Research Challenges in Modelling Ecosystems

Antonio Cerone[1](✉) and Marco Scotti[2]

[1] Dipartimento di Informatica, University of Pisa, Pisa, Italy
`cerone@di.unipi.it`
[2] GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany
`marcoscot@gmail.com`

Abstract. Ecosystems and their biodiversity have to be protected and preserved as sources of services and goods. The human population controls and modifies ecosystems to improve its health conditions and welfare. The consequences of human activities should be carefully monitored and ecosystems should be managed to protect all of the species and preserve their functioning. The development of strategies for ecosystem management benefits from the use of computational techniques to model the dynamics of species that interact with their abiotic and biotic environment. Life scientists and computer scientists need to work together to define and analyse ecosystem models. However, there is a multifaceted gap between the approaches used in life science and those used in computer science. Such gap is both cultural and technical, and results in a number of challenges. In this paper we identify these challenges and provide technical and cultural proposals for solving them.

1 Introduction

As human activity threatens the functioning of ecological systems by habitat destruction [26], fragmentation [69], climate change [12], and introduction of allochthonous species [27,64], we face the problem of understanding and managing the consequences of these impacts. The goal of environmental policy actions is to preserve biodiversity and ecosystem services. Then identifying the key features responsible for species survival (e.g. absence of a specific pollutant; level of fragmentation of the landscape network; genetic heterogeneity within the population) is the only viable long term solution for managing biodiversity loss.

In population ecology (which deals with the dynamic behaviour of populations, by focusing on the interactions with other species and the abiotic environment), there is a need to combine and coordinate information from different domains. The behaviour of each individual emerges from the complex interplay between social relationships within the population, trophic and non-trophic

---

We would like to thank Paolo Milazzo for inspiring discussions and for suggesting Dynamic I/O Automata as a possibly appropriate modelling formalism. Discussions with Matteo Pedercini contributed to clarify strengths and weaknesses of System Dynamics (SD). Finally, we also would like to thank the anonymous referees who provided alternative but equally valuable opinions reported in Sect. 5.

C. Canal and A. Idani (Eds.): SEFM 2014 Workshops, LNCS 8938, pp. 276–293, 2015.
DOI: 10.1007/978-3-319-15201-1_18

interactions (e.g. host-parasite and plant-pollinator) with individuals of other species and spatial movements (i.e. dispersal in the landscape network). Linking these organisational levels is still a challenge: an increasing need for hierarchical thinking is present in ecological stoichiometry (community-level patterns concerning the ratios of certain elements [30]), and community genetics (how genetic variance influences ecosystem functioning [38]). Traditional modelling, focusing on macroscopic patterns and adopting a deterministic approach based on average population behaviour (i.e. through the application of ordinary differential equations), is weak in several respects. The inherent stochasticity and variability and the large-scale patterns produced by local rules are important features that should be more thoroughly investigated. Although the importance of these aspects is recognised, novel approaches should be developed to incorporate stochastic dynamics in ecological modelling (e.g. stochastic processes are often modelled by deterministic equations with added random noise).

New tools should be implemented for better understanding how to preserve highly endangered species and plan actions of biodiversity conservation in complex ecological communities. There is a need to improve stochastic modelling for better understanding demographic noise and local interactions, especially in case of small populations. Stochasticity is not a source of unpredictability and randomness; rather, it represents a set of processes that are often neglected in the phase of model design, but that can produce higher-level patterns [23]. Such new tools would help in modelling the link between local and global processes, simulating density dependence [17] and dealing with several other challenges of ecology. Most likely the explicit modelling of hierarchical organisation will be one of the key contributions to ecological research [47,65]. Ecologists emphasised the importance of modelling demographic and environmental stochasticity in metapopulation dynamics [18], investigated fluctuations affecting the densities of populations in communities as a consequence of environmental variability [62], and analysed the effects of random perturbations on cyclic population dynamics [43]. Actions of conservation biology often aim to protect rare species, which are characterised by small population size, with individuals showing a highly heterogeneous behaviour. For these reasons, we argue that stochastic modelling can represent a step ahead in the domain of ecological research. Ecosystem management would benefit from novel computational tools that allow researchers to extend stochastic-based dynamics towards spatial and temporal simulations. Results extracted from these analyses could serve for suggesting best strategies of environmental sustainability and planning actions for biodiversity conservation [52]. In practice, they might aid in planning systems-based conservation strategies [15], defining optimum programmes for managing multispecies fisheries [71], creating sustainable agroecosystems [58], investigating the functioning of bio-geochemical cycles [19], predicting risks of secondary extinction [29], and ranking of conservation priorities [50].

Section 2 reviews the main modelling approaches used in ecosystem science. Section 3 discusses the multifaceted gap between the individual-based modelling used in life science and identifies challenges that emerge from such gap. Section 4 provides technical and cultural proposals for solving the identified challenges and Sect. 5 concludes with some considerations on opinions of other scientists that provide possible alternatives to our proposals.

2 State of the Art and Literature Review

2.1 Mathematical Modelling

Mathematical models [33] are essential in making precise theoretical arguments about the factors affecting observed phenomena. Once validated, mathematical models can be exploited to make predictions about the future evolutions of the system under study.

The use of mathematical models in population biology and ecology is nowadays common practice. Many books describe the basic concepts and the most well-established models [36,46,53,60]. Among the most successful modelling strategies we mention two approaches: age-structured population growth and spatial spread [36]. The first approach allows predictions concerning long-term changes in population numbers based on information about the age at which individuals have offspring and the probabilities of death at different ages. The second approach allows predictions about the future rate of spread of some populations from initial observations. However, the use of mathematical models has also some limitations: for example, such models are often based either on differential equations or on recurrence relations describing how the size of a population changes over time (with continuous or discrete time, respectively). Since differential equations and recurrence relations are deterministic, they are not suitable to model systems whose behaviour could be determined by choices between alternatives associated with probabilities.

