A Simulation Model for Integrating Multidisciplinary Knowledge in Natural Sciences

Heuristic and Application to Wild Rodent Studies

Jean Le Fur¹, Pape A. Mboup² and Moussa Sall³

¹Institut de Recherche pour le Développement (IRD), Centre de Biologie pour la Gestion des Populations (CBGP), Campus Baillarguet, CS 30016, F-34988 Montferrier-sur-Lez, France

²Dépt. de mathématiques et d’informatique, Univ. C. A. Diop/18522 Dakar, Senegal

³Dépt. Informatique, Univ. G. Berger/Saint-Louis, Sénégal and lab. IRD-BIOPASS, Campus Bel-Air, Dakar, Senegal

Keywords: Multidisciplinary Knowledge, Heuristic, Object-oriented Modelling, Case Study, Agent-based Model, Simulation Model, Robustness, Rodents.

Abstract: Knowledge about rodents has been obtained at multiple observation scales and covers many functional levels, from DNA to ecosystems. We have developed an object/agent-oriented simulation model to represent these elements and phenomena in an integrated manner. Given the diversity of domains, items, processes and scales to be considered, we used an incremental approach to model development, with contrasting case studies successively represented and connected within the same model. Each study enriches the model and benefits from previous developments. The results emerging from this compilation are reflected into a shared class tree composed of three broad domains of variability: (i) concrete agents, (ii) specific genomes that instantiate the characteristics of each type of agent and (iii) agent containers that can be described on several scales. The classification that appears is characterized by the triviality of the categories obtained. It resembles natural partitioning, which lends it certain robustness, facilitating its extension. The essentially transitory nature of the construction is discussed, together with its dependence on the formalisms used. The current model, built on a combination of eight case studies, appears to be sufficiently robust to address new aspects and to serve as a basis for the further construction of an integrated view of the complex dynamics associated here with rodents.

1 INTRODUCTION

Disciplinary specialization is, logically, seen as a necessity for deciphering the increasingly specific facets of natural systems (Newell, 2001). In the context of research on small rodents, for example, knowledge is obtained at all levels of life, from genes, through cells, organisms, populations and communities, and up to ecosystem level. Each level can itself be broken down into numerous departments corresponding to individual disciplines. This discipline-based diversification is observed in many domains of research on rodents, from biogeography to taxonomy, morphometry or epidemiology.

The increasingly acute compartmentalisation of approaches can sometimes become problematic (Roussel, 2002), particularly if it leads to a disconnection from concrete situations (Burger and Kamber, 2003) where the complex dynamics displayed integrate several levels and scales with a diversity of elements and relationships (Bar Yam, 1997). The development of an integrative model or at least of a suitable approach for achieving such a model, would facilitate characterization and comprehension of the complex systems studied (Newell, 2001). We focused on this issue of integrated representation of disciplinary knowledge in the ‘restricted’ domain of the biology of small rodents and their parasites with a view to improve our understanding of the phenomena observed in situ.

Object-oriented paradigms have long been considered a powerful approach for the modelling of complex systems (Clapham and Crosby, 1992). When associated with simulation models and ‘agent’
formalism (Casti, 1997), this approach constitutes a flexible and relevant tool (Goldspink, 2002) that has already been used for the formal integration of disciplinary knowledge.

The approach adopted is based on the empirical construction of a model by the successive aggregation of sub-models of concrete case studies. This heuristic aims to develop a joint formalization of the diversity of fields, questions and knowledge worked separately by each discipline, while remaining limited to the components strictly necessary for the modelling of the case studies. This approach aims to use object-oriented programming generalisation to reveal (“let emerge”) a generic architecture accounting for the composite knowledge relating to the rodents populations dynamics. Given the hypothetical outcome, this approach is implemented as a heuristic, to facilitate progress towards a better simulation of events or processes really occurring in the field.

This article presents the heuristic we have used and the case studies underlying the development of the model. We then present two examples and the overall class hierarchy obtained before discussing the results in the light of our objectives.

2 DESIGN OF THE MODELLING ARCHITECTURE

2.1 Presentation of the Heuristic

The approach adopted is based on the incremental accretion of case studies specific to each research discipline (ecology, genetics, geography, etc.). Each time a new study considered, it is selected on the basis of displaying as large a contrast as possible with the previous studies, so as to capture a diversity of components and processes, thereby testing the robustness of the approach. For each particular case study, modelling represents only the relevant elements. The general model ‘grows’ through progressive aggregation of the concepts resulting from the modelling of each case study and by seeking the lowest common denominator, by identifying similarities and then factorising them (see below). The resulting model must be systematically back-compatible with all the previous case studies represented, and capable of integrating others.

