Building Qualitative Models in Ecology

Paulo Salles
Institute of Ecology and Resource Management
University of Edinburgh, Mayfield Road
Edinburgh EH9 3JU, UK
psalles@holyrood.ed.ac.uk

Bert Bredeweg
Department of Social Science Informatics (S.W.I.)
University of Amsterdam, Roetersstraat 15
1018 WB Amsterdam (The Netherlands)
bert@swi.psy.uva.nl

Abstract

Building qualitative models is a difficult task. The construction of re-usable models, as well as the formalisation of the modelling process itself, are goals both to researchers in qualitative reasoning and ecology. This paper presents a library of model fragments for reasoning about the behaviour of ecological communities. We have developed a kernel of partial models that represents general knowledge about populations, which can be (re-)used in different situations. In addition, this paper discusses guidelines for the construction of qualitative models, based on our experience in representing the ecology of fire in the Brazilian cerrado vegetation. Both the guidelines and the library of model fragments are part of a larger research effort in trying to use qualitative techniques as a basis for Interactive Learning Environments (ILE) (see also (Bredeweg and Winkels, 1996)).

Our application domain is ecology. There is a growing concern about the world-wide destruction of natural resources. There is a need for educational tools that will contribute to the ecological awareness by learners. We believe that qualitative models can be the basis for some of these tools (Salles et al., 1997).

After many years of developing representations and formalisms to reason qualitatively about physical systems, the qualitative reasoning community starts to recognise the importance of the modelling process itself (Schut and Bredeweg, 1996). A related observation can be made within the community of ecological modellers. They are also trying to formalise the modelling process in order to recognise the principles that can be used by the modellers to explicate their intentions while modelling (Muetszfeldt, 1991). One important goal is to have a library of partial models, and a model construction environment, that can be used by ecologists to build their own models (Robertson et al., 1991).

A few models of ecological systems have been created using qualitative techniques. For example, (Guerrin, 1991) and (Heller et al., 1995) describe hydroecological systems, and (Hunt and Coke, 1994)
modelled the photosynthesis process. (Kamps and Gábor, 1995) describe the implementation of a model about the logistic equation applied to the organisational theory. However, none discusses the modelling process in itself.

The contents of this paper is as follows. The next section introduces the ecological domain that we are dealing with. This section also briefly discusses the need for education within this domain and the importance of qualitative models for that purpose. The third section contains a brief description of GARP (Bredeweg, 1992), the qualitative simulation environment that we use for implementing our models. The guidelines for the modelling process are described in the fourth section. The fifth section presents the qualitative simulation model that we have constructed of the cerrado vegetation. The last section presents our results and discusses ideas for further research.

Ecology in Brazilian Cerrado

The central region of Brazil is covered by a vegetation called cerrado. This huge area of almost 2 million square kilometres is characterised by a tropical climate, with two well marked seasons (wet and dry), and by soils that have low fertility. Within the cerrado vegetation it is possible to identify several types of cerrado physiognomies, spanning from open grasslands to more or less closed forests (see Figure 1). These physiognomies have well defined floristic composition and are mainly determined by fire, soil fertility and the amount of water available in the soil during the dry season. Researchers have investigated the effects of fire on the cerrado (e.g. Coutinho, 1990; Miranda et al., 1996). It has been shown that fire can affect both physical and biological factors: fire causes changes on

- the energy flux,
- water and nutrient cycles,
- the species composition of communities,
- the biomass in many plants, and
- stimulates flowering and germination of seeds in many species.

Fire can therefore be used as a management tool, for example to stabilise the vegetation in certain areas, and to reduce the risk of big fire events (cf. (Pivello, 1992)).

Cerrado is nowadays under great pressure due to farming and human occupation. Large areas of natural vegetation are being destroyed, causing concern about the future. Education is essential to increase ecological awareness.

Computer simulations are important tools for learning. They can complement and, in many cases, substitute field work. It is possible, for example, to control some environmental factors during simulations, and carry out experiments that cannot be done with real ecological systems. By doing so, learners can construct their own hypothesis and investigate them in the context of alternative scenarios. It is a generally held position that this 'learning by doing' will aid learning (cf. (de Jong (editor), 1991)).