Although mathematical models become difficult to be studied analytically when the complexity of the modelled system grows, a large increase in computational power and the development of high-level modelling methods now support the simulation of highly complex models. New methods and tools have been developed to ease the definition of models that are based on differential equations. One of these methods, System Dynamics (SD), developed at the Massachusetts Institute of Technology in the early 1960s, provides a powerful framework to build, simulate and analyse complex models, stressing the relationship between model structure and behaviour [61]. Moreover, the SD method enables a multidisciplinary approach to problems [68], and thus supports the development of comprehensive models for decision-making. Such comprehensive models can be built in a modular way, to allow for some flexibility and adaptability of model structure to different circumstances.

A fundamental characteristic of these models is that system descriptions are very high level, with populations represented as a whole and their dynamics defined top-down in terms of global laws. The internal dynamics is, therefore, a black box. Thus, models can describe neither biological aspects of individuals nor interactions among individuals that are not reflected in global laws controlling the dynamics of the population. A bottom-up definition of the population dynamics, in which population properties emerge through the interaction of individuals, is not possible using SD or methods based on differential equations. Although it is possible to play with the parameters of global laws and identify patterns of changes, such as growth, oscillation and decay, and how these patterns may respond to human intervention, there is actually no way to capture the impact of human intervention at a lower level, e.g. at individual level. For example, imagine that we introduce genetically modified plants that are resistant to chemical treatments in a natural ecosystem, with the purpose of using a given herbicide without leading to the death of the genetically modified individuals, and we know that a side effect (e.g. susceptibility to drought) of this genetic modification is a change in single individual's behaviour. Methods based on differential equations cannot capture the impact of the behavioural changes in single individuals on the population dynamics, unless such a situation has been observed in the past, which is not always the case.

2.2 Individual-Based Models Using Formal Notations

The notion of individual-based model in ecological modelling corresponds to the computer science notion of agent-based model, namely of a model in which there are multiple active entities (the agents), whose behaviour is governed by a set of usually simple rules, that are allowed to interact with each other and move in some virtual environment. A classical example of agent-based modelling notation is that of Cellular Automata [25,45], which consists in a regular grid of cells (usually one- or bi-dimensional), where each cell is associated with a finite number of states and can pass from one state to another depending on the states of adjacent cells in the grid. In the simplest versions of Cellular Automata agents are identified with cells, but there exist extensions in which cells represent positions in the environment, whereas agents are entities that can move from one cell to another, and behave in accordance with the state of the cell and of other agents in the same cell.

A class of modelling notations that are well-known in computer science and can be used to describe populations at the individual level is Petri Nets [59]. Petri Nets, in their most common formulation, are diagrammatic notations consisting of places and transitions, which have been defined with the aim of modelling concurrent systems sharing common resources. Places represent conditions and transitions represent events that may occur when there are agents that satisfy certain conditions. Agents are represented by tokens that can move from one place to another when transitions take place. Petri Nets are one of the simplest and most successful notations of computer science for the description and analysis of concurrent systems. They have also been applied to the modelling of ecological systems [66], also in combination with Cellular Automata [34].

Other recent definitions of individual-based models that exploit formal notations of computer science are based on membrane systems [54]. Membrane systems are distributed parallel computing devices inspired by the structure and the functioning of living cells. A membrane system consists of a hierarchy of membranes, each of them containing a multiset of objects, representing molecules, a set of evolution rules, representing chemical reactions, and possibly other membranes. A model of a population of Bearded Vultures based on membrane systems has been developed using P systems, a formalism belonging to the category of membrane system [20].

The individual-based approach [51] is computationally more expensive than the mathematical modelling approach based on differential equations described in Sect. 2.1, but allows life scientists to explore how the dynamics of a population or of an ecosystem arises from the ways in which individuals interact with each other and with the environment. Due to this computational cost a pure individual-based approach can only be used with populations consisting of a small number of individuals.

2.3 Stochastic Simulation and Individual-Based Models

The limitations of mathematical modelling, the effectiveness of computational models to deal with stochastic aspects, and the level of performances reached by computers in the last few years motivate the increase in the application of computational means in life sciences. In fact, the adjective "computational" is becoming widely used in life sciences to qualify disciplines such as biology, ecology, epidemiology, and so on. However, in these disciplines the adjective "computational" often simply means that stochastic simulation techniques are exploited in order to study properties of mathematical models of systems of interest.

Most stochastic simulation techniques usually consider a relatively small number of classes of individuals, and then use standard probability distributions (binomial, Poisson, etc.) to generate the number of individuals in each class, at each successive time step.

The current trend in the study of population dynamics is to enrich individual-based models with stochasticity, in order to attempt to follow each individual in the population from its birth, through growth, dispersal and reproduction, to death [10]. Such an "individual-based" stochastic approach [51] is computationally more expensive than stochastic traditional approaches, which make use of a small number of aggregate categories, but allows life scientists to explore how the dynamics of a population or of an ecosystem arises from the ways in which individuals interact with each other and with the environment. In this individual-based context, stochastic simulation algorithms such as one of the variants of the Gillespie Algorithm [31,32] generate statistically correct population evolutions.

3 Identification of Research Challenges

Formal analysis techniques of theoretical computer science, such as static analysis and model checking, can be applied to agent-based models in order to verify properties of the described systems. These techniques are well-established in computer science but are practically unknown to life scientists. The rest of this section is organised as follows. In Sect. 3.1 we identify the multifaceted gap between individual-based modelling and formal analysis techniques. Then in Sect. 3.2 we illustrate the research challenges that we encounter in order to fill in this gap.

3.1 A Multifaceted Gap

There is a mutifaceted gap between the individual-based modelling used in life science and the formal analysis techniques used in computer science. This gap can be characterised by the following aspects.

