was insignificant compared to the one following 6 years later. The H5N1 virus caused the recent "pandemic" that occurred in Asia, Europe and Africa **[Who10, Oie11b]** as illustrated in Figure 3.4 where dark red corresponds to countries in which the virus killed humans while bright red corresponds to countries in which the virus killed only birds.
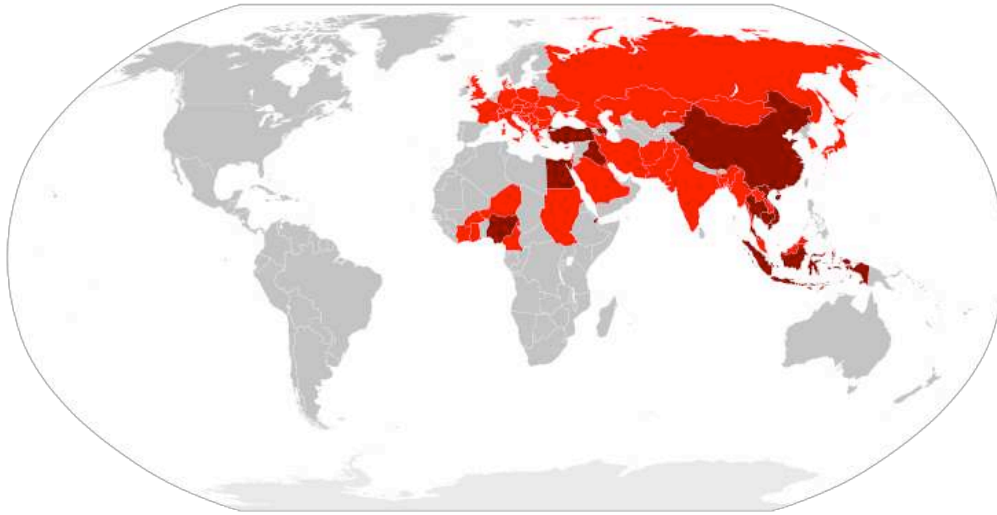
Fig.3.4 Spatial extends of avian influenza **[Wik03]**

In december 2003, the virus was reported in Korea and Thailand. In january 2004, it was isolated in Vietnam, Japan, Cambodia and Lao and Indonesia. The next year, 2005, the disease propagated west through Russia, Kazakhstan, Mongolia, Ukraine, Turkey, etc. It spread over Europe before the end of 2005. In 2006, numerous europeans countries reported presence of the virus in wild or domestic birds. Finally, the epidemics extended to Africa (Burkina Faso, Sudan, Côte d'Ivoire). Nowadays, some outbreaks still occur in european and african countries (e.g. Egypt). However, the disease remains much more prevalent in Vietnam and Indonesia, which are the two main endemic sites.

1.2.2.H5N1: the next pandemic?

The avian influenza virus (H5N1) is not (yet) circulating much in the human population. Figure 3.5 shows the geographical distribution of attested human cases (since 2003). As we can see in this figure, Asia bears the largest toll, especially Vietnam and Indonesia. Even if the number of death attributed to the H5N1 virus is low (less than 300 world wide), its human mortality rate is very high at around 60% **[Who11a]** whereas the *swine flu* (H1N1) one is "only" about 1% **[Fra09]**. This virus strain has a high human mortality rate and thus is called Highly Pathogenic Avian Influenza (HPAI). Concerning birds, the mortality rate is high for chicken but generally low for duck (depending on the strain) **[Pan07a, Pan07b]**.
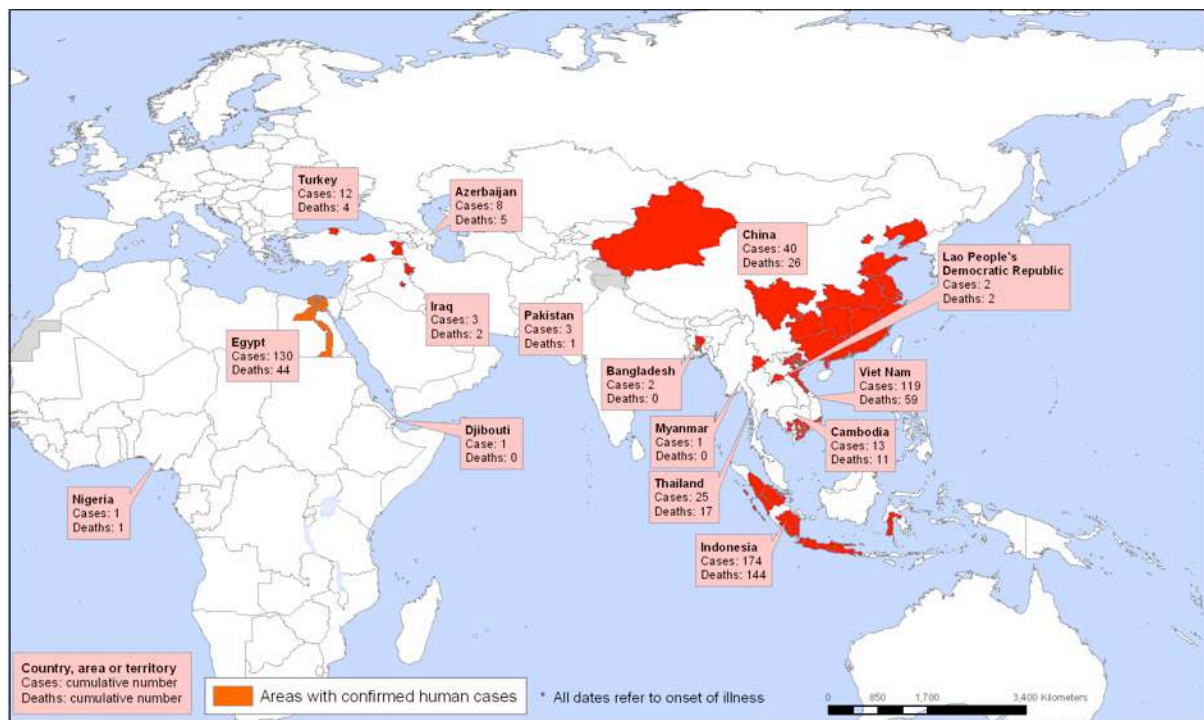
Fig.3.5 Geographical distribution of AI human cases **[Who11b]**.

A H5N1 pandemic has not occurred yet because the virus cannot be easily transmitted from humans to humans. However, the next influenza pandemic will probably be due to the recombination of the H5N1 with a benin human flu (easily communicable). With respect to a 60% mortality rate, the recombined virus could spread at the world-scale with tragic consequences (like the *spanish flu*).

1.2.3.Conclusion

Avian influenza is a major threat for humanity. In spite of a "weak" number of victims (around 300 human deaths), this disease is a "good candidate" for the next pandemic: world-scale propagation and high human mortality rate. The recombination with another virus strain more adapted to humans is possible, and is probably a matter of time. Moreover, the current economic impact of the epizootic part of avian influenza is massive. The world counts around 16 billions of chicken **[Bro06]** and a wide propagation of the virus would irreparably damage the economy of many developing countries like Vietnam. To prevent the massive propagation of the virus, different actions have been introduced: bio-control measures, culling of farms containing infected poultry, massive vaccination but also research which already cost billions of dollars **[Wik04]**.

1.3.Vietnam: poultry production and avian influenza

Vietnam is the easternmost country of the Indochina peninsula in South East Asia. It is a S-shaped country delimited by China on the north, Lao and Cambodia on the west and a long coastal border of 3,260km with the gulf of Tonkin (south) and the South China Sea (east). As shown in Figure 3.6, the country is decomposed in 8 regions, 58 provinces and has 2 important deltas: The Red River Delta (North) and the Mekong Delta (South) where 80% of the population lives and where the largest cities are located: Hanoi (the political capital) and Ho Chi Minh city (the "economical capital") **[Des08]**.

Fig.3.6 Map of the vietnamese regions **[Wik05]**.

Vietnam has a global tropical monsoon climate but North, Centre and South parts have clearly different climates. North has four seasons while the South as only two, dry and wet seasons.

Vietnam is the third largest country of South East Asia and has an overall population of almost 90 millions persons (july 2010). The country is experiencing a rapid urbanisation, 3.1% per year between 2005 and 2010, with a strong economical growth since several years, more than 5% GDP (Gross Domestic Product) growth per year since 2000 **[Cia11]**. Despite these economic and social evolutions, Vietnam is still a rural country where 70% of the population lives in rural areas **[Des08]**.

1.3.1.Poultry production in Vietnam

In spite of a decreasing part of the economical activities, Agriculture is still a crucial sector in Vietnam where more than 2/3 of the population works. The main production is rice whereas the main livestocks are pigs and poultry.

According to **[Agr07]**, four poultry production sectors can be distinguished according to the producers' size and the industrialisation level. First, *industrial farms* contain up to dozens of thousands of poultry and are fully integrated: All interactions with the outside are strictly limited and controlled, birds are always kept indoor to prevent any contact. Then, *commercial farms* are smaller and less integrated. They contain up to a few thousands birds, which are also kept indoor, but contacts are less limited or controlled. *Semi-commercial* farms contain from 50 to 150 birds at a time and are little integrated: Birds may have contact with the outside. Finally, *traditional farms* or *backyard farms* contain up to 50 birds without integration. Poultries freely wander in and out of the farms in mixed flocks, as shown in Figure 3.7. In rural areas, almost 90% of households have a backyard flock **[Des08]**. This production is important for many of these households as it represents 19% of their income.

Fig.3.7 Backyard poultry (chickens, ducks and gooses) **[Des08]**.

At the country scale, in 2006, the poultry population was composed by 150 millions chickens and around 60 millions ducks **[Des08]**. All farms produced more than 300,000 tons of poultry meat and almost 4 billions of eggs **[Gso07]**. Figure 3.8 illustrates the poultry density per province in Vietnam. This figure highlights the two main basins of production: Red River and Mekong deltas.
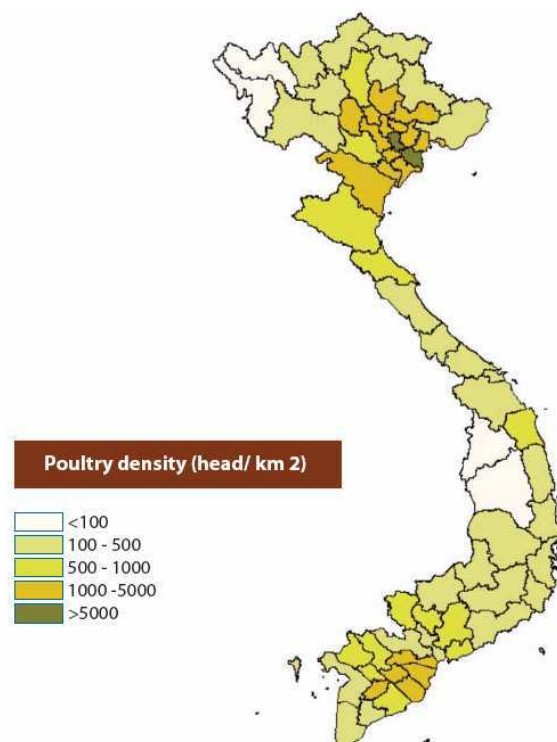
Fig.3.8 Poultry density per province in Vietnam in 2006 **[Gso07]**.

1.3.2.Avian influenza epidemics in Vietnam

Since 2003, several avian influenza epizootics have occurred, and Vietnam became the country with the highest incidence rate. Indeed, between 2003 and 2009, more than 2300 local outbreaks have been reported **[Oie11c]**. Despite strong (and sometimes desperate) bio-control measures including massive culling in infected farms and intensive vaccination, the avian influenza virus is still endemic in Vietnam **[Sch09]**. Figure 3.9 presents the location of these outbreaks and highlights two important build-up in both Red River and Mekong deltas.
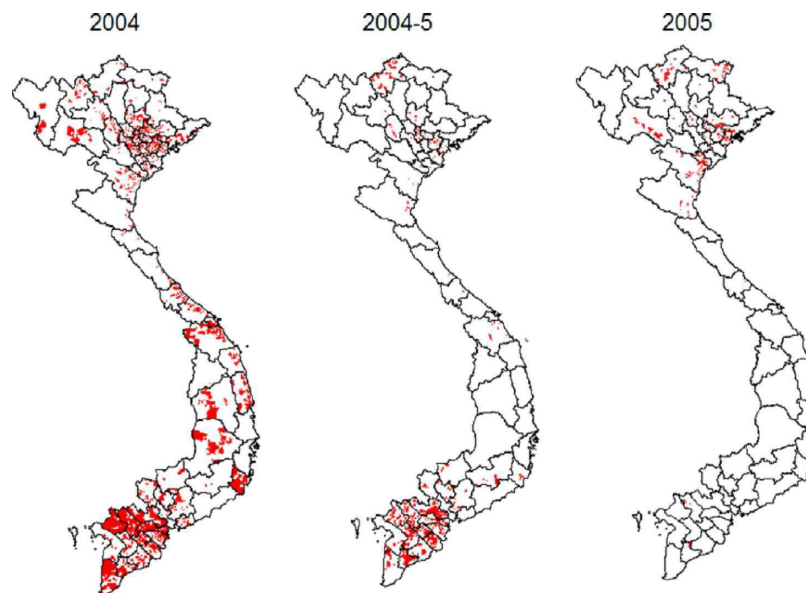
Fig.3.9 Location of the avian influenza outbreaks in Vietnam in 2004-2005 **[Pfe07]**.

In an optimistic point of view, the disease seems under control as the outbreaks frequency is decreasing. However, there is no guarantee that large epidemics will not burst out again. At best, such a re-emergence of the disease would lead to a massive blow to the country's economy. At worst, a pandemic similar to the Spanish flu one could happen with large human fatalities.

The cost of the bio-control measures is already important, especially in a developing country such as Vietnam. The culling and vaccination campaign are particularly costly. Furthermore, if the disease bursts again into a country wide epizootic, it would have major impact on the poultry production sector and in particular on households with low income which rely on this important side source of income (19%) **[Des08]**. For instance, the cost of the first epidemic waves was estimated over 100 millions USD **[Rus05]**.

1.4. The Gripavi project

Gripavi [9] is a collaborative project lead by the CIRAD institute that started in early 2007, when recurrent outbreaks of avian influenza were still frequent in Vietnam. At that time, the vietnamese government had already put in place bio-control measures to restrain the epidemics. In spite of massive poultry culling and a mass vaccination campaign that have been carried out since 2005 **[Sim07]**, the situation was not totally under control. For example, outbreaks occured in 2007 after about a year without reported case **[Min09]**. Moreover, outbreaks continued to be reported until late 2009 **[Bro10]** and are likely to continue to occur (even if not always reported).

In this context, many researchers of various domains (Epidemiology, Virology, Social sciences, etc.) are studying the avian influenza situation in Vietnam even today when the disease appears to be under control. In particular, the Gripavi project was designed to understand *how* and *why* the disease reappears, especially since poultry production sector is economically essential in Vietnam. The need for models was initially accounted for because field studies require important financial support, time and would be insufficient to evaluate domain experts hypotheses. Thus, Gamavi, the application model described in Chapter 4, has been conceived within this project.

1.4.1. Gripavi and other CIRAD's avian influenza projects

---

[9] Gripavi website: http://gripavi.cirad.fr/

CIRAD is a research institute strongly implicated in the study of animal emerging disease. Evidently, avian influenza is one of them thus the institute started two projects to study the disease and its epidemic: *RevAsia*[10] and Gripavi. The former focuses on evaluating the surveillance system to detect spreads of avian influenza virus. The latter focuses on understanding the ecology of the disease, assessing the risks and enhancing the knowledge of stakeholders. In addition, CIRAD was also involved in a project called RIVERS[11] whose objective is to characterise the virus survival in water environment and is also linked to Gripavi.

The global Gripavi project has three main objectives:

- Assessing the risk of avian influenza (mainly from an agronomical point of view)

- Evaluating and proposing surveillance and control technics

- Analysing the economical impact of the avian influenza epidemics and of the mitigation measures

Even though the three objectives are linked together (no cost analysis can be done without knowledge on the mitigation measures proposed which cannot be done without risk assessment), the objective of the present collaboration was only part of the risk assessment.

1.4.2. Gripavi in North Vietnam

Gripavi is a project with field of study in many different countries: Ethiopia, Madagascar, Mauritania, Zimbabwe and Vietnam. The vietnamese chapter of the Gripavi project was coordinated by Dr. Desvaux. Thus our collaboration was integrated in her project. The global project takes place in several countries (Ethiopia, Madagascar, Mauriatnia, etc.) but the vietnamese situation is peculiar. Indeed, this country is one of the few sustaining recurrent avian influenza (local) epidemics. This situation is causing important economical issues to poultry production stakeholders and is threatening public health.

As a part of the Gripavi project, the vietnamese chapter is specifically about investigating the local factors causing the avian influenza epidemics. More specifically to the vietnamese situation, experts already assumed that the interaction between natural environment, wild poultry and domestic poultry are likely to be the combined origin of these recurrent outbreaks.

This project started in 2006 by getting into contact with local authorities, searching for possible fields of study and the beginning of retrospective data collection. I initiated the collaboration with Dr. Desvaux at the opportunity of one of her preparatory visits in Vietnam. In addition to their statistical analysis and static models (e.g. network based) I proposed to design a world model of the avian influenza in North Vietnam to investigate the epidemiologists' hypotheses.

1.4.3. Epidemiologists' objectives and description of the reference system

Gripavi was the "umbrella project" for our collaboration and provided us with a clear research question: "*what are the local determinants and patterns influencing introduction, persistence and diffusion of the virus in North Vietnam among domestic poultry?*". Nevertheless, we limited our study to persistence and propagation (details on the collaboration objectives are given in Annex 1).

There are two main areas of poultry production in Vietnam, the Red River Delta in the north and the Mekong River Delta in the south. Even though they are structurally similar in terms of poultry production, sufficient differences exist to distinguished them: Climate (4 vs 2 seasons), different ratio of production sectors and different practices (e.g. transhumant ducks in the south). Consequently epidemiologists decided to focus

[10] Revasia website: http://avian-influenza.cirad.fr/fr/projets/recherche/revasia

[11] RIVERS website: http://avian-influenza.cirad.fr/fr/projets/recherche/rivers

on one region. They focus on the Red River Delta, more specifically, nine communes located in two provinces; Ha Tay[12] and Bac Giang as located in Figure 3.10.
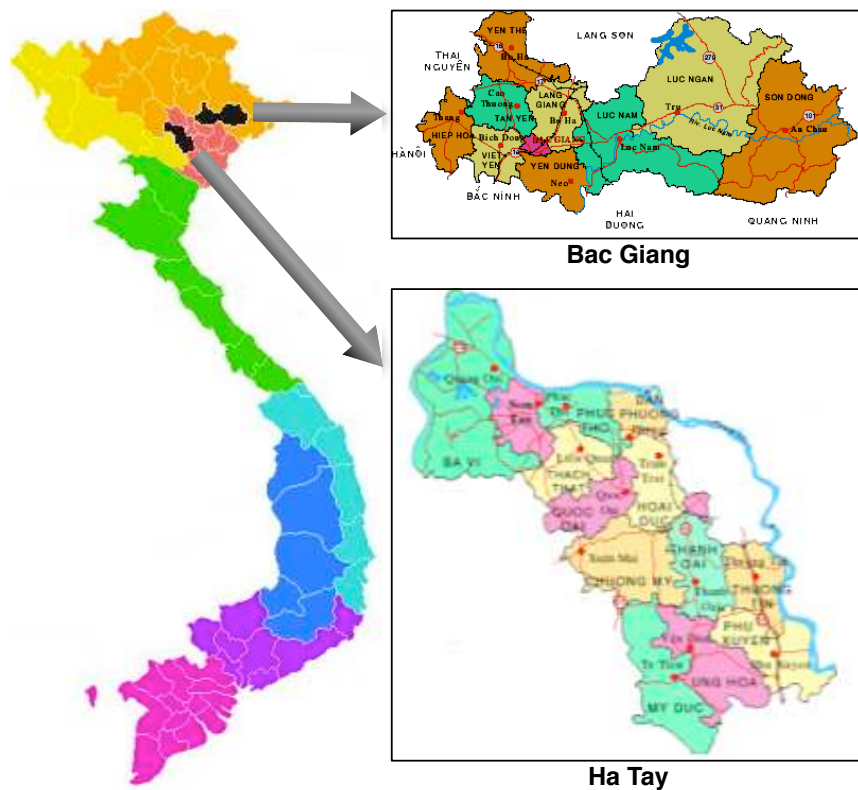
Bac Giang

Ha Tay

```
Fig.3.10 Bac Giang and Ha Tay provinces where the studied villages are
                               located
```

The communes of Bac Giang province can be considered to be representative of the Red River Delta agricultural system while Ha Tay ones are specific because they are strongly specialised in poultry production. The production in Ha Tay is mostly sold in Hanoi city and is also much more industrial than other Red River Delta provinces.

2. Avian influenza in Vietnam: associating classic models to create a world model

Avian influenza have been widely studied from numerous standpoints with very varied models. Many of these models are of interest as they characterise various elements of the reference system. Thus, I partially re-use most of the model presented.

In the current section, I start by sketching out what is my application: I detail the initial research question addressed to the model and present its reference system in Section 2.1. Even though these two elements belong to the modelling cycle presented in Chapter 4, it is important to introduce them here to allow the selection of models for the review presented in Section 2.2. Finally, Section 2.3 introduces the issues on integrating all the relevant information scattered in the model presented in Section 2.2.

2.1. Characterisation of the reference system

2.1.1. The reference from the domain experts' point of view

---

[12] Former province that has been integrated into the Hanoi province in 2008

Given the epidemiologists' research question, "*what are the mechanisms at play in North Vietnam allowing the local persistence and local propagation of the avian influenza virus among domestic poultry?*", I can detail it to sketch out the requirements for the design of a world model.

Poultry have to be represented at the "right" spatial scale which is the "local scale" which is at most a province. Concerning the time scale it is implicit in the "*persistence"* term because persistence (in an ecosystem) implies at least one year, so the time scale would be of 1 year.

Domain experts hypotheses focused on two elements: poultry and natural environment. Indeed, the virus might circulate among poultry or there could be a cycle between the poultry and the environment. Thus the interaction among poultry, organised as flocks, and with the environment, notably with the ponds and rice-fields, should be represent.

The dynamic of the domestic poultry is ruled by humans which they are interacting with them. There are, actually, 2 type of interaction: the day-to-day interactions (feeding, herding, etc.) and the commercial-related ones (flock renewal, trade, etc.). Thus humans, the production and the commercial system should be represented. The flock organisation is thought to have a strong effect on the virus epidemic dynamics because it aggregates individuals as cluster which facilitates transmission between them. In addition, interactions might also occurs with other domestic animals and commensal or wild poultry. Consequently, all these entities should be integrated in the model. All the previously presented entities can be infected by the virus thus we need a mechanism to represent the evolution of the infection at the individual level and its transmission. Finally, recent studies **[Tra10]** showed that the virus can survived in the environment for several days in general and even longer in the water. It will be necessary to represent the dynamic or persistence of the virus in the environment.

2.1.2. The reference system from the modeller point of view

The extended description of the reference system's constituents from the thematicians' point of view need to be translated into representation's requirements in order for the modeller to design the model and assess existing models that may be of interest.

The central entity of the system is the poultry which is related to four main dynamics modelling: poultry trading, poultry production, virus dynamic in the environment and in birds (pathobiology). These dynamics constrain the choice of the spatial and temporal scale and resolution of the system. Through consideration of the interactions of poultry within these dynamics, it also introduces the other entities we need to consider.

The spatial scale is 100km as we limit our consideration of the poultry trading to the province level. The spatial resolution needs to allow representation of the day-to-day dynamic of the poultry (i.e. production dynamic) which includes movement of poultry (*e.g.* farm to fields or backyard divagation within the village). Thus the spatial resolution has to allow representation of the building, rice-field lots and ponds which can be represented with a resolution of 10 meters. Domain experts first defined a time scale of 1 year but in order to account for inter-seasonal changes, it has been decided to consider an extended time scale of 2 years. The temporal resolution is imposed by the consideration of the poultry movements (day-to-day dynamic) which can be approximated by the hour.

Numerous entities have to be represented to be able to model the different dynamics. These entities operates at different scales (spatial and temporal). For instance, on the one hand, a backyard poultry may have a relevant behaviour defined by the hour, repeated everyday and limited to a village. On the other hand, a specific market may operate once a month and attracts traders from the whole province. Thus, it is required to enable representation of these very different scales and resolution.

The environment needs to be detailed for two reasons: virus persistence and entities movement. The latter needs to take into account the land-use (e.g. rice-field, dry culture, ground) while the former need to represent the soil without culture and the road system for trading. The within building topology may be accounted as well. Finally, the environment needs to have its own dynamic because, for example, of the seasonal changes, which influence the virus persistence, and the possible propagation of the virus.

In order to assess domain experts hypotheses a single formalism needs to answer all these requirements. It needs also to allow alternatives hypotheses representation and frequent revision of the model. More specifically, all the considered interaction between the envisioned dynamics are oriented towards epidemiological process thus the integration of these elements needs to focus on these processes.

Finally, a not clearly expressed, yet important, requirements of the domain experts is to be able to use field data. Indeed, persistence or propagation within the given context may emerge in the reference system from various causes. Domain experts have various hypotheses on them that can be expressed thanks to the representation of entities, interactions or specific valuation of parameters. Nevertheless, all these hypotheses may not contain the actual cause thus it would be of interest to be able to represent as descriptively as possible the reference system in order to integrate parameters or entities that should not impact the investigated phenomena but actually are.

Many of these requirements have been addressed by existing models. Nevertheless, as I show in the following section they usually focus on a single dynamics and none corresponds to the need of our application.

2.2.Review of prior avian influenza models

Since its initial discovery in 1996 **[Who10]**, the avian influenza disease and its causing agent, the H5N1 virus, have been studied in many ways **[Bro10]**: evolution of the virus (i.e. phylogenetic **[Sho98]**), spatial patterns at various scales **[Jou07, Gai08, Leb09]**, persistence of the virus in the environment **[Ren07]**, pathobiology **[Pan07a, Pan07b]**, etc.

In this section, I focus my review of avian influenza models on two elements: propagation of the disease at various spatial scales and persistence of the virus in the environment. Initially, the envisioned model of the application was defined at the province scale. Later on, we moved to decreasing scale (district then commune) thus, models localised at the inter-communal and intra-communal level are reviewed as well. This progressive decrease in scale is illustrated with the Vietnam country, Ha Tay (former) province, a commune and a village in Figure 3.11. In addition, the persistence of the virus in watered environment was stated initially to be an important mechanisms to evaluate thus two models proposing representation of this process are reviewed.
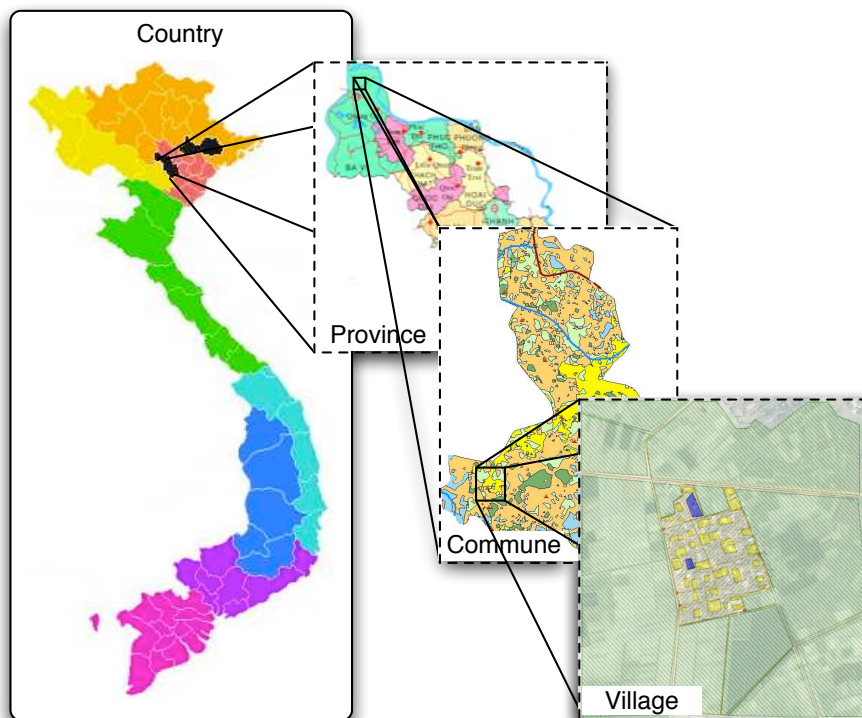
Fig.3.11 Illustration of the different spatial scales considered in the avian influenza models

This Avian influenza models review is informative for my application context as many concepts and data will be re-used and it highlights where are located the issues of the "classic" models. Unfortunately, the presentation made of these models in scientific publications do not present how the models have been built, the modelling method is not described nor even mentioned. Consequently, they do not provide hints or advices on how to conceive a world model in Epidemiology and more specifically in the Avian Influenza in South East Asia case.

2.2.1.Poultry trade and farm contacts at the national level

- **Objective**

Containment strategies or banning movement of potentially infected individuals is often an efficient bio-control measure to stop the spread of a disease. The model proposed in **[Kim08, Kim10]** focuses on evaluating such a strategy regarding a potential HPAI spread in South Korea.

- **Model**

The reference system considered are the farms of South Korea. At the time of the study, the number of large sized chicken farms were 2,231 (around 90 millions of poultry) and 808 for the duck ones (around 10 millions of poultry).

Chicken and duck are represented as (differentiated) flocks which are located over the map of South Korea (according to field data). A farm can be susceptible, infected or "activated" (SIA) as in classical compartmental model. The activated compartment represents the period where birds are actually dying of the disease and thus infection is visible while in the present infected state, the virus transmission occurs but the infection is not detected yet. The represented key parameters are: The infection probability (transmission from farm to farm), incubation rate (period of time when the disease is present in the flock but not detected) and the spreading distance (maximum distance between a infected farm and a susceptible one to be able to infect the latter).

Figure 3.12 gives a representation of the model. Farm agents are localised in a environment (left part) and according to a certain radius an infected farm agent can infected its surrounding susceptible ones (bottom right). The "SIA" compartmental model is shown on the top right.
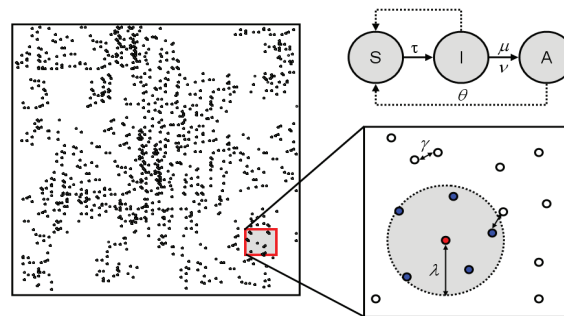
Fig.3.12 Location of the farms, zoom to illustrate the transmission
radius and compartmental model   **[Kim08]**

A simulated epidemic based on data from the 2008 South Korean avian influenza epidemic was conducted on the built multi-agent system. The simulation was populated with 500 agents actually representing chicken and duck farms (250 each). The simulation was executed for 50 simulated days (1 day = 1 step) with different values of infection rates and quarantine radii.

- **Discussion**

Based on the data from the 2008 outbreak, the authors proposed optimised quarantine radii that minimise both poultry depopulation and risk of avian influenza spread. The impact of incubation was also evaluated. The authors showed that the longer is the incubation period the larger will be the infected farms number in the end (due to delay of detection and thus quarantine enforcement). Finally, the authors claim that the level of disease control (quarantine) should be determined after the virus transmission probability (the more virulent the disease, the larger the quarantine).

This model makes use of the agent meta-model in order to represent one type of entities of the reference system: The farms. Consequently, they are able to generate the disease spatial propagation from farm to another and evaluate mitigation measures (*i.e.* quarantine). Unfortunately, this model limits its consideration to one dynamic even though it may be easily extended or adapted as it is an agent-based model. In addition, the modelling method is not described except for the use of geographical and epidemiological data at the farm level.

2.2.2.*Inter-province commercial system and avian influenza*

- *Objective*

The previous model considers the contact structure at the country level with data from South-Korea. The model proposed in **[Mag10a]** is also about avian influenza propagation but considers the commercial system at the regional level with data from North Vietnam. The model's objective is to characterise the live poultry flow in the trade network in order to evaluate its potential role in the diffusion of the H5N1 virus.

- *Model*

The data were collected over 131 communes of North Vietnam, their location is shown in Figure 3.13. The considered entities are : live bird markets (LBM) and wholesale markets interconnection by live poultry traders (LPT). Gathering of data focused on the greater Hanoi where 12 LBMs and 2 wholesale markets were locating at the time of the study (2007).
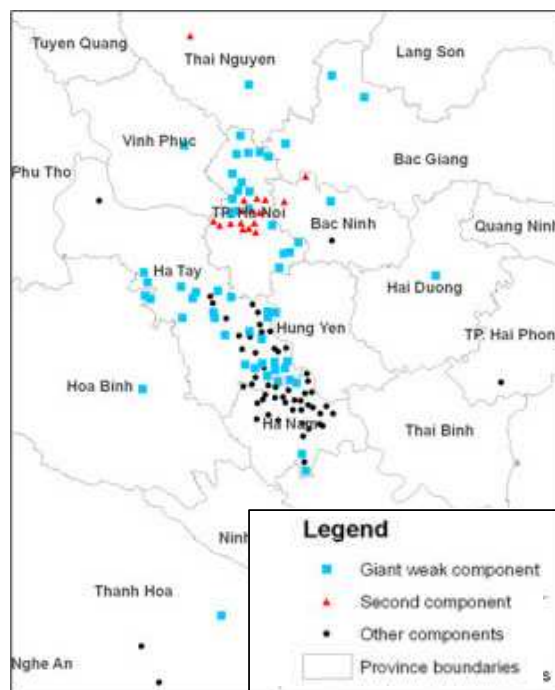
Fig.3.13 Communes where the investigated traders operate **[Mag10a]**

- *Discussion*

The authors find out that the network of communes can be decomposed as: A core component (strongly connected), a second component (less connected) and several lowly connected components as shown on Figure 3.13 (blue squares, red triangles and black dots respectively).

This model emphasis the need to improve bio-control measures. Less experienced traders (less than a year of activity) are specifically at risk because they are more likely to trade in areas where HPAI recently occurred (in comparison to experienced traders). Usually, a trader buy a whole flock and then sell it but (s)he can group several flocks. The study showed that mixing of flocks is indeed a common practice but should be discouraged as it increased the transmission risk. In addition, having some LPT of a given LBM limiting their catchment area facilitates the persistence of the disease in this group of communes.