Qualitative simulations are particularly important for ecological domains, such as in the cerrado vegetation. Not only are quantitative data often almost non-existent, qualitative models also provide many additional features that are important for having learners interact with the simulation (see also (Bredeweg and Winkels, 1994; Bredeweg and Winkels, 1996)).

Qualitative Simulation Tool

We use GARP (Bredeweg, 1992) as our qualitative simulation tool. GARP takes as input a set of scenarios, a library of model fragments representing the domain knowledge, and a set of transition rules. When running GARP a specific scenario can be selected for simulation. The user can control the simulation in terms of deciding which transition to explore and for how long (one or more successor states). GARP can also be run to produce a full simulation (i.e. total-environment).

GARP uses the following building blocks for constructing a model. First, representation of sim-
ple (physical) entities and their structural relations. Second, representation of time varying properties in terms of quantities and quantity spaces. Quantity spaces are represented independently from a specific model. When building a qualitative model, quantity spaces are assigned to quantities and constraints can be specified between them. Third, representation of all kinds of dependencies between quantities and values of quantities.

Using these building blocks, initial scenarios can be specified as well libraries of model fragments. Scenarios usually consist of a structural description of the system and some initial values for certain quantities. Model fragments are rule-like, in terms of having conditions and consequences. The former specifies the structural descriptions and the specific quantity conditions that must hold in order for the model fragment to be applicable. The givens of a model fragment specify the behavioural features that can be derived. An important part of this is the specification of the causal model underlying the behaviour.

GARP uses a set of rules to reason about state transitions. In normal situations GARP takes a set of general (domain independent) rules for reasoning about terminations and their precedence. However, rules have a similar structure as model fragments and allow the knowledge engineer (or teacher) to represent a rich set of conditions if required. The latter can be important for focusing the simulation and it provides additional leverage for generating explanations.

Guidelines for Modelling

The construction of qualitative models is a difficult and often much time consuming task. One way to support the process of modelling is to learn from decisions made by previous modelling events. It is in this respect important to understand the activities involved in the modelling process as well as the critical decisions to be taken within these activities. In this section we present an initial set of guidelines for the modelling process. Our aim is to explicate the decisions we took during the construction of our models for ecology and to reformulate them as (more) general guidelines for constructing qualitative models.

Purpose of the Model

Modelling requires a great deal of idealisation. The purpose of the model gives the perspective the modeler should take when conceptualising the system to be modelled. The purpose of the model also gives the ‘golden standard’ for evaluation.

Two critical factors in understanding the purpose of a model are:

- the type of user, and
- the role of the model.

Starting with the former, it makes a big difference whether the constructed model is to be used by experienced ecologists or by students in secondary schools. It also matters what kind of task the user will perform with the model. A qualitative model can be used as a tool for inspecting the dynamics of some complex system (e.g. (Yip, 1995)). In this situation the emphasis will be on correct, complete and advanced simulations. In other situations however, different aspects may become more important. Particularly, in educational settings the realisation of ‘articulated’ models is important (Falkenhainer and Forbus, 1991).

For the models described in this paper the objective is to construct an Interactive Learning Environment (ILE) with which students can create their own models using a library of model fragments, run simulations and receive assignments and explanations (Salles et al., 1997). As a result the following aspects are important:

Interactive Simulation It is important that students can change the conditions of the simulation or the values of certain quantities, in order to have a better understanding of the system being simulated. The students should also be allowed to focus the simulation into directions they prefer. It is therefore important that the simulator is fully controllable by the student, so that it can be run step by step if required.

Model fragments as knowledge chunks Model fragments should represent ‘stand-alone’ parts of the domain knowledge that students should master. The idea is that each relevant domain concept (e.g. small population, germination, etc.) is expressed in one model fragment. Model fragments will therefore be an important ingredient for deriving an explanation.