For each case study, modelling is carried out in collaboration with the researchers in biology, ecology, geography, and other relevant disciplines responsible for carrying out the study concerned. Each case study is approached as an agent-based model, in the ‘usual way’ (e.g., Macal and North, 2006), by reifying data and knowledge with objects, environments and processes required by the particular context of each of the studies considered. At each addition of a formalised study, new types of objects may need to be implemented, but some of the existing elements can be reused.

During the transition between studies, there is a return to the ‘generic’ model, during which the code is essentially revised through three types of operation: (i) if common structures emerge, classes are generalised by abstraction and interfacing. Shared methods are factorised, through polymorphism and inheritance. Packages may also be defined or rearranged; (ii) the method names are then refactored, to clarify the logic constructed, and time and space unit conversions are checked; (iii) tuning or encapsulation relating to new added features is performed, to ensure the back-compatibility of the code, which underlies the validity and robustness of the approach.

2.2 Simulation Environment

The architecture presented was developed in Java, with the Repast Symphony platform (North et al., 2005) and some of its primitives (schedulers, contexts, continuous space, and graphical interface). The representation of studies and the continual overhaul of the model also apply to its three dependencies, which benefit from this approach of continuous improvement and are also the subject of methodological studies relating to the taking into account of interdisciplinary knowledge.

A ’data’ package is responsible for collecting externalities to the case studies. A class system is built up progressively to process data (maps, tables etc.), constants, parameters and seeds of random number generators.

An ‘observation’ package aims to ensure multimodal rendering (display, data storage, connection to other software) of the functioning of the system. This module makes use of the principle of epiphytic or recommender systems (Richard and Tchoumikine, 2004) based on a set of specialized observers/inspectors that simulate the data collection (surveys, trapping, counting etc.) actually carried out in the field by scientists and compute output indicators.

Finally, each case study aggregated in the model is characterised, distinguished and implemented autonomously by a ‘protocol’ that realises the
Rodents are the subject of many scientific studies, because they are reservoirs of infectious diseases (Taylor et al., 2008) or involved in the degradation of goods and crops (Skonhoft et al., 2006). Rodents have many characteristics and lifestyles. They may be solitary, live in colonies, dig burrows or make nests, or associate with humans. Rodent research is extremely broad, ranging from laboratory experiments to in situ studies, at both small and large scales.

The boundaries of the study are defined by the work carried out by a multidisciplinary team of researchers working in the domain (the ‘Rodent’s group’ of the French Biology Center for Population Management, UMR 022 INRA-IRD-Cirad-SupAgro, France) We used eight selected case studies for model development and tests of the robustness of the approach. These case studies were selected on the basis of their being as different as possible, in terms of the questions tackled and their temporal and spatial extents, ranging from laboratory cages to the eco-climatic zone (Table 1).

Table 1: Case studies from particular disciplines successfully modelled (chronological classification) and diversity of the corresponding scales.

<table>
<thead>
<tr>
<th>Main Features</th>
<th>Space extent (m)</th>
<th>Time extent (year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Common voles in agricultural landscapes</td>
<td>566</td>
<td>10</td>
</tr>
<tr>
<td>2. Cage hybridisation of African rodents</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>3. Catch-mark-recatch experiment in an African reserve</td>
<td>441</td>
<td>20</td>
</tr>
<tr>
<td>4&amp;5. Epidemiology and transportation of black rats</td>
<td>817,810</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>471,432</td>
<td>40</td>
</tr>
<tr>
<td>6. House mouse invasion in Senegal</td>
<td>681,120</td>
<td>40</td>
</tr>
<tr>
<td>7. household habitats exploration by mice</td>
<td>138</td>
<td>1</td>
</tr>
<tr>
<td>8. Sahel invasion by a gerbil species</td>
<td>1,148,831+</td>
<td>15</td>
</tr>
</tbody>
</table>

In the first study, the notions of landscape, agricultural operations, crop rotation, rodent burrowing, reproductive and social behaviour were included. In the second study, the chromosomal and gene levels were represented, as well as the cellular processes of fertilisation (meiosis, fusion, etc.). In the third case study, the trapping devices and their manipulation have been formalised. The next three studies were devoted to the transportation of commensal (associated with humans) rodents at several spatial and temporal scales. Road networks, transport vehicles, human carriers, cities, markets, and economic zones were therefore added to the model. The seventh study led to the introduction of daily activity rhythms into the model.

The final study led to formalization within the model of vegetation, its growth and the impact of rodents on its dynamics, as well as the integration of data derived from remote sensing or taking the effects of predatory owls into account.