A1 — High-level vs. Low-level Descriptions. *Life scientists* use *high-level* notations that represent models in a visual way annotated with natural language descriptions. Such notations support an almost direct representation of biological and ecological processes. However, the semantics of such models is not formally defined and there is no guarantee that simulations really reflect the intended behaviour of the model. *Formal languages* are based on *low-level* primitives that are close to machine-readable operators, but have to be combined in a complex manner to define high-level biological and ecological processes. Their semantics can be unambiguously defined using mathematical transformations and tools, which, however, obscure the intuition and require deep mathematical skills to be used.

A2 — Extensive Simulations vs. Property Verification. *Life scientists* perform a large number of *simulations* of the same model and then use statistics and/or data mining techniques to extract patterns, oscillations and tendencies in the population dynamics. *Static analysis and model checking techniques* support the characterisation and verification of properties of a model of population dynamics without explicit recourse to simulations.

A3 — Ecological Problem vs. Mathematical Tool. *Life scientists* focus on the *ecological problem* and see mathematical notations and tools as mere instruments to solve their problem. *Computer scientists* normally focus on *mathematical notations and tools* and use simplified and often unrealistic versions of biological and ecological problems to investigate theoretical aspects of such notations and tools.

A4 — Field Data Collection vs. *Ad-hoc* Data. *Life scientists* collect *data in the field* and/or use *historical* data to calibrate their models and run simulations. *Computer scientists* often have to cope with the lack of data to be used for running their models. Thus they are often forced to define *ad-hoc data*, which may not be realistic, with the danger of being biased in choosing data that best illustrate the features and potential of their formal languages and analysis techniques.

A5 — Realistic Models vs. Abstract Models. *Life scientists* tend to include *realistic details* that facilitate the intuitive understanding of the model behaviour and make the model more appealing, but this often increases the computational complexity of the model itself with a negative impact on the efficacy of the analysis techniques. *Computer scientists* define *abstract* models that contain only the details needed for the intended analysis, thus possibly obscuring the understanding of the model behaviour.

A6 — Understand/Control vs. Replication. The final goal of *life scientists* is to be able to *understand* the functioning of the ecosystems and test possible *control* intervention on components of the ecosystem model, aiming to perform adjustments to the system behaviour and evaluate the impact of such intervention on the entire ecosystem. Although the final goal of life scientists has been supported by the mathematical modelling work performed using the SD approach, the use of formal models has been restricted to the *in-silico replication* of the ecosystem evolution without much attention to the evaluation of human intervention impact. To make it worse, as we have seen in Aspect **A4**, since computer scientists typically do not use real data, the ability of replicating reality is mostly just potential and is seldom documented in the literature.

3.2 Research Challenges

A lot of efforts have been devoted during the last decade to the attempt of filling in the gap between individual-based modelling and formal analysis techniques. In spite of such large efforts there are still no conclusive results in this direction and a number of challenges have emerged in the process. In this section we illustrate such challenges and match each of them to the aspects of the gap identified in Sect. 3.1, by using the same top-level sequential number **n** for challenges (**Cn**) as we used for their corresponding aspects (**An**).

C1 — Define an Appropriate Common Language. The definition of a common language that allows life scientists and computer scientists to cooperate in the definition and analysis of models requires the selection of basic biological and ecological processes and their implementation using a formal language. The main challenges in this task are:

- **C1.1 — Language Expressiveness.** There is a need to define a set of high-level primitives that is sufficiently expressive for life scientists.
- **C1.2 — Intuitive Semantics.** Each primitive should be associated with a simple semantics that addresses intuition and can be understood without a deep mathematical knowledge.
- **C1.3 — Correctness of the Implementation.** There is a need to guarantee that the implementation faithfully captures the behaviour resulting from the selected biological and ecological processes.
- **C1.4 — Modelling Ease.** The use of the primitives in the modelling process should be facilitated through the use of templates, defined operators and modelling frameworks.

- **C2 — Limitations of Simulation and Verification**
  - **C2.1 — Analysis of Simulations.** Simulations provide only a sample of possible behaviours rather than a characterisation of all possible behaviours. Moreover, extracting global information from a set of simulations is not an easy task and the outcome of this process may vary depending on the techniques used (i.e. various statistical methods and data mining techniques) and on the assumptions and choices made in applying such techniques (choice of simulation parameters and pieces of information to data mine).
  - **C2.2 — State Space Explosion Problem.** The use of verification techniques in a stochastic individual-based approach results in state spaces that grow exponentially with the size of the population; the use of spatiality makes the exponential growth even faster.
- **C3 — Right Model for a Given Ecosystem.** A cultural challenge is to urge computer scientists to shift the primary focus of their research investigation from the mathematical tool to the ecological problem. The challenge for computer scientists is therefore to be able to define "the right model for a given ecosystem" rather than "the appropriate ecosystem for their own model".
- **C4 — Data Collection.** There is a need to create multidisciplinary research teams in which life scientists and computer scientists collaborate in all phases of the research: field work planning, data collection, data analysis, model design and implementation, in-silico experiments and their interpretation. In particular, computer scientists cannot use data that have been collected by life scientists for other purposes, but they have to design new field work for collecting appropriate data for their research. The challenge here is for life scientists and computer scientists to define the appropriate form of collaboration, in which the field work is planned by a multidisciplinary research team, is carried out by life scientists and produces data to be analysed by a multidisciplinary research team.
- **C5 — Right Level of Abstraction.** The model must be defined at an abstraction level sufficiently informative to keep alive the intuition about the system behaviour without including irrelevant details that may have a negative impact on the computational complexity.
- **C6 — Addressing Policy Support.** The use of formal models to address policy support is a challenging task. SD has been successful in exploring the impact of policy implementation on behaviour of ecosystems [68]. In particular, the use of SD in the T21 modelling framework [1], developed by the Millennium Institute, integrates economic, social and environmental factors to support comprehensive and participatory development planning. However, the T21 approach, and in general any approach based on SD, does not support the modelling of the impact of policy implementation at individual level.