Subsystem Selection

Selecting the subsystem to which a set of equations can be applied is standard practice in physics education (e.g. (Mettes and Roossink, 1981). When building qualitative models we face a similar problem. We have to decide upon the system that will constitute the heart of the model. The subsystem selection will set the focus on what should be modelled and what will be left out.

For the model described in this paper we decided to represent the dynamics of the cerrado communities. Communities are complex entities consisting of many types of plants and animals. We have to abstract from the enormous diversity of organisms and define a finite set of representative entities. We applied the notion of ‘functional group’, commonly used by ecologists to describe communities in cerrado as groups of trees, shrubs and herbaceous-graminoid plants. A functional
group is a set of plants that have some common features and that display similar predictable behaviour when exposed to certain environmental factors. Each functional group can be seen as a population:

**Population as a key concept**

Reasoning about changes in communities requires knowledge about populations (Salles et al., 1996). We have therefore developed a kernel of partial models that represents general knowledge about populations, which can be used in different situations.

**Building an Ontology**

**Individuation** How to characterise the basic concepts that constitute the model? It is important to recognise permanent and temporary properties of the individuals with respect to the purpose of the model. When thinking about populations in ecology, decisions should be made whether to include features such as: sex, age classes, etc. In general, entities must be defined, as much as possible, on the basis of their permanent characteristics (Hayes, 1978).

Our choice was to represent populations as sets of individuals that can be affected by flows of 'born', 'dead', 'immigrated' and 'emigrated' individuals. This approach can be compared to the 'contained stuff' ontology used by (Collins and Forbus, 1989) to build models of thermodynamic processes.

**Quantities and Quantity Spaces**

Properties that change over time are typically represented by quantities and quantity spaces. Crucial for the accessibility and understandibility of a model by students is the amount of variation that is represented within these modelling primitives (too much variation will become confusing for a student). We refer to this as the:

**Minimum required variation** Build quantity spaces such that they facilitate the generation of all the qualitative distinct states that are important for the system at hand.

This rule can be operationalised in different ways. First, we can point out one or more critical quantities in a simulation and assign to these larger quantity spaces, and a more restricted one for the other quantities. Second, we can focus on deviations from a certain 'normal' value. In situations where the notion of equilibrium of the system is important, the quantity space (below-normal, normal, above-normal) can be used. Notice that 'relative' values can be linked to the other quantity spaces by different types of correspondences (see (Bredeweg, 1992)). Third, we can have simulations being based on inequality statements between quantities which can have only a single value. For example, two 'populations sizes' are both 'plus', but one is bigger/equal/smaller than the other. In all the cases described here, the resulting models capture a vocabulary that is focused on the relevant quantities and how these may vary. This makes a model easier to understand for students.

**Processes and Actions**

Following the idea of processes (Forbus, 1984) we represent changes as starting from direct influences which then propagate via indirect proportionalities. The notion of a causal model is crucial in an educational setting. However, additional vocabulary is required in order to capture many complex and intertwined processes in ecological systems. We explicitly use: subtype and consist-of hierarchy between processes. In addition we use the notion of agent models to account for human intervention (Bredeweg, 1992). The subtype hierarchy is important in order to generate utterances such as: 'colonisation is a kind of immigration process that occurs when there's no population in a certain area'. Aggregated processes consists of the sum of a number of processes at a lower level. Sometimes the aggregated process has an ecological meaning and different vocabulary exists for reasoning about that (e.g. notion of population growth consisting of natality, mortality, immigration and emigration processes). Finally, in order to represent human actions that affect some ecological system we used the notion of an 'agent model'. Usually an agent model sets the value of a derivative, for example, the notion of 'conservation' is represented by decreasing fire frequency (see also next section).

**Simplifying the Simulation**

In an educational context, there is a limit to the number of states that can be dealt with by a student (both understanding and motivation become problematic when the set of states is too high). We employed two important mechanisms to simplify the simulation.

**Model fragments as simplifying assumptions**

In order to reuse detailed model fragments in complex scenarios, it is often necessary to take a more abstract view of the system. In our model we realised more abstract views by defining model fragments that summarise certain variations at a lower level.