This model provides a characterisation of the North Vietnam poultry trading network, which is a profitable data for our application. It also demonstrates that this trading network is likely to play an important role in the avian influenza propagation. Nevertheless, the model itself cannot be adapted as it is a global and static analytical model. Finally, the data collection process is introduced in the publication but no other methodological concerns are expressed.

2.2.3. Commune Commercial system and avian influenza

• **Objective**

In **[Pay09]**, the reference system is also the commercial system but at the commune level with data from the Red River Delta. Compared to the reference system of the previous model, it includes explicitly the farms in addition to the traders and markets (and add the commune markets). More specifically, the authors wanted to define the poultry flows of this system at the commune level which can then be used to characterise the risk of H5N1 virus diffusion within it.

• **Model**

In this study **[Pay09]**, the reference system is a theoretical commune defined after field study conducted in the Bac Giang province (North Vietnam). The actors considered are backyard and semi-commercial farms, small scale traders, commune markets and consumers.

In order to create this model, the authors had to get a insightful knowledge of the production and commercial dynamics. Indeed, they characterised the general production process: band type (group of chicks that are bought all at once and the adult poultry are sold altogether), dynamic of traders and seasonal changes.

• **Discussion**

The chicken (meat) production dynamic is a band one where chicks are bought altogether and adults are sold on a short period of time (for example10 days). The sales through traders occur all year long, except in march. The largest sales are observed in may, july and in the pre-lunar-new-year season (november to january) but smaller peaks are observed every 2-3 months. Figure 3.14 shows the evolution of chicken stocks within the different actors of the system: Farms, traders and markets in a generated network.

Traders are selling ducks all year long but the volume is much lower in January and February. Stronger sales occurs in april to june and september-october which correspond to the rice harvesting periods. Actually, it is a very commune practice to let ducks feed on harvested rice-fields at that time.
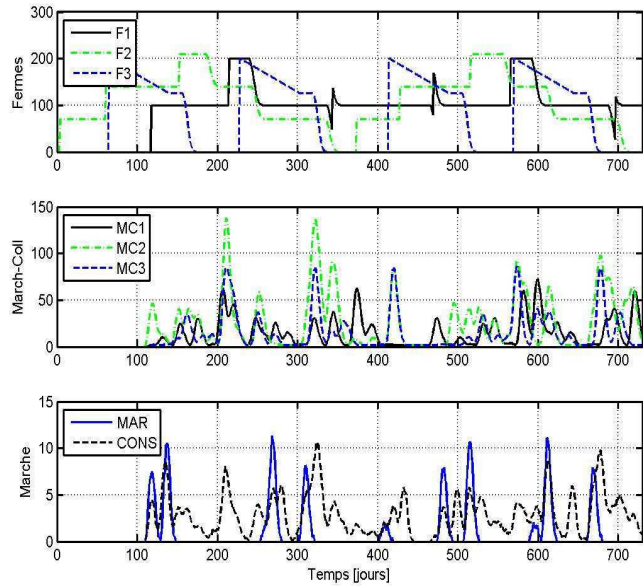
Fig.3.14 Generated evolution of the number of chicken in the different
actors: 3 farms, 3 traders, 1 market (and consumers)   **[Pay09]**

Similarly to the previous model, this one characterises a poultry trading network of North Vietnam. However, the considered reference system is a commune not a province. Moreover, it also generates (temporal) dynamic of the poultry flows within the system, flows that present clear seasonalities. Another interest of this model is the characterisation of several stakeholders of the poultry production sectors: Farms (different types), wholesale and small traders, clients and markets. Nevertheless, this model cannot be re-used directly for our application but its results (network structure and flows temporality) can be used as data-sources. Finally, the only methodological concerns mentioned are related to the way data were collected (i.e. snowball analysis) and processed.

2.2.4. Environment and avian influenza

- *Objective*

In **[Roc09]**, the model's objective (already introduced in Chapter 1) is to determine the most plausible road of avian influenza virus (AIV) transmission: direct contact or through the (watered) environment for wild waterfowl.

- *Model*

The reference system is the Camargue area (South of France) where many migratory birds come into contacts. The reference system is represented as a "host community" (birds) and the (global) environment. To characterise the host community demographic and AIV prevalence surveys were conducted on the field.

The model is a classic compartmental model with the three SIR compartments to which is added an environmental one (B) as defined by the set of equations below **[Roc09]**. The population dynamic is taken into consideration by two parameters describing the birth, b_m, and the death, d_m. There is also an immunity loss represented through the ε parameter. The environment compartment dynamic is governed by two antagonist processes: an increase depending on the number of infected individuals (parameterised by γ, the individual daily viral shedding rate) and the viral particle inactivation rate in water (π).

$$\frac{dS}{dt} = b_m N - (\lambda + d_m)S + \varepsilon R$$

$$\frac{dI}{dt} = \lambda S - (d_m + \sigma)I$$

$$\frac{dR}{dt} = \sigma I - (d_m + \varepsilon)R$$

$$\frac{dB}{dt} = \gamma I - \pi^* B$$

- *Discussion*

Comparison of the model outputs with actual data showed that the most probable transmission process actually involved two roads: inter-individual and water-borne. Moreover, the authors showed that the most adequate representation of the inter-individual transmission road is the density-dependent one (in opposition to frequency-dependent). Figure 3.15 illustrates such considerations by comparing the disease prevalence computed through the model with actual data. We can see that the density-dependent inter-individual transmission road combined with water-borne fits better the data. This model supports the hypothesis of the interplay between direct transmission and environment-mediated one to explain the avian influenza epidemics.
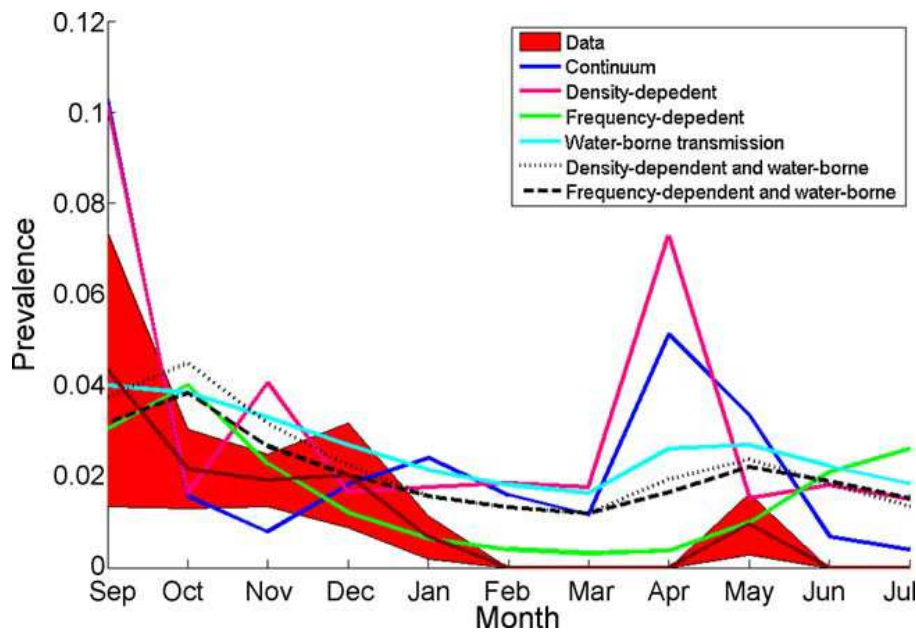
Fig.3.15 Generated and observed prevalence [Roc09]

This model demonstrates the role of the (watered) environment in the persistence of avian influenza. More specifically, the interaction between the environment and the wild duck population. It also provides several parameter valuation for epidemiological parameters. This model is also interesting because it allows users to test different hypotheses even if the authors do not describe the modelling process. They wanted to compare the different transmission roads (*i.e.* direct, through environment and combined) and two representations of the environmental road (*i.e.* density and frequency dependent). To attain this objective they designed a general model where a simple change of parameters affects the considered transmission road(s) and its representation. Nevertheless, this model remains at a global level and cannot be adapted easily. The considered entities are wild ducks whereas Gamavi focuses on domestic poultry. Similarly, the environment is a pond (or the concatenation of the ponds of an area) in France whereas Gamavi is designed to represent descriptively the natural and cultivated environment (including ponds) of North Vietnam.

2.2.5.H5N1 persistence in water

- **_Objective_**

The model proposed by René and Bicout in **[Ren07]** is similar to the previous one but has for objective to assess the risk of transmission of two types of avian influenza through watered environment. More precisely, the authors wanted to determine if both LP (low pathogenic) and HP (high pathogenic) avian influenza can be transmitted among wild ducks through the watered environment and if it allows long persistence of the virus, possibly over a season.

- **_Model_**

The reference system selected is the Dombes area (Eastern France) where many ponds and swamps can be found. The watered environment is represent as a whole while the individuals taken into consideration are the wild duck as they have no relevant interaction with other animals in the present study. Indeed, the investigated phenomena is the "oral-faecal transmission" of the virus between wild ducks through water.

The proposed model is composed of three sub-models as summarised in Figure 3.16. More precisely, it consists of: a representation of the wild ducks including the evolution of the epidemic within, a representation of the environment focusing of the virus dynamic within and a representation of the season which force variation in terms of temperature (affecting virus persistence) and bird density (population size).

The first sub-models is a compartmental model representing the whole wild duck population. Its role is to represent the evolution of the disease within the population and allow viral interactions with the environment. The authors chose to decompose the population in five compartments as a "SEIIR" model. It is well-known that duck contaminated by avian influenza viruses are showing no clinical signs for a short period of time (exposed status) then they excrete the virus and show clinical signs and possibly mortality (infected status) and finally they recover or die (recovered status). More precisely, the LP avian influenza (LPAI) model has only one infected status (I_2 in Figure 3.16) whereas the HPAI model has effectively 2 infected statuses. Such a difference exists because HPAI induce a high mortality and a stronger viral excretion represented by status I_1.

The second sub-model represents the viral dynamic in a pond. It has only one dynamic, it decrements its only attribute: the viral particle concentration. The depletion rate is parameterised with the temperature and a constant determined through _in vitro_ experiments. More precisely, the depletion rate (k) is computed as: $k = k_0 \ e^{a(T-T0)/T}$. k_0 and a are parameters evaluated in laboratory and given in a paper by Stallknecht **[Sta90]** while T is the actual temperature and T_0 the 0°C temperature, both expressed in kelvin degree.

The Figure 3.16 shows the whole conceptual models (population representation, classical compartmental model), the pond (or water environment) and their relationships. Initially, an individual[13] is part of the susceptible compartment, it might it get infected through ingestion of viral particles in water (depending on the concentration of virus and a probability). In such case, it will become exposed and then infected. If the considered virus is a LPAI it will not go through I_1 but directly to I_2, conversely if it is a HPAI virus it will go through I_1 first. During this stage, the bird can also die (according to a probability μ). Finally, the individual is recovered though an exposed individuals does not experience infected status necessarily (according to a probability _(1-x)v)_.

---

13 I detail the evolution of an individual in a clarification purpose but a compartmental model represents proportions of the population
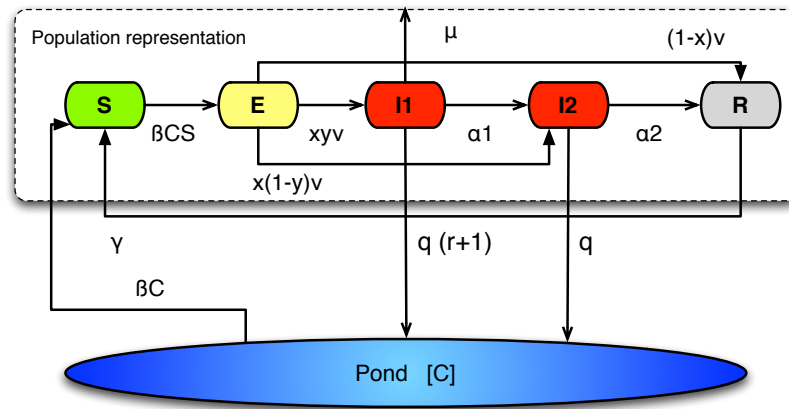
Fig3.16 Diffusion of the avian influenza among a population and through water (adapted from **[Ren07]**)

- *Discussion*

According to the authors, this model shows that the LPAI virus is likely to persist inter-annually (e.g. 2 years). In addition, the disease introduction date has an important impact on the infection risk while the disease prevalence is (surprisingly) less important. Concerning HPAI, whatever the combination of parameters are (mortality or none, introduction date, population renewal) the risk of infection is low according to the authors. In a second time, the authors study the influence of the renewal of the population (introduction of susceptible individuals). They claim that even in such conditions, the virus can persist at a low level.

This model provides a characterisation of the interactions between the watered environment and a wild ducks population. In addition, it proposes a parameterised formula to compute the virus depletion rate in water. The authors also claim that a specific compartmental model, SEIIR, is more appropriate to represent the HPAI effects on ducks. Unfortunately, this model is a compartment based one and cannot be re-used directly in an agent-based model. Moreover, it is difficult to integrate other dynamic into it. Similarly to the previous model, the reference system it considers is relatively different to the Gamavi one. The methodological aspects of the model are described through data collection consideration only.

2.2.6. Conclusion

Many more questions related to avian influenza diffusion have been investigating. For instance, the influence of the vaccination is assessed in **[Wal10, Mag10b]**, dynamic within live bird markets **[Iau07,Fou10]**, spread of the virus through wild birds **[Leb09, Gai10]** , etc.

The described models investigate several important mechanisms related to my application: propagation of the disease through trading network, virus persistence in the environment and influence of the production dynamics. The first model **[Kim08]** considers a larger scale reference system than the Gamavi one and investigates the effects of parameters instead of providing a way to compute them. The second model **[Mag10a]** proposes a description of one transmission mechanism: contacts at markets which are connected to each other through poultry traders. This model focuses on the commercial network at the inter-commune level while the subsequent one **[Pay09]** considers this network within the commune and introduce different stakeholders.

The two last models **[Roc09]** and **[Ren07]** investigates another transmission route: the indirect one through the water environment. Both of them show that the environment is, indeed, another transmission road and is likely to be important in the persistence mechanism of the avian influenza virus. In addition, **[Ren07]** provides a way to compute virus depletion in watered environment.

Integrating all these concerns together would certainly enhance the understanding of how all these mechanisms interact and produce the global dynamic. Nevertheless, coupling all these models together is probably an impossible task due to the diversity of parameters, level of representations of the different

97

models and unknown interactions between them. If the reference system is considered at a high resolution scale (the commune or even higher), it is possible to focus on local dynamics in order to represent the global dynamic. Unfortunately, the problem of the choice of the formalism remains: Which type of model would be able to integrate such diversity of concerns? It is exactly the point of the World models and the model type chosen for them is the agent-based one as described in Chapter 2. Thus I present, in the next section how it is possible to attain such objective.

2.3. Using these information to create a world model

I state in Section 2.1 what are the important elements of the reference system to consider according to domain experts and their translations from a modeller point of view. The entities envisioned are poultry, humans, other domestic animals and natural environment. Their interactions and specific dynamics, poultry trade and production in particular, must be represented. Finally, the selected time scale is 2 years while the spatial one is of the province size, more or less 100km. In the previous Section I review several models that propose representations for the trade dynamics and virus persistence in the environment while the production dynamics have been textually (or informally) described. Nevertheless, many issues are still pending, especially the consideration of individuals and the integration of the different representations. In the current Section, I demonstrate the relevancy of the world model representation for the Gamavi project.

2.3.1. Representing the various dynamics

Section 2.2 presents a few models representing the trading dynamics and two models of the virus survival in water environment. Integrating trading, production and epidemiological (including virus persistence in the environment) dynamics is not a straightforward process, especially due to their different formalisms. Coupling the represent dynamics as an equation system would certainly prove to result in an intractable model (see the last model of Chapter 1, Section 3.2.3, for an example).

A descriptive representation of the reference system entities would ease the simultaneous representation of both dynamics. Each entity represented with its meaningful characteristics, its dynamics and its interactions with other represented entities in a similar fashion as the real entities. The most adequate formalism is certainly the agent-based model. For instance, it would be easy to consider a duck agent having daily interactions with other duck agents and the environment or being "sold" to a trader agent and processed within the production chain.

In addition, ABMs do not require necessarily a formal mechanisms to couple entities or dynamics taking place at different scales and resolution. The process needs only to be rigorous and can be conducted in a conceptually simple way by defining each interaction between two entities, one by one. For instance, a duck agent can have epidemiological interaction with the environment on an hourly basis and be sold on a monthly based to a trader. The only moment it is an issue is in the implemented model where the scheduler is responsible to execute the simulation as defined in the conceptual model.

2.3.2. Representing the environment

The environment representation has two main purposes in general: locate the agents and interact with them (environment mediated interaction is actually a specific type of agent-environment interaction). In the current objectives, poultry can be located in various building, bare soil and natural environment which constrain the poultry movement. The considered agent-environment interactions relate only to epidemiological process, namely disease transmission. In practice, it means the environment representation has to allow: excretion of the virus onto it, stock of the virus and infection of susceptible poultry. The second process is the persistence mechanism which implies that this representation has its own dynamic to correctly represent the depletion of the virus according to environmental parameters. Thus it requires that the environment representation has a topology, a set of characteristics, a dynamics and allow interaction with it. A very efficient tools to represent the environment topology is the Geographical Information System (GIS). The other requirements can be addressed by defining the environment representation as a set of agents which are specific due to their location-related services (provided by the GIS).

2.3.3.Enabling hypotheses assessment

In the present application, thematicians require to be able to test different hypotheses and use alternative representation. Amending an element of the model is likely to impact the representation of other reference system elements. It is not an issue for agent-based models thanks to their lineage to object-oriented programming (OOP). If an hypothesis changes only the internal dynamic or characteristic of an agent, only this agent is impacted. On the contrary, an hypothesis may question the representation of an entity, for instance poultry are not represented individually but as a group (or flock). This is likely to impact the whole model in many formalism. In an agent-based model the agent is redefined and other agents which interact with it need to be adapted but only marginally in the general case. Moreover, it is facilitated by the modelling cycle and conceptual tools introduced in Chapter 2.

Another advantage of the agent-based models over many formalism is that they allow a descriptive representation of the reference system, they do not rely on its characterisation through aggregated indicators. Nevertheless, they do not enforce themselves such descriptive representation which is the purpose of the modelling cycle introduced in Chapter 2.

Integrating the representation of elements acting at various scale is not an issue for the implemented model thanks to the OOP lineage of agent-based models with the adequate methodological tools and concepts. For instance, one may rely on the focal systems as defined in **[Aum07]** (see Chapter 2 Section 2.3.2) to link in a simple but rigorous manner different representation levels.

2.3.4.Integrating field data

Using "raw" field data is another strong requirement of the domain experts for this application in particular. Field data are always used in a model but they are usually processed first. Several parameters can be aggregated into one parameter, like the R_0 in Epidemiology. Another approach is to measure the values of a parameter over the population and use simply its mean value though its statistical distribution can be also accounted for (*i.e.* variance, skewness, kurtosis). The most radical orientation is to use the measured value on a one-to-one basis. Each set of parameters values is used by one model constituent to represent the specific reference system entity.

In general, the more extensive are the data the more complicated it is to integrate them into a model. For instance, using the mean and variance of a parameter adds more variable to an equation system, complicating its resolution. Moreover, using data on a way one-to-one basis is usually even more complicated if not impossible except if the model integrates a descriptive representation of the constituent of the system.

In an agent-based model, having extensive data is not a strong issue as pre-processing is not necessarily required in contrary to many other meta-model. It is possible to integrate "raw" field-data by attributing a value to the parameter from the field data on a one-to-one basis (one agent uses the values of one individuals of the reference system). While the aggregated situation does not rise any problem (the same value of parameter is used for all the agent), the intermediate situation requires some processing. In practice, it requires to generate individual value from the parameter distribution parameters, namely mean, variance and so on which results in a generated population complying with the known characteristics.

ABM do not constrain how a parameter can be represented in the model: aggregated, statistical distribution-based or one-to-one valuations can be used. In the current application context it means that different levels of data aggregation can be used simultaneously: The R_0 parameter may be used at the global level, the age of poultry can be described by a distribution and each farm can be instantiated according to field data on a one-to-one basis.

Using spatial field data is another issue which is treated in a similar way. Spatially variable parameters are at best rigidly discretised. In an agent-based model it is possible to use as accurately spatially defined values as the data collection furnish it thanks to the use of a GIS.

All these concerns and requirements lead to the need of representing the Gamavi's reference system as a world model. Indeed, world model characteristics, the associated modelling cycle and support tools, introduced in Chapter 2, address such requirements.

3. Conclusion

In this Chapter, I start by presenting the general application context which is summarised in the research question of the epidemiologists: what are the mechanisms explaining the persistence and propagation of the avian influenza virus at a local scale in North Vietnam. I first introduce the influenza disease which has a long history of epidemic and pandemic in the human population. Nowadays, this disease and more specifically the avian influenza remains a threat for the world population.

The origin of this possible large epidemic, if not pandemic, might be in Vietnam as it has a high density of poultry and humans with limited bio-control measures which are good conditions for the virus to become more easily communicable among humans. To assess the risk, it is important to know how the disease is circulating among domestic poultry which is one of the objective of the Gripavi project. In this context, I collaborated with different domain experts in a subproject called Gamavi. The objective of this project is to provide domain experts with a model which allow them to evaluate their hypotheses about local persistence and propagation of avian influenza in North Vietnam.

I review meaningful examples of models that characterise various aspects of the avian influenza, keeping in mind that these models are possible sources of data to integrate in my model. In contrary, these models provides little to no information on how they were designed. Consequently, they cannot be used as a reference for the design of the Gamavi mode. Therefore, I raise general concern on how to build the application model and presented the collaboration, its thematic objective and the reference system.

The next chapter presents the application of this thesis, the modelling process that occurred to investigate the epidemiologists' hypotheses. The application model was designed after the methodological tools and concepts introduced in the previous chapter.

Chapter IV : The GAMAVI model

The previous chapter introduced the collaboration context: the avian influenza disease and its causing agent the H5N1. This disease arouses a lot of interest due to the important impact on the economy of developing countries. Moreover, the H5N1 virus is a "good candidate" for the next pandemic. For these reasons, many research programs have been initiated, like the CIRAD's Gripavi project whose of objective is to study the ecology of the disease, assess the propagation risk and enhance stakeholders knowledge (Chapter 3). An element of this project was the GAMAVI project which started in early 2007 until late 2010. The objective of GAMAVI was to create collaboratively models to evaluate hypotheses about avian influenza local persistence and propagation in North Vietnam.

The conception of the GAMAVI model followed the KIMONO modelling cycle introduced in Chapter 2. (and detailed in Chapter 5) The reference system has been informally defined after the research question. This description has been detailed in a domain model constituted of an ontology, textual clarification and sub-models expressed in domain experts formalism as presented in Section 1. After this domain, the modelling cycle iterates over a sequence of design, operational and implemented models called *iterative modelling* which is presented in Section 2. Each iteration of this sequence has been oriented thanks to the definition of the sub-models. In addition, several iterations forced the model contributors to re-define the domain model due to issues that appeared during these iterations (lack of data, unforeseen incoherency, etc.). In Section 3, I present the final implemented and I highlights the key methodological elements used during the development of this application.

1. Initial steps

In this section I remind the research question and the reference system description introduced in Chapter 3. I introduce the domain model in terms of ontology, documentation and sub-models. Even though the definitions of these elements did not evolved frequently, some amendment to issues that emerged during the iterative modelling had to be made. Consequently, the initial research question, the reference system description and the domain model ontology conceived initially and their final counterpart are presented in this section.

1.1. Introduction

Dr. Desvaux was the coordinator of the Gripavi project in Vietnam. On behalf of the project, she was the person at the origin of the GAMAVI model whereas I acted as the modeller and the implementer. She was also the most involved domain expert though several other experts intervened in the project according to their area of expertise. Dr. Bicout is an epidemiologist from the TIMC-IMAG laboratory[14] and worked on the virus persistence in the environment. Ms. A. Payne conducted a study of the commercial system under the supervision of Dr. Desvaux and Dr. D. Bicout. Mr. G. Fournié was a PhD student in epidemiology who studied the commercial system as well but more specifically the markets dynamics. Several veterinarians of NIVR, including Dr. Dung, a virologist who collaborated more closely with Dr. Desvaux. Finally, Dr. R. Duboz, who was involved in the Gripavi, and Pr. A. Drogoul provided support as senior modellers and implementers.

1.2. Research question and reference system description

The initial research question of the GAMAVI model was expressed as two objectives in the collaboration agreement (Annex 1, in french): "*The first objective of the epidemiological study is to identify natural environmental and human factors related to the introduction and dissemination of the avian influenza in North Vietnam. The second objective is to characterise the the flow of avian influenza virus between domestic and commensal animal compartments.*". The research question evolved all along the modelling process and its final formulation is :"*Can the interplays between the poultry*

---

[14] http://www-timc.imag.fr/

Chapter IV : The GAMAVI model

The previous chapter introduced the collaboration context: the avian influenza disease and its causing agent the H5N1. This disease arouses a lot of interest due to the important impact on the economy of developing countries. Moreover, the H5N1 virus is a "good candidate" for the next pandemic. For these reasons, many research programs have been initiated, like the CIRAD's Gripavi project whose of objective is to study the ecology of the disease, assess the propagation risk and enhance stakeholders knowledge (Chapter 3). An element of this project was the GAMAVI project which started in early 2007 until late 2010. The objective of GAMAVI was to create collaboratively models to evaluate hypotheses about avian influenza local persistence and propagation in North Vietnam.

The conception of the GAMAVI model followed the KIMONO modelling cycle introduced in Chapter 2. (and detailed in Chapter 5) The reference system has been informally defined after the research question. This description has been detailed in a domain model constituted of an ontology, textual clarification and sub-models expressed in domain experts formalism as presented in Section 1. After this domain, the modelling cycle iterates over a sequence of design, operational and implemented models called *iterative modelling* which is presented in Section 2. Each iteration of this sequence has been oriented thanks to the definition of the sub-models. In addition, several iterations forced the model contributors to re-define the domain model due to issues that appeared during these iterations (lack of data, unforeseen incoherency, etc.). In Section 3, I present the final implemented and I highlights the key methodological elements used during the development of this application.

1. Initial steps

In this section I remind the research question and the reference system description introduced in Chapter 3. I introduce the domain model in terms of ontology, documentation and sub-models. Even though the definitions of these elements did not evolved frequently, some amendment to issues that emerged during the iterative modelling had to be made. Consequently, the initial research question, the reference system description and the domain model ontology conceived initially and their final counterpart are presented in this section.

1.1. Introduction

Dr. Desvaux was the coordinator of the Gripavi project in Vietnam. On behalf of the project, she was the person at the origin of the GAMAVI model whereas I acted as the modeller and the implementer. She was also the most involved domain expert though several other experts intervened in the project according to their area of expertise. Dr. Bicout is an epidemiologist from the TIMC-IMAG laboratory[14] and worked on the virus persistence in the environment. Ms. A. Payne conducted a study of the commercial system under the supervision of Dr. Desvaux and Dr. D. Bicout. Mr. G. Fournié was a PhD student in epidemiology who studied the commercial system as well but more specifically the markets dynamics. Several veterinarians of NIVR, including Dr. Dung, a virologist who collaborated more closely with Dr. Desvaux. Finally, Dr. R. Duboz, who was involved in the Gripavi, and Pr. A. Drogoul provided support as senior modellers and implementers.

1.2. Research question and reference system description

The initial research question of the GAMAVI model was expressed as two objectives in the collaboration agreement (Annex 1, in french): "*The first objective of the epidemiological study is to identify natural environmental and human factors related to the introduction and dissemination of the avian influenza in North Vietnam. The second objective is to characterise the the flow of avian influenza virus between domestic and commensal animal compartments.*". The research question evolved all along the modelling process and its final formulation is :"*Can the interplays between the poultry*

---

[14] http://www-timc.imag.fr/

Chapter IV : The GAMAVI model

The previous chapter introduced the collaboration context: the avian influenza disease and its causing agent the H5N1. This disease arouses a lot of interest due to the important impact on the economy of developing countries. Moreover, the H5N1 virus is a "good candidate" for the next pandemic. For these reasons, many research programs have been initiated, like the CIRAD's Gripavi project whose of objective is to study the ecology of the disease, assess the propagation risk and enhance stakeholders knowledge (Chapter 3). An element of this project was the GAMAVI project which started in early 2007 until late 2010. The objective of GAMAVI was to create collaboratively models to evaluate hypotheses about avian influenza local persistence and propagation in North Vietnam.

The conception of the GAMAVI model followed the KIMONO modelling cycle introduced in Chapter 2. (and detailed in Chapter 5) The reference system has been informally defined after the research question. This description has been detailed in a domain model constituted of an ontology, textual clarification and sub-models expressed in domain experts formalism as presented in Section 1. After this domain, the modelling cycle iterates over a sequence of design, operational and implemented models called *iterative modelling* which is presented in Section 2. Each iteration of this sequence has been oriented thanks to the definition of the sub-models. In addition, several iterations forced the model contributors to re-define the domain model due to issues that appeared during these iterations (lack of data, unforeseen incoherency, etc.). In Section 3, I present the final implemented and I highlights the key methodological elements used during the development of this application.

1. Initial steps

In this section I remind the research question and the reference system description introduced in Chapter 3. I introduce the domain model in terms of ontology, documentation and sub-models. Even though the definitions of these elements did not evolved frequently, some amendment to issues that emerged during the iterative modelling had to be made. Consequently, the initial research question, the reference system description and the domain model ontology conceived initially and their final counterpart are presented in this section.

1.1. Introduction

Dr. Desvaux was the coordinator of the Gripavi project in Vietnam. On behalf of the project, she was the person at the origin of the GAMAVI model whereas I acted as the modeller and the implementer. She was also the most involved domain expert though several other experts intervened in the project according to their area of expertise. Dr. Bicout is an epidemiologist from the TIMC-IMAG laboratory[14] and worked on the virus persistence in the environment. Ms. A. Payne conducted a study of the commercial system under the supervision of Dr. Desvaux and Dr. D. Bicout. Mr. G. Fournié was a PhD student in epidemiology who studied the commercial system as well but more specifically the markets dynamics. Several veterinarians of NIVR, including Dr. Dung, a virologist who collaborated more closely with Dr. Desvaux. Finally, Dr. R. Duboz, who was involved in the Gripavi, and Pr. A. Drogoul provided support as senior modellers and implementers.

1.2. Research question and reference system description

The initial research question of the GAMAVI model was expressed as two objectives in the collaboration agreement (Annex 1, in french): "*The first objective of the epidemiological study is to identify natural environmental and human factors related to the introduction and dissemination of the avian influenza in North Vietnam. The second objective is to characterise the the flow of avian influenza virus between domestic and commensal animal compartments.*". The research question evolved all along the modelling process and its final formulation is :"*Can the interplays between the poultry*

---

[14] http://www-timc.imag.fr/

production dynamic, the natural and cultivated environment can explain the persistence and the propagation of the avian virus influenza within a village or a commune?".

Similarly, the description of the reference system given in Chapter 2 has evolved and the last description is as follows:

The reference system is a vietnamese village of the Red River Delta (from Ha Tay or Bac Giang provinces). The environment contains the inner village (building and open field) and a 1 km belt of culture and water around it. The considered entities are flocks (many types) and farms (backyard and semi-commercial). The epidemiological interactions are: individual pathobiology, direct transmission within and between flocks and environment mediated infection. Consequently, watered environment can sustain the virus which is parameterised by physical characteristics. The production dynamic is represented (but not the trading one): farms' flocks are renewed according to observed pattern. Flocks have varied daily dynamics according to their types. In addition, the considered hypotheses are: (1) the virus can propagate directly from flock to flock all year long, (2) the environment act as a long term reservoir, (3) repeating virus introduction are required (through trading dynamic).

The evolution of the research question and reference system description took place all along the modelling process. The most important change was the redefinition of the spatial scale (and resolution). Initially, the spatial extent of the system was the province where communes were interconnected through trading links. Data initially available were not relevant anymore due to the strong re-organisation of the commercial system after the avian influenza epidemics of 2005 **[Agr07]**. A study by A. Payne **[Pay09]** has been planned, addressing the defects of this commercial system data. Unfortunately, these data were insufficient to represent this system within the GAMAVI model. Consequently, we considered the village as the spatial extent of the reference system. Indeed, according to the domain experts, we considered the village as the atomic epidemiological unit because very limited trading occur between villages. It was also a better spatial scale because the interplay between the environment and the poultry dynamic occurs at this scale.

1.3. Domain model

Similarly to the research question and the reference system description, the domain model evolved during the whole modelling process. Consequently, I present in the current section the initial and the final domain models.

Domain experts, with my support as the modeller, represented the reference system with different ontologies. The early iteration of the ontology considered the province as the reference system. The final ontology is set at the village level because the trading dynamics is not considered anymore leaving no relevancy to the province as no other (considered) dynamic occurs at this level.