The challenge in using a formal approach for individual-based modelling is the integration of economic, social and environmental factors within the same model.

4 Addressing Challenges

In this section we address the challenges identified in Sect. 3.2 and propose possible strategies and research questions to bridge the gap between individual-based modelling of populations and formal analysis techniques.

The final aim is to develop new theories of population dynamics based on theoretical computer science means. New theories should be based on well-established computer science notations, such as rewrite systems, finite state automata and Petri Nets, adapted and extended to describe population individuals. Moreover, such theories should deal with both deterministic and stochastic behaviours of individuals and take into account spatial movement and landscape dispersal.

4.1 Formal Notation (Addressing Challenge C1)

The events in the life of an individual that are usually of interest for the construction of a population model are birth, death and interactions with other individuals (either conspecific or belonging to different species) and with the environment. Examples of relevant interaction events are those that have some influence on the population size (e.g. mating, predation) or on the life-conditions of the individual itself (e.g. nutrition, migration). All these events are often discrete and may cause new individuals to appear (e.g. to be born, to come from another population in the neighbourhood), and current individuals to disappear (e.g. to die, to leave the local population) or change their states (e.g. from “available to mate” to “pregnant”, from “egg” to “larva”).

Rewrite Systems. A possible way to model such events is by using rewrite systems [11]. In the rewrite systems approach events may be modelled as rewrite rules, such as $egg \longrightarrow larva$, that is the rule that rewrites a term (egg) into a new one ($larva$). In this way the set of rewrite rules of the model of a population (or a category of individuals in a population) predicts all events that may occur to that population (or to that category of individuals). For example set $\{egg \longrightarrow larva, egg \longrightarrow \epsilon\}$, where $egg \longrightarrow \epsilon$ describes the death of an egg, with ϵ denoting the empty term, predicts all possible events that may occur to an egg.

The occurrence of these kinds of events, however, depends not only on the current state of the individual (e.g. an egg may change to larva, but an adult cannot) but also on the current state of the environment in which the individual lives (e.g. an egg dies if the temperature goes below a specific threshold). In general, in ecological systems we need to deal with a variety of environmental events, whose cause is often unknown or depends on a very complex combination of factors, which are external to the system itself. For example the dynamics of a population of a given species depends not only on the interaction with other species within the same ecosystem, such as predators, prey and competitors, but also on the occurrence of environmental events such as climatic events (i.e. variation of temperature and rainfalls) and events related to habitats (i.e. tree clearing, bushfires, desiccation of a water container, pollution, hunting and human settlement). Therefore, we have to associate a representation of the environment Env with the current term and include in the rule the representation of the environmental condition $cond(Env)$ that enables the rule. Thus the rule that models the death of an egg becomes $cond(Env) : [egg]_{Env} \longrightarrow [\epsilon]_{Env}$. Finally, in order to introduce stochasticity in an individual-based model defined as a rewrite system, rewrite rules are associated with a rate k that describes the frequency with which the rule is used in the computation. Thus the rule that models the death of an egg becomes $cond(Env) : [egg]_{Env} \longrightarrow_k [\epsilon]_{Env}$.

In previous work [11] we assume the existence of a list of external events, with information about the time when these events occur. The occurrence of an external event may modify some environmental information that affects ecosystem evolution, such as temperature, vegetation density, volume of water, level of pollution, size of a population, human density. Moreover, the list of external events may change dynamically. For instance, a bushfire event, which decreases the vegetation density, will be removed from the list of external events after the occurrence of a rainfall event, and will be replaced with a new bushfire event with a later occurrence time. Similarly, a desiccation event, which decreases the volume of a water container will be removed from the list of external events after the occurrence of a rainfall event, and will be replaced with a new desiccation event with a later desiccation time. Lists of external events that contain historical data or data collected through field work are used to calibrate the model. Once calibrated, the model is run together with a new list of external events that describe human intervention and policy implementation.

This approach addresses Challenges **C1.1**–**C1.3**: the rule construct is sufficiently expressive to describe relevant high-level events such as birth and death (**C1.1**); the semantics of rewriting is quite intuitive even for the stochastic version in which rules are associated with rates (**C1.2**); variants of Gillespie algorithms ensure a correct implementation of the rules (**C1.3**). However, the approach does not address Challenge **C1.4**. This is due to the following two issues: (1) when the number of the details needed for describing the complexity of an organism increases, the set of rules associated with the term that describes a state of that organism also increases and each rule of such set may be affected by a complex combination of environmental conditions and interactions with other individuals; (2) the rate of a rule is not constant but often depends on a complex combination of environmental conditions.

Although the number of different states in which an individual might be is usually quite small, the large number of rules that describes the state transition associated with each state and the complexity of functions describing rule rates make the task of the modeller difficult.

Automata, Process Algebras and Petri Nets. The fact that the number of different states in which an individual might be is usually quite small suggests that some extensions of finite state automata and process algebras with appropriate parallel composition and interaction capabilities could be exploited.

Some interesting work has been carried out in systems biology using modelling languages based on process algebras. Ciocchetta and Hillston developed Bio-PEPA [22], a language for the modelling and analysis of biochemical networks, which is based on PEPA (Performance Evaluation Process Algebra). Although Bio-PEPA can successfully handle some features of biochemical networks, such as stoichiometry and different kinds of kinetic laws, the operators that describe interactions of reactants, products and enzymes do not address intuition and, therefore, do not appeal life scientists. Moreover, Bio-PEPA does not support the modelling of external events, thus lacking an essential feature for modelling ecosystems. Kahramanoğulları *et al.* [41,42] developed LIME (Language Interface for Modeling Ecology), a language tool for stochastic dynamic simulation in ecology. LIME supports model definition using a narrative style that facilitates the analysis of parallel, multiple ecological interactions in metacommunities. LIME translates the model description into the BlenX programming language for stochastic dynamical simulation [28]. In BlenX, the propensities of interactions between individual entities can be modelled either as simple rates or in terms of more complex functions (e.g. Holling's type functional responses), and the spatial distribution is described in terms of membership to discrete locations in space (e.g. landscape patches). This discrete description of space might impair the chance of modelling individual dynamics for which exact spatial coordinates need to be known and traced (e.g. fish schooling [48]).