**Focused state transitions**

State transitions have different probabilities of occurrence. Termination rules (Bredeweg, 1992) can be used to reduce the number of possible states, by means of removing terminations with low probability. Using this approach the number of possible terminations the qualitative simulator considers at each state transition can be reduced considerably.

Both mechanisms implement a kind of simplifying assumption although the realisation in our model is rather different from (Falkenhainer and Forbus, 1991).
Instead of using the assumptions to select the required model fragments they explicitly limit the possible branching of the simulator. The set of used model fragments remains the same (for examples see last subsection below).

**Modelling Cerrado Vegetation**

Following important research on the Brazilian cerrado vegetation (Coutinho, 1990; Pivello, 1992; Moreira, 1992) our models should support explanations about the relation between fire frequency and the structure of the vegetation, expressed as follows:

- If the fire frequency decreases, for example because of human actions, then succession will occur and as a consequence the vegetation will evolve and become denser, with more trees and shrubs and less grass.
- If the fire frequency increases, then a degradation process is active and the vegetation tends to become more open, with less trees and shrubs and more grass.

Given a certain scenario or initial problem, the models should allow the students to make predictions and postdictions about the system (cf. (Forbus, 1984)). Typically, the models should facilitate derivations such as "the campo sujo changed to campo cerrado because fire frequency decreased". They should also provide access to the underlying causal models that represent how these changes follow from different responses of populations of trees, shrubs and grass to environmental influences such as light, humidity and temperature. As mentioned before, reasoning about changes in communities requires knowledge about populations.

**Models of Populations**

Populations consist of groups of individuals of the same kind, living in a certain place, during a certain period of time. The size of the population is an important factor, because it is an indication of the balance of the forces acting on the individuals. The quantity introduced to express this is: Number-of. It can take on values from different quantity spaces, depending on the objectives of the model. We used mainly the quantity space \{zero, low, medium, high, maximum\} to describe the 'absolute' qualitative values, and make comparisons between populations.

Following our assumption that each relevant concept should be represented by a model fragment, we defined fragments for the concepts: small, medium, large, maximum sized, non-existent and extinct populations. Behaviour is expressed in model fragments representing the notions of increasing, decreasing and steady populations. We included also definitions of populations of trees, shrubs, and grass, the most used types of organisms represented in these models (see also Figure 3).

**Figure 3: Model fragments hierarchy of populations**

In order to predict changes in populations and to build explanations about the results of simulations, we need a vocabulary to express the basic processes affecting the individuals. They are being born (natality), they die (mortality), they arrive from elsewhere (immigration), and they may leave (emigration). Changes in the size of a population depend on the balance of these processes. The four basic processes introduce the following quantities (rates): Born-flow, Dead-flow, Immigrated-flow and Emigrated-flow. These flows have the quantity space \{zero, plus\}, because they cannot be negative. Subtypes of the basic processes were defined to take into account some particularities. For example, instances of natality and mortality processes were used to describe how environmental factors can influence the Born-flow and Dead-flow in trees, shrubs and grass. Also we defined the colonisation process as a subtype of immigration.

Note that these processes are independent of each other and as such do not (individually) define the final direction of change in the size of the population. We need the additional notion of population growth to express how the basic processes combine in a particular situation. The population growth process is defined as an aggregation of the four basic processes and represents a unique flow. It introduces the quantity Growth-rate with the quantity space \{minus, zero, plus\}. Growth-rate can be calculated by the addition of the amount of individuals represented in each flow. The hierarchy of processes is shown in Figure 2.

**Models of Communities**

Communities are groups of populations. Communities in cerrado can be classified according to the Number-of trees, shrubs and grass. Typically, researchers (cf. (Coutinho, 1990)) classify the cerrado into: campo limpo, campo sujo, campo cerrado, cerrado sensu stricto and cerrado. In order to model this classification we created model fragments representing each of
them. Some intermediate communities were included, to help the understanding of the transitions (see also Figure 4). The values used to characterise the main Figure 4: Model fragments hierarchy of communities types of communities are presented in Table 1.