1.3.1. Elicitation of the initial domain model

The first produced ontology is specifically oriented by the domain experts habits and is presented in Figure 4.1. Indeed, this representation described the general poultry production system at the province level. Many interactions are accounted for, making the diagram almost inextricable. However, three levels can be distinguished: Poultry, trading-related entities and other entities that may be infective. The different types of poultry production constitute the core of the diagram: They interact with the environment (and other animals) and the commercial system. Despite their importance regarding the avian influenza epidemics, 3 entities can be set apart: veterinarian, consumer and the international trade. The *regional trade* is clearly out of the scope of this study (the province is the largest envisioned scale). *Consumer* is an epidemiological dead-end (it can be infected by the virus but cannot transmit it back to the system entities) and interactions with *Veterinarian*, represented through a vaccination process, can be considered by simple parameters of the model.
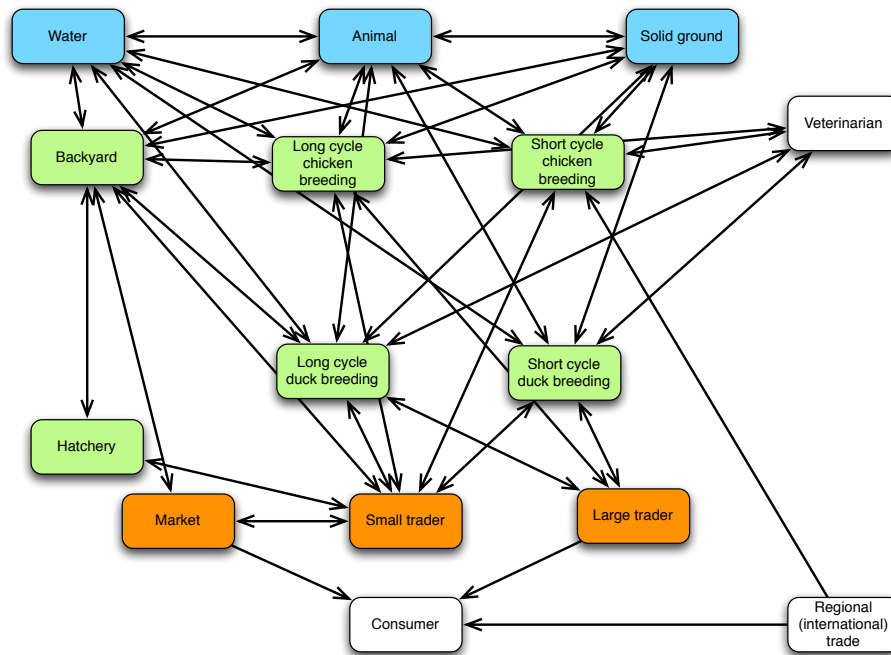
Fig.4.1 Entities of the reference system and their interaction (adapted from Dr. Desvaux and Dr. Duboz)

Based on the previous ontology, I proposed an hierarchy of entities as described in Figure 4.2, focused on three main types of entities: human, animal and commercial system entities (including farms). We considered more types of human in this ontology: Breeder, neighbour and para-veterinarian. Several types of animal have been distinguished and classified: wild and peri-domestic poultry and some domestic animals (that may propagate the virus mechanically). Finally, we considered the poultry transports as entities because they can spread (mechanically) the virus.
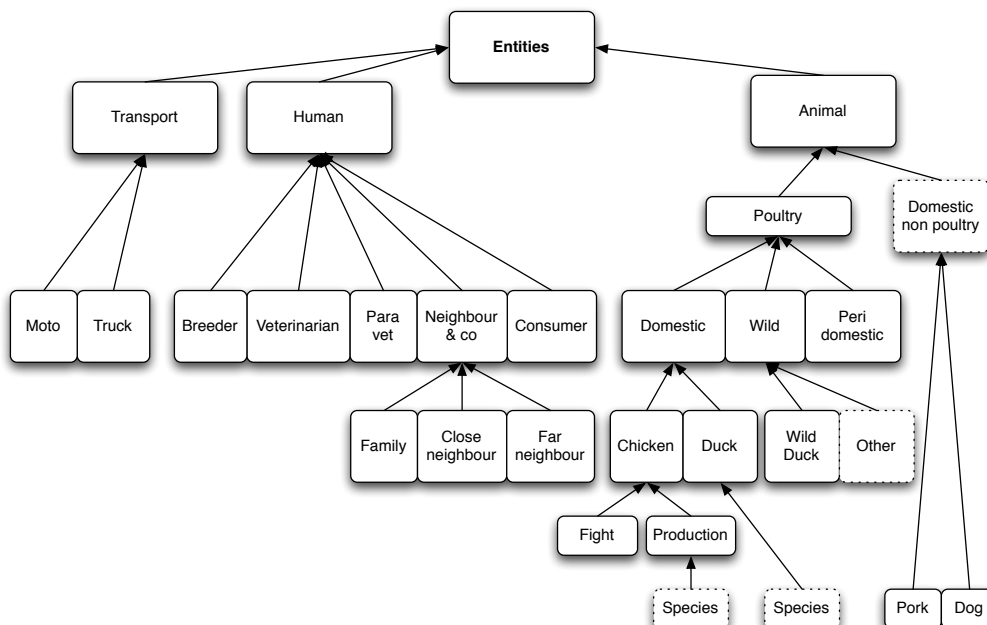
Fig.4.2 Entities as initially envisioned (19/07/07)

Using the first ontology and the entity hierarchy, I proposed a new organisation as illustrated in Figure 4.3. It decomposes the representation of the reference system in four administrative levels. There is no entity acting at the district or commune level, thus only the province (in orange) and the village (in blue) are actually considered. Finally, the different types of environment are represented (in green). As one may see, several

entities have been discarded, such as domestic animal, wild birds, etc., because no study characterises their (potential) actions within the avian influenza representation context.
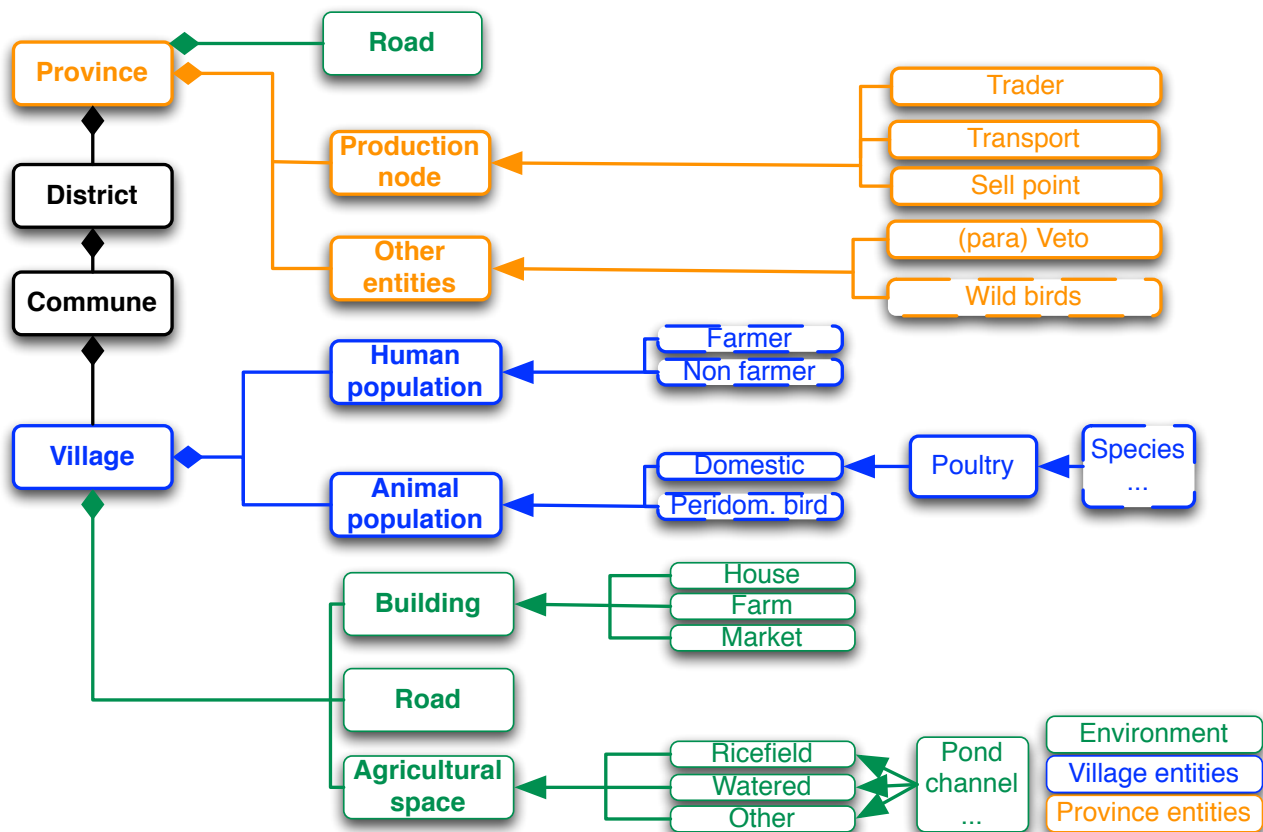
Fig.4.3 Hierarchies of entities

1.3.2. Final domain model

The final domain model results from numerous modelling iterations detailed in Section 2. For instance, the behaviour of humans within our reference system appeared too badly characterised as soon we considered them. In contrary, the representation of the trading network was considered unachievable after the review of a field study results.

- **Ontology**

The final domain model abandons any consideration of the province due to the lack of adequate data concerning the trading dynamic. Similarly, no data could be found to represent the dynamic of humans or peri-domestic birds. Consequently, the final domain, illustrated in Figure 4.4, represents a village. As show in this diagram a village includes two types of entities: Farm and flock. The overall environment is decomposed in spatial units ("environmental lot" in the figure) characterised by their type. For each of these three elements, a typology has been defined. The definition of these typologies has been conducted through several modelling iterations.

Fig.4.4 Village entities and their interactions

- **Documentation**

The documentation of the domain model includes a description of the previous ontology, an updated set of hypotheses and a listing of the sub-models that can be extracted from the domain model ontology.

Ontology description

In order to document the domain model ontology, we re-used the O.D.D. *Overview* section structure (Chapter 3): (1) *Purpose*, (2) *Entities, states variables and scales* and (3) *Process overview and scheduling*. This documentation is completed by the hypotheses list and the set of sub-models extracted from the domain model.

The purpose of the model is already stated by the research question and the reference system description (Section 1.2). The two other sections (2) and (3) are detailed within Annex 2, thus I only provide a summary here:

- *Entities, states variables and scales :*The variables characterising a flock are: Individuals, location, species, type and farm. A farm is characterised by its: flocks and location. An environment lot is characterised by its: type, location and shape. In addition, a pond or watered rice-field also include a volume and a virus concentration. All these entities are located within a village which is characterise by its topology and its farms and flocks population. The spatial scale depends on the village which is represented (as we used GIS field data) but is generally about a few kilometres wide whereas the spatial resolution is 10m. The temporal scale is 2 years and the time step is 1 hour.

- *Process overview and scheduling:* Flocks have two main processes: Movement and infection. The non-confined flocks can go out during the day, their movement are computed at each time step. If a flock include one or several infected individuals, the flock excretes a computed level of viral particle in the environment at each time step. Similarly, the flock collects viral particle from the environment but the computation of the infection is considered on a daily basis. Farms can renew their flock on a daily basis. The length between depends on the flock type: From 45 days to 2 years (See Annex 3 for the data source). The potential infection between flock of the same farm or among farms, a farm with an infected flock computes on a daily the basis if such infection occurs.Ponds and watered environment compute their virus concentration daily by accounting the flocks' excretions and the depletion of the previous viral concentration.At the global level, the seasonalities are accounted for a daily basis and affect: flock size and renewal period and temperature of water.

Hypotheses

High level hypotheses have been expressed in the reference system description. Thanks to the detail provided in the domain model, it is possible to precise them here. These hypotheses are summarised as follows:

- Can the poultry population with its production dynamics sustain the virus for a year?
    - the composition of the poultry population has an influence
    - more ducks means a more silent circulation of the virus
    - longer production flocks favour the circulation of the virus
    - vaccination can increase the silent circulation of the virus
- Does the addition of commercial dynamics affect virus persistence and propagation?
    - local commercial system (commune markets)
    - province commercial system (commune and district markets, traders, etc.)
- Is the environment required to model the virus persistence and propagation?
    - persistence of the virus in ponds during winter is sufficient
    - persistence all years long is possible

For each hypothesis, a high level evaluation scenario is defined. Such a scenario includes the required entities (and environment) to be represented, the parameters that may vary (to evaluate their influence), those considered fixed and the expected outputs. An early scenario example produced by domain experts (after the definition of the domain model) have been included in Annex 4.

Sub-models

In order, to facilitate the data assessment and collection but also to structure the iterative modelling, we decomposed the domain model in sub-models. The initial domain model included the commercial system, the human dynamics, etc., but the final domain model includes only the following ones:

- Farm
- Flock
- Spatial environment
- Village population
- Epidemiological model (individual and environment)
- Epidemiological interactions

1.4. Global data assessment

During this collaboration, the modelling process and the data collection were initiated simultaneously. Initially a field study (the commodity chain survey), several collaborations (RIVERS project, **[Pay09]**, **[Ren07]**) and literature review were envisioned. The domain model evolved due to the iterative modelling and the required data evolved as well. Especially, modelling the environment required more data than initially envisioned whereas some data considered to be collectible were finally not available.

1.4.1. Commodity chain survey

Several field studies were conducted within the *Gripavi* project but one of them was used as the main data source: "The commodity chain survey" which resulted in **[Des09]** and annex 3. The aim of this field study was to identify factors related to humans, to the environment or related to the poultry production system. In practice, it resulted in a list of all farms (location, production sector and main production type) of the studied area but it also aggregates farms from the village level to the province level. It also categorised production types using statistical data, in terms of species, outcomes, flock size and (breeding) duration in the different communes. Concretely, this study provided us the required information to model flocks and farms.

An interesting output of this field study is a precise the definition of production prototypes based on species (duck, muscovy and chicken) and outlets (egg, meat, etc.). For each pair species-outlets, it defines the proportion per housing (i.e. closed building, free-range, etc.), the duration of production and the flock size. It also characterises communes and villages according to the proportion of outlets production and to the land-use of the environment. This characterisation helped to establish five villages prototypes..

1.4.2. Other data sources

The commodity chain survey was complemented by a large literature review (conducted by Dr. Desvaux and me), other surveys and collaboration with other projects. A specific publication, **[Agr07]**, was of strong interest as it extensively describes the vietnamese poultry production system (i.e. commercial network and production practices). The description of the commercial network provided by this publication could not be used because this study had been conducted during the reorganisation of the commercial network that occurred after the 2003-2005 avian influenza epidemic in vietnam. Consequently, domain experts planned a complementary study **[Pay09]**. Unfortunately, the results of this study were insufficient to produce a satisfactory representation of the commercial system in the GAMAVI context. Despite the detailed description of the commercial network, no evaluation of the epidemic flows within it was proposed.

Representation of the environment topology would rely on two sources of data. The first one would be a geo-referenced representation of the North Vietnam communes, villages, roads and land covers at a sufficient resolution. The second one was the topologies of village collected during the "commodity chain survey". Finally, these data were not collected and thus, we had to rely on literature review (to have a prototypical topology) **[Gou37, Fon99, Tes03]** and publicly available satellite images of the village with limited details.

The representation of the disease-related (persistence, transmission, etc.) characteristic of the environment should have come mainly from the RIVERS project (Chapter 3). Unfortunately, the conducted experiments were not insightful in time.The valuation of the environmental parameters would come either from field study, literature review and/or data from local institutes.

Finally, the infectious model of the individuals would re-used data from literature review, especially **[Ren07, Roc09]** and a collaboration with R. J. Magalhães **[Mag10a, Mag10b]**.

2. Iterative modelling

After the definition of the domain model and the general data assessment we focused on the different sub-models in order to obtain a complete implemented model. Each sub-model iteration was decomposed in different steps: specific data assessment, design model, operational and finally implemented model. Each of iteration improved the models from the previous iteration. For instance, the design, operational and implemented models of iteration *i+1* re-used the respective models from iteration *i*. In addition, we decided to create first a model where no epidemiological dynamics was accounted for. We used this model as a basis on which the epidemiological dynamics was integrated and hypotheses evaluated afterward. It allowed us to focus on the representation on well-known process that can be evaluated through domain experts macro-knowledge.

This section is dedicated to the description of different sub-models and the justification of modelling choices made I, as the modeller and the implementer, made in collaboration with domain experts. The order according to which sub-models are described in this section does not correspond exactly to the chronological order followed in practice. Indeed, the development of several sub-models had to be stopped or re-conducted due to lack of data. Figure 4.5 presents the chronological order and the reasons of regressions.
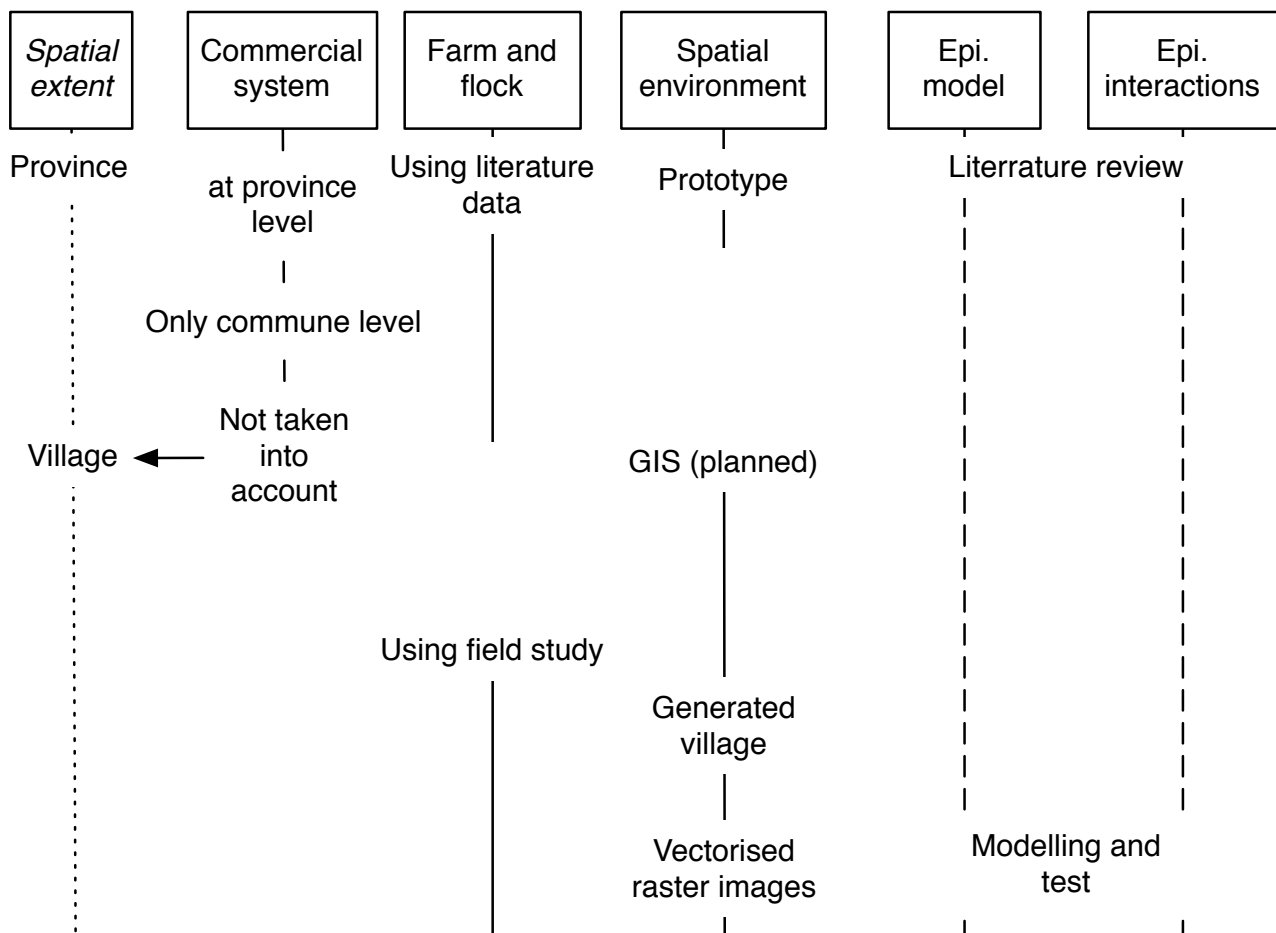
Fig.4.5 Chronology of the sub-models (iterations) development

2.1. Commercial system

The first iteration I present is the representation of the commercial system proposed in **[Pay09]**. This sub-model has been considered in the initial model but had to be abandoned due to the incoherency between the available data and the objective of the GAMAVI model. It is interesting to introduce this sub-model to show that the modelling method we used can cope with unforeseen lack of data. Obviously, such a lack of data leads to a redefinition of a the domain model and the associated hypotheses.

2.1.1. Payne network model(s): province and commune

As the GAMAVI reference system, the field study was conducted on the Ha Tay and Bac Giang province. After analysing the raw data, the authors decomposed the commercial system into two sub-networks according to their scale: the commune and the province networks depicted in Figure 4.6. Within a commune, most of the production is sold to traders who sell it to final consumers on local markets. At the province level, assemblers sell the poultry production to province markets, which then sell it to final consumers. Retailers selling directly to consumers also exist. The two networks are inter-connected through farms (one farm can sell to different individuals) and some small-scale traders sell most of their poultry to larger retailers and wholesale traders.

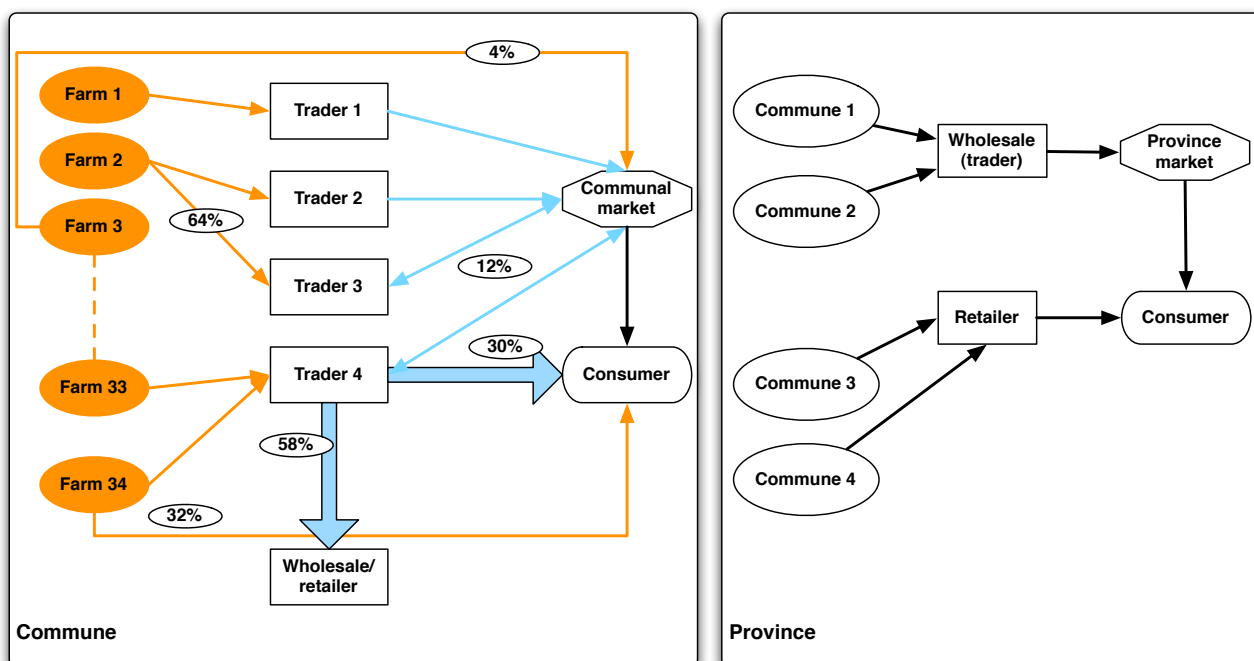Fig.4.6 Poultry commercial system **[Pay09]**

2.1.2.Limitations

In **[Pay09]**, the authors focused only on one poultry production sector: The semi-industrial one. In contrary, the GAMAVI domain model considers "traditional" sector (small scale poultry producers) as well. Moreover, the authors focused the meat outlet production only because it includes movements of living birds, which implies a much more important infectious risk than eggs. Nevertheless, the GAMAVI domain model also takes into account the eggs production. The authors focused on the Bac Giang province because Ha Tay is a province adjacent to Hanoi and most of its production is consumed directly in the city. In contrary, the GAMAVI domain model have been conceived to be able to represent both province. Finally, this study does not provide any transmission probability for the commercial flows thus it prevents any proper modelling of the virus propagation within the commercial system.

2.1.3.Conclusion

The commercial system is not considered in our model because of the limitations previously expressed and that no other fitted data could be found in time. Consequently, the domain model had to be redefined and the main consequence was the change of spatial scale, from the province level to the village level. Indeed, the other considered dynamics can be completely represented at the village scale. Nevertheless, this study still produced results (seasonality of farms' sales) that we integrated as presented in Section 2.2.

2.2.Modelling the farm entity

The characterisation of farms and flocks are related to each other because a given type of farm have a limited set of possible flocks types. Consequently, the iterations that lead to the representation of both of these entities referenced each other. In this section, I present the farm characterisation and its dynamics whereas the flock modelling is considered in the next iteration described in Section 2.3.

2.2.1.Farm characterisation

The definition of our farm typology was made with respect to three definitions of the "production sector" provided by the GSO (General Statistical Office), the MARD (Ministry of Agriculture and Rural Development) and the FAO (Food and Agriculture Organisation, a UN agency). Based on these classifications, domain experts proposed their own typology **[Des08]** which was amended by the modeller/implementer.

109

- **MARD, GSO and FAO classifications**

The GSO focuses only on the livestock size of the farm. It separates the production in four sectors: (1) more than 2000 birds; (2) 150 to 2000 birds; (3) 40 to 150 birds; (4) less than 50 birds. This classification is not insightful as no farms characteristics are derived from this classification.

The MARD **[Mar06a, Mar06b]** classification considers four different types of poultry farms differentiating by the production practices. The "*village farming system*" which corresponds to a small scale farm, owned by an household and where birds are free-ranged (they can wander out of the farm or house). The "*duck transhumant farming system*" is specific to South Vietnam; ducks are brought to newly harvested rice-field everyday (which occurs for a limited time). The *semi-industrial farming system* represents farms of 200 to 500 birds that are kept indoor. The *industrial chicken farming system* represents very large farms (up to dozen of thousands chickens) where chickens are kept indoor and high bio-security controls are in place. These controls prevent any kind of epidemiological with the outside world.

The FAO classification is commonly used at the international level which allows comparison of different poultry production sectors organisation. It is based on the production organisation, which depends on several characteristics of farms.

- **Sector 1**: very large production of one outlet, less than 100,000 (usually a few dozens of thousands). This sector, called the **industrial sector**, represents around 20% of the chicken production (but only 0.1% of the overall number of farms). Each farm of this class produces only one type of outlet (e.g. eggs or for meat). The organisation of such farm is very rigorous and bio-security is very high, similarly to western country. The lone outside contact is the sales of poultry in supermarkets of big cities.

- **Sector 2**: smaller production of multiples outlets, less than 10,000 (usually a few thousands). This sector is also part of the **semi-industrial sector** as considered by the MARD. Different types of outlet or different species can be produced in the same farm. The production is band-based (a group of poultry bred altogether), and each band is separated from others. The organisation of such farms is less rigorous and bio-security are less important. Such farms always have an indoor production.

- **Sector 3**: smaller production of multiples outlets, less than 1000. This sector is also a part of the **semi-industrial sector**. The whole semi-industrial sector represents around 25-30% of the chicken production, 10-15% of the duck production and 10-15% of production of vietnamese farms. The differences (in comparison to Sector 2) are the size, the organisation and the bio-security level. In addition, production practices in such farms is not necessarily indoor (e.g. free-range duck).

- **Sector 4**: informal, not "band-based" and very small production, less than 100 (usually a few dozens). This sector is also called the **traditional sector**. It represents the majority of the farms (65%) but only 15-20% of the duck production. This sector is very informal and most families (90%) own few poultry and thus can be considered as belonging to this sector. There is no bio-security measures: no vaccination, different species are mixed and there is no veterinarian check. The production is not band-based in contrary to other sectors.

- **Domain experts classification**

According to the domain experts, the previously described classifications are not well adapted to characterise expressively enough the North Vietnam poultry production. Some distinctions could not be considered while some arbitrary distinctions were made (e.g. the size of the livestock is not discriminating). Thus, domain experts proposed a new classification based on only three sectors:

- **Sector 1**: Fully integrated chicken farms. It corresponds quite directly to the FAO's *sector 1* or the MARD's "Industrial chicken farming system";

- **Sector 2**: Semi-commercial farming. It corresponds to the FAO's *sectors 2 and 3* or the MARD's *Semi-industrial farming system* and *Duck transhumant farming system*. The size of the livestock can go up to few thousands.

- **Sector 3**: Backyard farming. It corresponds to FAO's *sector 4* or the MARD's *village farming system*. The size of the livestock can go up to an hundred.

• *Relevant sectors to consider*

In the GAMAVI context, fully integrated chicken farms (i.e. farms belonging to Sector 1) are clearly beyond the scope of our study. Indeed, in such farms, an AIV infection would be of large consequences. Veterinarian would detect it very early, especially due to the high bio-security level. Consequently, only farms from *semi-commercial* and *backyard* sectors are considered.

- **Semi-commercial farming**: Each farm has one or several bands of possibly different species. Each band is bred separately. Such farms can produce ducks and/or chickens for meat and/or egg outlet (one band for one specific outlet). These farms can contain from few hundreds to few thousands birds.

- **Backyard farming**: Farms belonging to this sector are usually households producing poultry eggs and meat, mainly for their own consumption. Birds are kept altogether and can (usually) wander freely around. Such farms may contain from few birds to almost 50 birds.

• *Conclusion*

Thanks to literature review, to field studies, and to the modelling process itself (at a conceptual level), domain experts proposed a farm typology. Indeed, through discussions and diagrams (ontology, UML class, etc.) about the elements constituting a farm, domain experts ideas evolved. They realised that the sector to which farms belongs is not of such an important matter according to our (epidemiological) concerns. Certainly, it specifies many aspects economically (business or household oriented) and commercially (with whom it interact). However, from an epidemiological point of view, a farm is defined by its constituting flocks. The next section shows the variety of flocks and highlights the fact that a simple classification between semi-industrial and backyard farms is not sufficient to represent all distinct poultry production dynamics in the model.

2.2.2. Farm dynamic

The farm entity has two dynamics: Poultry population dynamics (flock renewal) and avian influenza transmission among flocks. The poultry population dynamics can be characterised by the production duration, the seasonality (the variation of the volume production according to the season **[Pay09]**), the time required to complete a sale (in one time or over several days) and by the building quarantine (the time elapsed between the sale of a band and the arrival of the next one). These parameters depend on the poultry type (duration-species) according to the data provided by domain experts after the field data collection **[Des09]**. At the opposite, the production volume of backyard farms does not vary. Indeed, we consider that the renewal is continuous (a new bird is added when one is discarded) because of the self-consumption of this production.

The transmission of the avian influenza virus from one flock to another is managed by the farm entity. It allowed us to take into account the effect of humans (who can mechanically carry the virus easily) without representing them. For instance, a breeder has contacts with his/her flocks everyday or the para-veterinarians can visit several farms within a day. Concretely, we defined a probability of transmission within a farm and among farms. It has the benefit to be adapted to existing data **[Mag10a, Mag10b]**.

2.3. Modelling the flock entity

2.3.1. Flock characterisation

The definition of the flock typology is crucial according to domain experts while the definition of flock behaviour is crucial according to the modeller. Hopefully, these definitions are tightly related as the flock

outlet constrains its behaviour. Thus, the outlet typology is first described in this section, then the housing typology and finally the flock typology we defined after them.

- **Outlet-based typology**

Domestic poultry can be classified straightforwardly using their species and production outlets. Data collected during the commodity chain survey are summarised in **[Des09]**. These data confirmed that the most largely represented species are: chicken, duck and muscovy. Other poultry like goose, quail, turkey are also bred but at a much lower scale. Thus, the field study did not considered them.

According to **[Agr07]**, the poultry outlets can be classified according to: (1) grand-parents, (2) parents, (3) pullets & duckling, (4) egg layer, (5) meat but also (6) mixed parents and egg layer and (7) mixed grand-parents and meat. The core of the production relies on egg layer, meat and parents (also called breeder). Even though other outlet types are needed to breed poultry, for meat and self-consumption, the number of farms producing (1,3,6,7) outlets is quite limited. For instance, grand-parents and parents are usually bred by extremely large farms, which have an influence area beyond the province level. Such large farms often have a high bio-security level, and thus do not have any impact on our epidemiological concerns. On the contrary, smaller farms with one more specific outlet must to be considered: backyard farms and their self-consumption outlets. The resulting poultry classification of poultry is presented in Table 4.1.

Species / Outlet	Meat	Egg	Breeder	Self-consumption
Chicken	Chicken broiler	Chicken layer	Chicken breeder	Ø
Duck	Duck broiler	Duck layer	Duck breeder	Ø
Muscovy duck	Muscovy broiler	Muscovy layer	Muscovy breeder	Ø
Mixed	Ø	Ø	Ø	Backyard

Tab4.1 Poultry types according to species and outlet

This classification allows us to define two important characteristics of a flock: duration of production and number of individuals (details in Annex 3).

- **Housing consideration**

Depending on the outlet and the species, several types of housing are possible. Housing defines how the poultry is bred. The different housings define by domain experts are the following ones:

(1) Birds in a close building all day long;

(2) Birds in an outdoor close pen with open house without access to river or channel;

(3) Birds in an outdoor close pen with open house with access to river or channel;

(4) Birds in an outdoor close pen with open house with access to a pond;

(5) Birds in channel, river and rice field all day and in a close pen at night;

(6) Scavenging birds, all day in side and outside the farm.

Considering species, outlets and housing provides more information on the production duration and on the number of individuals, but the resulting number of possibilities is too large (as detailed in Annex 3). Figure 4.7 summarises the different factors and shows the too large number of possibilities, while direct products are not even represented.
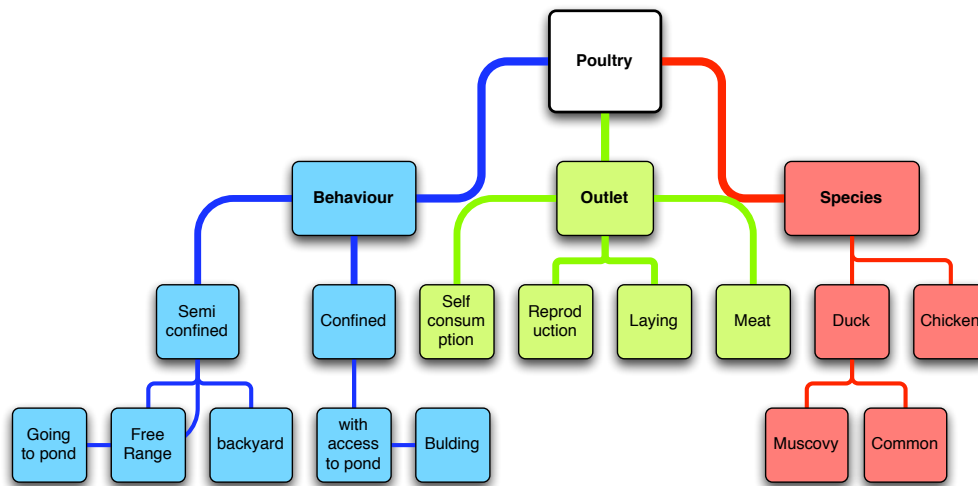
Fig.4.7 Possible differential factors (revised 10/09/09)

- **Selected typology**

This classification provides too much details with respect to the epidemiological objectives. Outlets are not relevant in the GAMAVI context. The production duration is related to the outlets, and is then sufficient for our classification. Thus, I proposed a simplification of the differential factors. The production duration is redefined as *long* and *short*, and the housing is simplified as: (1) confinement, (2) confinement with a shared access to water, (3) free-range. The backyard flock is considered apart because it is a mixed flock, which has a scavenging behaviour and a production duration very variable. Table 4.2 summarises our poultry classification. Backyard housing is not represent because it has a specific behaviour: wandering in the village at day and confined at night.