Since births and deaths of individuals must always be described while modelling ecosystems, it would be useful to have a formalism that supports dynamic creation and destruction of components of a parallel composition. A formalism that presents this feature is the Dynamic I/O Automata proposed by Attie and Lynch [4]. An automata-based formalism can be suitably used to build a population model by starting from the description of the events that may happen in the life of each kind of individual. We would need to define an automaton for each kind of individual and compose in parallel as many copies of such automata as individuals of the corresponding kind are present in the initial population. Another important characteristic that makes Dynamic I/O Automata useful in modelling populations of individuals is the ability to dynamically change the signature of an automaton, that is, the set of actions in which the automaton can participate. In this way an automaton describing an individual can change its signature to mimic the evolution of that individual through its maturation stages (e.g. from "egg" to "larva" to "adult"). This is definitely more natural than the destruction of the old term and the creation of a new term that is used to model maturation with a rewrite system.

We might also think of translating a process algebraic model or a constructed parallel composition of automata into a Place/Transition Petri Net. Since the number of kinds of individuals that belong to a population is finite, as well as the number of states of every process/automaton modelling a single individual, a Petri Net could be constructed by considering one place for each state of each process/automaton modelling a kind of individual, and one transition for each transition in any of such process/automaton (by taking into account synchronisations between processes/automata). The translation into Petri Nets would allow some properties of the population dynamics to be verified statically by computing the invariants of the obtained net.

In order to be used for modelling ecosystems, formalisms based on automata, process algebras and Petri nets must be extended aiming to the definition of a complete modelling framework in which also quantitative [9] and spatial aspects [7,8] of population dynamics are taken into account. Quantitative aspects of population dynamics are related with duration, frequency and probability of the events that may happen in the population. Spatial aspects consist of the description of the topology of the population environment, the positions of the individuals in the environment and the movement from one position to another. Several probabilistic, stochastic, timed and spatial extensions of automata and Petri Nets have already been defined and are now well-established in computer science [3,14,44,70]. Similar extensions have also been proposed for other kinds of formalisms such as process algebras [13,37,40,57] and rewrite systems [6,11, 16,21,55,67].

Automata-based and process algebraic formalisms as well as Petri Nets have the advantage that verification techniques, such as model checking, can often be applied easily to them. Moreover, they are usually associated with friendly graphical notations, which make them immediately understandable also to non-specialists. These advantages with respect to rewrite systems clearly address Challenge **C1.4**. However, up to now, we could not identify any approaches based on automata, process algebras or Petri nets that address Challenges **C1.1**–**C1.2**.

4.2 Analysis Methodologies (Addressing Challenge C2)

Simulation is nowadays one of the most common analysis techniques for models of biological and ecological systems. Simulators can be implemented quite easily by following standard approaches (e.g. Monte Carlo simulation and numerical integration) and can give useful information on the dynamics of the modelled systems with acceptable computational costs. Moreover, simulators for some standard ecological models are available to be used by ecologists and wildlife managers without the need of knowing model details. Furthermore, simulation may characterise the most probable system behaviours and be used for calibration purposes, that is, to validate models against available data.

Concerning formal analysis, model checking and abstract interpretations are well established techniques that can potentially be used to analyse biological and ecological systems. Efficient probabilistic model checkers, such as PRISM [49], are the most promising tools in this sense; modelling notations for ecosystem modelling can be translated into the input language of a model checker. Important work in this direction has been carried out by Romero-Campero *et al.* [63] and by Philippou, Toro and Antonaki [56].

However, in order to deal with quantitative and spatial aspects of population biology and ecology, formalisms must express notions such as position, age, probability and duration, which all together could make the translation into the input language of the model checker not feasible. Consequently, it would be reasonable to translate into the model checker language only fragments of the formalism that are suitable to describe particular classes of biological and ecological systems, whereas new verification techniques should be developed, in which all the quantitative and spatial notions are handled.

Interesting model checking methodologies for stochastic processes have been developed in the last decade. Quantitative properties of stochastic systems are usually specified in logics that explicitly compare the measure of executions satisfying certain temporal properties with thresholds. The model checking problem for stochastic systems with respect to such logics is typically solved by a numerical approach that iteratively computes (or approximates) the exact measure of paths satisfying relevant subformulae; the algorithms themselves depend on the class of systems being analysed as well as the logic used for specifying the properties. Hansson and Jonsson [35] introduced the Probabilistic Computation Tree Logic (PCTL) for specifying properties of Discrete-Time Markov Chain (DTMC) while Baier *et al.* [5] carried out extensive work on model checking of Continuous-Time Markov Chains (CTMC), by defining the Continuous Stochastic Logic (CSL) and developing the proofs of theoretical foundations as well as the model checking algorithms. In general, these model checking approaches, called numerical model checking, have a number of limitations: (1) numerical algorithms work only for special systems that have certain structural properties (e.g. Markov Models); (2) numerical algorithms require a lot of time and space, thus scaling to large systems is a challenge; (3) the logics for which model checking algorithms exist are extensions of classical temporal logics, which are often not the most popular among life scientists.

One way to overcome these weaknesses of numerical model checking could be the search for the right compromise between simulation and model checking. An interesting approach in this direction is *statistical model checking*, which overcomes the disadvantages of numerical model checking at the cost of being less accurate. In this approach the system is simulated for finitely many runs, using hypothesis testing to infer whether the samples provide a statistical evidence for the satisfaction or violation of the specification. Statistical model checking was first introduced by Younes [72] in 2005. The idea underlying statistical model checking is to perform the model checking analysis on a sample of the population rather than the entire population. Although the use of a sample causes a loss in accuracy, statistical model checking provides a mechanism to calculate the size of the sample that ensures the satisfaction of the property with a given probability.