The model obtained and described hereafter is constituted of 81 classes, 132 attributes, 143 relations, and 562 operations. As we focus here on abstracting and interfacing, we have chosen to prefix interfaces with ‘I_’, abstract classes with ‘A_’ and standard classes with ‘C_’ in the source code and the article. Moreover, to ensure the integrity of the multiple scales and units (more than 30 are used in the current version) dealt with in all case studies, we have suffixed most methods or properties with ‘Uxxx’ where xxx is the unit (e.g., meter, day, cell, tick, gramPerDay) of the method or property.

3 RESULTS

The model makes it possible to represent contrasting simulations, addressing diverse aspects of dynamics, for several rodent species, over various spatial and temporal scales, within different simulation contexts. For instance, in the example shown in Figure 1a, only aspects relating to the cross-breeding of rodents were considered, whereas, in the example shown in Figure 1b, studies of the diffusion of black rats over a century required the simulation of a rich historical and geographical environment, including all forms of commercial transport in the country concerned.

In this study, the main result lies in the model’s structure that emerges from the compilation of case studies. This model has three necessary and sufficient domains of diversity: simulated concrete entities, genomes associated with living organisms and different types of substrate in which objects and agents can be located.
Figure 1: Displays of simulations for two extreme case studies from the eight considered in this work (table 1). a) case n°2: simulation of an animal room protocol for the crossing of rodents from sibling species in different configurations. The challenge here is to study the barriers to fertilisation linked to chromosomal differences between the species. b) case n°4: simulation of the colonisation of Senegal (West Africa) by the black rat, over a century. The biological, historical and geographical aspects relate to the likelihood of invasion through the transport of rodents in commercial vehicles (trucks, trains, boats).

For each of these domains, various classes have appeared successively, corresponding to separate functions compatible with several case studies. Class methods were aggregated, abstracted, repositioned or refined when integrating successive case studies. The first domain constitutes the principal tree (Figure 2). It describes the agents that can intervene in the model.

At the root of the tree, any element within any system represented in the model is considered to be a nearly decomposable system (NDS) (Simon, 1962).

Figure 2: UML-based class diagram of the agents of the model, as determined from the case studies and engineered according to natural classifications. For the sake of clarity, only the relevant methods, properties or relationships are presented. Legend: A_: abstract class, I_: interface, C_: Class, _Uxxx: unit of the method or property, NDS: nearly decomposable system (Simon, 1962), see text.

This notion is used as a proxy for complex systems organised into a hierarchy (Le Fur, 2013). It is operative in the sense that a system, if it can be identified as an emerging entity resulting from the interaction of its components, automatically acquires individuality. Once its existence is revealed, the entity can be identified immediately (i.e., named), and the beginning of its existence can be noted. Finally, its existence necessarily has a duration that is formalised by ageing, a method that can then be recursively overloaded by the daughter classes, i.e., agents belonging to the leaf classes accumulate ‘skills’ for ‘growing older’ all over the steps of the specialisation chain. They thus acquire sophisticated competencies for action or capacity to respond to their environment. These basic and minimalist notions outline the definition of the model over time: they make it possible to encapsulate here the dynamic aspect of the model.
The first specialisation (A_VisibleAgent) relates to space and makes it possible to distinguish between objects located in their environment and to identify the fundamental properties relating to this characteristic. Visible objects and agents implement the I_SituatedThing interface with abstracted procedures for localisation. In this model, as soon as an object is located, the notion of environment takes on a meaning. Perception then establishes a faculty attributed to all agents. Beyond this class, NDS are either the agents (organisms) or their containers (Figure 4).

The second specialisation (A_Organism) is linked to an essential characteristic of living organisms: the determination of the properties of the organism by the genes it carries, to which values can be assigned. Each agent from this class or its daughters bears a genome (see Figure 3) with 'genes' that can be transcribed into properties or parameters ('phenotypes' or 'life traits' in Biology) defining how agents carry out their action in their environment. An agent belonging to the organism hierarchy also implements I_ReproducingThing, enabling it to reproduce with other organisms of the same species and to transmit part of its DNA to its descendants. Various agent properties from these three superclasses were gradually introduced into the model. Any function identified within a class was matched systematically to a biological justification within the hierarchy of life sciences for its qualification: animals (A_Animal) are able to move, burrowing rodents (C_RodentFossorial) dig burrows, colonial rodents (C_RodentFossorialColonial) have social interaction, and so on. The agents operate according to a PDE (perception-deliberation-execution) behavioural scheme (e.g., Macia Perez et al., 2014).