New housing / Species	Confined	Confined with access to water	Free range
Duck	Long and short	Long and short	Short
Muscovy duck	Long and short	Long and short	ø
Chicken	Long and short	ø	ø

Tab.4.2 Production duration per simplified housing and species

2.3.2. Flock dynamics

Flocks have two types of dynamics: Epidemiological one and movements. The first one has been defined through a specific iteration because it strongly relates several entities and the modelling choices would impact them all at once.

The movement dynamics can be derived from the housing types. The confined flocks (with or without access to water) are kept indoor and thus have no movement. The free-range flocks are scavenging in rice-field during the day. The backyard flocks are wandering within the village.

Concretely, Free-range flocks are either located in a building without interaction, at a field or moving to the field. Backyard flocks can be located in a building or wandering within the village only (according to domain experts). Within the operational model, I envisioned three ways to represent these movements: (1) a poultry can move *randomly* within an area centred on the farm; (2) *tropism* : for example free-range poultry are attracted to go to rice-field during the day and back home in the evening; (3) *fixed*: for example free-range poultry goes to the rice-field following the shortest-path and stay there until the evening. Figure 4.8 illustrates these possibilities.
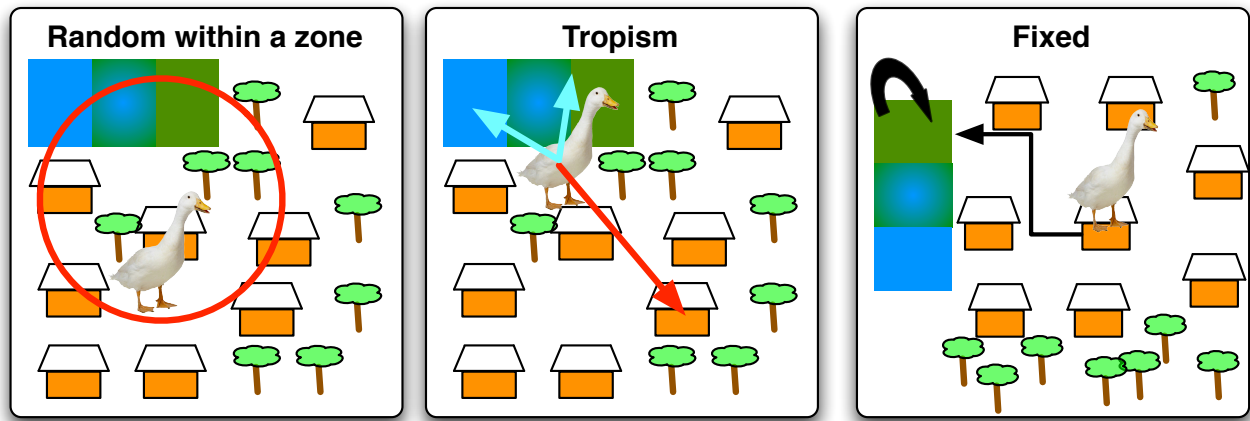
113

Fig.4.8 Three movement representations

The initially planned commodity chain survey did not consider how a free-range or a backyard flock would move. Consequently, we designed an unforeseen but brief field study, which I conducted, to question farmers owning free-range ducks or backyard flocks. According to these farmers, backyard poultry have some routine that includes point of interests which advocates for movement type (2) but no general nor specific pattern could be elicited. This poultry stays close to its farm and the straying movement are fairly random. Thus, movement type (1) is an acceptable representation. Free-range poultry are more oriented and patterns could be defined. Generally, they go to a random but nearby rice-field using the shortest possible path (or an approximation) and come back similarly in the evening. Thus, the use of a defined pattern is plausible: go out, select a nearby rice-field, stay there, come back in the evening. Although, it is also common for free-range poultry to come back around lunch time due to feeding from the farmer. Finally, free-range duck usually avoid to go to a rice-field where other birds already are. I summarise the movement dynamic of the free-range duck in Figure 4.9. In the morning, a free-range duck flock selects a rice-field free from other flocks, go to it, wander within it and comes back in for the night.

Fig.4.9 Free-range duck behaviour

2.3.3.Representation: population, flock or individual

A specific iteration was conducted to decide how to represent the flocks. From the initial domain model it was obvious that the population representation, as in a compartment-based model, was not adapted. Indeed, it would not allow spatially constrained interaction. Consequently, we evaluated two other flock representations: A set of spatially constrained individuals or a single group entity. The three representations are illustrated in Figure 4.10: A compartment-based, a group of individuals spatially constrained, or a single entity having a similar location.

According to domain experts **[Des09,Pay09]**, most of the poultry production is organised as band: a whole group of "day old chicks" (one day old chicks identified as DOC) is bought in one time, bred together and finally sold altogether. From these information, it is reasonable to consider the whole flock to be homogeneous. In addition, poultry are also gregarious animals, the whole flock stays within a given radius. Thus representing the flock by a single located agent appears to be a marginally simplified representation.

$$\frac{dS}{dt} = b_m N - (\lambda + d_m)S + \varepsilon R$$

$$\frac{dI}{dt} = \lambda S - (d_m + \sigma)I$$

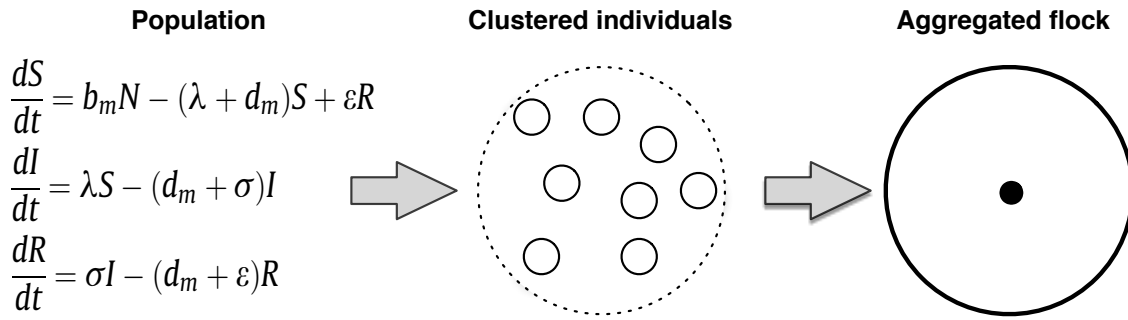$$\frac{dR}{dt} = \sigma I - (d_m + \varepsilon)R$$

Fig.4.10 Different representations of poultry

Nevertheless, two issues remain: heterogeneity in a backyard flock (in terms of age and species) and the envisioned disease-related interactions (precisely described in Section 2.7). As the implementer, I proposed to combine a single flock agent to represent a simplified location of the flock, and a matrix of individuals (as in micro-simulations) to represent the heterogeneity of the flock. This representation addresses the heterogeneity issue while diminishing the execution cost (of the implemented model). Nevertheless, an issue still remained: does the location simplification alter significantly the interaction with the environment?

To assess the relevancy of this simplified representation, I develop two implemented models representing a theoretical reference system. Both models represent flocks moving randomly in an homogeneous environment (using a grid) where a virus can persist temporarily. If an individual is infected, it excretes virus onto the environment. If an individual is located on an infected part of the environment, it gets infected according to a predefined probability. The first model represents explicitly individuals, which move randomly but are maintained grouped within a given radius. The second model represents the aggregated flock as a single agent. This agent perceives a circular area of a given radius. For each infected cell of this area, the flock agent determines if an individual is located on it (according to the density of individual in the defined area). In Figure 4.11, red areas represent the infected cell of the environment. On the left side of the figure, the other colours represent individuals grouped in flocks. On the right side, the green dots represent the centre of flocks (aggregation of individuals).
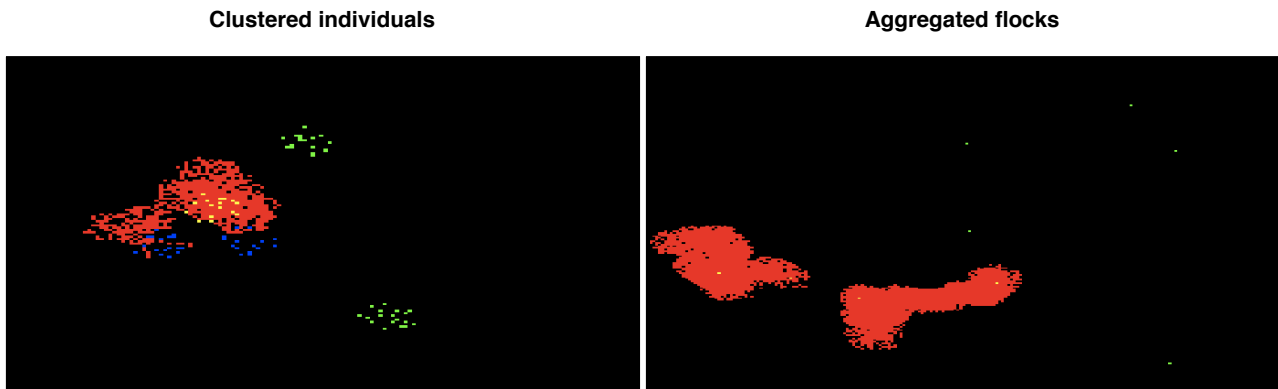
Fig.4.11 Visualisation of the two implemented model

After numerous simulations and statistical comparisons, no significant difference could be found between the two models thus the aggregated flock representation was used as it greatly fasten the execution of the implemented model.

2.4. Spatial environment

According to the domain model, flocks can infect other directly or through the environment. Thus, the representation of the environment had been considered in details within a dedicated iteration. The initially envisioned data were not available at first and we had to rely on a very schematic representation. Once, the commodity chain survey was conducted and the GIS data from local authorities was obtained they appeared

to be clearly insufficient. Consequently, we conducted a new specific data assessment through literature review **[Tri09]**. Once more, the data was insufficient. Consequently, we finally complemented the partial data already obtained with raster satellite images of villages from the Bac Giang province.

2.4.1. Prototypes: transect-based topologies

Early models relied exclusively on the transect presented in Figure 4.12 which proposes a 2D topology of a village. We combined this information with domain experts knowledge. Usually, vietnamese village are close to a river and bordered by rice-fields (which is in accordance to the transect as well). Thus, between the village and the river, the environment follows the transect topology: inner village, channel, dike, dry culture (high land), maize (low land) and the river. Several representations have been tested as presented in Figure 4.13, with the hope that GIS data would be available later. Pictures 1-5 are hand drawn schematic villages used to test different topologies while 6 is based on a generic map generator
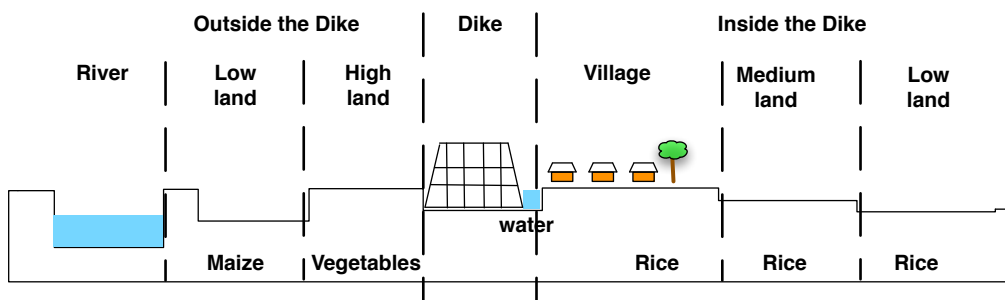
Fig.4.12 North Vietnam villages transect **[Amo08]** (adapted from **[Ben99]**)
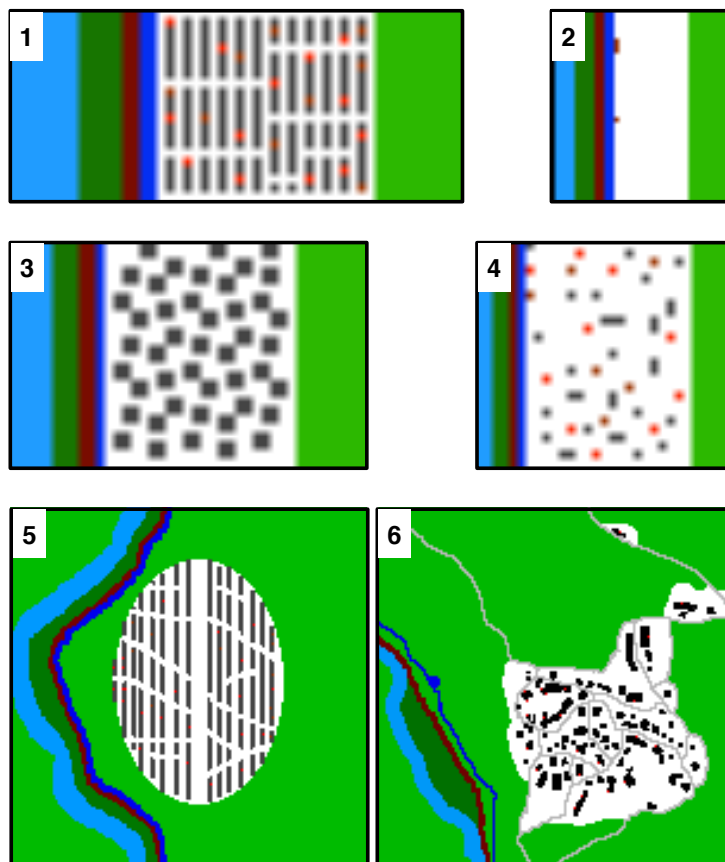
Fig.4.13 Village schematic topologies

2.4.2. GIS data issue and workaround

Unfortunately, no GIS data of the village topology were obtained, only crude data at the commune level. As shown in Figure 4.14, the villages topologies are not represented. Moreover, the available shape-files appeared to be inconsistent with other domain experts data collecting their field studies (*e.g.* too few rice-fields compared to the "populated areas" which actually include rice-fields).
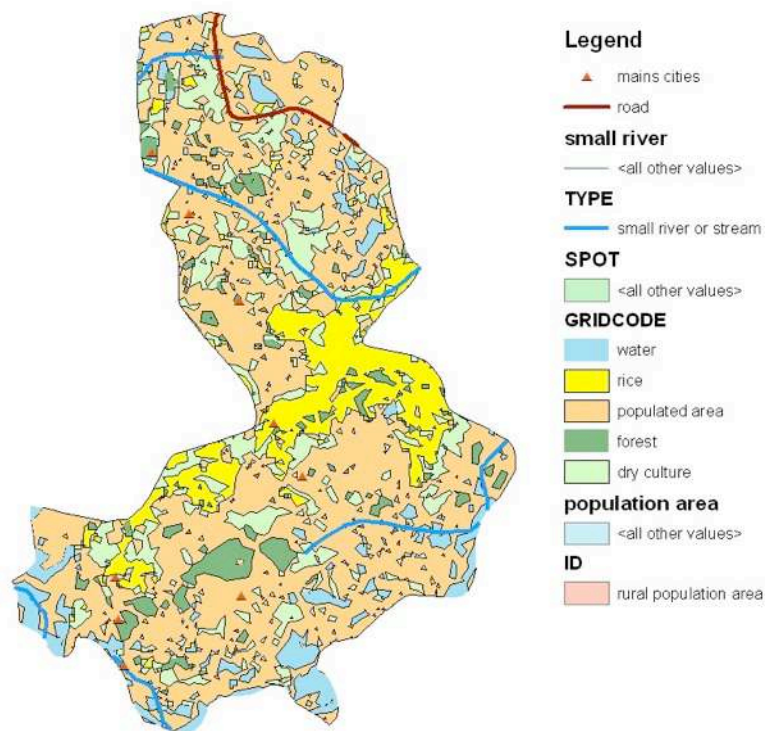
Fig.4.14 GIS data of a Bac Giang commune

To overcome this situation, a study of typical vietnamese village topology was initiated **[Tri09]** in order to create a village generator that would be suitable for our study. Most of the studies on the structure of the vietnamese villages refer to **[Gou37]**. The author defines numerous different types of villages but only two types can be commonly found in the Red River Delta.

"*Peal necklace shaped*" villages are located along a river. In these villages, the houses are aligned to the river and cultures are beside the main road. Thus the built area is surrounded by water (channels and ponds).Villages "*river's main bed*" are located next to the Red River, on the upper side of the river's bed. The built area includes dry cultures. Using the knowledge from **[Gou37, Fon99, Tes03]**, Trillaud established global characteristics of interest for our work:

- Each household has a pond;

- Dry crops are within the built area while rice-field lots are outside;

- There is usually one or two main roads with many connected lanes (few dead ends);

- House parcels have no predefined shapes (in contrary to **[Gou37]** where all parcels were rectangles);

- A house parcel measures around 600 square metres with sides between 20 to 60 metres;

From all these information, Trillaud created a village generator suited to generate vietnamese villages. Figure 4.15 presents a generated village with 3 different types of land-cover and roads.
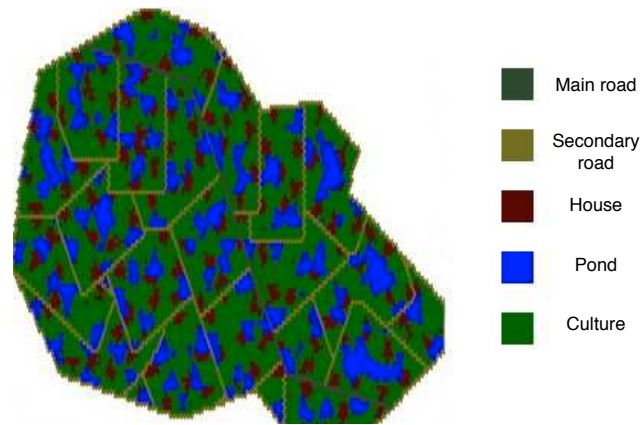
Fig.4.15 Generated village **[Tri09]**

The prototypes (especially with a realistic built area topology) interested a lot domain experts, who concluded that the differentiated areas should be: *building*, "*ground*", *rice-field lots*, and *ponds.* Rivers and channels could not be generated because of a lack of data and Trillaud's generator was not used in the end.

Initially, we intended to use GIS data provided by domain experts coupled with field reports (including the GPS localisation of the inner village's buildings). The GIS files would have contained the precise locations and representations of the different elements (house, roads, crops, etc). Unfortunately, the best data we could get was the centroid of the village along statistical data about it.

2.4.3.Raster satellite images based representation

Real GIS data were not available and Trillaud's generator was not adapted because it focuses only on the village cluster. Another way to generate a realistic representation of the inner village and its surrounding must be designed. Surprisingly, google maps images were decent enough to envision such a representation, as illustrated in Figure 4.16.

Fig.4.16 Image of a Bac Giang village from google maps ©

This image has been vectorised as shown in Figure 4.17. The inner village, which contains some buildings (farms) and two ponds, is surrounded by rice-fields which are cut into different areas by roads.

118
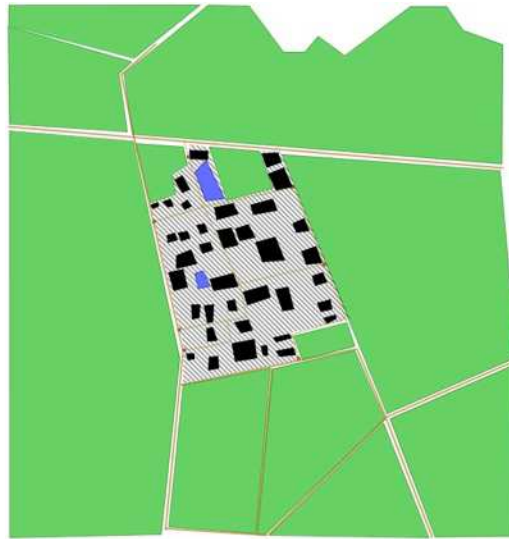
Fig.4.17 Vectorised image of the previous village in Bac Giang

The vector image of Figure 4.18 still lacks the decomposition of rice-field area into lots. Some built-in primitives of the simulation platform were then used to discretise rice-field area in rice-field lots as shown in Figure 4.17.
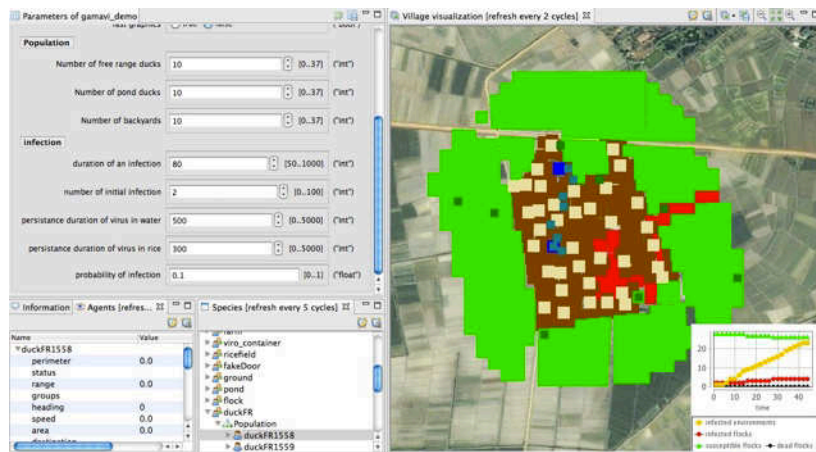
Fig.4.18 Visualisation of the final (implemented) GAMAVI model

Initially, the representation of the environment topology was envisioned to be based on precise geo-referenced data. These data were not available from the beginning thus we defined very schematic environment topologies. When it became obvious that this data could not to be collected, we relied on literature review to improve the schematic topology. Unfortunately, there was not enough information available. Thus, we finally "vectorised" publicly available satellite images of Bac Giang villages. It had the advantage to provide not only a workaround for this lack of data but also allow future integration of GIS data when they are available.

2.5. Generating the village

The separate modelling of the flocks, farms and the environment is integrated within the village model. Indeed, a village is both a spatial environment and a population of farms and flocks. The environment has already been considered in the previous section except for the location of farms. No geo-referenced data was recorded concerning their location in a village. My specific study on backyard and free-range flocks behaviour provided only a basic statement by farmers that "semi-commercial farms with free-range flocks are generally located on the outside belt of the village". Without a clear characterisation, we decided that farms would be located randomly within the builds of the environment topology.

119

In collaboration with domain experts, I extracted the parameters from their field studies data (Annex 3) on which can rely the farms and flocks population characterisation. Two sets of data were available: village-dependent data and general statistics. The village-dependent data have been clustered into five stereotypes of villages thanks to a Principal Component Analysis (ACP).

Villages have been characterising according to the following parameters: (1) number of households (90% have backyard flocks) and (2) number of flocks from each type (Section 2.3.1). General statistics are related to (a) the distribution of flock number per farm (from 1 to 4), (b) the flock size per flock type and (c) the production duration per flock type.

In the operational model, I added the consideration of mean value, standard deviation, the minimum and the maximum of each parameters in order to be able to generate a variety of village representation for each cluster. Finally, I organised the generation of a village as the following sequence:

1. Selection of the village stereotypes

2. Computation of the number of households

3. Computation of the number of backyard flocks (and farms)

4. Computation of the different semi-commercial flocks (global statistic)

    4.1. Computation of their number per flock type

    4.2. Computation of their duration

    4.3. Computation of their production duration

5. Generation of the farms

    5.1. Random determination if the farm is associated to an household with backyard or not

    5.2. Computation of the number of flock for the current farm (minus -1 if it has a backyard)

        5.2.1. Associate randomly a flock to the farm

        5.2.2. Iterate until the number of flocks is reached

    5.3. Iterate until all the flocks are associated with a farm

This iteration provided us with a model on which we integrated the epidemiological dynamics as described in the two following sections (2.6 and 2.7).

2.6.Epidemiological model

In order to represent the different aspect of the individual epidemiological sub-model, I decomposed it in three parts: (1) individual epidemiological model, (2) persistence of the virus in the environment and finally the (3) infectious interactions. The first part represents the evolution of infectious status of an individual. The second part considers how to model the virus dynamic in the environment. The third part describes the interaction among flocks and between flocks and the environment.

2.6.1.Individual infectious evolution

As stated in Section 2.3.3, each poultry is considered through its internal state thanks to the individuals matrix of each flock, representing the individuals location. Thus, I represent the infectious status of each individual poultry (i.e. birds) in the model.

- *Introduction*

Poultry differ greatly in their reaction to the avian influenza. Even different strains of the avian influenza virus have different effect on the poultry. Actually, ducks are often considered to be "*carrier*" because they are able to carry the virus without any symptom **[Stu05]**. On the contrary, chicken die easily from the disease (it depends on virus strains, some were less lethal). Thus, chicken and ducks had to be differentiated **[Spi08]**. As stated in Section 2.3.1, three species were considered: chicken, duck and muscovy duck. However, we did

not find any data that differentiate duck and muscovy duck reaction. Consequently, we decided to consider them as identical for all epidemiological concerns. The chicken infectious statuses evolution was represented according to a classic SIR. In practice, we considered a 100% mortality rate within a few days (according to domain experts knowledge and supported by **[Spi08, Bou09]**.

- *Duck infectious evolution*

Dr. Bicout collaborated on the Gripavi project at the commercial system assessment, introduced in Section 2.1. Prior to this study, he also proposed a SEIIR model of the avian influenza infectious evolution **[Ren07]**. This model has been adapted to be used in conjunction with a model of the virus persistence in water as described next.

The first infectious evolution model I designed is a close translation of the model presented in **[Ren07]**. Figure 4.19 presents the evolution: a *susceptible* duck can get the infection and become *exposed* . After 1 or 2 days it becomes *infected* (status I1 for 6 days and I2 for 0 to 5 days) then it becomes *recovered* (immune) or *dead*.
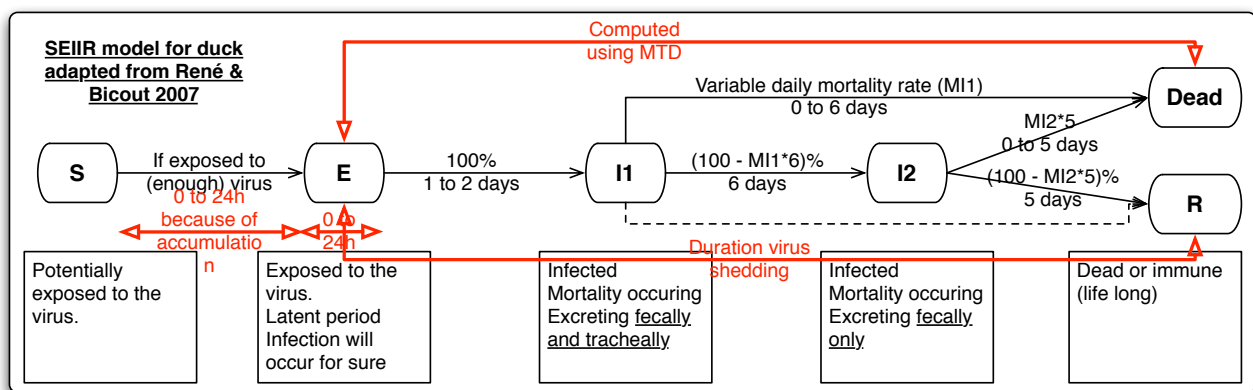
Fig.4.19 First ducks infectious statuses representation, adapted from
[Ren07]

The model is, actually, slightly more complicated. According to **[Ren07]**, excretion rates are different in I1 and I2 stages and mortality can also occur during I1. All parameters presented in Figure 4.19 are based on **[Ren07]** except the duration of I1 and I2 and the mortality rate which come from **[Hul05, Stu05]** as required by domain experts. Finally, the immunity is considered to be life-long as it is much more longer than any production duration.

After an extensive literature review and accordingly to domain experts, I decided to simplify the model. It appeared that the variability of the parameters was much larger than envisioned. For instance, some strains of HPAI induce a high mortality for ducks in contradiction to the previous hypothesis.

Figure 4.20 presents a simplified model. All the parameters (status duration, mortality and excretion level) can vary according to the scenario defined from data of **[Hul05,Stu05,Ren07,Spi08]** among many others. The only fixed parameter is the duration of the exposed stage as defined by domain experts with respect to **[Spi08]**.
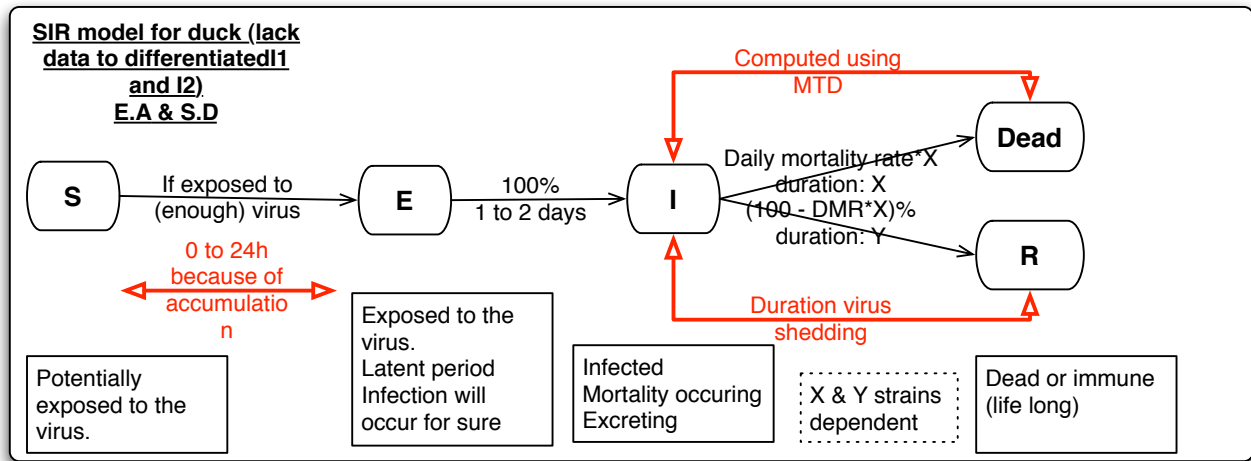
Fig.4.20 Updated ducks infectious statuses representation, derived from **[Ren07]**

Parameters of the SIR model (for chicken) and SEIR model (for duck) are mean values for a given strains. Depending on the hypothesis to evaluate, it could be of interest to consider more parameters. Especially, the vaccination effect could be integrated in order to evaluate its efficiency but also the possibility of a "silent circulation" of the virus among vaccinated poultry **[Sav06]**.

2.6.2. *Virus persistence in the environment*

The persistence of the virus should have been investigated in various environment: stagnant water, stream water, dry culture and in manure. However, due to the lack of data, only stagnant water environments have been considered. Two types of stagnant waters exist: ponds and rice-field which are flooded only temporarily.

In this purpose, I developed a model based on **[Ren07]**. This model considers that the persistence of viral particles decreases exponentially according to the temperature: $k = k_0 e^{a(T-T_0)/T}$ where k_0 and a are parameters evaluated in laboratory and given in **[Sta90]**, while T is the actual temperature in Kelvin degree and $T_0 = 0°K$. Domain Experts collected data about water temperature from local authorities in Ha Tay and Hanoi provinces (Ha Dong district).

In addition, pH and organic matter concentration affect the virus survival **[Doq05, Son06, Ben07, Bro08]**. pH data have been obtained from Ha Tay and Hanoi provinces but we could not integrate. These information have not been integrated in the model because no survey describes the pH impact precisely enough. The Rivers project (also conducted by CIRAD) aims to evaluate the persistence of the virus in water and humid environments according to natural conditions (in opposition to common practices such as **[Sta90, Bro80]**). Unfortunately, no data were available before the completion of our collaboration. This model is applied at each pond of the model. During the rice-field flooded season, each rice-field is linked to such a model.

2.7. Epidemiological interactions

The influenza viruses can be transmitted airborne, through direct contact or indirect contact, i.e. through the environment. The main transmission way is the direct interaction (airborne and direct contact) and thus has been most studied a lot while indirect transmission ways have usually been considered negligible.

Two main categories of entities interact within the disease cycle: individuals and the environment. The whole dynamics can be translated into a set of interactions and actions. Interactions are summarised in Figure 4.21: (1) an infected individual infects directly a non-infected one, (2) an infected individual excretes virus in the environment, (3) a non-infected individual becomes infected through the environment. Actions are summarised as follows: (1) an infected individual becomes non-infected (immune, dead, recovered), (2) the

122

environment infectivity decreases or disappears, (3) a new non-infected individual enters the system, (4) an individual quits the system (infected or not).
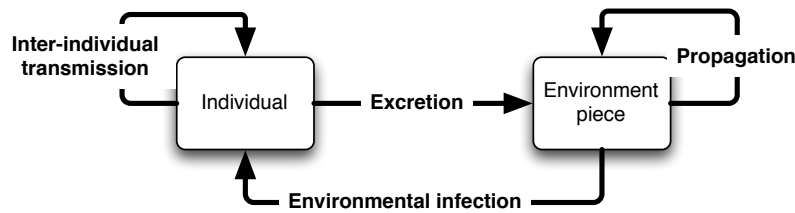
Fig.4.21 Interactions related to the disease transmission

At the beginning, many different entities were envisioned: domesticated poultry, humans, domesticated animals, production chain actors (transporters, collectors, etc). They could infect each other according to their infectious status and a given probability. Such a diversity of entities was given up due to the complexity of representation and to the lack of data. For example, it is possible that a transporter's truck would be contaminated by the virus and thus infect poultry in a farm, but how would that happened? Probabilities of such events have not been studied, which would also be an significant issue. Consequently, the epidemiological models are presented next only representing the environment and the poultry (as individuals).

We first chose to focus only on indirect transmission routes. Only infections through the environment were considered since this study focuses on the environment's influence on the disease propagation and persistence. Even if direct transmission were not considered, the persistence of the virus could still be observed in the model. It highlighted the importance of indirect transmission in our model. However, it was considered not realistic at all. Direct transmission had to be considered, even abstractly using transmission rates when a contact between two flocks occurs.

Consequently, we decided to differentiate clearly direct transmission and environment mediated one. The direct transmission was considered at three levels: Within flock, within farms and between farms. At each of this level we considered the population as a whole which allowed us to use transmission rates given in compartment-based models. We used the data published from an external collaborator, Dr. R. J. Magalhães **[Mag10a, Mag10b]**.

Initially, it was envisioned to consider the number of viral particles an individual excretes per hour and the number it could collect per hour in a watered environment. To attain such objective, we conducted an extensive literature review to obtain infection probability after a given amount of ingested viral particles (**[Stu05, Hul05, Ren07]** for instance) and the excretion amounts. Unfortunately, collecting all the required parameters appeared to be impossible as most of the epidemiological study are conducted at the population level. Consequently, we parameterised the environment mediated transmission to fit observed level of transmission given by the global R_0. This allowed domain experts to design theoretical experiments to evaluate the interactions between the direct and the environment-mediated transmissions.