The most basic statistical model checking algorithm considers the probability α of false positive and β of false negative with respect to a given property φ. Then, given a probability p, the algorithm computes two natural numbers, c and n, such that property φ has to be satisfied by c simulations of a Stochastic Discrete Event System $\mathcal{M}$ out of a total of n performed simulations to ensure that φ is satisfied in $\mathcal{M}$ with probability p.

Interest in statistical model checking has been growing during the last five years and a workshop explicitly devoted to statistical model checking has been held for the first time in 2013 [2]. The limited number of applications of statistical model checking to biological systems that have been carried out up to now include the verification of temporal properties of rule-based models of cellular signalling networks [24] and a sophisticated statistical model checking algorithm that uses Bayesian sequential hypothesis testing. This requires fewer system simulations and has the ability to incorporate prior biological knowledge about the model being verified [39].

We believe that statistical model checking has the potential to address Challenge **C2** by realising an optimal compromise between simulation and verification. As a model checking technique it supports the verification of a property, but drastically reduces the number of system simulations, thus overcoming the state explosion problem (Challenge **C2.1**).

4.3 Filling in the Cultural Gap (Addressing Challenges C3–C6)

In Sects. 4.1 and 4.2 we have dealt with the most technical Challenges (**C1**–**C2**) in filling the gap between the individual-based modelling used in life science and the formal analysis techniques used in computer science. In this section we globally address Challenges **C3**–**C6** using a *cultural* rather than *technical* perspective.

Challenge **C3** is the most representative aspect of the *cultural gap* between life scientist and computer scientist. Here the need to change culture only involves computer scientists, who should shift their research focus from mathematical tools to ecological problems.

Only after this cultural challenge is solved, the other three challenges (**C4**–**C6**) can be properly addressed. Moreover, establishing the technical basis of a common language (Challenge **C1**) is the prerequisite that can facilitate the creation of multidisciplinary research teams and their collaboration throughout all research phases from field data collection to interpretation of in-silico experiments (Challenge **C4**). Throughout this continuous collaboration process, multidisciplinary teams should be also facilitated in agreeing on the right level of abstraction for the considered problem (Challenge **C5**) and on the choice of the factors to consider in evaluating and comparing policy implementations and their impact on the ecosystem (Challenge **C6**).

Finally, we must mention that it is opinion of some researchers from both computer science and ecology that cultural differences between the two communities are slowly disappearing. Worldwide there are efforts in proposing multidisciplinary projects, and universities are developing new multidisciplinary educational programmes. We can optimistically expect that future generation of scientists will have the necessary multidisciplinary culture to successfully address Challenges **C1-C6**.

5 Final Considerations

In this paper we have taken an ecology-driven perspective and claimed that it is essential to address all Challenges **C1**–**C6** in order to be able to define an effective framework for modelling ecosystem dynamics.

Some computer scientists, instead, have as their main concern the challenge of designing the appropriate mathematical notations for capturing ecological systems, while dealing with the state-space explosion problem and other technical challenges. In this perspective, the main aim is that of refining frameworks in order to better capture aspects and properties of ecological systems. The use of simplistic or even unrealistic *ad hoc* data, therefore, would be justified by a need to first address Challenges **C1**–**C2**, without taking Aspects **A3**–**A6** into account, with the expectation that, once the theory matures enough, researchers will naturally turn to address Challenges **C3**–**C6**.

In Sect. 4.1 we have surveyed a number of formal notations used in modelling biological and ecological systems and identified which, in our opinion, may be appropriately extended to successfully address Challenges **C1.1**–**C1.4**.

An alternative approach, which has both computer scientists and life scientists among its supporters, favours the adoption of graphical languages similar to the ones typically adopted by ecologists. The main challenges of such approach would be to give a formal semantics to the graphical language and, based on such a semantics, define a translation to a formal language or tool to be used for analysis.

References

1. T21. http://www.millennium-institute.org/integrated_planning/tools/t21/
2. Workshop on statistical model checking. http://rv2013.gforge.inria.fr/workshop.html
3. Alur, R., Dill, D.L.: A theory of timed automata. Theoret. Comput. Sci. **126**, 183–235 (1994)
4. Attie, P.C., Lynch, N.A.: Dynamic input/output automata: a formal model for dynamic systems. In: Larsen, K.G., Nielsen, M. (eds.) CONCUR 2001. LNCS, vol. 2154, pp. 137–151. Springer, Heidelberg (2001)
5. Baier, C., Haverkort, B., Hermanns, H., Kaoten, J.-P.: Model-checking algorithms for continuous-time markov chains. IEEE Trans. Softw. Eng. **29**(7), 524–541 (2003)
6. Barbuti, R., Cerone, A., Maggiolo-Schettini, A., Milazzo, P., Setiawan, S.: Modelling population dynamics using grid systems. In: Cerone, A., Persico, D., Fernandes, S., Garcia-Perez, A., Katsaros, P., Ahmed Shaikh, S., Stamelos, I. (eds.) SEFM 2012 Satellite Events. LNCS, vol. 7991, pp. 172–189. Springer, Heidelberg (2014)
7. Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Pardini, C.: Spatial calculus of looping sequences. Theor. Comput. Sci. **412**(43), 5976–6001 (2011)
8. Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Pardini, C., Tesei, L.: Spatial P systems. Nat. Comput. **10**(1), 3–16 (2011)
9. Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Troina, A.: A methodology for the stochastic modeling and simulation of sympatric speciation by sexual selection. J. Biol. Syst. **17**(3), 349–376 (2009)