The root classes were rapidly identified and then progressively refined by displacement or refactoring of the methods or properties. Between case studies, several rodent agent classes were successively added, on the basis of functionality criteria. The classification obtained (bottom of Figure 2) was ultimately shown to correspond to the various known social statuses in rodents (commensal, fossorial, etc.). Once the nature of the classification was identified, it was strengthened.

The second hierarchy used (Figure 3) represents all aspects relating to the genetics of living organisms. It includes the mechanical elements (genes, alleles, etc.) required for genomic operations such as meiosis, segregation, fertilisation, mutation, and recombination.

The natural 'gene' - 'chromosome' - 'pair of chromosomes' - 'genome' composition is adapted from the work of Shaw and Wagner (2008) on locust genomes. In this sequence, any genome aggregates various pairs of 'chromosomes' that can be recombined and inherited, in part, during 'reproduction'. Each agent is associated with a genome (I_DiploidGenome) that corresponds to its species. This branch makes use of the inheritance principle of the object paradigm to instantiate the characteristics (called 'traits') of living agents in a inheritable and cumulative manner up to species-specific values. By exploiting the analogy between the biological approach (phylogeny) and object-oriented programming (inheritance), this part of the tree can also be considered as an object-oriented knowledge base (right-hand part of Figure 3), making it possible to value the movement speed, litter size, age at sexual maturity, etc., as genetically encoded characteristics for differing species of agents. Here, too, the mother classes are established by generalisation: when a biological trait corresponds to a natural classification, the corresponding class is created. For example, the duration of gestation is coded as a property of...
amniotes, which constitute the branch of animals in which a foetus is formed, or eukaryotes are characterised by the possession of a genome consisting of pairs of chromosomes.

The third variable domain concerns space. Most agents are located at a particular point in space, the status of which may be associated with diverse contexts, depending on the case studies considered. Successive rearrangements of the objects characterising the substrates and elements of the substrate in which agents evolve led to the development of the arbitrary concept of a 'container' (Figure 4). I_Container is implemented by all structures that can contain agents. Containers are defined as a recursive system, in that they may themselves contain other containers.

![UML class diagram](image)

**Figure 4 UML-based class diagram of the types of ground organisation, as determined from the aggregated case studies. Every instance in this hierarchy shares the I_Container features. Three successive levels of composition have been distinguished: landscape, land plot, soil cell. Each level of the composition is itself the root of a hierarchy that can be further extended (on the right) to describe various features of the worlds simulated.**

A_Container implements I_Container and allows containers to contain ISituatedThing objects. Containers are visible agents and, as NDS (Figure 2), they also have an age and an ability 'to age', or, in other words, to change over time according to specifications. Within this model, environment is discretised into elementary cells of the same size that describe the type of terrain (road, river, crop, etc.). Cell size depends on the level of detail chosen. A set of contiguous elementary cells of the same type is a C_LandPlot. This class provides information about the delimitation and identification of different areas within the simulated environments (e.g., crop, city, region, road, river). Finally, for each case study, an object of type C_Landscape is defined to contain the set of visible objects within the simulation: this class contains one continuous space representing the topology over which the agents move and a grid (matrix) of I_Container, the elements of which constitute a discretisation of the environment and a topology used by the agents to perceive their neighbourhood (Moore neighbourhood).

As the model was gradually refined, it became increasingly clear that almost all the relevant objects of the represented systems (burrow, nest, vehicle, trap) could be assimilated as containers for agents, a single class (A_SupportedContainer) facilitating their integration into this scheme.

### 4 DISCUSSION

The model developed here encompassed all the components linked to the different domains modelled between case studies. Every case study was simulated using the same model. Through time, each of them both enriched, and benefited from, the mechanisms and components included in the simulator and hence, evolved following the model’s development. This led, for example, to the progressive inclusion of cell processes, social behaviour, spatial structures, and several types of movement.

According to the heuristic, only classes that provide functions specific to their rank are implemented. Several characteristics of the architecture of the model emerge from this approach. The first of these characteristics is the partitioning of the model into three main areas of diversity: the diversity of agents (concrete objects), specific traits (genomes) and spaces (containers). Other elements were gradually defined and confirmed, such as the encapsulation of time-related aspects in the root of the class tree (that may appear convenient or appropriate for a simulation model), the partitioning of rodent types according to a behavioural classification (fossorial, commensal, etc.) similar to that used in natural sciences, the functional breakdown of genomes (eukaryotes, amniotes), which also follows the natural tree, the composition of space around the concept of a container or the factorization of most of the objects of the domains modelled via the notion of a supported container (Figure 4).
Two salient features emerge from the results presented. The first relates to the evidence-based or trivial nature of the categories resulting from the heuristic used. The second concerns the constantly unfinished and perfectible aspect of the architecture, which adapts and evolves over time.