2.8. Conclusion

In this section, I presented the different modelling iterations conducted to develop the GAMAVI model. This model cannot completely answer to the initial research question due to numerous lack of data. The iterative modelling can be continued when new data is available. Nevertheless, the modelling process allowed domain experts to improve their knowledge of the reference system. Especially, they finally considered that the virus can persist within the village without frequent virus re-introduction from the commercial system. Indeed, they focused their research on the interplay between the poultry population dynamics and the virus persistence in the environment during winter. Nevertheless, the final implemented model that was used by domain experts to evaluate their hypotheses is presented in the next section.

3. Conclusion

3.1.Final implemented model

Most of the insights produced during the modelling process came from the collaborative modelling at the conceptual level presented in the previous sections. Nevertheless, a modelling iteration is complete only when the implemented model, coded by the implementer, has been delivered to the domain experts as a ready-to-use tool. This implemented model has also been used by domain expert to evaluate their hypotheses through simulation scenarios. Thus, I present the final implemented model of the application in this section. This model is presented in Figure 4.22 and it contains different classes gathered around four entities: flocks, farms, "viro_container" and environmental areas.

Flocks are implemented thanks to 6 classes. Basic flock defines the structure of a generic flock (individuals, species, housing, etc.). Viro_flock adds the epidemiological internal model. All confined poultry flocks are defined thanks to this class. Interacting_flock defines a general flock that can excrete virus in the environment. Moving_flock extends the previous class to allow flocks to move. Finally, the two specific free-range duck flock and backyard flock have specific sub-classes. This decomposition allows to adapt each dynamics of the flock without impact on others.

Farms are defined with three classes. The first one, differentiated_farm implements the production dynamics of the model (flocks renewal). The viro_farm adds the implementation of the infectious contacts among flocks and farms, which is a part of the transmission model. Finally, the mover_farm implements the production dynamics of farms that integrate non confined poultry.

The **environment** is implemented in two parts. Its topology is defined thanks to a GIS environment class which manages several areas: rice-field, village_heart, open ground and buildings. For each rice-field and village_heart area, the simulator manager (GAMA *world* class) discretises the areas in grids (rice or village ground) because rice-field lots are not represented in GIS data. This discretisation allows to represent the virus in the environment. For each cell and pond (of the GIS environment), a **viro_container** is adjoined to manage the virus depletion dynamics. In the present model, the virus can persist in ponds and watered rice-field lot. We also linked other cells to viro_containers to allow domain experts to conduct theoretical experiments to envision the transmission through dry environment (as faeces can remain infective up to a few days) and to prepare its representation if new data about the persistence of the virus in the dry environment is available later.
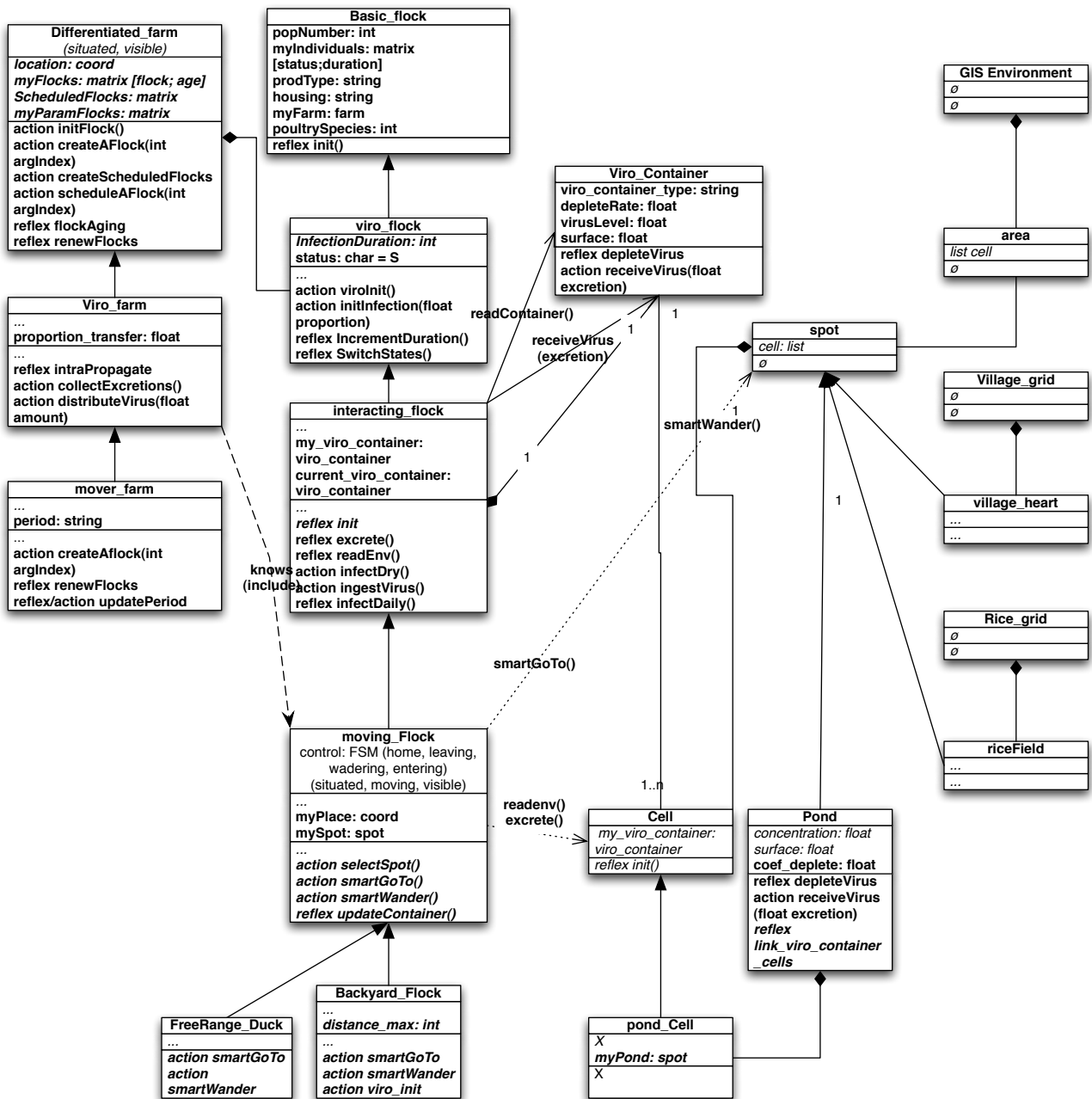
Fig.4.22 UML diagram of the implemented model

The final implemented model has been coded in GAML and executed on the GAMA simulation platform. Figure 4.23 presents a visualisation of the execution of the implemented model. In this figure, 4 panels can be distinguished: parameter, monitor, inspector, graph and visualisation. The parameter panel allows domain experts to change a parameter before launching a simulation. The monitor panel contains the value of global variables at each time-step. The graph panel shows the evolution of global variables, for instance the number of flocks fully susceptible, with infected individual, with recovered individual and those totally decimated by the virus. The visualisation panel presents the spatial evolution of the different entities: red represents virus in the environment, light green the rice-fields, dark green duck flocks, blue ponds, dark brown inner village ground, light brown farms.

In the example illustrated by Figure 4.22, the virus is much more presents on the village grounds. The number of infected flocks is globally stable but the number of dead flock is slowly increasing. It means that the epidemic is ravaging each flock but the transmission between flocks is much slower.

Fig.4.23 GAMAVI simulation in GAMA

This implemented model has been used by domain expert to evaluate their hypotheses through simulation scenarios. They compared the infectious dynamics produced by the different transmission mechanisms. For instance, within the same simulated village, a first simulation was conducted considering only the direct transmission whereas a second simulation was dedicated to the environment mediated transmission. Other simulations were also conducted to assess the influence of the flocks population or the village typology on the different transmission and persistence mechanisms.

3.2.Synthesis

In this Chapter, the modelling process of the application, investigation of the local avian influenza dynamics, has been detailed. I started by presenting its initialisation, including the presentation of involved domain experts, the research question and the description of the reference system. I also presented the collaborative elicitation of the initial domain through different ontologies. Similarly, to the research question and the reference system description, the domain model had been amended due to issues and choices made during the iterative modelling. Thus, I presented the final domain model in terms of ontology, documentation based on the O.D.D protocol, and sub-models. These sub-models allowed the model contributors to structure the general data assessment and the iterative modelling.

Each sub-model we tried to integrate in the model has been presented. During this process, it appeared that the commercial system influence on local persistence and propagation of the avian influenza could not be correctly represented due to lack of data. Nevertheless, domain experts assumed that this system would only marginally affect the local disease dynamics. Consequently, a redefinition of the reference system (its scales and entities) had to be made: The village level was considered instead of the province one. The representation of the reference system during the iterative modelling can be decomposed in two phases. First, we conceived a model of the village without any disease-related mechanism. Then, we used the previously conceived model to test different representations of the disease-related mechanisms and to evaluate domain experts hypotheses. The final implemented model is introduced in Section 3.1 and its description is provided as an O.D.D document in Annex 2.

The modelling process of this application was specific as it was applied in a given domain, Epidemiology. Many *ad hoc* solutions were proposed though formal and general tools remained as the centre of the process: the O.D.D. communication protocol. Its benefits were prominent in the process because I, as the modeller

126

and implementer, relied on it to formalise the communication with the domain experts (and the initiator). At each important step of the modelling process, the model was described according to the O.D.D. formalism in order to allow an equivalent model understanding from the initiator and the domain experts. Its general structure (*Overview* then *Design concepts* then *Details*) and especially the recursive *sub-models* section allows a progressive complexity in the model's description, which facilitates its understanding. Conversely, it allows to use a presentation as detailed as required (*details* section). It also forces the description of modelling specificities that are often overlooked (*design concepts* section).

The final model, presented in Section 3.1, may or may not be re-used in the near future, but this was not the main objective of the collaboration. The main interest of this modelling process was to allow Dr. Desvaux, the initiator and a domain expert, to orient her research on avian influenza local propagation and persistence in North Vietnam. It was a way to assess the relevancy of her hypotheses in order to prepare further field studies to effectively validate her hypotheses. More precisely, the description of the reference system with other domain experts widen her perspectives. In opposition, design and implementation of the model allows to assess the interplay between the different dynamics and, finally, to evaluate which hypotheses are more likely to define the local persistence and propagation of the avian influenza in North Vietnam.

Even though the presented modelling process was applied only in Epidemiology and that many *ad hoc* solutions have been proposed, this process can be generalised to other situations. The only condition is that the model is defined and the modelling process is conducted in order to explore or explain a partially characterised reference system. I present such generalisation in the following Chapter.

Chapter V : KIMONO

The modelling process of the application presented in Chapter 4 highlighted that O.D.D. can be used as a central landmark during the modelling process. It can be used to assess the evolution state of the model under construction by both modellers and domain experts. In addition to its inherent advantages presented in Chapter 2 (Section 3.2), O.D.D can already be considered a standard: The original paper **[Gri06]** has been cited 167 times according to Web of science. O.D.D., complemented with UML diagrams (Chapter 2, Section 3.1), constitutes an adapted tool to describe a world model and document its design. Nevertheless, O.D.D is not intended to structure the modelling process in any way. It is the purpose of a complete modelling method including support tools.

I propose the *KIMONO* modelling method, *Keep It Modular, Open-ended, and Neatly Organised*, to attain such objective. More precisely, KIMONO is an iterative and collaborative modelling method which structures the design of exploratory and explanatory model and specifically world models. KIMONO proposes a modelling cycle and definition of the contributor roles adapted to the creation of world models. It also uses several existing tools: (1) O.D.D for documentation, (2) UML diagrams for expression of the conceptual models and relies on: (3) GAMA (and GAML) for implementation.

1. Introduction

The purpose of the KIMONO modelling method is to structure and, more generally, to facilitate the creation of models addressing exploratory or explanatory research questions. Such models have been characterised as *World models* in Chapter 2. To attain this objective, KIMONO requires a strong and continuous implication of domain experts. Indeed, a close collaboration between domain experts and modellers is essential to be able to design world models.

KIMONO's modelling cycle proposes a way to incrementally design a model. Each increment (or iteration) is a model prototype that can be used right away by the domain expert as a research support tool to generate and evaluate hypotheses. Each of these model prototypes has a specific purpose which can be: Testing a domain expert's hypothesis, evaluating a representation alternatives or increasing the represented subset of the reference system (incrementing the model).

The GAMAVI model presented in Chapter 4 is an explanatory oriented model designed after the KIMONO method. I derive simplified examples from this study case in order to illustrate and clarify the method.

2. Model contributor roles

Before initiating a modelling process, it is important to state the role of each model contributor. World models usually rely on numerous sources of data, from different domains, which are integrated in an agent-based model (Chapter 2). Moreover, the KIMONO modelling cycle includes several levels of description of the reference system (from the initial informal description to implemented model) which are designed by different contributors. The KIMONO method considers four contributor roles: *initiator*, *domain expert*, *modeller* and *implementer* to conduct such tasks.

The *initiator* provides an exploratory or explanatory oriented research question that is not addressable directly through literature review nor classic models. Thus, (s)he requires the design of a world model to address the question. In addition, (s)he is the only person able to decide whether or not the initial question is answered by the model. Consequently, this role is specifically involved in the early steps of the modelling cycle (*i.e.* until domain model) and at the ending of the modelling cycle. Due to the exploratory and explanatory models objective of the model designed through KIMONO, the initiator is also often a domain expert. Nevertheless, the initiator can also be a decision-maker (*e.g* politician) who needs a better understanding or an evaluation of the possible evolution of the reference system in order to make a decision.

A *domain expert* is a role acted by one or several persons knowledgeable about a domain related to the reference system. This role is similar to the "*thematician*" role proposed in **[Dro02]**. Due to the open nature of world models, a given modelling process may involve different domains thus several domain experts may be required. The domain expert micro-knowledge (Chapter 2) is used to create the model **[Dro02]** whereas his/her macro-knowledge is used to verify and validate it (Chapter 2). Even though (s)he is involved in all the steps of the modelling cycle, his/her role is prominent in the early steps (from research question to domain model) and during the iteration assessments. Due to the open nature of a world model, a specific model is likely to require the involvement of several domain experts from different areas of expertise.

The *modeller* is a specialist of agent-based models because the agent meta-model has been selected to describe the reference system as a world model. (S)He supports (or leads) the integration of the various domain experts' knowledge into the model. In addition, (s)he collaborates with the implementer to translate conceptual models into an executable one.

The *implementer* is a role similar to the "*computer scientist*" one described in **[Dro02]**. This role conceives the operational model, conduct the implementation and facilitate its evaluation. In practice, (s)he ensures the exact correspondence between the implemented and the operational models (e.g. correct any simulation artefact). The implementer is thus a specialist of the selected implementation tools (i.e. *ABMS* platform) which is GAMA in my applicative context.

For the GAMAVI model, the initiator was impersonated by Dr. Desvaux but more generally it was the Gripavi project. She also acted as a domain expert among many others but on specific domains: Dr. Bicout for the virus persistent in water, Dr Dung for the characterisation of avian influenza viruses (present in Vietnam) and G. Fournié for the commercial dynamics. I acted as the modeller and the implementer.

3. Modelling cycle

KIMONO proposes a modelling cycle with two main phases: Initialisation of the process and iterative modelling. During the modelling process, the evolution of the model is recorded in the documentation (an O.D.D. formatted document) as detailed in Section 4.

Figure 5.1 illustrates the modelling cycle with an emphasis on the beginning of the cycle which is decomposed in five steps. The cycle starts by the expression of a research question formulated by the model initiator. Then, the reference system and hypotheses are described textually at first and then more formally as a domain model. This model is then used to assess the knowledge required to design the subsequent intermediary models created during the iterative modelling phase.
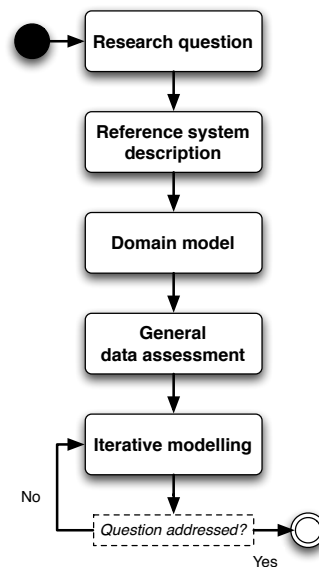

Fig.5.1 KIMONO General modelling cycle

3.1.Initial steps

3.1.1.Research question

The definition of the initial research question is the first step of any modelling process which even precedes any assessment of the reference system. This question is expressed by the initiator but is much likely to evolve during (and thanks to) the KIMONO modelling process. It is especially true during the initial steps but it can also evolve during the iterative modelling as hypotheses are evaluated or data lacks force to redefine the question.

For instance, the original question of Dr. Desvaux was: "*what are the local determinants and patterns influencing persistence and diffusion of the virus in North Vietnam among domestic poultry?*" **[Amo07]** (personal translation), while her final question can be summarised as: "*what are the mechanisms, in conjunction with the environment, which enables the persistence and the propagation of the H5N1 virus at a local scale (village or district)* ".

3.1.2.Reference system description

A first description of the reference system is derived from the research question by the initiator and domain experts. They list the relevant entities (including the environment)   and their interactions. Then, a first definition of the spatial and temporal scales and resolution can be derived. In addition, hypotheses to evaluate during the modelling process should be included. The objective of KIMONO is to allow evaluation of these hypotheses but also to help to generate them all along the modelling process. Thus the hypotheses list can be updated at each modelling step. This description of the reference system can be made in an informal text.

For instance, the first description of the Gamavi reference system is presented in Chapter 3, Section 2.1. It loosely defines the extends of the reference system (temporal:2 years and spatial: a province). It lists the entities considered at first (poultry, humans, animals and the spatial environment) and some of their relationships (e.g. poultry bred by humans, poultry organised as a flock, etc.). The main hypothesis was an interplay between poultry population and the environment as a cyclic persistence mechanism.

3.1.3.Domain model

The domain model details and structures the previous reference system description. In particular, previously overlooked elements of the reference system are likely to be added. It can be designed as an ontology completed by a textual documentation for clarification. Domain experts construct this model with support from the modeller. The initiator is also involved to ascertain the coherency of the domain model with his/her question.

The ontology represents the different entities and the environment considered to be relevant by domain experts and the initiator. Entities (and the environment) are linked to each other through interactions. Elicitation of the ontology can be made iteratively. For instance, a first ontology integrating as many as possible entities can be defined and then each entity relevancy is assessed. It is also possible to consider different ontologies: For instance, an ontology based on relationships and one based on inheritance link (as a simplified UML class diagram).

During the elicitation of the ontology entities may be added or discarded and required justifications are expressed in the documentation. Any entity that remains not linked to others should be discarded. Finally, a complete list of the relevant entities specifies further spatial scales and resolutions previously defined.

In addition to the ontology and the textual documentation, a formal representation of the reference system in the domain experts' formalisms should be provided. In a world model context, the representation of the reference system often involves different domains that focus on different subsets of the system. Consequently, several representations, or sub-models, in different domain expert formalisms can be designed. Later on, this sub-models decomposition can be used to orient the iterative modelling steps or data collection.

The ontology and the textual documentation can be combined to obtain the *Overview* section of the O.D.D document. Sub-models expressed in specific domain formalism are logically included in the *Sub-models* sub-section of the *Details* section of the O.D.D.

3.1.4.Global data assessment

In general, it is necessary to assess if sufficient data are available before starting the modelling process according to **[Aum07]** who call this step the "*knowledge synthesis*". In the KIMONO context this synthesis is slightly different: The required data is likely to be incomplete. Assessing which data is available and which one is not allows to parallelise the data gathering and the subsequent modelling steps. Gathering data can rely on three processes: literature review, laboratory experiments and field studies. Experiments and field studies, even more, are long process thus it is important to plan them as soon as possible. The sub-models defined along the domain model can be used to decomposed the data gathering by domain or reference system sub-set.

Again, the O.D.D. description can be used as a base: the *initialisation* and *input data* can be used to list all the parameters. Each parameter is detailed with: units, domain value, known value(s) and what remains with unknown valuation.

3.2.Iterative modelling

After the definition of the domain model, the modelling cycle contains an iterative sequence. Each sequence is documented by an O.D.D. document which re-used much of the previously defined O.D.D. sections (and the O.D.D. document of the previous iteration). This sequence is constituted of five steps: *iteration objective, data assessment, design, operational and implemented model* as illustrated in Figure 5.2. The three intermediary models (design, operational and implemented) are adapted from the method proposed in **[Dro02]**. Moreover, the feedback loops introduced in this method are conserved (red arrows in the figure).

Another loop (not represented in the figure) is added to consider the feedback of an iteration onto the domain model. Indeed, the need to represent new entities (and new dynamics) can emerge during an iterative modelling step. Thus the domain model may be either incremented or simplified (confirmation or negation of an hypothesis, lack of data, etc.).
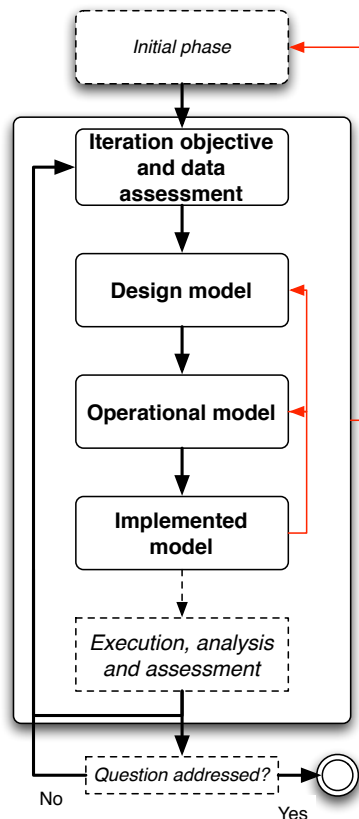

Fig.5.2 Detailed iterative modelling

3.2.1.Iteration objective and data assessment

I consider three types of *iteration*:

131

(1) Incrementation of the model: a non-represented sub-set of the domain model (representing a sub-set of the reference system) is added to the previous model. More precisely, the current models (i.e. design, operational and implemented) are complemented with a representation of the considered sub-set. The definition of a coherent sub-set can be made after the sub-models (expressed in domain's formalism) defined at the same time as the domain model was.

(2) Evaluation of a representation alternative: Two (or more) models are designed to evaluate which representation is the most adequate to represent a given element of the reference system. This objective requires characterisation of both representations and a evaluation mean in terms of simulation outputs, generally. Depending on the situation, the evaluation can be made within the existing model or can be conducted externally (e.g. in a theoretical model).

(3) Evaluation of an hypothesis: The model is used to confirm or negate an hypothesis. This steps requires two elements: All relevant entities are represented and the definition of a scenario **[Dro02]**. If some relevant entities are not represented (in the models of the previous iteration), they are added similarly to the incrementation iteration. The scenario is a way to test the hypothesis. In practice, initial conditions of the model (inputs) to observe the considered phenomena (output) have to be define. In a world model context, such a phenomena is likely to be the emergent consequences of the interactions between entities usually considered in separate domain.

The definition of the iteration objective is not sufficient, domain experts have to make sure all required data are available. If not, a different objective has to be planned and the lacking data collected before going back to this iteration. This specific data assessment is required because of the importance of a parameter, an entity, an interaction re-evaluated after the conduction of a previous iteration. Such assessment requires that all data that were not presented in the previous domain model. Thus, a modelling iteration is preceded by a data collection, possibly a field study which requires an important time. Nevertheless, another iteration can be conducted during this time.

3.2.2.Design model

The modeller is responsible for the creation of the *design model* though (s)he is supported by domain experts (and the initiator). The design model purpose is to provide a first formal representation of the reference system using conceptual agent. It can be expressed as an UML class diagram where each entity is represented along its characteristics, actions and interactions. A textual description should be added in the O.D.D. document for clarification. Indeed, this model has to remain accessible to domain experts (and the initiator).

3.2.3.Operational model

The *operational model* prepares the translation between conceptual models to the implemented one. More precisely any implementation issues or concerns have to be addressed through the conception of this model. In addition, the simulation management is accounted for in this model for the first time. Consequently, it is the responsibility of the implementer to conceive this model but other contributors are still involved. In particular, the domain expert and modeller can be asked to clarify some elements of the design model. In practice, three types of UML diagram should be conceived: Object, activity and state-transition diagrams which formalise, respectively, the representations of the interactions, the schedule of activities and the life-cycle of entities.

3.2.4.Implemented model

The *implemented model* is a direct translation of the operational model which allows execution (simulation) of the model. The implementer conceives this model with support of the other contributors. In particular, domain expert can review the source code with support of the implementer in a similar fashion as the "extreme programming" from the Agile methods introduced in Chapter 2.

3.2.5.Execution and model assessment

Assessment of the generated model depends on numerous parameters (knowledge available, objective of the iteration, etc), thus it is left to the choice of the contributors. At least, the model can be evaluated thanks to domain experts macro-knowledge (Chapter 2). Actually, several methods such as "experimental frames" **[Zei00, Uhr09]**, or "Pattern-oriented modelling" **[Grim05]** can be used.

Nevertheless, I advice to rely of the assessment process detailed in **[Dro02]**. The authors propose a "verification" of the operational model through (unit) test conducted on executions of the implemented model. The design model is evaluated through analysis of simulation in a process called "internal validation". Finally, the "external validation" is conducted by comparing simulation interpretations and domain experts macro-knowledge.

New hypotheses, data lack or discrepancies may appear during the conduction of the incremental modelling at any step. Such events are likely to appear during the creation of the design model as it is the first formal representation of the reference system. These events may also appear with the implemented model as domain experts (and the initiator) can explore the dynamic of the reference system through this model. Consequently, the domain model has to be amended by domain experts (and the initiator). It also means that the macro-knowledge of domain experts is questioned.

For instance, an iteration of the GAMAVI model focused on representation of the virus persistence in water. During, the creation of the design model it appeared clearly that the virus could not persist in water during summer. The viral excretion of ducks in water was dramatically insufficient to compensate the fast viral depletion rate (due to high temperature).

The case of the hypothesis evaluation iteration is peculiar. In general, its objective is to increase the (macro-) knowledge of domain-experts. Thus, the model cannot be completely assessed as in other types of iteration. It requires new data collection on the field (or in laboratory experiments) which are oriented by the evaluation of the hypothesis (confirmation or invalidation) within the model.

Finally, the newly created models (i.e. design, operational and implemented) are used as the base on which the next iteration will start from. Although, the change or addition made during an hypothesis evaluation are kept in the model only if the domain experts consider the hypothesis to be sufficiently confirmed (by field study especially).

3.3. End of the process and possible use of the final model

The end of the iterative modelling can occur for three reasons:

(1) New knowledge is required but it will never be available, at least in a foreseeable future;

(2) The research question is addressed: All the hypotheses have been evaluated;

(3) The research question is still left unanswered but no new influencing entities can be envisioned. It is, obviously, the worst case scenario which lead to a complete reassessment of the conducted work. A new point of view needs to be used: a new definition of the reference system and even of the research question may be necessary, thus new domain experts will be implicated. Nevertheless, all the achieved work is not lost as much of it can be re-used during the new KIMONO driven modelling process.

4. Modelling guidelines

According to the methodological requirements (G1-4) expressed in Chapter 2, a method has to provide support during all the modelling steps: From the definition of the research question to the implementation of the model. KIMONO allows to develop exploratory and explanatory oriented models (i.e. world models).

In Chapter 2, I select the agent meta-model to express world models. Consequently, the formalism used to express the conceptual (and implemented) models has to be compatible with the agent meta-model. UML satisfies such criterion because agent-based model can be seen as an improvement of the object paradigm.

4.1. Conceptual models

4.1.1. Domain model

• *Description*

Conception of the domain model is the responsibility of the domain expert but ought to be kept accessible to all model contributors. An ontology representing all the entities accounted for and their interactions is an adapted representation is accessible enough. Nevertheless, such a basic representation cannot be sufficiently

informative thus it is complemented by two types of description. First, a textual description following the *Overview Section* of the O.D.D. protocol should be written. Second, descriptions of the reference system in domain experts own formalisms should be included (potentially as references). Many different domains are likely to be involved in the representation of the reference system. Thus, instead of one model in one domain expert's formalism, several sub-models in different formalisms may be included. These three elements (ontology, textual description and domain specific sub-models) constitute references for the subsequent models (*i.e.* design, operational and implemented).

- ***Example***

Figure 5.3 presents an ontology example derived from the application (Chapter 4). In this diagram, there are three entities (market, trader, farm and poultry) and three types of environment (pond, rice, inner village) where flocks wander. In addition, two entities initially considered have been finally declared as irrelevant (red crosses).
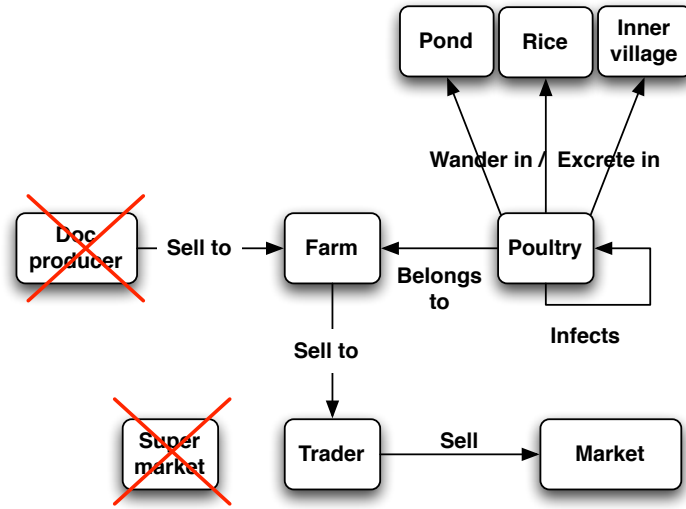
Fig.5.3 A ontology example derived from the application

From this ontology, a sub-model example could be the epidemiological one (transmission from a poultry to another). To represent it a domain expert, an epidemiologist in the present case, would use a compartment based model (Chapter 1) and expresses it as a ordinary differential equation system as follows:

$$
\begin{cases}
\dfrac{dS}{dt} & = -\beta SI \\[2ex]
\dfrac{dI}{dt} & = \beta SI - \nu I \\[2ex]
\dfrac{dR}{dt} & = \nu I
\end{cases}
$$

An example of short textual documentation structured as a *Overview section* of the O.D.D. document would be as follows.

- •*Purpose*: Purpose of the model is to understand the interaction between direct transmission of the virus and environment mediated-transmission.

- •*Entities, state variables, and scales*: Poultry have an infectious status and belongs to a farm. Traders own flocks temporarily. The environment has a virus concentration. The reference system spatial scale is set after a village, a few square kilometres wide, and the resolution is 10 meter. The temporal scale is 1 year and the time step is 1 hour.

- •*Process overview and scheduling*: Poultry can infect each other and recover from the disease. They excrete and get infected through the environment. They stay at night in the farm and wander in the environment

134

during the day. A trader buy poultry from a farm and sell it at the market once a week. The environment depletes its virus concentration daily.

4.1.2. Design model

• **Description**

According to **[Dro02]**, the design model is the first formal model and is designed by the modeller with support of the domain expert. It can be represented as an UML class diagram. Each entity considered in the domain model is represented with its behaviour (*i.e.* element of a dynamic), their internal states (variables and constants), their interactions (including multiplicity). All these elements are derived from domain expert micro-knowledge expressed in his/her own domain formalism (as included in the domain model). If required, a textual description of each conceptual agent (and its interactions) can be integrated in the *details Section* of the O.D.D. document.

• **Example**

Figure 5.4 provides a example of a UML class diagram of the reference system of the previous example (Section 4.1.1).
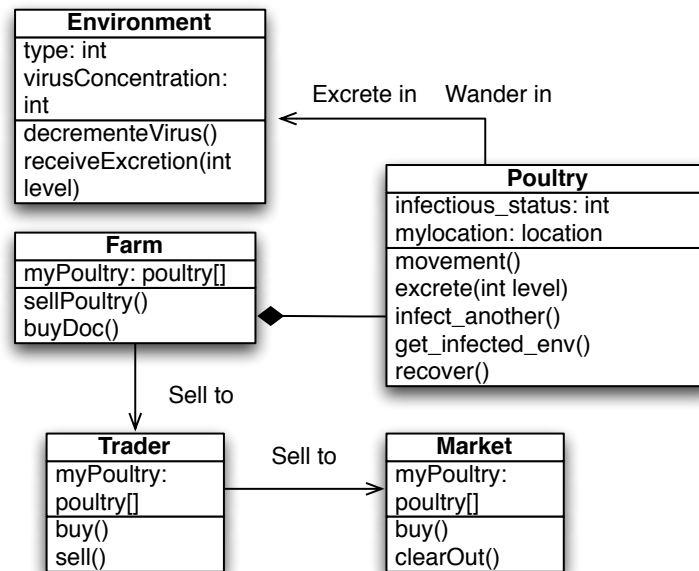
Fig.5.4 UML class diagram representation example

4.1.3. Operational model

• **Description**

The operational model can be expressed with three UML diagrams: object, activity and state transition diagrams. The object diagram provides examples of relationships between entities (including the environment). It allows clarification of specific situations of the reference system that have to be represented in the model. The activity diagram represents the exact schedule of action (and interaction) of one agent at a time. The state transition diagram represents the "life-cycle": The agent has a set of states and transitions between them that are triggered by internal and/or external conditions. Finally, the class diagram (previously defined) should be extended in order to integrate the relation with data files used by the simulation. Although, it is largely dependent on the selected ABMS platform, the data (e.g. aggregated or not), and the data file formats.

• **Example**

Figure 5.5 provides a example of object diagram representing a situation where a farm belonging to "Nguyen" which has three poultry presently located in three different environments.

135
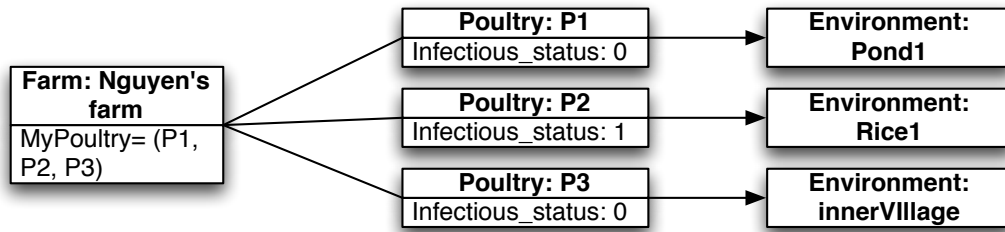
Fig.5.5 UML object diagram representation example

The activity diagram is illustrated in Figure 5.6 with the movement behaviour of a poultry. As we can see, the poultry wanders during the day, comes home at night and stay there (until it is day again).
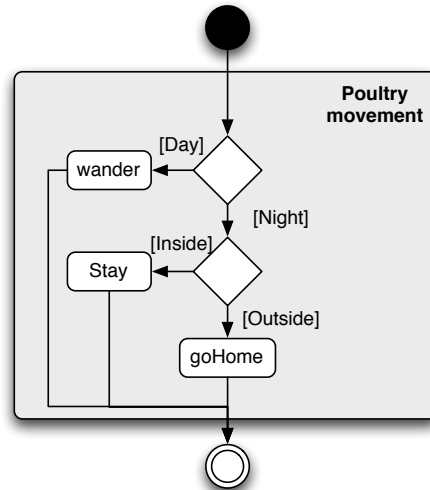
Fig.5.6 Activity diagram of a poultry movement

4.2.Implemented models

Most of the ABMS platforms allows the translation of conceptual agents to implemented agents in all their complexity except for the environment. Indeed, the environment implementation is generally simplified and its complexity unaccounted for. I propose to consider the implementation of the environment in two steps: Its internal state (and structure) can be represented thanks to a Geographical Information System (GIS) and its dynamic as a set of agents (each one representing an area). Practically, an ABMS platform instantiates "environmental agents" using the data contained in a GIS whereas their dynamics are defined in the implemented model (source-code). In addition, the GIS can be used for any spatial operation (e.g. translation, deformation, shortest path computation, etc.)