10. Barbuti, R., Mautner, S., Carnevale, G., Milazzo, P., Rama, A., Sturmbauer, C.: Population dynamics with a mixed type of sexual and asexual reproduction in a fluctuating environment. BMC Evol. Biol. **12**(1), 49 (2012)
11. Basuki, T.A., Cerone, A., Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Rossi, R.: Modelling the dynamics of an aedes albopictus population. In: Proceedings of AMCA-POP 2010. Electronic Proceedings in Theoretical Computer Science, vol. 33, pp. 18–36 (2010)
12. Bawa, K.S., Markham, A.: Climate change and tropical forests. Trends Ecol. Evol. **10**, 348–349 (1995)
13. Beaten, J.C.M., Bergstra, J.A.: Real-time process algebra. Formal Aspects Comput. **3**, 142–188 (1991)
14. Beauquier, D.: On probabilistic timed automata. Theoret. Comput. Sci. **292**, 65–84 (2003)
15. Berkes, F.: Rethinking community-based conservation. Conserv. Biol. **96**, 5066–5071 (2004)
16. Bistarelli, S., Cervesato, I., Lenzini, G., Marangoni, R., Martinelli, F.: On representing biological systems through multiset rewriting. In: Moreno-Díaz Jr., R., Pichler, F. (eds.) EUROCAST 2003. LNCS, vol. 2809, pp. 415–426. Springer, Heidelberg (2003)
17. Björnstad, O.N., Fromentin, J.M., Stenseth, N.C., Gjøsæter, J.: Cycles and trends in cod populations. Proc. Nat. Acad. Sci. U.S.A. **96**, 5066–5071 (2009)
18. Bonsall, M.B., Hastings, A.: Demographic and environmental stochasticity in predator-prey metapopulation dynamics. J. Anim. Ecol. **73**, 1043–1055 (2004)
19. Botter, G., Settin, T., Marani, M., Rinaldo, A.: A stochastic model of nitrate transport and cycling at basin scale. Water Resour. Res. **42**, 404–415 (2006)
20. Cardona, M., Colomer, M.A., Pérez-Jiménez, M.J., Sanuy, D., Margalida, A.: Modeling ecosystems using P systems: the bearded vulture, a case study. In: Corne, D.W., Frisco, P., Păun, G., Rozenberg, G., Salomaa, A. (eds.) WMC 2008. LNCS, vol. 5391, pp. 137–156. Springer, Heidelberg (2009)
21. Cavaliere, M., Sburlan, D.: Time–independent P systems. In: Mauri, G., Păun, G., Jesús Pérez-Jímenez, M., Rozenberg, G., Salomaa, A. (eds.) WMC 2004. LNCS, vol. 3365, pp. 239–258. Springer, Heidelberg (2005)
22. Ciocchetta, F., Hillston, J.: Bio-pepa: a framework for the modelling and analysis of biochemical networks. Theoret. Comput. Sci. **410**(33–34), 3065–3084 (2009)
23. Clark, J.S.: Beyond neutral science. Trends Ecol. Evol. **24**, 8–15 (2009)
24. Clarke, E.M., Faeder, J.R., Langmead, C.J., Harris, L.A., Jha, S.K., Legay, A.: Statistical model checking in *BioLab*: applications to the automated analysis of T-cell receptor signaling pathway. In: Heiner, M., Uhrmacher, A.M. (eds.) CMSB 2008. LNCS (LNBI), vol. 5307, pp. 231–250. Springer, Heidelberg (2008)
25. Codd, E.F.: Cellular Automata. Academic Press, New York (1968)
26. Coleman, F.C., Williams, S.L.: Overexploiting marine ecosystem engineers: potential consequences for biodiversity. Trends Ecol. Evol. **17**, 40–44 (2002)
27. Da Silva, J.M.C., Tabarelli, M.: Tree species impoverishment and the future flora of the Atlantic forest of northeast Brazil. Nature **404**(6773), 72–74 (2000)
28. Dematté, L., Priami, C., Romanel, A.: The BlenX language: a tutorial. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) SFM 2008. LNCS, vol. 5016, pp. 313–365. Springer, Heidelberg (2008)
29. Ebenman, B., Jonsson, T.: Using community viability analysis to identify fragile systems and keystone species. Trends Ecol. Evol. **20**, 568–575 (2005)