4.1 Obviousness of Emerging Categories and Associated Robustness

A posteriori, the arborescence that has emerged is trivial. This is the case for the root categories, which sequentially encapsulate time, space, biology, motion, and reproduction, for any domain agent. Similarly, the typology of the diversified behaviour of rodents obtained (Figure 2 bottom) reconstitutes a widely accepted classification of known types of behaviour that is common to zoology (e.g., Nowak, 1999). The same is also true for the representation of genomes, for which, thanks to the similarity between natural phylogeny and OOP inheritance, the characterization of the agents follows a universal biological classification.

The a posteriori evidence of the classifications obtained can be seen as a sign of robustness, an issue of prime importance for the proposed approach. This is the case for the ‘time / space / biology / movement / reproduction / behaviour’ hierarchy, which may, a priori, be applicable to a wide variety of domains. This is also the case for classifications converging on the hierarchies recognized in Nature, which are inherently the most robust. The approach would be thus analogous to those used in bio-inspired modelling (Egan, 2015), but in this case dedicated to the representation of living systems.

4.2 Transitory and Perfectible Nature of the Construction

The incremental approach adopted assumes both continuous re-engineering of the code and control over back-compatibility. With this approach, all the case studies modelled, even the oldest ones, remain active and constantly updated as the model is built. They, thus, continually generate new results, enhanced by the new functions added.

The evidence of robustness for this heuristic makes it possible to consider diverse extensions and improvements to the model. For example, coupling agents with species-specific genomes would make it possible to simulate diverse species, with a view to addressing community ecology issues (Chesson, 2000), a key element in the understanding rodent spread. The model could also be adapted to consider the representation of viruses or parasites in the transmission or non-transmission of diseases to humans, the perception of odours by rodents or the taking into account of energy balances. This would open up as many new possibilities for this approach as there are functionalities to be added.

The class tree can, therefore, continually be tuned step-by-step, by taking new studies into account, provided that these new studies are compatible with the previous ones, or that the previous studies can be rendered compatible with the constraints of the new ones. Consequently, the construction is, by nature, unbounded or perpetually transitory (as long as it remains robust), with methods and properties that can be modified or repositioned.

4.3 Influence of Formalisms on Results

Despite its generic nature, this model may not provide a canonical representation of these diverse worlds. Indeed, each of the choices made in computer modelling, such as the choice of a Perception-Deliberation-Execution paradigm for agent activities or the choices made for the discretisation of space, is only one of the many possibilities that could have been proposed. For example, it would have been possible to base the model on other formalisms, such as ‘Agent / Group / Role’ (Ferber et al., 2004) or ‘Belief / Desire / Intention’ (Caillou et al., 2015). Logically, even using the heuristic presented here, the choice of one of these other possibilities could have resulted in a different architecture.

5 CONCLUSION

The proposed approach is based on the incremental articulation of contrasting case studies within a single model. The continuous consolidation and questioning of the model through new case studies, including studies based on other disciplinary approaches, appears to satisfy the robustness requirements for long-term integration.

The results obtained from the accretion of case studies show that each approach adds to the others already included in the model, to yield an integrated system. The challenge, however, is articulating the model so that it can reveal new processes or dynamics through the multidisciplinary integration of items, concepts or processes (McMurtry, 2009).
The chosen approach must, by definition, be considered to be in a state of continual improvement. However, it can be used to identify the components that are shared, could be shared, or are irreconcilable, and, possibly, the ways to achieve the mutually beneficial integration of points of view specific to different disciplines.

ACKNOWLEDGEMENTS

We would like to thank J.F. Cosson, J.P. Quéré, C. Berthier, B. Gauffre, J.M. Duplantier, L. Granjon, G. Ganem, J. Britton, O. Ninot, J. Lombard, P. Handschumacher, S. Piry, the scientists who kindly agreed to decipher their disciplinary expertise for the formalization of thematic case studies. This study owes much to the work done by Q. Baduel, A. Realini, J.E. Longueville, A. Comte and M. Diakhate, as part of their student internships. The SimMasto project (http://simmasto.org) was supported, through the various case studies of the Chancira projects (grant IRD-ANR-11-CEPL-0010), Cerise (grant IRD-FRB no.AAP-SCEN-20B III), and with occasional support from the French National Research Institute for Sustainable Development (IRD) and the “Centre de Biologie pour la Gestion des Populations” (CBGP, UMR no.22 INRA/IRD/Cirad/Supagro). We also wish to thank J. Sappa and L. Granjon for their help during the revision process.

REFERENCES