Another important advice for the implementation of the model is the use of field data, geo-referenced (as GIS) ones in particular. Using aggregated (and/or normalised data) is a common modelling practice. Unfortunately, such data already suffers from a first processing (and/or analysis) of the system made by a "specialist" for a given purpose before it can be integrated in the model. Consequently, I advise to use raw field data or statistical distribution if possible as they are usually not processed yet.

For instance, the infectivity of the avian influenza virus was defined after laboratory experiments which defined the number of viral particle to infect 50% of a population of birds (BID50). A probability parameterised by a number of viral particle ingested would have been much more relevant. Although, in this case it is not possible. Moreover, the recourse to R0 is even more emblematic of this situation: an single aggregated variable define the outcome of an epidemic (in a compartmental model).

In practice, I relied on the GAMA simulation platform **[Amo09, Tai10a]** because it is the only one addressing all the implementation requirements (P1-4, expressed in Chapter 2, Section 3.3.1) though upcoming adapted ABMS platform may be available later. Such a platform would only have to address requirements (P1-6) defined in Chapter 2 (Section 3.3.1). P2, facilitating the domain expert manipulation

136

(and access) to the implemented model is generally not considered in the "ABMS platform designer community", but is very important in a world model context. Nevertheless, KIMONO takes a great care of diminishing any uncertainty about a reference system at any modelling step because initiator's and domain expert's confidence in the model is crucial. In order to ease the domain expert's understanding of the simulator, having an unaccessible implemented model is already a wrong move as it increases the overall uncertainty level. In practice, this has been addressed through the use of GAML which is a modelling language designed to ease all model contributors access to the implemented model (Chapter 2).

For instance, I conducted a few sessions of GAML code reading with Dr. Desvaux which resulted in implemented model's small correction, in particular the behaviour scheduling of the flocks. In addition, her comprehension of the implemented and thus her confidence in it increased.

5. Support tools for the documentation

Model documentation facilitates access of all the contributors to the domain model and every models of each modelling iteration. In practice, each model iteration is described in an O.D.D formatted document. This document includes the ontology and UML diagrams representing the conceptual models. The class diagram is included in the *purpose* subsection. Object diagrams are included in the *Entities, state variables, and scales* subsection. Activity and state transition diagrams are included in the *Process overview and scheduling*. Similarly, the model source-code or implemented model should be included in the *Details* section.

In addition to the documentation of each model iteration, a global document, which can be called a "model road-book", keeps track of the model evolution. This document references each significant modification of the model in terms of hypothesis and modelling alternative evaluation but also in term of increased or deceased representation completeness. KIMONO includes incremental design of the model but also allows decrement of the model due to lack of data, for instance, thus this evolution have to be accounted. A model clade diagram, similar to the one presented in Chapter 1, should be sufficient to represent the chronological evolution of the model.For each iteration a description and justification is included as well as a reference of the iteration's O.D.D. document. Any documentation of the (latest) model is constituted of the latest O.D.D. and the model road-book.

6. Conclusion

KIMONO is a modelling method facilitating conception of world models and more generally exploratory and explanatory oriented models. It is composed of four main elements: Definition of the contributor roles, a modelling cycle, modelling guidelines and documentation. The conception of a world model is required by the initiator research question because it belongs to exploratory or explanatory objectives. Domain experts provides micro and macro-knowledge required to design the model whereas the modeller and the implementer support them for the design of conceptual and implemented model respectively. The modelling guidelines list the tools that facilitate the model conception: ontology and UML for conceptual models and an adapted ABMS platform (GAMA in my context) for the implemented models. The documentation is constituted of a set of O.D.D. documents complemented by a road book keeping track of the model evolution. Finally, the modelling cycle is decomposed in two phases: Initialisation of the process and iterative modelling. These two phases are decomposed as illustrated in Figure 5.7. This figure also presents the contributor involved in each step and which tool is used during the step. The "*" denotes the most involved contributor of the step. The iterative phase and the modelling cycle itself end when the initiator's research question have been addressed. Nevertheless, the produced model can be re-used. If a new research question is defined with a similar reference systems the adaptation of the model is possible. Indeed, having expressed the model in the agent meta-model facilitate its adaption as any object-oriented and program. In addition, the O.D.D. documentation provides all the complementary information to adapt the model. In particular, the representation choices are justified in this document.

Conrtibutor	Modelling step	Tool
Initiator	**Research question**	*Informal text*
Initiator *Domain experts*	**Reference system description**	*Informal text*
Domain experts* *Initiator* *Modeller*	**Domain model**	*Ontology* *O.D.D overview* *Submodels*
Domain experts* *Modeller*	**General data assessment**	*Listing* *(+ references)*
Domain experts *Modeller*	**Iteration objective and data assessment**	*Informal text*
Modeller* *Domain experts*	**Design model**	*Class diagram* *(+ explanation)*
Implementer* *Modeller*	**Operational model**	*Object, activity and state-transition diagrams*
Implementer* *Modeller*	**Implemented model**	*GAML*
Domain experts* *Implementer* *Modeller*	*Execution and analysis*	*GAMA*
Initiator	*Question addressed?*	

No

Yes

Fig.5.7 KIMONO complete modelling cycle, contributors and tools

138

This method has been applied (and refined), during the development of the GAMAVI project. Even though this application does not prove the generality of the method, the modelling evolution towards "world models" (Chapter 1) tend to support the generality of the method. Nevertheless, several issues remains such as the required implication of possibly numerous domain experts and the maintenance of a large ODD documentation. Finally, the method application has been permitted by the use of a specific ABM simulation platform, GAMA **[Amo09, Tai10a]**, but it can be generalised to other platforms. In addition, the GAML language is XML based which slows done the reading process and impedes domain experts access to it. Although, an improved modelling language based on Xtext (a domain specific language generator) **[Eff06]** is under preparation for the next version of GAMA (1.4).

Discussion

Elicitation of the domain model is a crucial step in the KIMONO method. This model is considered to be a landmark all along the modelling process but it may be amended through feedbacks from the iterative modelling. Presently, the domain model elicitation rely on the collaboration between domain experts and the initiator with support of the modeller. The process itself is the progressive design of an ontology that have to be consensual. Proposing a more precise organisation and more detailed practices would certainly ease the process. In addition, a way to derive very quickly an implemented model prototype from the domain model would certainly enhance the stability of the domain model by putting into the light incoherencies not easily observable within the domain model.

Even though GAMA provides a modelling language, GAML, accessible to domain experts, the implemented support is usually required. Moreover, this " proxy access" to the implemented model doest not completely the objective of having an implemented model accessible to all model contributors, domain experts especially. Indeed, programming the implemented model in GAML requires minimal knowledge of object-oriented programming. A more adapted modelling language or even a different mean of programming would certainly be profitable.

Chapter VI : Conclusion and perspectives

1. General conclusion

Since a few decades, the modelling activity is evolving in terms of objective assigned to models, modelling approach and representation. I have characterised this modelling evolution, and illustrated it with three thematically different examples in Chapter 1. The initial point of this evolution can be named the "*KISS*" modelling principle as proposed by Axelrod **[Axe97]**, while the latest characterised state have been defined by Edmonds *et al.* and named the "*KIDS*" modelling approach **[Edm04]**. I demonstrated that these shifts of objective, approach and representation are currently taking place in Epidemiology and can be foreseen to have large impact in this domain especially. Indeed, as shown by several recent examples **[Bar08, Lin09, Lap09]**, these new models allows the epidemiologists to address new questions. Unfortunately, no method to design and implement these models have been proposed yet and, even worse, these models have not been even clearly characterised.

In Chapter 2, I proposed a definition of these innovative models as "*world models*": They are oriented towards an exploratory or explanatory objectives. Their modelling process is incremental and results in a model which has a structure equivalent to the reference system one (at a given observation scale). I have also shown that these world models are the results of the co-evolution of the modelling technics and the research questions. Nevertheless, characterisation of these models is not sufficient, a modelling method had to be proposed. Thus, I elicited the methodological requirements to conceive world models after their own characterisation. A complete modelling method requires the definition of the model contributor roles, modelling cycle, modelling guidelines and support tools. The innovative model examples, presented in Chapter 1, appeared to have some insights on the matter: They are all expressed as agent-based models, the environment is represented in details thanks to the use of G.I.S. and the O.D.D. protocol is used to facilitate access to the model.

I reviewed several methods from classical software engineering, agent-oriented software engineering (AOSE) and, finally, modelling methods specifically dedicated to complex system. The first ones are relevant as the agent paradigm can be considered as an improved of the object paradigm. Nevertheless, only the recent agile methods appeared to have practical insights. Even though, AOSE methods are more adapted to the design of agent-based models but they only provide only general modelling guidelines. Finally, the complex system modelling methods, provide basis for the definition of contributor roles, the modelling cycle and introduce several support tools. After highlighting the elements that could be re-used and adapted, I presented the set of selected tools the modelling method would profit from. Concretely, conceptual models can be expressed using ontologies and more formal UML diagrams. while the implemented model can expressed in GAML because it is a modelling language accessible to domain experts. The execution of this model can be conducted on the GAMA platform **[Amo09,Tai10a]** which allows the implementation of any world models thanks to its primitive library, its GIS support and extensibility. Finally, I selected the O.D.D. communication protocol **[Gri10]** as the best way to document the modelling process and the model itself. It also has the advantage of becoming a standard tool to describe models in the individual-based model community (IBM), especially in the ecological IBM community.

In Chapter 3, I introduced the applicative context of my thesis. The development of a model to investigate avian influenza local dynamics in North Vietnam. This project was called GAMAVI and has been used to evaluate the modelling method I propose in this theses. This application has been conducted in close collaboration with Dr. Desvaux who provided the initial research question. Consequently, she acted as the initiator of the modelling process but also as a domain expert among other domain experts supporting the development of the model. I acted as the modeller and implementer with support from Dr. Duboz and Pr. Drogoul. The thematic background introduces the avian influenza disease and the Vietnam situation are described. The general objective of the modelling collaboration is expressed: investigating the local persistence and propagation of the avian influenza in the North Vietnam context. Then, existing avian influenza models have been reviewed in order to assess their adaptation to the present application.

The modelling process of the GAMAVI project have been introduced in Chapter 4. It started by refining the thematician initial research question, originally it can be summarised as "*what are the environmental and human factors determining the introduction and dissemination of the Avian influenza in North Vietnam?*" **[Amo07]** (personal translation). At the end of our collaborative modelling it has been rephrased as: "*the mechanisms, in conjunction with the environment, that enables the persistence and the propagation of the H5N1 virus at a local scale (village or district)*" (Annex 2). I detailed the research question but also the description of the reference system. From this informal description, domain experts elicited a first domain model, presented as on ontology, with my support. This domain model evolved greatly with feedbacks from the subsequent modelling step: iterative modelling. Consequently, I presented the final domain model, its documentation and its sub-models. These sub-models were used to structure the general data assessment and the iterative modelling. Each iteration has been presented. The presented iteration sequence allows a progressive description of the model. For instance, the commercial system initially envisioned as a key element of the reference system could not be represented. In addition to the modelling cycle, two key elements allowed the development of the GAMAVI model: The GAMA platform for the model implementation and the O.D.D. protocol for the model documentation.

This application has been used to verify the applicability of the modelling method I propose in this thesis, but also to collect feedbacks from domain experts. Nevertheless, a clear and generalised presentation as yet to be made. Consequently, I detailed the proposed modelling method, called KIMONO (*Keep It Modular, Open-ended, and Neatly Organised*) in Chapter 5. This method is constituted of four elements: contributor roles, modelling cycle, modelling guidelines and support tools which I consider to be mandatory to propose a complete modelling method.

- The *contributor roles* are: the initiator (who provides the research question), domain experts (who provides knowledge about the reference system), the modeller (who transforms experts knowledge in an agent-based model) and the implementer (who translates a conceptual ABM into an executable one and ensures their coherency).

- The *modelling cycle* is decomposed in two phases: Initial representation and iterative modelling. The initial representation requires a formalisation of both the research question and the reference system which result to the elicitation of a domain model. From this representation, a general data assessment is conducted to list all the data sources. The domain model is stable but can be questioned through feedbacks of the subsequent phase. The iterative modelling allows the incremental creation of the target model. Each iteration is decomposed in four steps. The first step is to define the objective of the iteration and conduct a specific data assessment to verify the actual accessibility of the required data. Then, a design model is conceived using the (global) domain model and the data. This model formally represents a subset of the reference system. The operational model is a translation of the previous one that allows to take into account all technical limitations that may appear in the implemented model. Finally, the implemented model is defined in a given modelling language.

- The *Modelling guideline*s is a set of tools and practices that facilitate the first description of the reference system, conceptual models and the implemented model. Especially, conceptual models can be expressed using different UML diagrams whereas the implemented model can be expressed in GAML and executed thanks to the GAMA platform.

- *Documentation* of the model is structured as defined by the O.D.D. protocol. Each modelling iteration should have its own O.D.D. formatted document which includes conceptual models and references to the implemented model source and to previous iterations documents. The chronology of modelling iteration should be recorded in a textual document keeping track of all modelling iterations and feedbacks to the domain. In particular, this document contain the justification of any modification of the domain model.

2. Perspectives

2.1. Application perspectives

The GAMAVI modelling process was used by the Dr. Desvaux to formalise her reference system in details and to orient her research and especially the field study. Nevertheless, the model could be improved in order if more data become available. In particular, the environment-mediated transmission is not yet well parameterised. It would require much field study and laboratory experiments to obtain adequate data. Moreover, the commercial system that was initially envisioned to be represented could not due to lack of data.

This model can also be re-used in as a basis to represent a different reference system. For instance, the south vietnamese poultry production sector is not so different from the north one. The village topology are largely similar whereas the village farms and flocks population would mostly differ only in terms of flock type proportions and size. Nevertheless, the (climatic and production) seasonalities only differ in terms of parameter values. The most important difference would certainly be the representation of the transhumant ducks (i.e. duck flocks are carried from village to village during the rice harvesting period). In addition, the model may be adapted to represent similar reference system (poultry production at the village scale) in other areas of South East Asia depending on the production practices.

2.2.KIMONO enhancements

2.2.1.Domain model elicitation

The conception of the domain models is a crucial step. For this reason, I proposed to precede it with a textual description of the reference system but improvements could be proposed. It is advisable to re-use the elicitation process proposed in ARDI **[Eti11]** (reviewed in Chapter 2) to iteratively define the ontology representing the reference system. The ARDI method considers four elements to represent a reference system: *Actors*, *Resources*, *Dynamics* and *Interactions*. In the KIMONO context, the concept of resources is not relevant. Consequently, only three steps of the ARDI modelling cycle are kept and adapted:

(A) During the "*Actor*" step, a list of all the entities and of the environment (spatial units) of the reference system is collaboratively proposed. Elements can be added or discarded through consensus of the model contributors.

(D) During the "*Dynamics*" step, all the processes that drive change in the reference system has to be listed. If there are too many of them, the model contributors can limit the list to those they consider the most important ones.

(I) During the "*Interactions*" step, "actor" are connected to each other through their interaction. If an "actor" remains without any interaction, it should be discarded because it shows that this actors actions are not affecting the reference system.

The ARDI method proposes to organise a workshop for each steps of the modelling cycle. Instead, I consider that a workshop should go through the three phases and also address the listing of the sub-models and the data assessment of the KIMONO modelling cycle. Nevertheless, several workshops should be conducted to allow incremental definition of the domain model and data assessment in order to obtain a definition as stable as possible of the domain model. To structure the workshops ARDI proposes to consider three roles: facilitator, observer and (reference system) stakeholders. The facilitator manages the process and facilitate general agreements whereas the observer record the collaborative modelling process (e.g. keep tracks of modelling choices, hypotheses, justification, etc). I consider that both these roles should enacted by the modeller. Finally, the stakeholders are providing the knowledge about the reference system. In the KIMONO context, they are replaced by the domain experts.

2.2.2.Easing domain experts access to the implemented model

Access to the implemented model by domain experts requires the support of the implementer. Thus, GAML does not satisfy fully the general requirement: The domain experts are at the centre of the modelling process. The active collaboration of domain experts on the development of the implemented model is still impossible unless they have strong object-oriented programming skills.

Several attempts to allow the programming of the implemented model by domain experts have been made. The NetLogo simulation platform **[Wil99]** has been conceived with an educational purpose in mind . It proposes a simplified modelling language (based on Logo) which limits the amount of training required for a domain expert to be able to develop its own implemented model on this platform. Unfortunately, the technical capabilities of this platform are limited **[Rail06]**. For instance, the management of GIS data have been recently added but is still very limited. Other Logo based simulation platforms **[Gau11]** made another choice to facilitate the development of the implemented model: Graphical modelling. StarLogo TNG is a Logo based simulation platform which proposes to developed the implemented model by concatenating graphical blocks representing structure controls or statements "*But this graphical programming becomes quickly heavy compared to classical free textual programming*" according to **[Gau11]** and its technical capabilities are limited. Finally, the latest version of Repast (Repast Symphony **[Nor06]**) proposes to ease the implemented model development thanks to an "agent-based modelling domain specific language" (ABM DSL) called "ReLogo". Unfortunately, the platform was not fully functional at the time of this study.

2.2.3. Collaborative O.D.D and Δ-ODD

O.D.D. can be considered to be a heavy protocol to describe every modelling iteration. Moreover, the frequent update that should be made to these descriptions during the iterative modelling requires extensive time to be completed. Grimm *et al.* already highlighted this issue in **[Grim10]** but only characterised the need for a "Δ-ODD" which would contains the change from a previous ODD description.

No proposition for a Δ-ODD has been made yet but a workaround can be proposed. A text editor could be adapted to generate O.D.D. formatted documents, its structure and the relationships between the different sections. For instance, defining a new entity would generate a characterisation stub including its variables and processes in the *Overview* section and the related initialisation in the *Details* section. Updating, a previous O.D.D. would only require to modify the required section, keeping references to other files such as sub-models. Finally, an automated coherency check could be envisioned.

Fortunately, a specific O.D.D. editor is under development by the authors of **[Ngu09]** (personal communication). This editor is based on *etherpad*[15] and allow the collaborative writing of an O.D.D document. In addition, it would be possible to extend the software to allow the definition of a Δ-ODD as this editor already supports versioning. For even more convenience, this editor should be extended to facilitate the O.D.D. description of the global domain model. It could automatically, list the parameters, entities, spatial units, etc. listed in each sub-models and add them to the global description. Although, a mechanism to "intelligently" filter the relevant elements should be provided, for instance some sub-models parameters values are defined by other sub-models. Such editor with a convenient versioning system would even ease the process of the re-assessment of the whole modelling if the outcome of the process result in such situation.

2.2.4. Distant collaboration

In this thesis, I did not consider the specificities of the distant collaboration. Nevertheless, it is a common and challenging situation. For instance, the development of the Sworm model (Chapter 1 Section 1.3) expressed the need of online collaborative tools. KIMONO does not consider this issue but, as the adaptation of the ARDI workshop organisation does not alter the method, considering the distant collaboration should not modify it. Moreover, specificities of distant collaboration oriented (modelling) methods should not conflict with a "non-distant" methods as both address very different issues.

The Sworm original modeller proposed recently a collaborative platform called PAMS **[Ngu09]**. This platform is not a complete distant modelling collaboration platform yet. It focuses on executing a simulation on a server and, analyses and amends it remotely in collaboration with other model contributors. For instance, the figure 6.1 presents its interface where we can see two scientists analysing a simulation through

---

[15] http://code.google.com/p/etherpad/

various outputs (visualisation, graph, monitor, etc) and using different communication tools (video conference, white boards, etc.).
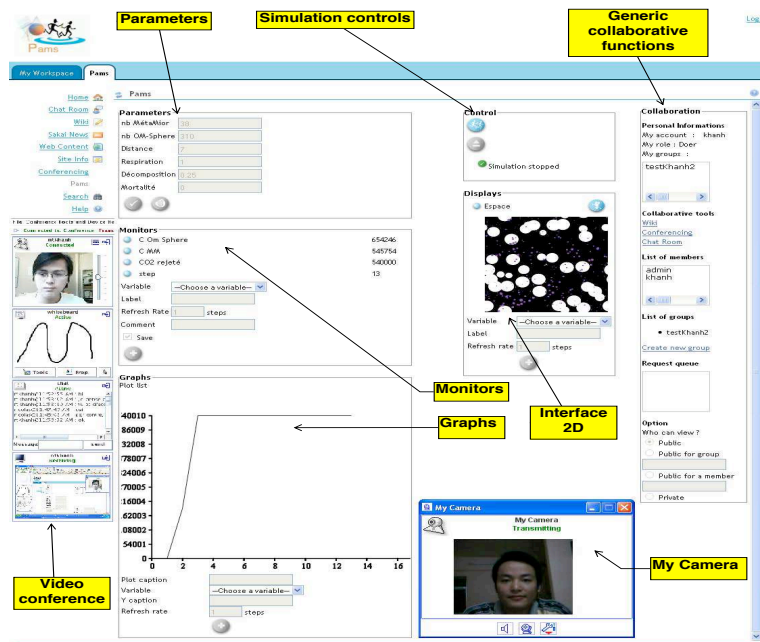
Fig.6.1 A collaborative simulation using the PAMS platform [Ngu09]

PAMS, despite its simulation analyse orientation, provides generic tools that can be used during other stages of the modelling process: synchronous tools (chat, whiteboard, desktop sharing) and asynchronous tools (forums, wikis). PAMS was built over a generic distant collaboration platform, SAKAI [Yan07]. It seems reasonable to think that a similar modelling platform specifically targeting distant collaborative modelling could be built similarly. Extending the PAMS platform may be a better solution than developing a distant collaboration modelling platform from scratch because the simulation analyses is already developed and would be interesting for the KIMONO method. Actually, the authors are the same as the collaborative O.D.D. editor prototype and integration of both projects would lead to to a complete distant modelling collaboration platform.

Bibliography

1. [Ack05] Ackleh A.S., Allen L.J.S.. *Competitive exclusion in SIS and SIR epidemic models with total cross immunity and density-dependent host mortality*. Discrete and Continuous Dynamical Systems Series B (2005) vol. 5 pp. 175–188

2. [Agr07] Agrifood consulting international. *The Impact of Avian Influenza on Poultry Sector Restructuring and its Socio-economic Effects*. (2007) pp. 78

3. [All09] Allan R. *Survey of agent based modelling and simulation tools*, Technical Report, DL-TR-2010-007, Science and Technology Facilities Council (2010)

4. [Amo07] Amouroux E., Boucher A., Drogoul A., Desvaux S. *Collaboration CIRAD – MSI pour l'étude et la modélisation de la grippe aviaire au nord Vietnam*, collaboration convention, pp.4, 2007

5. [Amo08] Amouroux E., Desvaux S., Drogoul A., *Towards virtual epidemiology: an agent-based approach to the modeling of H5N1 propagation and persistence in North-Vietnam*. Intelligent Agents and Multi-Agent Systems. Eds.: The Duy Bui, Tuong Vinh Ho, Quang Thuy Ha. LNAI 5357. Springer-Verlag (2008) pp. 26-33

6. [Amo09] Amouroux E., Chu T.Q., Boucher A., Drogoul A. *GAMA: An Environment for Implementing and Running Spatially Explicit Multi-agent Simulations*. Agent Computing and Multi-Agent Systems. Eds.: Aditya Ghose, Guido Governatori, Ramakoti Sadananda. LNAI 5044. Springer-Verlag (2009) pp.359-371

7. [Amo10a] Amouroux E., Taillandier P., Drogoul A. *Complex environment representation in epidemiology ABM: application on H5N1 propagation*. Journal of Science and Technology – Vietnamese Academy of Science and Technology (2010) pp. 1-12

8. [Amo10b] Amouroux E., Gaudou B., Desvaux S., Drogoul A. *O.D.D.: a Promising but Incomplete Formalism For Individual-Based Model Specification*. Proceedings of RIVF'10 international conference, IEEE (2010) pp. 1-6

9. [Art01]Artzrouni M., Gouteux J.P. *Population dynamics of Gambian sleeping sickness: A micro-simulation*. Simulation & Gaming (2001) vol. 32 (2) pp. 215-227

10. [Aum07] Aumann C.A. *A method for developing simulation models of complex systems. ecological modelling*. Ecological modelling. (2007) vol. 202 (3-4) pp. 385-396

11. [Axe97] AXELROD R. *The Complexity of Cooperation: Agent-based models of conflict and cooperation*. Princeton University Press, 1997.

12. [Bau01] Bauer, B., Mller, J.P., Odell, J.: *Agent UML: A formalism for specifying multiagent interaction*. In: AOSE, pp. 91–103. Springer, Berlin, Germany (2001)

13. [Bar03] Barabasi A.L, Bonabeau, E., *Scale-free networks*. Scientific American, 288, 60-69 (2003).

14. [Bar04a] Barrett C., Eubank S., Kumar V. *et al. Understanding large-scale social and infrastructure networks: a simulation-based approach*. SIAM news (2004) vol. 37 (4) pp. 1–5

15. [Bar04b] Barry J.M.. *The site of origin of the 1918 influenza pandemic and its public health implications*. Journal of Translational Medicine (2004) vol. 2 (1) pp. 3

16. [Bar05] Barrettt C., Eubank S., Smith J. *If Smallpox Strikes Portland...* Scientific American (2005) vol. 292 (3) pp. 54-61

17. [Bar08] Barrett C., Bisset K., Eubank S., *et al. EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks*. Proceedings of the 2008 ACM/IEEE conference on Supercomputing (2008) pp. 1-12

18. [Bau01] Bauer B., Müller J.P. , Odell J. *Agent UML: A Formalism for Specifying Multiagent Interaction*. AOSE, Eds.: Ciancarini P. and Wooldridge P.., Springer-Verlag, Berlin, 2001, pp. 91-103.

19. [Bec01] Beck K. . *Manifesto for Agile software development*. Agile Alliance. (2001)

20. [Bel05] Belshe R.B. *The origins of pandemic influenza, lessons from the 1918 virus*. N Engl J Med. 2005, Vol. 353, 2209-2211.

21. [Ben85] Ben-Akiva, M., Lerman S.R. . *Discrete Choice Analysis, Theory and Application to Travel Demand*. Cambridge, MA: MIT Press (1985).

22. [Ben99] Bénédicte, S,J.: *Atlas Bac Hung Hai*, Gret (personnal communication) (1999)

23. [Ben05] Bennett S., McRobb, S., Farmer R. *Object-Oriented Systems Analysis and Design Using UML 3rd ed.* London: McGraw-Hill. ISBN: 0077110005 (2005)

24. [Ben07] De Benedictis P., Beato M.S., Capua I. *Inactivation of avian influenza viruses by chemical agents and physical conditions: a review*. Zoonoses and Public Health (2007) vol. 54 (2) pp. 51-68

25. [Ber03] Bernon C., Gleizes M.P., Peyruqueou S., Picard G. *ADELFE: A method for adaptive multi-agent systems engineering*. Engineering Societies in the Agents World III (2003) pp. 70-81

26. [Ber68] Bertalanffy L. V., *General Systems Theory, Foundation, Development, Applications*. G. Braziller. 1968

27. [Bjo02] Bjornstad,O.N., Finkenstädt, B., Grenfell, B.T. *Endemic and epidemic dynamics of measles: estimating epidemiological scaling with a time series SIR model*. Ecol. Monogr. Vol.72, pp.169-184. 2002

28. [Bla09] Blanchart E. Marilleau N., Chotte J.L. *et al. SWORM: an agent-based model to simulate the effect of earthworms on soil structure*. European Journal of Soil Science (2009) vol. 60 (1) pp. 13-21

29. [Boe86] Boehm B. *A Spiral Model of Software Development and Enhancement*, ACM SIGSOFT Software Engineering Notes, ACM, Vol. 11(4), pp. 14-24, 1986

30. [Bou98] Bousquet F., Bakam I., Proton H., Le Page C. *Cormas: common-pool resources and multi-agents systems*. IEA/AIE. Vol. 2, pp. 826--837 (1998)

31. [Bou05] Bousquet F., Trébuil G. *Introduction to companion modeling and multi-agent systems for integrated natural resource management in Asia*. Companion Modeling and Multi-Agent Systems for Integrated Natural Resource Management in Asia (2005) pp. 1-20

32. [Bou09] Bouma A., Claassen I., Matih K. *et al. Estimation of transmission parameters of H5N1 avian influenza virus in chickens*. PLoS Pathogens (2009) vol. 5 (1). pp1-13

33. [Bow10a] Bowong S., Tewa J.J. *Mathematical analysis of a model for hepatitis B-malaria co-infection*. UMMISCO research team seminar (2010)

34. [Bow10b] Bowong S., Emvudu Y, Moualeu D.P., Tewa J.J.. Mathematical properties of a tuberculosis model with two differential infectivity and N latent classes. J. of Nonlinear Sys. and App. (2010) pp.13- 26

35. [Box02] Box, P. *Spatial Units as Agents. Integrating GIS and Agent-Based Modelling Techniques*. ed. Oxford. 2002.

36. [Bre02] Bresciani P., Giorgini P. Giunchiglia F. *et al. Tropos: An agent-oriented software development method*. Technical Report # DIT-02-0015 2002

37. [Bro06] Brown C.. *Avian influenza: Virchow's reminder*. The American j. of pathology (2006) vol. 168 (1) pp. 6

38. [Bro08] Brown J.D., Goekjian G., Poulson R. *et al. Avian influenza virus in water: Infectivity is dependent on pH, salinity and temperature*. Veterinary Microbiology (2008) pp. 1-7

39. [Bro10] Brown I.H. *Summary of avian influenza activity in Europe, Asia, and Africa, 2006-2009*. Avian diseases (2010) vol. 54 (1) pp. 187-193

40. [Cas06] Castle C. et Crooks A. *Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations* . CASA working papers series (110) (2006)

41. [Cer05] Cernuzzi L. , Cossentino M., Zambonelli F. *Process models for agent-based development*. Engineering Applications of Artificial Intelligence. Vol.18 (2005) pp.205–222

42. [Cho08] Choisy, M. Guégan J.F., Rohani P. *Mathematical Modeling of Infectious Diseases Dynamics*. In Encyclopeida of infectious diseases: modern methods (2008) pp. 26

43. [Chu08] Chu et al. Interactive Learning of Expert Criteria for Rescue Simulations. Intelligent Agents and Multi-Agent Systems (2008) pp. 127-138

44. [Cia11] CIA. *CIA world factbook: Vietnam* . https://www.cia.gov/library/publications/the-world-factbook/geos/vm.html accessed 08/04/11

45. [Cro06] Crooks A.. *Exploring cities using agent-based models and GIS*. Proc. of the Agent 2006 Conf. on Social Agents: Results and Prospects (2006)

46. [Cro08] Crooks A., Castle C., Batty M. *Key challenges in agent-based modelling for geo-spatial simulation*. Computers, Environment and Urban Systems (2008) vol. 32 (6) pp. 417-430

47. [Cro10] Crooks A. *Constructing and implementing an agent-based model of residential segregation through vector GIS*. Int. J. of Geographical Information Science (2010) vol. 24 (5) pp. 661-675

48. [Cro89] Crosby A.W. *America's Forgotten Pandemic. The Influenza of 1918*. Cambridge: Cambridge University Press. (1989)

49. [Cro94] Cross S.S., Cotton D.W.K. *Chaos and antichaos in pathology*. Human Pathology Journal., 25:630–637, 1994

50. [Dep02] De Palma A., Marchal F. *Real cases applications of the fully dynamic METROPOLIS tool-box: An advocacy for large-scale mesoscopic transportation systems*. Networks and Spatial Economics (2002) vol. 2 (4) pp. 347-369

51. [Der75] De Rosnay J., *Le macroscope: vers une vision globale*. Seuil. 1975

52. [Des08] Desvaux S., Ton V.D., Thang P?D., Hoa P.T.T.. *A General Review and a Description of the Poultry Production in Vietnam*. Agricultural publishing house, Hanoi, 2008

53. [Des09] Desvaux S. *Poultry Production system description of the Ha Tay and Bac Giang provinces, North Vietnam*. Technical report. pp.41. 2009

54. [Doq05] Do Quy P. *Seroprevalence study and virus isolation of Avian Influenza in rural poultry of Thai Binh province*. MSc thesis (personal communication). (2005) pp. 74

55. [Dro92] Drogoul A., Ferber J. *Multi-Agent Simulation as a Tool for Modelling Societies: Application to Social Differentiation in Ant Colonies*. Proceeding MAAMAW '92 (1992)

56. [Dro02] Drogoul A., Vanbergue D., Meurisse T. *Multi-agent based simulation: Where are the agents?* Proceedings of MABS'02. (2002)

57. [Dun05] Dunham J.B. *An Agent-Based Spatially Explicit Epidemiological Model in MASON*. JASS (2005) vol. 9 (1) pp. 3

58. [Ecd10] ECDC "Daily Update: 2009 influenza A (H1N1) pandemic", 18/10/10, http://ecdc.europa.eu/en/healthtopics/Documents/100118_Influenza_AH1N1_Situation_Report_0900hrs.pdf

59. [Edm74] Edmonds E.A. . *A Process for the Development of Software for Nontechnical Users as an Adaptive System*. General Systems Vol.19, pp. 215–18. (1974)

60. [Edm04] Edmonds B, Moss S. *From KISS to KIDS: An" anti-simplistic" Modeling Approach*. Proceedings of MABS 2004 (2004) pp. 130-144

61. [Eff06] Efftinge S., Völter M. *oAW xText: A framework for textual DSLs*. Workshop on Modeling Symposium at Eclipse Summit (2006) vol. 32

62. [Eti11] Etienne M, Du Toit D.R. Pollard S. *ARDI: a co-construction method for participatory modeling in natural resources management*. Ecology and Society (2011) vol. 16 (1) pp. 44

63. [Eub04] Eubank S. Guclu H., Kumar A., *et al. Modelling disease outbreaks in realistic urban social networks*. Nature (2004) vol. 429 (6988) pp. 180-184

64. [Fer03] Ferber J., Michel F., Gutknecht O. *Agent/Group/Roles : Simulating with organizations*. Agent Based Simulation workshop 4, Montpellier, 28-30 Avril 2003.