30. Elser, J.J., Sterner, R.W., Gorokhova, E., Fagan, W.F., Markow, T.A., Cotner, J.B., Harrison, J.F., Hobbie, S.E., Odell, G.M., Weider, L.W.: Biological stoichiometry from genes to ecosystems. Ecol. Lett. **3**(6), 540–550 (2000)
31. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**(4), 403–434 (1976)
32. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**(25), 2340–2361 (1977)
33. Giordano, F.R., Weir, M.D., Fox, W.P.: A First Course in Mathematical Modeling. Brooks/Cole, Cengage Learning, Belmont (2009)
34. Gronewold, A., Sonnenschein, M.: Event-based modelling of ecological systems with asynchronous cellular automata. Ecol. Model. **108**, 37–52 (1998)
35. Hansson, H. Jonsson, B.: A logic for reasoning about time and reliability. Research report SICS/R(0013, SICS) (1994)
36. Hastings, A.: Population Biology: Concepts and Models. Springer, New York (1997)
37. Hennessy, M., Regan, T.: A process algebra for timed systems. Inf. Comput. **117**, 221–239 (1995)
38. Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N., Vellend, M.: Ecological consequences of genetic diversity. Ecol. Lett. **11**, 609–623 (2008)
39. Jha, S.K., Clarke, E.M., Langmead, C.J., Legay, A., Platzer, A., Zuliani, P.: A bayesian approach to model checking biological systems. In: Degano, P., Gorrieri, R. (eds.) CMSB 2009. LNCS, vol. 5688, pp. 218–234. Springer, Heidelberg (2009)
40. Jonsson, B., Larsen, K.G., Yi, W.: Probabilistic extensions of process algebras. In: Bergstra, J.A., Ponse, A., Smolka, S.A. (eds.) Handbook of Process Algebra. Elsevier, New York (2001)
41. Kahramanoğulları, O., Jordán, O., Lynch, J.F.: A language interface for stochastic dynamical modelling in ecology. Environ. Model Softw. **26**(5), 685–687 (2011)
42. Kahramanoğulları, O., Lynch, J.F., Priami, C.: Algorithmic systems ecology: experiments on multiple interaction types and patches. In: Cerone, A., Persico, D., Fernandes, S., Garcia-Perez, A., Katsaros, P., Ahmed Shaikh, S., Stamelos, I. (eds.) SEFM 2012 Satellite Events. LNCS, vol. 7991, pp. 154–171. Springer, Heidelberg (2014)
43. Kaitala, V., Ranta, E., Lindstroem, J.: Cyclic population dynamics and random perturbations. J. Anim. Ecol. **65**, 249–251 (1996)
44. Kartson, D., Balbo, G., Donatelli, S., Franceschini, G., Conte, G.: Modelling with Generalized Stochastic Petri Net. Wiley, New York (1994)
45. Kier, L.B., Seybold, P.G., Cheng, C.: Modelling Chemical Systems Using Cellular Automata. Springer, Dordrecht (2005)
46. Kingsland, S.: Modelling Nature: Episodes in the History of Population Ecology. University of Chicago Press, Chicago (1995)
47. Kolasa, J.: Complexity, system integration, and susceptibility to change: biodiversity connection. Ecol. Complex. **2**, 431–442 (2005)
48. Kunz, H., Hemelrijk, C.K.: Artificial fish schools: collective effects of school size, body size, and body form. Artif. Life **9**, 237–253 (2003)
49. Kwiatkowska, M., Norman, G., Parker, D.: PRISM: probabilistic symbolic model checker. In: Field, T., Harrison, P.G., Bradley, J., Harder, U. (eds.) TOOLS 2002. LNCS, vol. 2324, pp. 200–204. Springer, Heidelberg (2002)
50. Lande, R., Engen, S., Swether, B.E.: Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press, Oxford (2003)
51. Lomnicki, A.: Population Ecology of Individuals. Princeton University Press, Princeton (1988)

52. Mace, G.M., Collar, N.J.: Priority setting in species conservation. In: Norris, K., Pain, D.J. (eds.) Conserving Bird Biodiversity. Cambridge University Press, Cambridge (2002)
53. McCallum, H.: Population Parameters: Estimation for Ecological Models. Wiley-Blackwell, New York (2000)
54. Paun, G.: Membrane Computing: An Introduction. Springer, Heidelberg (2002)
55. Pescini, D., Besozzi, B., Mauri, G., Zandron, C.: Dynamical probabilistic P systems. Int. J. Found. Comput. Sci. **17**, 183–204 (2006)
56. Philippou, A., Toro, M., Antonaki, M.: Simulation and verification for a process calculus for spatially-explicit ecological models. Sci. Ann. Comput. Sci. **23**(1), 119–167 (2013)
57. Priami, C.: Stochastic pi-calculus. Comput. J. **38**, 578–589 (1995)
58. Rasmussen, P.E., Goulding, K.W.T., Brown, J.R., Grace, P.R., Janzen, H.H., Körschens, M.: Long term agroecosystem experiments: assessing agricultural sustainability and global change. Science **282**(5390), 893–896 (1998)
59. Reisig, W.: Petri Nets: An Introduction. Springer, Heidelberg (1985)
60. Renshaw, E.: Modelling Biological Population in Space and Time. Cambridge Univerity Press, Cambridge (1991)
61. Richardson, G.P.: Introduction to System Dynamics Modeling with Dynamo. MIT Press, Cambridge (1981)
62. Ripa, J., Ives, A.R.: Food web dynamics in correlated and autocorrelated environments. Theor. Popul. Biol. **64**, 369–384 (2003)
63. Romero-Campero, F.J., Gheorghe, M., Bianco, L., Pescini, D., Jesús Pérez-Jímenez, M., Ceterchi, R.: Towards probabilistic model checking on P systems using PRISM. In: Hoogeboom, H.J., Păun, G., Rozenberg, G., Salomaa, A. (eds.) WMC 2006. LNCS, vol. 4361, pp. 477–495. Springer, Heidelberg (2006)
64. Schea, K., Chesson, P.: Community ecology theory as a framework for biological invasions. Trends Ecol. Evol. **17**, 170–176 (2002)
65. Scotti, M., Ciocchetta, F., Jordán, F.: Social and landscape effects on food webs: a multi-level network simulation model. J. Complex Netw. **1**(2), 160–182 (2013)
66. Seppelt, R., Temme, M.M.: Hybrid low level petri nets in environmental modelling - development platform and case studies. In: Matthies, M., Malchow, H., Kriz, J. (eds.) Integrative Systems Approach to Natural and Social Science. Springer, Heidelberg (2002)
67. Setiawan, S., Cerone, A.: Stochastic modelling of seasonal migration using rewriting systems with spatiality. In: Counsell, S., Núñez, M. (eds.) SEFM 2013. LNCS, vol. 8368, pp. 313–328. Springer, Heidelberg (2014)
68. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw Hill Higher Education, New York (2000)
69. Tigasa, L.A., Vurena, D.H.V., Sauvajot, R.M.: Behavioral responses of bobcats and coyotes to habitat fragmentation and corridors in an urban environment. Biol. Conserv. **108**, 299–306 (2002)
70. Wang, J.: Timed Petri Nets: Theory and Applications. Kluwer Academic Publisher, Boston (1998)
71. Yodzis, P.: Must top predators be culled for the sake of fisheries? Trends Ecol. Evol. **16**, 78–84 (2001)
72. Younes, H.L.S.: Verification and Planning for Stochastic Processes with Asynchronous Events. PhD thesis, Carnegie Mellon University (2005)