![Figure 2: Model fragments hierarchy of Processes](image)

![Figure 4: Model fragments hierarchy of communities](image)

tors are always present in any scenario described by the models. Thus, their quantity space is \{plus\}.

The canopy of the trees has an important influence on the factors mentioned above. In our models this is represented by the quantity Cover, with the same quantity space as used for the population of trees. It is assumed that there exists a direct correspondence between the Number-of trees and the amount of Cover: the value taken by the former is also assigned to the latter. For example, if the value of Number-of trees is low, then Cover is also low.

All the above mentioned factors are influenced by fire. Fire frequency is a component of the so called 'fire regime' (Whelan, 1995). It expresses how often a vegetation is burned. In the model this is represented by the quantity Fire-frequency, which can take on values from the quantity space \{zero, plus\}. Fire frequency changes as a consequence of human actions. This is modelled by using agent models.

The influence from fire frequency on the community is indirect: it propagates through the described network of environmental factors that finally influences the basic processes of plant populations. Altogether, 16 direct influences (I) and 32 indirect influences (P) affecting 33 quantities constitute the full structure of the causal model, as shown in Figure 5.

**Results from Simulations**

We ran several simulations with the model. One of the simulations produced the environment graph depicted in Figure 6. It shows the successional changes in cerrado predicted by the hypothesis presented in the beginning of this section.

In order to reduce the number of ambiguities and possible states in the full simulation, we added a few assumptions to the model. The most important one was redefining the campo cerrado as a typical community with the values for trees, shrubs and grass

<table>
<thead>
<tr>
<th></th>
<th>Grass</th>
<th>Shrub</th>
<th>Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>campo limpo</td>
<td>\textgreater \textit{med}</td>
<td>\textit{zero}</td>
<td>\textit{zero}</td>
</tr>
<tr>
<td>campo sujo</td>
<td>\textit{high}</td>
<td>\textit{low}</td>
<td>\textit{med}</td>
</tr>
<tr>
<td>campo cerrado</td>
<td>\textit{zero} &amp; \textit{max}</td>
<td>\textit{zero} &amp; \textit{max}</td>
<td>\textit{med}</td>
</tr>
<tr>
<td>cerrado sensu s.</td>
<td>\textit{low}</td>
<td>\textit{med}</td>
<td>\textit{high}</td>
</tr>
<tr>
<td>cerradoao</td>
<td>\textit{zero}</td>
<td>\textit{high}</td>
<td>\textit{med}</td>
</tr>
</tbody>
</table>

Table 1: Classification of typical cerrado communities

The cerrado communities are related to a general ecosystem, the cerrado sensu lato. The model fragment that defines the cerrado sensu lato specifies relevant properties of the micro-environment at the surface of the soil. It introduces the quantities Nutrient, Humidity, Light and Temperature. These quantities are related to the amount of Litter, the dead material that covers the ground (leaves, small pieces of wood and other parts of plants). We assume that these factors are always present in any scenario described by the models. Thus, their quantity space is \{plus\}.

The canopy of the trees has an important influence on the factors mentioned above. In our models this is represented by the quantity Cover, with the same quantity space as used for the population of trees. It is assumed that there exists a direct correspondence between the Number-of trees and the amount of Cover: the value taken by the former is also assigned to the latter. For example, if the value of Number-of trees is low, then Cover is also low.

All the above mentioned factors are influenced by fire. Fire frequency is a component of the so called 'fire regime' (Whelan, 1995). It expresses how often a vegetation is burned. In the model this is represented by the quantity Fire-frequency, which can take on values from the quantity space \{zero, plus\}. Fire frequency changes as a consequence of human actions. This is modelled by using agent models.

The influence from fire frequency on the community is indirect: it propagates through the described network of environmental factors that finally influences the basic processes of plant populations. Altogether, 16 direct influences (I) and 32 indirect influences (P) affecting 33 quantities constitute the full structure of the causal model, as shown in Figure 5.