65. [Fis95] Fishwick P. *Simulation model design and execution*. Prentice Hall (1995)

66. [Fon99] Fontenelle D., Hai M.V. *Transformation et permanence de l'espace habité dans un village du delta du fleuve Rouge (Việt-nam)*. Nature Sciences Sociétés (1999), Vol 7 (1), pp.5-20

67. [Fou10] Fournie G., Guitian F.J., Mangtani P., Ghani A.C. *Impact of the implementation of rest days in live bird markets on the dynamics of H5N1 highly pathogenic avian influenza*. Journal of The Royal Society Interface (2010) pp. 1-11

68. [Fra09] Fraser C., Donnelly C.A., Cauchemez S., *et al. Pandemic potential of a strain of influenza A (H1N1): early findings*. Science (2009) vol. 324 (5934) pp. 1557

69. [Fu03] Fu S. Milne G. *Epidemic modelling using cellular automata*. Proc. of the Australian Conference on AI (2003)

70. [Gai08] Gaidet N., Newman S.H, Hagemeijer W., *et al*. *Duck migration and past influenza A (H5N1) outbreak areas*. Emerging Infectious Diseases (2008) vol. 14 (7) pp. 1164

71. [Gai10] Gaidet N., Cappelle J., Takekawa J.Y., *et al*. *Potential spread of highly pathogenic avian influenza H5N1 by wildfowl: dispersal ranges and rates determined from large-scale satellite telemetry*. Journal of Applied Ecology 2010, Vol.47 (5), pages 1147–1157

72. [Gam11] GAMA project. Online collaborative documentation http://code.google.com/p/gama-platform/wiki/Documentation retrieved 01/03/11

73. [Gau07] Gaud N. Systèmes multi-agents holoniques : de l'analyse à l'implémentation. PhD thesis (2007) pp. 1-255

74. [Gau09] Gaudou B., Ho T.V., Marilleau N. *Introduce Collaboration in Methods of Modeling and Simulation of Complex Systems*. Proc. Proceedings of INCOS'09 (2009)

75. [Gau11] Gaudou B., Marilleau N., Ho T.V. *Toward a Method of Collaborative Modeling and Simulation of Complex Systems*. Proceedings of INCOS'11 (2011) pp. 27-53

76. [Gil99] Gilbert N., Troitzsch K.G. *Simulation for the Social Scientist*. Open University Press (1999)

77. [Gio04] Giorgini P., Kolp M., Mylopoulos J. *et al*. *The Tropos Method: An overview*. Multiagent Systems, Artificial Societies, And Simulated Organizations, 2004, Vol 11(2), pp.89-106

78. [Gli07] Glisse. *Exploration scientifique par la conception collaborative de systèmes multi-agents.* (2007). PhD thesis pp. 1-185

79. [Gra62] Graunt J. *Natural and Political Observations Made upon the Bills of Mortality* 1662

80. [Gre35] Greenshield, B.D. *A study of traffic capacity*. Highway research board proceedings Vol. 14, pp. 448-477, 1935

81. [Grim05] GRIMM. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science (2005) vol. 310 (5750) pp. 987-991

82. [Grim06] Grimm V., Berger U., Bastiansen F. *et al*. *A standard protocol for describing individual-based and agent-based models*. Ecological Modelling (2006) vol. 198 (1-2) pp. 115-126

83. [Grim09] GRIMM V. Berger U., De Angelis D. *et al*. *The ODD protocol: A review and first update*. Ecological Modelling (2010), Vol 221 (23), pp. 2760-2768

84. [Gso07] GSO, General Statistics Office. *Statistical yearbook of Vietnam 2006*. Statistic publishing house. Hanoi, Vietnam. 2007

85. [Hak01] Haklay M., O'Sullivan D., Thurstain-Goodwin M., Schelhorn T. *So Go Downtown": Simulating Pedestrian Movement in Town Centres*. Environment and Planning B: Planning and Design, Vol. 28 (3), pp. 343-359. 2001

86. [Har11] Harder T. C., Werner O. *Avian Influenza report* http://www.influenzareport.com/ir/ai.htm Accessed 18/02/11

87. [Hea10] Heath B.L. *The History, Philosophy, and Practice of Agent-Based Modeling and the Development of the Conceptual Model for Simulation Diagram*. PhD thesis. (2010) pp. 1-300

88. [Hen77] Henderson D.A. *Smallpox Eradication*. Proceedings of the Royal Society B: Biological Sciences (1977) vol. 199 (1134) pp. 83-97

89. [Het76] Hethcote. Qualitative analyses of communicable disease models* 1. Mathematical Biosciences (1976) vol. 28 (3-4) pp. 335-356

90. [Hil08] Hill D. et al. Design of a spatial and stochastic simulator for bird flu spreading in Corsica. International Simulation Multi-Conference. Grand Modeling Challange

91. [Hil09] Hill D. et al. Decision support system for a regional spreading of A/H1N1 influenza virus. IEEE/Eurosis : European Simulation Multi-Conference.

92. [Hul05] Hulse-Post D.J., Sturm-Ramirez M., Humberd J. *et al. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia.* PNAS (2005) vol. 102 (30) pp. 10682-10687

93. [Joh02] Johnson N.P.A.S, Mueller J. *Updating the accounts: global mortality of the 1918-1920" Spanish" influenza pandemic.* Bulletin of the History of Medicine (2002) vol. 76 (1) pp. 105-115

94. [Jon07] Jones. *Notes on R0.* (2007) pp. 19 http://www.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf (accessed 21/08/2010)

95. [Jou07] Jourdain E., Gauthier-Clerc M.G., Bicout D.J. *et al. Bird migration routes and risk for pathogen dispersion into Western Mediterranean wetlands.* Emerging infectious diseases (2007) vol. 13 (3) pp. 365

96. [Kam06] Kamps B., Hoffman C., Preiser W. *Influenza report 2006.* (2006) pp. 225 Flying publisher ISBN: 3-924774-51-X

97. [Ker27] Kermack W.O., McKendrick A.G. *A Contribution to the Mathematical Theory of Epidemics.* Proc. R. Soc. Lond. A. Vol.115 (772): pp. 700-721. 1927

98. [Ker32] Kermack W.O., McKendrick A.G. *Contributions to the mathematical theory of epidemics*, part. II. Proc. R. Soc. Lond. B Biol. Sci., Vol. 138, pp. 55-83, 1932

99. [Kim08] Kim T., Hwang W., Zhang A., *et al.. Multi-agent model analysis of the containment strategy for avian influenza (AI) in South Korea.* In proceedings of IEEE International Conference on Bioinformatics and Biomedicine (2008) pp. 335-338

100. [Kim10] Kim T., Hwang W., Zhang A., *et al. Multi-agent modeling of the South Korean avian influenza epidemic.* BMC Infectious Diseases (2010) vol. 10 (1) pp. 236

101. [Klo05] Klopfer, E., Begel, A. *StarLogo TNG: An Introduction to Game Development.* Journal of E-Learning (2005)

102. [Koo04] Koopman J. *Modeling infection transmission.* Annual review of public health (2004) vol. 25 pp. 303-26

103. [Kos96] Kosonen I.. *HUTSIM: A Simulation Tool for Traffic Signal Control Planning.* Helsinki University of Technology, Laboratory of Transportation engineering. Licentiate Thesis. 1996. 150 p.

104. [Kyr05] Kyrychko Y., Blyuss K.B. *Global properties of a delayed SIR model with temporary immunity and nonlinear incidence rate.* Nonlinear Analysis: Real World Applications (2005) vol. 6 (3) pp. 495-507

105. [Lap09a] Laperrière V. *Apport de la modélisation individu-centrée spatialement explicite à la compréhension de l'expression d'une maladie transmissible : la peste bubonique à Madagascar.* PhD thesis, (2009) pp. 1-296

106. [Lap09] Laperrière V., Badariotti, Banos A., Müller J.P. *Structural validation of an individual-based model for plague epidemics simulation.* Ecological Complexity (2009) vol. 6 (2) pp. 102-112

107. [Lau07] Lau, E. H. Y., Leung, Y. H. C., Zhang, L. J., Cowling, B. J., Mak, S. P., Guan, Y., Leung, G. M. & Peiris, J. S. M. *Effect of interventions on influenza A (H9N2) isolation in Hong-Kong's live poultry markets, 1999–2005.* Emerging Infectious Diseases. Vol.13, pp. 1340-1347. 2007

108. [Leb09] Lebarbenchon C., Albespy F., Brochet A.L., Grandhomme V., *et al. Spread of avian influenza viruses by common teal (Anas crecca) in Europe.* PloS one (2009) vol. 4 (10) pp. e7289

109. [Lem77] Le Moigne J.L., *La théorie du système général.* P. U. F.. 1977

110. [Lig55] Lighthill, M,J., Whitham G.B. *On kinematic waves: II. a theory of traffic flow on long crowded roads.* Proceedings of the royal society 229 (Series A) : 317-345 (1955)

111. [Lin09] Linard C., Ponçon N., Fontenille D., Lambin E. *A multi-agent simulation to assess the risk of malaria re-emergence in southern France.* Ecological modelling (2009) vol. 220 (2) pp. 160-174

112. [Liu99] Liu, R., Van Vliet D., Watling D. *DRACULA: A Microscopic, Day-to-Day Dynamic Modelling of Traffic Assignment and Simulation.* In Applications and Advanced Technology in Transportation Engineering. New York, ASCE Press, pp. 444-448. (1999)

113. [Luk05] Luke S., Cioffi-Revilla C., Panait L., Sullivan K.. *MASON: A multiagent simulation environment.* Simulation (2005) vol. 81 (7) pp. 517

114. [Lyn08] Lynch P. *The origins of computer weather prediction and climate modeling.* Journal of computational physics (2008) vol. 227 (7) pp. 3431-3444

115. [Mag10a] Magalhães R.J., Ortiz-Pelaez A., Thi K.L, *et al. Associations between attributes of live poultry trade and HPAI H5N1 outbreaks: a descriptive and network analysis study in northern Vietnam.* BMC veterinary research (2010) vol. 6 pp. 10

116. [Mag10b] Magalhaes R.J., Pfeiffer D., Otte J. *Evaluating the control of HPAIV H5N1 in Vietnam: virus transmission within infected flocks reported before and after vaccination.* BMC Veterinary Research (2010) vol. 6 (1) pp. 31

117. [Mah01] Mahmassani H.S., Pillai R., Stevens J. . *DYNASMART-P for Work Zone Planning.* Presented at the 80th Transportation Research Board Annual Meeting, Washington, D.C. (2001)

118. [Mar06a] MARD, Department of Livestock. *Duck production in 2001-2005 and its orientation in 2006-2015.* Hanoi, Vietnam, 2006

119. [Mar06b] MARD, Department of Livestock. *Chicken production in 2001-2005 and its orientation in 2006-2015.* Hanoi, Vietnam, 2006

120. [Mar06c] Marilleau N. *Méthodologie, formalismes et outils de modélisation-simulation pour l'étude des systèmes complexes: application à la mobilité géographique.* PhD Thesis. (2006)

121. [Mar08] Marilleau N., Cambier C., Drogoul A., *et al. Multiscale MAS modelling to simulate the soil environment: application to soil ecology.* Simulation Modelling Practice and Theory vol. 16 (7) pp. 736-745

122. [Mcc96] McConnel S. (1996). *Rapid Development: Taming Wild Software Schedules.* Microsoft Press Books, ISBN 978-1556159008. (2008)

123. [Min96] Minar N., Burkhart R., Langton C., Askenazi M. *The Swarm Simulation System: A Toolkit for Building Multi-Agent Simulations*, SFI Working Paper 96-06-042. 1996.

124. [Min09] Minh, P. Q., Morris R. S., Schauer B., *et al. Spatio-temporal epidemiology of highly pathogenic avian influenza outbreaks in the two deltas of Vietnam during 2003-2007.* Prev. Vet. Med. Vol.89(1-2), pp.16-24. (2009)

125. [Mos08] Moss S. *Alternative Approaches to the Empirical Validation of Agent-Based Models.* JASS (2008) vol. 11 (1) pp. 20

126. [Mul04] Muller G., Grébaut P., Gouteux J.P. *An agent-based model of sleeping sickness: simulation trials of a forest focus in southern Cameroon.* Comptes rendus Biologies. vol. 327 (1) pp. 1-11 (2004)

127. [Mul05] Müller J.P., Ratsé C., Gillet F., Stoffel K. *Modeling and simulating hierarchies using an agent-based approach.* in proceedings of MODSIM'05, (2005)

128. [Mur89a] Murray J. *Mathematical Biology: an introduction.* Springer-Verlag. ISBN 0387194606, pp. 767 (1989)

129. [Mur89b] Murata T. *Petri nets: Properties, analysis and applications.* Proceedings of the IEEE, Vol.77 (4). pp. 541-580, 1989.

130. [Nag92] Nagel K., Schreckenberg M. *A cellular automaton model for freeway traffic.* Journal de Physique I vol. 2 (12) pp. 2221-2229. (1992)

131. [Nei89] Neimark A.V. *Multiscale percolation systems*, Soviet Physics. JETP Vol.69 pp. 786-791. (1989)

132. [Ngu09] Nguyen T.K., Gaudou B., Ho T.V., Marilleau N. *Application of PAMS collaboration platform to simulation-based researches in soil science: The case of the micro-organism project.* In proceedings of IEEE-RIVF'09. (2009).

133. [Nor06] North M.J., Collier N.T., Vos J. R. *Experiences Creating Three Implementations of the Repast Agent Modeling Toolkit.* ACM Trans. on Mod. and Com. Sim. Vol. 16 (1). pp. 1-25, (2006)

134. [Nor07] North M.J., Tatara E., Collier N., T., Ozik J. *Visual Agent-based Model Development with Repast Simphony*, Proceedings of Conference on Complex Interaction and Social Emergence (2007)

135. [Oie11a] OIE (World Health Organisation for Animal Health). *Highly Pathogenic Avian Influenza outbreak map* http://web.oie.int/wahis/public.php?page=disease_outbreak_map . Accessed 07/04/11

136. [Oie11b] OIE (World Health Organisation for Animal Health). *Disease timelines: Highly pathogenic avian influenza* http://web.oie.int/wahis/public.php?page=disease_timelines Accessed 07/04/11

137. [Oie11c] OIE (World Health Organisation for Animal Health). *Detailed avian influenza incidence in Vietnam.* http://web.oie.int/wahis/public.php?page=disease_status_detail Accessed 07/04/11

138. [Pad01] Padgham L., Winikoff M. *Prometheus: A pragmatic method for engineering intelligent agents.* Proceedings of the OOPSLA 2002 Workshop on Agent-Oriented Methods (2002) pp. 97–108

139. [Pan07a] Pantin-Jackwood M.J., Swayne D.E. *Pathobiology of Asian highly pathogenic avian influenza H5N1 virus infections in ducks.* Avian Diseases (2007) vol. 51 (1) pp. 250-259

140. [Pan07b] Pantin-Jackwood M.J., Suarez D.L., Spackman E., Swayne D.E. *Age at infection affects the pathogenicity of Asian highly pathogenic avian influenza H5N1 viruses in ducks.* Virus research vol. 130. pp. 151-61. (2007)

141. [Pat91] Patterson K.D., and Pyte G.F. *The geography and mortality of the 1918 influenza pandemic.* Bull Hist Med. Vol .65 (1), pp.4-21. 1991

142. [Pav05] Pavon J., Gomez-Sanz J., Fuentes R. *The INGENIAS Method and Tools.* In Agent-Oriented Methods. pp. 236-276. Eds.: Henderson-Sellers, B., Giorgini, P.. IGI Global. 2005.

143. [Pay09] Payne A., Desvaux A., Chalvet-Monfray K., *et al. Flux de volailles et propagation de l'influenza aviaire dans la filière avicole au Vietnam.* Epidémiologie et santé animale, vol. 55 pp. 137-152 (2009)

144. [Pea25] Pearl R. *The Biology of Population Growth.* Alfred A. Knopf, New York, 1925.

145. [Per02] Perrier E., Bird N.R.A., *Modelling soil fragmentation: the pore solid fractal approach,* Soil Tillage Research Vol. 64 pp. 91–99. (2002)

146. [Pfe07] Pfeiffer D.U., Minh P.Q., Martin V., *et al.. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data.* Veterinary. Journal, Vol. 174 (2), pp. 302-309. (2007)

147. [Phi04] Phillips H. Killingray D. *The Spanish influenza pandemic of 1918-19: new perspectives.* Routledge editions. pp.380 ISBN: 0-203-77661-5 (2004)

148. [Pol04] Polhill G., Izquierdo L., Gotts N. *The ghost in the model (and other effects of floating point arithmetic),* JASSS vol. 8 (1). 2004

149. [Pol08] Polhill G., Parker D. Brown D., Grimm V. *Using the ODD protocol for describing three agent-based social simulation models of land-use change.* JASSS (2008) vol. 11 (2)

150. [Pot01] Potter C.W. *A history of influenza.* Journal of Applied Microbiology vol. 91 (4) pp. 572-579 (2001)

151. [Que09] Quesnel G., Duboz R., Ramat, E. *The Virtual Laboratory Environment–An operational framework for multi-modelling, simulation and analysis of complex dynamical systems.* Simulation Modelling Practice and Theory vol. 17 (4) pp. 641-653. (2009)

152. [Rai06] Railsback et al. *Agent-based simulation platforms: Review and development recommendations.* Simulation (2006) vol. 82 (9) pp. 609

153. [Rai10] Railsback S.F., Grimm V. *A Course in Individual-based and Agent-based Modeling.* Preprint (personal communication) pp.224 (2010)

154. [Ren07] René M., Bicout D. *Influenza aviaire : Modélisation du risque d'infection des oiseaux à partir d'étangs contaminés.* Epidémiologie et santé animale vol. 51 pp. 95-109 (2007)

155. [Res96] Resnick M. *StarLogo: an environment for decentralized modeling and decentralized thinking.* In proceedings of CHI '96. (1996)

156. [Ric56] Richards, P.I. *Shock waves on the highway.* Operation research Vol.4, pp. 42-51. (1956)

157. [Roc09] Roche B., Lebarbenchon C., Gauthier-Clerc M., *et al. Water-borne transmission drives avian influenza dynamics in wild birds: The case of the 2005-2006 epidemics in the Camargue area.* Infection, Genetics and Evolution vol. 9 (5) pp. 800-805 (2009)

158. [Ros16] Ross R. *An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I.* Proceedings of the Royal Society of London. Series A, Vol. 92, pp. 204–230. (1916)

159. [Rus05] Rushton J, Viscarra R., Guernebleich E., Mcleod A. *Impact of avian influenza outbreaks in the poultry sectors of five South East Asian countries (Cambodia, Indonesia,, Lao PDR, Thailand, Viet Nam) outbreak costs, responses and potential long term control.* Proceedings of the Nutrition Society . Vol. 61 (3) pp. 491-514 (2005)

160. [Sar04] Sarnak, P., *What is an expander?* Notice of the AMS, Vol. 51:7, pp. 762-763 (2004)

161. [Sar05] Sargent R. *Verification and validation of simulation models*. Proceedings of the 37th conference on Winter simulation (2005) pp. 130-143

162. [Sav06] Savill N., St Rose S., Keeling M., Woolhouse M.. *Silent spread of H5N1 in vaccinated poultry*. Nature vol. 442 (7104) pp. 757-757 (2006)

163. [Sha98] SHANNON R.E. *Introduction to the art and science of simulation*. Proceedings of Winter Sim. Conf. IEEE Computer Society Press Vol. 1 : 7-14 (1998)

164. [Sch02] Schadschneider A. *Traffic flow: A statistical physics point of view*. Physica A: Statistical Mechanics and its Applications Vol. 313 (1-2) pp. 153-187 (2002)

165. [Sch71] Schelling T.C. *Dynamic models of segregation*. The Journal of Mathematical Sociology vol. 1 (2) pp. 143-186 (1971)

166. [Sch09] Schuft E. *La filière du canard de chair au Vietnam et son rôle dans la propagation de l'Influenza aviaire hautement pathogène à H5N1*. PhD Thesis 2009.

167. [Sho93] Shoham Y. *Agent-oriented programming*. Artificial Intelligence, Vol. 60 (1), pp. 51-92, 1993.

168. [Sho98] Shortridge K.F., Zhou N.N, Guan Y. *et al. Characterization of avian H5N1 influenza viruses from poultry in Hong Kong*. Virology vol. 252 (2) pp. 331-42 (1998)

169. [Sim06] Simoes J. *An Agent-Based Approach to Spatial Epidemics through GIS*. PhD thesis. pp. 211 (2006)

170. [Sim07] Sims L. *Lessons learned from Asian H5N1 outbreak control*. Avian Diseases Digest vol. 51 pp. 174-181 (2007)

171. [Son06] Songserm T., Jam-on R., Sae-Heng N., Meemak N. *Survival and stability of HPAI H5N1 in different environments and susceptibility to disinfectants*. In: Schudel, A., and M. Lombard (eds), Developments in Biologicals, Vol. 124. pp. 254. 2006

172. [Spi08] Spickler A.R, Trampel D.W, Roth J.A. *The onset of virus shedding and clinical signs in chickens infected with high-pathogenicity and low-pathogenicity avian influenza viruses*. Avian pathology : journal of the WVPA vol. 37 (6) pp. 555-77 (2008)

173. [Sta90] Stallknecht J.D, Swayne D.E., Cooper, R.J. *Persistence of avian influenza viruses in water*. Avian Diseases vol. 34 (2) pp. 406-11 (1990)

174. [Str07] Stroud P., Del Valle S., Sydoriak S. *et al. Spatial Dynamics of Pandemic Influenza in a Massive Artificial Society*. JASS Vol. 10 (4) pp. 18 (2007)

175. [Stu05] Sturm-Ramirez K., Hulse-Post D., Govorkova E., *et al. Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia?*. Journal of virology vol. 79 (17) pp. 11269-11279 (2005)

176. [Tai10a] Taillandier P., Vo D.A., Amouroux E., Drogoul A. *GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control*. Proceedings of PRIMA'10 pp.8 (2010)

177. [Tai10b] Taillandier O., Drogoul A., Amouroux E. *Formation à la plate-forme GAMA, introduction à GAMA*. GAMA training workshop, 01/10/10, Bondy, France.

178. [Tau06] Taubenberger J.K, Morens D.M. *1918 influenza: the mother of all pandemics*. Emerging Infectious Diseases vol. 17 (1) pp. 69-79 (2006)

179. [Tra10] Tran A., Goutard F., Chamaillé L., *et al. Remote sensing and avian influenza: A review of image processing methods for extracting key variables affecting avian influenza virus survival in water from Earth observation satellites*. International Journal of Applied Earth Observation and Geoinformation. Vol. 12, pp 1-8 (2010)

180. [Tre08] Treuil J., Drogoul A., Zucker J. *Modélisation et simulation à base d'agents*. Dunod Editions. pp.352. 2008

181. [Tri09] Trillaud F. *Intégration SIG et simulation*. MSc thesis (2009) pp. 1-74

182. [Uhr09] Uhrmacher A., Weyns D. *Multi-Agent Systems: Simulation and Applications*. CRC Press, pp. 566. ISBN: 1420070231 (2009)

183. [Urb08] Urbani D., Delhom M. *Analyzing Knowledge Exchanges in Hybrid MAS GIS Decision Support Systems, Toward a New DSS Architecture*, LNCS Vol. 4953, pp.323-332 (2008)

184. [Ver38] Verhulst, P. (1838). Notice sur la loi que la population pour- suit dans son accroissement. Correspondance mathematique et phyisque, 10:113–121. 94

185. [Vin03] Vinten-Johansen P. . *Cholera, Chloroform, and the Science of Medicine: A Life of John Snow*. Oxford University Press. (2003)

186. [Wal92] Waldrop, M. . *Complexity : the Emerging Science at the Edge of Order and Chaos*. Simon & Schuster. 384p. (1992)

187. [Wal10] Walker P., Cauchemez S., Métras R., *et al*. *A Bayesian approach to quantifying the effects of mass poultry vaccination upon the spatial and temporal dynamics of H5N1 in Northern Vietnam*. PLoS Computational Biology (2010) vol. 6 (2)

188. [Wan09] Wang H., Li J., Chen Q., Ni D. *Speed-Density Relationship: From Deterministic to Stochastic*. Pre-print from the 88th Transportation Research Board (TRB) Annual Meeting, Washington, DC (2009)

189. [Web92] Webster R.G., Bean W.J., Gorman O., *et al*. . *Evolution and ecology of influenza A viruses*. Microbiol Rev Vol. 56 (1), pp. 152-179. (1992)

190. [Web98] Webster R.G. *Influenza: an emerging disease*. Emerging infectious diseases (1998) vol. 4 (3) pp. 436-441

191. [Who10] World health organisation. *H5N1 avian influenza: Timeline of major events* (09/12/10).

192. [Who11a] World health organisation. *Cumulative Number of Confirmed Human Cases of Avian Influenza A/ (H5N1) Reported*. http://www.who.int/csr/disease/avian_influenza/country/cases_table_2011_04_06/en/index.html

193. [Who11b] World health organisation. *Areas with confirmed human cases of H5N1* (01/01/2003 to 16/03/11) http://gamapserver.who.int/mapLibrary/Files/Maps/Global_H5N1inHumanCUMULATIVE_FIMS_20110316.png

194. [Wik01][16] Wikipedia. *The KISS principle* http://en.wikipedia.org/wiki/KISS_principle. Accessed 10/02/11

195. [Wik03] Wikipedia. *Influenza virus H5N1* http://en.wikipedia.org/wiki/Influenza_A_virus_subtype_H5N1 Accessed 18/02/11

196. [Wik04] Wikipedia. *Influenza Social impact of the H5N1 virus* http://en.wikipedia.org/wiki/Social_impact_of_H5N1 Accessed 18/02/11

197. [Wik05] Wikipedia. *Regions of Vietnam* http://en.wikipedia.org/wiki/Geography_of_Vietnam Accessed 08/04/11

198. [Wil98] Wilensky U. *NetLogo Virus model* (http://ccl.northwestern.edu/netlogo/models/Virus) (1998)

199. [Wil99] Wilensky U. *NetLogo*. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL 1999

200. [Wil00] Williams L.A., Kessler R.R. *All I really need to know about pair programming I learned in kindergarten*. Communications of the ACM vol. 43 (5) pp. 108-114 (2000)

201. [Woo01] Wood M.F., DeLoach S.A. *An overview of the multi-agent systems engineering method*. LNCS Vol.1957, pp. 207-221,. Springer Verlag. 2001

202. [Woo00] Wooldridge M., Jennings N.R., Kinny D. *The GAIA method for agent-oriented analysis and design*. Journal of Autonomous Agents and Multi-Agent Systems Vol. 3, pp. 285–312 (2000)

203. [Yan07] Severance C., Hardin J., Golden G., *et al.*. *Using the Sakai collaborative toolkit in e-Research applications*. Concurrency and Computation: Practice and Experience Vol. 19 (12), pp.1643-1652 (2007)

204. [Yu95] E. Yu. *Modelling Strategic Relationships for Process Reengineering*. PhD thesis, Univ. of Toronto, Dep. of Computer Science, 1995.

205. [Yor79] Yorke J.A., Nathanson N., Pianigiani G., Martin J. *Seasonality and the requirements for perpetuation and eradication of viruses in populations*. American journal of epidemiology vol. 109 (2) pp. 103-123 (1979)

206. [Zei89] Zeigler B.P. *DEVS representation of dynamical systems: event-based intelligent control*. Proceedings of the IEEE. Vol. 77 (1) pp.72-80. 1989

207. [Zei00] Zeigler B.P., Kim T.G., Praehofer H. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*, pp. 510, Academic Press, ISBN 0127784551, 2000.

---

[16] Wikipedia references are not indexed by year, they are numbered after their access date

Annex 1: Collaboration CIRAD – MSI pour l'étude et la modélisation de la grippe aviaire au nord Vietnam

Cette collaboration s'inscrit, d'une part, dans le cadre du projet Gripavi, financé par le Ministère des Affaires Étrangère Français, mis en œuvre par le CIRAD au Vietnam et dans 6 pays d'Afrique et, d'autre part, dans le cadre du projet GAMA du laboratoire MSI.

1. **Présentation des partenaires**

CIRAD

Le thématicien basé au Vietnam impliqué dans ces travaux est rattaché à l'Unité Epidémiologie et Ecologie des maladies animales, UR 16, du CIRAD. D'autres agents de cette unité pourront également, indirectement, être associés à ce travail.

Le Cirad est un institut français de recherche agronomique au service du développement des pays du Sud et de l'outre-mer français (http://www.cirad.fr). Le Cirad intervient par des recherches et expérimentations, des actions de formation, d'information et d'innovation, et des expertises. Ses compétences relèvent des sciences du vivant, des sciences humaines et des sciences de l'ingénieur, appliquées à l'agriculture et l'alimentation, à la gestion des ressources naturelles et aux sociétés.

Les principales questions de recherche de l'UR 16, appliquées aux maladies majeures en termes économiques et de santé publique, s'articulent autour de la description de l'occurrence et de la distribution des maladies des populations animales et de l'environnement dans lequel elles évoluent, de l'identification et de la quantification des déterminants et de l'évaluation des programmes de lutte et de surveillance.

L'UR16 s'intéresse à la modélisation des dynamiques infectieuses comme outils pour la compréhension de cycles et de patterns épidémiologiques et outils d'aide à la décision pour améliorer la surveillance, le prévention et le contrôle des maladies.

MSI / UR GEODES

Les activités de recherches informatiques seront réalisées dans le cadre du projet GAMA au sein du laboratoire MSI à l'IFI à Hanoi.

L'Institut de la Francophonie pour l'Informatique (IFI), créé en 1993 par l'Agence Universitaire de la Francophonie (AUF) à la demande du gouvernement du Vietnam pour assurer la formation de cadres supérieurs en informatique et aider à l'émergence de professeurs d'informatique de calibre international, est rattaché à l'Institut Polytechnique de Hanoi (IPH) selon une décision du Ministère de l'Education et de Formation du Vietnam (arrêté no 1549 / GD-DT du 13 mai 1997). Le programme de Master aux normes européennes comporte deux options : "Systèmes et réseaux" et "Intelligence artificielle & multimédia". En 2006, plus de 230 Masters auront été délivrés, dont plus

de 60 sont doctorants et 37 ont terminé leur thèse (à l'étranger).

L'équipe de recherche, MSI (Modélisation et Simulation Informatique des systèmes complexes) a été créée en mars 2006 au sein de l'IFI. Cette jeune équipe vise à accompagner le programme d'enseignement Master et à développer la recherche à l'IFI. Elle est associée à l'UR 079 GEODES de l'Institut de Recherche pour le Développement (IRD) dans le cadre du programme Jeune Equipe Associée à l'IRD (JEAI) pour la période 2007-2010. Les principaux axes de recherche sont : 1) Environnement de travail collaboratif pour la recherche en Modélisation et simulation ; (2) Plate-formes de simulation pour la modélisation des systèmes complexes ; (3) Observation distribuée et aide à la décision pour la sécurité environnementale. L'équipe est composé de 6 enseignants-chercheurs et de 12 doctorants en co-direction.

L'UR GEODES du département Milieux et Environnements (DME) de l'IRD, est spécialisée dans la recherche sur la modélisation et la simulation de systèmes naturels complexes. Elle explore et développe, de ce fait, un ensemble conséquent d'outils mathématiques et informatiques originaux ainsi que les démarches et méthodologies, de conception comme de validation, qui leurs sont associées. L'UR, du fait de la diversité de ses membres, s'intéresse particulièrement à l'articulation, sur des terrains d'application spécifiques (Maroc, Sénégal, Vietnam, Cameroun) de différentes méthodes de modélisation, depuis les EDP jusqu'aux systèmes d'évolution artificielle, en passant bien entendu par des simulations multi-agents, dans le but, d'une part, de vérifier leur pertinence en fonction des objectifs assignés aux modèles et, d'autre part, d'en mesurer la complémentarité.

<u>Chercheurs impliqués</u> : Edouard Amouroux (VI IRD, MSI/ GEODES), Alain Boucher (Prof AUF, MSI), Alexis Drogoul (DR IRD, MSI/ GEODES)
Ce projet constituera une grande partie du sujet de thèse d'Edouard Amouroux, qui a trait à la modélisation et simulation individu centré dans des contextes où la localisation et les informations environnementales jouent un rôle déterminant.
Edouard Amouroux, qui sera donc le principal intervenant de MSI dans le projet, est informaticien (Master Info P6), a travaillé en Thaïlande et au Vietnam sur une première architecture de SMA + SIG avec pour application une modélisation de la propagation de la dengue en milieu urbain en Thaïlande.

2. <u>Problématiques, enjeux et résultats attendus des recherches</u>

1. <u>Dans le domaine de l'épidémiologie</u>

Le premier objectif de l'étude épidémiologique sera d'identifier des déterminants de l'environnement naturel et humain liés à l'introduction et la dissémination de la grippe aviaire au nord Vietnam.
Le second objectif sera de déterminer les échanges de virus influenza entre les différents compartiments d'animaux domestiques et commensaux.
Cette démarche ayant pour objectif à court terme de pouvoir proposer des moyens d'action adaptés localement. Mais aussi de mettre en place des moyens de surveillance ou d'anticipation adaptés à la réalité locale.

En parallèle il est envisagé de travailler sur des méthodes de surveillance ciblées adaptées à la

grippe aviaire.

Pour ce faire la démarche qui suit est envisagée.
Une récolte de données dont les types sont a priori connus mais qui peuvent évoluer selon les besoins, en particulier pour ceux de la simulation individu-centrée (voir plus loin). Dans un premier temps des données épidémiologiques chronologiques relatives aux foyers de grippe aviaire des années précédentes sont en cours de récupération auprès des services vétérinaires vietnamiens. Ces données seront utilisées dans le cadre d'une étude transversale rétrospective au niveau local qui visera à recueillir également des données géoréférencées, sélectionnées essentiellement selon l'expérience des thématiciens, avec par exemple la position des marchés, des lieux de vente, les comportements de migration et de déplacements, etc. Ces données seront limitées à l'échelle de quelques communes ou l'équivalent d'un canton.
Dans un second temps une étude longitudinale sur une période de 1an ou 18 mois sera mise en place pour suivre la diffusion des virus infuenza entre différents compartiments au niveau local (compartiments des canards domestiques, des poulets domestiques, des mammifères domestiques…)

2. **Dans le domaine de l'informatique**

Ce projet a de nombreux intérêts du point de vue de la recherche purement informatique.
Tout d'abord la réalisation préalable d'une plateforme de simulation consistant en un système multi-agent couplé à un système d'information géographique. Cette plateforme permettant dès lors des simulations individu-centrées géolocalisées dans un environnement rendant compte de manière précise de la réalité du terrain. La mise en place et l'utilisation d'un "laboratoire virtuel" avec des besoins de résultats concrets. La recherche de techniques générales d'optimisation multicritères. Ces techniques ayant pour principale originalité de devoir être réalisées dans un environnement réaliste continu sur des données géoréférencées; techniques qui seront validées ensuite par la simulation. Ces techniques auront pour application concrète l'aide à la création d'un réseau de surveillance ciblée.

Enfin il est clairement envisagé de réaliser une plateforme permettant un travail de modélisation itératif du thématicien en collaboration avec l'informaticien mais aussi de la réutilisabilité du résultat.
Plus précisément, la simulation de l'environnement hors maladie peut très bien être utilisée pour une toute autre maladie, à quelques ajouts près. Enfin il est souhaité d'obtenir une plateforme qui permette son utilisation par le thématicien sans intervention de l'informaticien dans le cas où l'ensemble des acteurs du système étudié auront déjà été modélisés au préalable. C'est à dire que le thématicien pourra réaliser de manière autonome des simulations sur un même système, avec des problématiques tout à fait différentes.

3. **Agenda des activités**

1. *Modélisation du contexte écologique et humain (d'ici à Septembre 2007)*
    Tout d'abord nous nous proposons de réaliser un modèle "neutre" en collaboration avec le thématicien, c'est à dire rendant compte de la réalité du système de production hors maladie. Ce modèle individu-centré, qui s'appuiera sur un système multi-agent relié à un système d'information

géographique pour son exécution, sera construit à partir des données déjà disponibles et très largement basé sur l'expérience du thématicien.

2. *Première validation (Septembre – octobre 2007)*

Dès que ce premier modèle est terminé il sera possible, en cas de besoin, d'orienter la récolte d'information au niveau local mais aussi de vérifier l'adéquation de l'échelle du modèle avec la problématique envisagée. À la suite de quoi un nouveau modèle plus abouti pourra être à nouveau réalisé.

3. *Modélisation de la maladie (Octobre 2007 – Début 2008)*

Une fois ce modèle neutre ayant atteint un niveau satisfaisant, le même processus itératif de construction d'un modèle prenant en compte la grippe aviaire commencera. Cette démarche débutera certainement en même temps voire avant l'étude longitudinale. La validation de ce modèle passera par la comparaison avec les données épidémiologique déjà connues et l'expérience du thématicien. Par la suite, une fois le modèle satisfaisant, il pourra être utilisé de manière plus prospective dans le cadre de ce que l'on peut appeler un "laboratoire virtuel". Il sera possible de simuler diverses situations (épidémies, pandémies, mutation du virus, changement des comportements, etc) et de voir l'évolution par rapport au moyen de lutte actuel. Mais aussi et surtout de tester, sans les conséquences potentielles en cas de tests grandeurs réelles, de nouveau moyen de lutte et/ou de surveillance étant donné que les modèles individus centrés permettent d'influer sur chaque paramètre individuel (contrairement aux modèles mathématiques classiques) au cours de simulations.

4. *Utilisation envisagée du modèle (fin 2008)*

Un des enjeux de la réalisation de ce modèle est de dépasser le cadre de la simple analyse/ prévision, et d'envisager de l'utiliser pour réaliser des recherches algorithmiques sur la façon d'optimiser le placement de point de contrôle d'un réseau de surveillance en fonction des données recueillies. Diverses techniques d'optimisation sont envisageables comme l'apprentissage artificiel, des techniques d'aide à la décision ou d'optimisation multi-critères. Du point de vue applicatif ces recherches informatiques permettront d'utiliser l'environnement comme banc d'expérience pour déterminer un placement optimal des points de contrôle d'un réseau de surveillance épidémiologique ciblé. Une épidémie sera simulée grâce à la plate-forme et diverses possibilités d'organisation de réseau de surveillance et de placement des points de contrôles seront testées dans cet environnement artificiel. Cette optimisation virtuelle du réseau de surveillance pourra également être guidée ou modulée, selon des mécanismes qu'il nous reste à déterminer, par les thématiciens du projet. Enfin il sera possible d'utiliser ce laboratoire virtuel pour simuler des évolutions du virus (paramètre de transmissions, virulence, etc) et estimer les performances des systèmes de contrôles existants et envisagés précédemment.

Annex 2: ODD description of the final GAMAVI model

3. Overview

3.1.Purpose

Epidemiologists want to study mechanisms of local spread and persistence of H5N1 in the context of semi-industrialised and traditional poultry sectors in North Vietnam. Their hypotheses focus on the agro-system, the social environment (village and culture) and their interactions. The purpose of GAMAVI is to investigate and evaluate the importance of various factors, including poultry production, environments (especially aquatic ones), topography, etc, on the persistence and spread of H5N1 within a village or a commune in the Red River delta. Specifically, the model is about investigating the relationships between environments (as virus reservoirs) and the traditional or semi-commercial poultry production systems.

3.2.Entities, state variables and scales

3.2.1.Poultry flock

Poultry flocks represent groups of poultry. They are the main type of entities in our system. These groups hold homogeneous individuals (with respect to their species, age, behaviour and location) so it is not mandatory to represent explicitly the individuals. Nevertheless, the infectious status of each individual can be different. It is held, within the flock, in an "infection matrix". In general, production occurs all year long, flocks are kept indoors and the whole flock is renewed when needed.

In addition to this general type, the model includes two types of flocks called "free range duck" and "backyard". The first ones represent a seasonal production (warm season) because they go and graze in rice fields. A "backyard" flock (90% of households own one) is a mix of different species; (taken into account in the individual matrix), with a continuous renewal process (private consumption).

Variable name	Brief description
Matrix of individuals	Represents individuals' infection status (SIR) and elapsed time in this status. Species information is added in the backyard case.
Location	Which environment it is in
Species	It can be backyard, chicken, duck, Muscovy duck
Number of individuals	Determined at creation and implies the matrix dimension
Production type	Outlet (meat/egg), determines the length of production and seasonality
Housing	Whether the flock is kept indoors all the time or not
Within-flock infection probability	A daily probability for an infected individual to infect each susceptible individuals

Variable name	Brief description
BID50*	Quantity of virus to ingest to get 0.5 probability of getting infected
Excretion rate*	Quantity of virus excreted per time-step
Renewal transmission probability	If the flock is infected it might leave virus within the building after its renewal and thus infect the new flock (0 means there is a building quarantine procedure)

3.2.2. Farm

Farms aggregate several flocks and are responsible for the dynamics of the production.

Variable name	Brief description
Location	Coordinates
List flock types	Which flock to create first and at renewal
Renewal dates	List of dates at which each flock should be renewed
Intra-farm infection probability	Daily probability for an infected flock to infect a susceptible one (compound with the number of infected individuals)
Inter-farm infection probability	Daily probability for an infected farm to infect a susceptible flock of another farm (compound with the number of infected individuals)

3.2.3. Environment

The environment is constituted of GIS data (spot images + "vectorised" Google maps). It allows the creation of the different spatial units.

3.2.4. Spatial units

- **Village, rice fields, ponds and farm buildings**

The village is a (multi) polygon that contains other polygons representing farms and ponds. Ponds can be located outside the village as well. Rice fields are obtained by a tessellation, into parcels, of the rice field area surrounding the village. Only ponds and rice fields hold state variables.

- **Virus reservoir**

The role of reservoirs is to receive virus from flocks, manage its depletion and allow flocks to be infected by it. We could have numerous sub-classes based on the environment type. However, we chose to focus on the most important ones: (watered) rice fields and ponds, as the aquatic environment is more favourable to virus persistence [10][11]. Other sub-classes can be added in the future when additional data are available. Nevertheless, dry environment is accounted for by using transmission rates (see infection processes).

Variable name	Brief description
Location	Coordinates

Variable name	Brief description
Volume	Water volume, to compute the concentration of virus
Virus concentration	
Depletion Rate basis	The basic depletion rate
Temperature	Used to make the actual depletion rate vary over time

3.2.5.Scales

The system modelled is the village, which is a few kilometres wide. The smallest elements are farm buildings, (a dozen of meters wide), while the surface of rice fields is a few hundred squared meters. As the model deals with persistence of the virus for more than a year, the simulated time can be up to two years. The discrete time step between two states of the model is set to one hour, in order to represent the movement of flocks, and a daily time-step is used for production and infectious status updates.

3.3.Process overview and scheduling

3.4.Flock

- Infection processes

Flocks' main processes are related to infection and are parameterised with species-dependent rates (see sub-models for a detailed description). Excretion and Virus collection (if located in a reservoir) happen hourly.

The daily-based processes involve the computation of direct transmission (within flock), the computation of environment mediated infection (in ponds and watered rice fields), the update of the individuals' infectious status, and the natural death computation.

- Movement processes

Backyard flocks wander randomly within the village, and go back at night to their farm by the shortest path.

Free-range flocks go graze in the rice fields following a predefined schedule: (1) Selection of the rice field: it must be free (not used by another flock) within a defined range from the farm; (2) Depart from farm by the shortest path to the nearest point of the village's outskirt to the selected rice field; (3) Grazing: stay in the rice-field (no movement) (4) Return to farm following the reversed scheme.

3.5.Farm

Farms are responsible for the dynamics of production. Farms renew flocks according to a schedule that depends on the production type. This renewal can be instantaneous or include a building quarantine (a few days). The renewal schedule is not necessarily regular and is drawn from input data. Backyard flocks are handled apart because of their continuous consumption: one individual is "consumed" (and renewed) at a time.

3.6.Virus reservoir

These agents can receive virus from a flock; they compute the new virus concentration from the flock's excretion. They change this concentration according to temperature dependent equations [11]. The change of season can impact their behaviours; for instance, rice fields only act as reservoirs in the warm season (when they are watered).

All processes related to infection are detailed in the sub-models section.
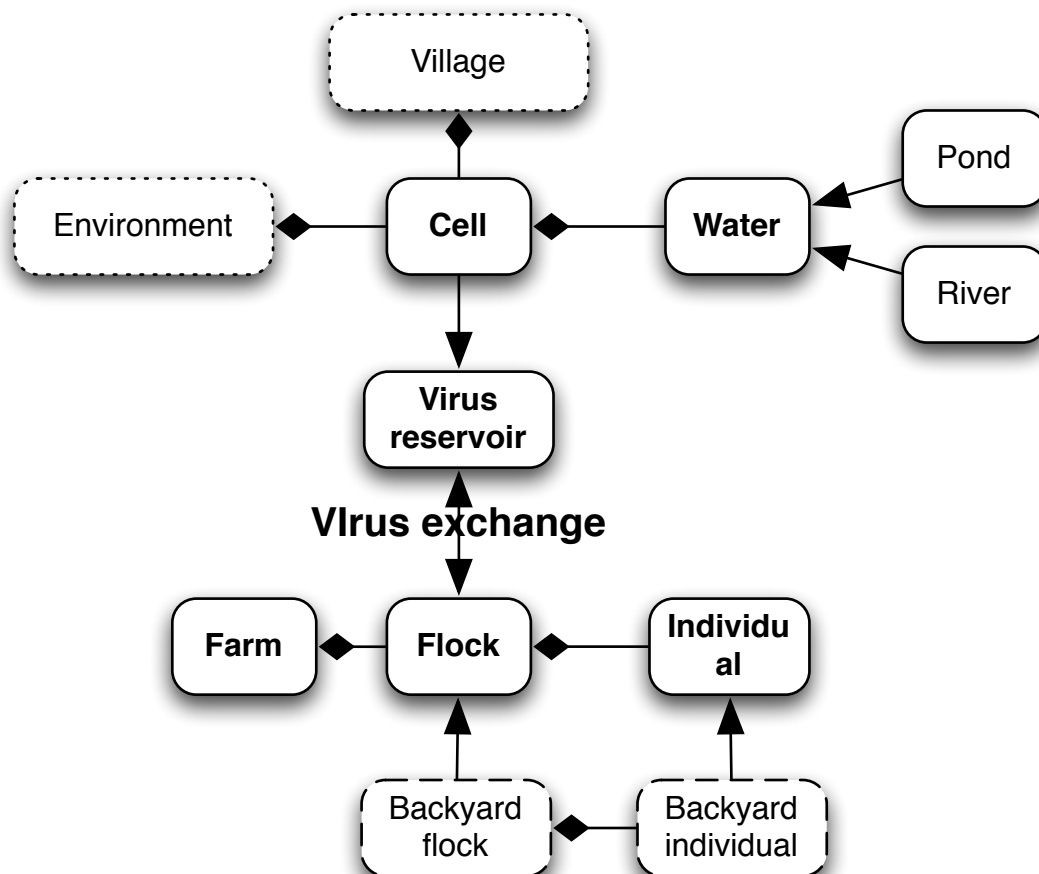
GAMAVI's conceptual organisation

4. Design Concepts

4.1. Basic principles

The data used to construct this model comes from field study by Stéphanie Desvaux (and a brief field study by Edouard Amouroux focused on free-range ducks and backyards farmer daily practices) and extensive literature review (epidemiological parameters). The dynamics taken into considerations are the production dynamics and the buy/sell process of the different flock types. The representation of the environment is derived from vectorised raster image of villages where the inner and outer village are separated; it respects the traditional organisation as presented by [Fon99] and [Ben07]. Briefly, each farms has a few flocks of different types that may stay inside or go outside and wander in different areas. While outside, they can excrete into watered environments and get infected from it. The excretion in watered environment deplete over time based on the model from [Ren07]. Infection through direct transmission are also considered at the level of the flock, the farm and between farms.

4.2. Emergence

Experts have not obtained all the necessary data yet, but several realistic dynamics have been observed with plausible data. Depending on the scenario chosen, we observe:

- The virus circulating among poultry flocks the whole year long

- The virus circulating among poultry flocks during the warm season and "stocked" in virus reservoirs during the cold season

- Faster epidemics with a higher proportion of chicken but less likely to last a whole year (chicken are easily infected but have 100% probability to die from it)

- A significant correlation between the number of ponds/rice fields, free-range ducks and the probability for epidemics to last more than a year.

4.3.Adaptation

Agents do not adapt to anything.

4.4.Objectives

There is no concept of goals. Although, the selection of a specific rice-field where to go is done by free-range ducks at the beginning of day but this selection is random.

4.5.Learning

There is no learning

4.6.Prediction

There is no uncertainty in the model thus no prediction.

4.7.Sensing

Flocks manage their individuals. Backyards know their immediate surrounding and the shortest path back to their farm. Free-range flocks have access to the location of within range rice fields, the shortest path to it within the village and the virus level of the virus reservoir where they are located.

Farms have access to their flocks and the date. Virus reservoirs know the current temperature, and rice-fields the current date.

4.8.Interaction

Flocks excrete virus in a virus reservoir and "collect" virus from it. An infected flock can infect a fully susceptible flock whether it is within the same farm or not (different probabilities for intra-farm and inter-farm infection).

4.9.Stochasticity

Backyard flocks' movement is made towards a randomly selected coordinates within a limited radius. Free-range flocks randomly select a free rice field within a radius.

The renewal of flocks concern a random number of individuals (truncated Gaussian), and a semi-random (with a probability dependent on production type) flag for the use of building quarantine.

Whenever a flock gets infected, an infected individual is chosen at random. When a flock collects more than a predetermined level of virus, each individual has a probability of 0,5 to get infected.

4.10.Collectives

Farms are collectives of flocks, which are collectives of individuals. Village also can be seen as a collective of rice- fields, ponds, farms and a "common ground".

4.11.Observation

The model offers have a graphical representation of the system where the environments, the moving flocks and the spread of the virus are represented. The evolution of the number of individuals per infectious status, flocks or restriction per production type, average concentration of virus in the reservoirs, number of infected reservoirs, instantaneous global transmission rate are all available. They can be visualise as instantaneous rates (monitors) or plotted on a graph update every X time-step.

We plan in a future version to add a representation of the graph of infection.

5. Details

5.1.Initialisation

At first, the beginning date of the simulation has to be defined in order to set which seasonal flocks to create. Then, flocks get an "age" (randomly set because of the renewal process and to avoid side-effects due to an artificial synchronicity). Depending on the season, rice fields are watered or dry. Flocks are all located within farms (we assume it is the beginning of the day at 00am) and their individuals are all susceptible except some (according to parameters). The number of flocks of farms is drawn from data. Farms are localised within the village thanks to the use of GIS data (or vectorised raster image).

The generation algorithm:

1. Selection of the village stereotypes

2. Computation of the number of households

3. Computation of the number of backyard flocks (and farms): 90% of the total household

4. Computation of the different semi-commercial flocks (global statistic)

   4.1. Computation of their number per flock type

   4.2. Computation of their duration

   4.3. Computation of their production duration

5. Generation of the farms

   5.1. Random determination if the farm is associated to an household with backyard or not

   5.2. Computation of the number of flock for the current farm (minus -1 if it has a backyard)

5.2.1.   Associate randomly a flock to the farm

        5.2.2.   Iterate until the number of flocks is reached

    5.3.   Iterate until all the flocks are associated with a farm.

5.2.Input data

5.2.1.Environment

A vectorised raster image of a village is used. It contains the topology of the inner village and (bare soil, houses, ponds) and the rice-field areas outside the village. Each element of the inner village is representing by a polygon whereas a grid decomposes the rice-field areas into lots.

The adaptation for real GIS data would be minimal (no grid anymore).

5.2.2.Poultry production system

The data presented below are input to the model, although they can be changed in order to represent other villages or test different hypotheses.

Name	Description
Number of households	Determines the number of backyard flocks (90% of households have a backyard)
Number of farms	
Distribution of number of flocks per farm	
Distribution of flocks per production type	i.e. X% of free-range ducks, Y% of short time chicken
Configuration of flocks per production type	Includes: size, duration, renewal pattern, seasonality
Seasonality	Warm season only or not (include flooding of rice-field)
Water consumption	Duck only
Wandering time	For free-range and backyard flocks
Natural death rate	Per species

5.2.3. Time series

The mean month temperature of water and air is derived from data of the Bac Ninh meteorological station.

5.3.Sub-models

5.3.1.Viral cycle

There are two transmission routes for the H5N1 virus: environment-mediated and inter-individual. According to [Roc09]: "The models combining an inter-individual and water-borne transmission are significantly better than all models involving only one transmission route". In GAMAVI, we focus on the watered-environment-mediated and inter- individual transmission at three levels: within the flock, the farm and the village. The latter route is modeled using three rates (obtained from literature and expert knowledge in the case of scenario based experiments).

Using rates for direct inter-individual transmission is a classical technique in epidemiological modelling. As the rates are constant, we fall into the frequency-dependent category, in opposition to the more classical density-dependent category of models.

Concerning environment-mediated transmission, we consider that the dry environment can be neglected or considered as part of the inter-individual transmission. This is justified as we focus on the persistence and long term circulation of the virus and that the watered-environment processes operate at a much longer time scale than those affecting the dry environment and the individuals.

Due to the scarcity of data, we had to propose a simple model for environmental infection. Consequently we base this process on the BID50, which is the dose of water a bird has to ingest to have 50% probability to become infected. Practically, free-range flocks collect virus particles (hourly water consumption multiplied by the time spent in a reservoir) and the amount is computed and compared to BID50 at the end of the day. Reciprocally, infected individuals of free-range flocks excrete a constant amount of viral particles in the reservoirs. Finally, the persistence of the virus within a reservoir is computed using the René and Bicout [11] equation within a day time-step:

$$\frac{dC}{dt} = -kC \ where k = k_0 e^{\alpha(T-T_0)/T}$$

$k0$ being the depletion speed at $T_0 = 293°K$ and a a coefficient obtained by linear regression from experimentations.

5.3.2. Parameters

Name	Description
Pathogenecity level	Includes death proportion, duration of an infection and mean time of death
Rice fields depletion rate	
Ponds depletion rate	
Excretion level	Number of viral particles excreted per hour per species
Probability of within flock infection	
Probability of within farm infection	
Probability of inter farm infection	
Probability of infection from renewal	In the case of no building quarantine
BID50	Quantity of virus to ingest to have a 0.5 probability of infection
Building quarantine proportions	

Name	Description
Tessellation size	mean size of a rice-field lot
Initial infection	How many individuals are initially infected

Annex 3: Poultry Production system description

240 semi-commercial farms interviewed

Description of the dataset by PRODSYST

```
PRODSYST1 |    Freq.    Percent     Cum.
-----------+---------------------------------
        1 |      1       0.16       0.16
        2 |      2       0.33       0.49
        3 |    425      70.02      70.51
        4 |    178      29.32      99.84
        5 |      1       0.16     100.00
-----------+---------------------------------
    Total |    607     100.00
```

Il y a 607 troupes réparties dans les 240 élevages enquêtés.

PRODSYST	PRODSYST
1	Villageois traditionnel / production continue
2	Villageois traditionnel / production par lots
3	Semi-industriel fixe
4	Semi-industriel / nomade ou transhumant

Pour la suite, on élimine les troupes de backyard, pour caractériser uniquement les élevages semi-commerciaux

```
PRODSYST1 |    Freq.    Percent     Cum.
-----------+---------------------------------
        3 |    425      70.36      70.36
        4 |    178      29.47      99.83
        5 |      1       0.17     100.00
-----------+---------------------------------
    Total |    604     100.00
```

Nombre de troupes par élevage

```
     Percentiles    Smallest
1%         1          1
5%         1          1
10%        1          1      Obs              240
25%        2          1      Sum of Wgt.      240

50%       2.5                Mean          2.558333
              Largest    Std. Dev.      1.15201
75%        3          5
90%        4          5      Variance      1.327127
95%        4          6      Skewness      .3755668
99%        5          7      Kurtosis       2.85687
```

```
.Total troupes |   Freq.    Percent     Cum.
------------+----------------------------------
      1 |     51      21.16     21.16
      2 |     69      28.63     49.79
      3 |     64      26.56     76.35
      4 |     50      20.75     97.10
      5 |      4       1.66     98.76
      6 |      1       0.41     99.17
      7 |      1       0.41     99.59
    613 |      1       0.41    100.00
------------+----------------------------------
  Total |    241     100.00
```

On garde de 1 à 5 troupes.
1 troupe = 21%
2 troupes = 29%
3 troupes = 27%
4 troupes = 21
5 troupes = 2%

Pour la repartition des troupes de différent prod type

Description of the dataset by PRODTYPE

```
ESPRPODTYPE |
          1 |     Freq.     Percent      Cum.
------------+-----------------------------------
         C3 |       7        1.14        1.14
         C5 |     157       25.61       26.75
         C9 |       1        0.16       26.92
         D5 |     112       18.27       45.19
        MD5 |      71       11.58       56.77
        MD9 |       1        0.16       56.93
        OT2 |       3        0.49       57.42
        OT5 |       2        0.33       57.75
        OT9 |       1        0.16       57.91
  coq combat |      8        1.31       59.22
 longchicken |    108       17.62       76.84
    longduck |    103       16.80       93.64
 longmuscovy |     39        6.36      100.00
------------+-----------------------------------
      Total |     613      100.00
```

Attention; dans les élevages il y a des troupes considérées comme backyard (en opposition à semi-commercial)⇨ on élimine ttes les troupes dont la taille des lots est inférieure à 40 animaux sauf pour la catégories de housingED 4 = free ranging birds

Dans ce cas il reste 530 troupes, qui se répartissent comme suit ;

```
esprpodtype |
          1 |     Freq.     Percent      Cum.
------------+-----------------------------------
         C3 |       7        1.32        1.32
         C5 |     142       26.79       28.11
         C9 |       1        0.19       28.30
         D5 |      98       18.49       46.79
        MD5 |      63       11.89       58.68
        MD9 |       1        0.19       58.87
        OT2 |       3        0.57       59.43
        OT5 |       1        0.19       59.62
        OT9 |       1        0.19       59.81
  coq combat |      8        1.51       61.32
 longchicken |     73       13.77       75.09
    longduck |     95       17.92       93.02
 longmuscovy |     37        6.98      100.00
------------+-----------------------------------
      Total |     530      100.00
```

.

PRODTYPE	PRODTYPE	Catégories ED
1	Grands parentaux	long
2	Parentaux	long
3	Production poulettes/cannette	Non pris en compte
4	Pondeuses	long
5	Chair	short
8	mixed parentaux, pondeur	long
9	mixed parentaux chair	Non pris en compte

Durée du cycle par espèce et type de production

Voir tableau

Housing

HOUSING	HOUSING	Categorie ED
1	Birds in a close building all day long	Enfermement simple = 1
2	Birds in an outdoor close pen with open house without access to river or chanel	Enfermement simple = 1
3	Birds in an outdoor close pen with open house with access to river or chanel	Enfermement et eau partagée = 2
4	Birds in an outdoor close pen with open house with access to a pond	Enfermement et eau partagée = 2
5	Birds in channel, river and rice field all day and in a close pen at night	Free range = 3
6	scavenging birds, all day in side and outside the farm	Backyard = 4 (quand moins de 50 adultes) Free range autre (quand plus de 50 adultes) = 5

Avec backyard

```
   HOUSING |    Freq.    Percent      Cum.
------------+-----------------------------------
        1 |      19       3.09        3.09
        2 |     214      34.85       37.95
        3 |     103      16.78       54.72
        4 |     124      20.20       74.92
        5 |      80      13.03       87.95
        6 |      70      11.40       99.35
        7 |       4       0.65      100.00
------------+-----------------------------------
    Total |     614     100.00
```

Sans backyard

```
   housing |    Freq.    Percent      Cum.
------------+-----------------------------------
```

```
        1 |     18      3.40      3.40
        2 |    162     30.62     34.03
        3 |     94     17.77     51.80
        4 |    110     20.79     72.59
        5 |     71     13.42     86.01
        6 |     70     13.23     99.24
        7 |      4      0.76    100.00
------------+-----------------------------------
    Total |    529    100.00
```

avec backyard
```
 housingED |    Freq.    Percent      Cum.
------------+-----------------------------------
        1 |    233     38.20     38.20
        2 |    227     37.21     75.41
        3 |     80     13.11     88.52
        4 |     56      9.18     97.70
        5 |     14      2.30    100.00
------------+-----------------------------------
    Total |    610    100.00
```

Sans backyard
```
housingED |    Freq.    Percent      Cum.
------------+-----------------------------------
        1 |    180     34.29     34.29
        2 |    204     38.86     73.14
        3 |     71     13.52     86.67
        4 |     56     10.67     97.33
        5 |     14      2.67    100.00
------------+-----------------------------------
    Total |    525    100.00
```

Catégories Prod et Housing ED

.**Taille des troupes**
Avec backyard
```
               BIRDFLOCK
-----------------------------------------------------------------
     Percentiles     Smallest
 1%         5          2
 5%        10          2
10%        12          3      Obs              599
25%        50          3      Sum of Wgt.      599

50%       150                 Mean        193.8114
```

		Largest	Std. Dev.	221.6294
75%	250	1200		
90%	460	1500	Variance	49119.57
95%	600	1500	Skewness	2.886419
99%	1000	2000	Kurtosis	16.20245

Sans backyard

birdflock

	Percentiles	Smallest		
1%	5	2		
5%	10	2		
10%	40	3	Obs	514
25%	90	4	Sum of Wgt.	514
50%	150		Mean	222.9241
		Largest	Std. Dev.	226.4126
75%	300	1200		
90%	500	1500	Variance	51262.68
95%	600	1500	Skewness	2.846554
99%	1000	2000	Kurtosis	15.71904

Comparaison taille troupe entre les 2 provinces

Avec backyard

prov = BG

	Percentiles	Smallest		
1%	5	2		
5%	8	2		
10%	10	3	Obs	407
25%	30	4	Sum of Wgt.	407
50%	100		Mean	172.7592
		Largest	Std. Dev.	225.8124
75%	200	1000		
90%	400	1200	Variance	50991.25
95%	500	1500	Skewness	3.263974
99%	1000	2000	Kurtosis	18.87698

-> prov = HT

	Percentiles	Smallest		
1%	5	3		
5%	20	5		
10%	50	5	Obs	191
25%	100	10	Sum of Wgt.	191

50%	200		Mean	238.1152
		Largest	Std. Dev.	206.5315
75%	300	700		
90%	500	750	Variance	42655.26
95%	600	1200	Skewness	2.247085
99%	1200	1500	Kurtosis	11.41874

T-Sided, Two-Sample T-Test (H0: M1 = M2 versus H1: M1 <> M2)

Variance Assumption	Power DF (Alpha=0.05)	Power T-Value (Alpha=0.01)	Prob Level	Conclude H1 at 5.0% Significance?
Equal	599.00	3.391	0.0007	Yes
	0.9238	0.7925		
Unequal	405.32	3.501	0.0005	Yes
	0.9373	0.8187		

Il y a une différence significative de la taille des troupes entre les 2 provinces⇨ on teste les différences pour toutes les catégories. (différence non testées pour les catégories)

Sans backyard

prov = BG

birdflock

Percentiles		Smallest		
1%	4	2		
5%	8	2		
10%	20	3	Obs	334
25%	50	4	Sum of Wgt.	334
50%	150		Mean	206.515
		Largest	Std. Dev.	236.1858
75%	280	1000		
90%	450	1200	Variance	55783.71
95%	650	1500	Skewness	3.108326
99%	1000	2000	Kurtosis	17.29704

-> prov = HT

birdflock

Percentiles		Smallest		
1%	15	12		
5%	50	15		
10%	100	20	Obs	179
25%	100	20	Sum of Wgt.	179
50%	200		Mean	253.1117
		Largest	Std. Dev.	204.7578
75%	300	700		
90%	500	750	Variance	41925.76
95%	600	1200	Skewness	2.329724
99%	1200	1500	Kurtosis	11.80651

Two-Sided, Two-Sample T-Test (H0: M1 = M2 versus H1: M1 <> M2)

H1 Variance Assumption	DF	T-Value	Prob Level	Conclude at 5.0% Significance?
	Power (Alpha=0.05)	Power (Alpha=0.01)		
Equal	511.00	-2.251	0.0248	Yes
	0.6146	0.3728		
Unequal	411.70	-2.352	0.0191	Yes
	0.6505	0.4078		

Lower, One-Sided, Two-Sample T-Test (H0: M1 = M2 versus H1: M1 < M2)

H1 Variance Assumption	DF	T-Value	Prob Level	Conclude at 5.0% Significance?
	Power (Alpha=0.05)	Power (Alpha=0.01)		
Equal	511.00	-2.251	0.0124	Yes
	0.7279	0.4701		
Unequal	411.70	-2.352	0.0096	Yes
	0.7603	0.5108		

Upper, One-Sided, Two-Sample T-Test (H0: M1 = M2 versus H1: M1 > M2)

H1 Variance Assumption	DF	T-Value	Prob Level	Conclude at 5.0% Significance?
	Power (Alpha=0.05)	Power (Alpha=0.01)		
Equal	511.00	-2.251	0.9876	No
	0.0000	0.0000		
Unequal	411.70	-2.352	0.9904	No
	0.0000	0.0000		

Si on exclue les troupes backyard, il reste une différence significative de la taille des troupes entre les 2 provinces (différence non testées pour les catégories)

Taille des troupes / espèces / protype / housing (voir tableau)

Annex 4: Simulations GAMAVI

Idée générale

Dans quelle(s) condition(s) liées à la survie du virus dans l'environnement, à la composition et dynamique de la population aviaire domestique (et à sa couverture vaccinale) peut-on observer un maintien du cycle viral à l'échelle d'un village ou de plusieurs villages.

Paramètres clés a faire varier :

- la composition de la population aviaire (en terme d'espèce, de taille et de durée de cycle et donc implicitement de saisonnalité de production)
- la souche virale et donc les taux de mortalité et la durée d'excretion par espèce
- la survie du virus dans les « environnements »
- le taux de couverture vaccinale des populations (cf influence sur la durée d'excrétion, les taux de mortalité et la dose infectieuse nécessaire pour causer une infection)

Qu'entend-on par maintien ?
Le maintient se définit dans notre cas, par la possibilité pour le virus, une fois introduit dans une population, de pouvoir entrainer un état d'endémicité.

Endémicité = la présence constante d'une maladie ou d'un agent infectieux au sein d'une zone géographique ou d'un groupe de population donné

Première série d'expérimentations

Question générale
Le maintien du virus influenza H5N1 hautement pathogène est-il possible dans les conditions du Nord Vietnam en absence de vaccination des populations aviaires sensibles ?

Différentes expérimentations à mettre en place
Les différentes expérimentations testeront successivement l'influence de groupe de paramètres.

- **Influence de survie du virus dans l'environnement** :

Hypothèse à tester : le virus peut se maintenir dans l'environnement et les populations avicoles grâce à la présence d'une saison froide.

Scénarios :
Paramètres fixes
- composition de la population avicole (on prend un village type du Delta)

Paramètres à faire varier

- taux de mortalité / durées d'excrétion : on prend les paramètres liés au virus H5N1 clade 2 (souche 2.3.4) et ensuite au virus H5N1 clade 1
- présence d'une saison froide versus température constante sur l'année (la température de référence étant la température de la saison chaude).

- **Influence de la composition de l'élevage du village** : utiliser différents village types.

Hypothèse à tester : le maintien du virus, en absence de vaccination, ne peut s'opérer qu'avec la présence d'une population à cycle long de canards suffisamment importante.

Scénarios :
Paramètres fixes
- la survie du virus dans l'environnement (on prend les valeurs moyennes pour les durées de survie et la présence de saison.
- Taille du système : on prend la taille moyenne de population avicole d'un village du Delta
Paramètres à faire varier :
- on teste différents type de villages avec ou sans population de canards à cycle long et avec un équilibre entre les populations de cycles longs et cycles courts variables.

- **Influence de la taille globale du système** représenté
  - dans un premier temps, jouer seulement sur la taille du village en testant
  - dans un second temps, lorsqu'on aura pris en compte les liens entre villages, jouer le nombre de villages.
L'hypothèse à tester : le maintien ne peut s'opérer qu'au-delà d'une taille critique des populations avicoles.

Expérimentation 2

Le maintien du virus influenza H5N1 hautement pathogène est-il possible dans les conditions du Nord Vietnam avec une vaccination des populations aviaires sensibles ?

- **Influence à la composition de l'élevage** : utiliser différents village types.

Hypothèse à tester : la vaccination permet le maintien du virus dans des conditions relatives à la population qui normalement ne le permettent pas.

Paramètres fixes :

- la survie du virus dans l'environnement (on prend les valeurs moyennes pour les durées de survie et la présence de saison.)

Paramètres à faire varier :
- le taux de couverture vaccinale = % de la population qui est vaccinée
- composition de l'élevage ; on teste successivement des compositions favorables et défavorables au maintien du virus en absence de vaccination (cf résultats expérimentation 1)