**Results from Simulations**

We ran several simulations with the model. One of the simulations produced the environment graph depicted in Figure 6. It shows the successional changes in cerrado predicted by the hypothesis presented in the beginning of this section.

In order to reduce the number of ambiguities and possible states in the full simulation, we added a few assumptions to the model. The most important one was redefining the campo cerrado as a typical community with the values for trees, shrubs and grass
being equal to 'medium'. The effect can be seen in the graph: there is some branching for the campo limpo and campo sujo communities, but the environment then moves in a straight line from campo cerrado up to cerrado. We also removed the influences from Humidity and Nutrient on the population of grass in order to reduce ambiguity. For all simulations we employed the 'Minimum required variation' rule when assigning quantity spaces to quantities. Finally, we adapted some termination rules in order to remove many 'impossible' transitions that the simulator was trying. For example, as long as there exists a population, the natality process will be active. However, due to environmental factors the Born-flow may decrease, triggering a termination to zero. An assumption specifies that this termination can only happen when the population becomes extinct. Adding assumptions such as these speed up the simulation process, and more important, make the result transparent and therefore easier to understand.

The resulting qualitative models offer several possibilities for tutoring. Using different initial scenarios, it is possible to explore selected parts of the causal path. For example, we can analyse the effect of each factor on the populations, or the effects of a group of factors on a specific population, etc.

**Discussion**

The construction of reusable domain models is an important goal both to researchers in qualitative reasoning and ecology. In this paper we have presented a library of model fragments for reasoning about the behaviour of ecological communities. We have developed a kernel of partial models that represents general knowledge about populations, which can be (re-)used in different situations. The models are implemented in GARP (Bredeweg, 1992), a qualitative simulation environment implemented in Prolog.

The construction of qualitative models is a difficult and often much time consuming task. Supporting the modelling process requires an understanding of the activities involved as well the critical decisions to be taken within these activities. In this paper we have discussed a set of initial guidelines for the construction of qualitative models, based on our experience in representing the ecology of fire in the Brazilian cerrado vegetation.

The purpose of a model is an important overall factor in determining how to conceptualise and represent a certain system. Our domain of application is ecology and the models are used as the basis for educational tools for teaching ecological awareness. In this paper we discuss how notions such as, (1) fully interactive simulation, (2) model fragments as knowledge chunks,
(3) population as a key concept, and (4) basic processes, aggregated processes, and 'agent models' are important for the construction of qualitative models that can be used in an 'guided discovery' oriented educational setting. Being concerned with teaching also requires that simulation models have a limited size, otherwise they become intractable for students. We have discussed how simplifying assumptions can be employed for this. Our approach differs from (Falkenhainer and Forbus, 1991) in that we limit the possible variations of the simulator, while the set of used model fragments remains the same.

We are currently improving the prototype of the ILE and its capacity of generating explanations (Salles et al., 1997). The work includes:

- expanding the library, in order to reason about fuel dynamics, climatic changes and other aspects of fire on the cerrado vegetation;

- describing the life cycle of cerrado plants, that is describing flowering, fruit production, seed production, germination, and the survival of young plants. Each of these stages can be affected by fire and there are several interesting points to be explored (particularly from the educational point of view);

- creating different simulations to produce explanations in specific contexts. We are creating tasks and problem solving situations for the students to explore the learning environment.

Acknowledgements This work has been done during a visit of Paulo Salles at the University of Amsterdam. We are grateful to the FUOS (96B20) for the financial support for this visit. Paulo Salles is also grateful to the Brazilian Conselho Nacional do Desenvolvimento Científico e Tecnológico (CNPq) for the scholarship to pursue a PhD program at the University of Edinburgh (process 201823/92-6).

Paulo Salles is on leave from: Laboratory for Research on Biology and Science Teaching, Department of Genetics and Morphology (GEM), University of Brasilia, 70.910 - Brasilia - DF, Brazil. E-mail: psalles@guarany.unb.br.

We would like to thank Joost Breuker, Kees de Koning, Radboud Winkels and Bob Muetzelfeldt for their support and comments on the research presented in this paper.

References


