Ecological Informatics 20 (2014) 58–66

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A machine learning approach to investigate the reasons behind species extinction

Morteza Mashayekhi [a,*], Brian MacPherson [b], Robin Gras [a,b]

[a] *School of Computer Science, University of Windsor, Canada*
[b] *Department of Biology, University of Windsor, Canada*

ARTICLE INFO

Article history:
Received 26 October 2013
Received in revised form 6 February 2014
Accepted 8 February 2014
Available online 12 February 2014

Keywords:
Individual-based model
Machine learning
Species extinction
Rule extraction

ABSTRACT

Species extinction is one of the most important phenomena in conservation biology. Many factors are involved in the disappearance of species, including stochastic population fluctuations, habitat change, resource depletion, and inbreeding. Due to the complexity of the interactions between these various factors and the lengthy time period required to make empirical observations, studying the phenomenon of species extinction can prove to be very difficult in nature. On the other hand, an investigation of the various features involved in species extinction using individual-based simulation modeling and machine learning techniques can be accomplished in a reasonably short period of time. Thus, the aim of this paper is to investigate multiple factors involved in species extinction using computer simulation modeling. We apply several machine learning techniques to the data generated by EcoSim, a predator–prey ecosystem simulation, in order to select the most prominent features involved in species extinction, along with extracting rules that outline conditions that have the potential to be used for predicting extinction. In particular, we used five feature selection methods resulting in the selection of 25 features followed by a reduction of these to 14 features using correlation analysis. Each of the remaining features was placed in one of three broad categories, viz., genetic, environmental, or demographic. The experimental results suggest that factors such as population fluctuation, reproductive age, and genetic distance are important in the occurrence of species extinction in EcoSim, similar to what is observed in nature. We argue that the study of the behavior of species through Individual-Based Modeling has the potential to give rise to new insights into the central factors involved in extinction for real ecosystems. This approach has the potential to help with the detection of early signals of species extinction that could in turn lead to conservation policies to help prevent extinction.

1. Introduction

One of the most fundamental questions in population biology and conservation biology relates to species persistence and the risk of extinction where one species cannot survive because its individuals are unable to reproduce, or they simply cannot tolerate the environmental conditions. Species extinctions result from a variety of biotic and abiotic factors, such as population size (Rai, 2003; Schueller and Hayes, 2011), habitat destruction and degradation, human intervention, infectious disease, reproduction rate, migration rate (Drake and Griffen, 2010), invasive species (Sax and Gaines, 2008), environmental variation (Drake and Lodge, 2004), habitat fragmentation (Joshi et al., 2006), habitat quality and size (Griffen and Drake, 2008b), the Allee effect (Dennis, 1989), genetic inbreeding (Reed et al., 2003), genetic diversity (Markert et al., 2010), initial population size (Drake et al., 2011), patch size (Collins et al., 2009), age (Doran et al., 2006), and energy (Evans et al., 2005). These factors increase the probability of extinction, and can be classified into three broad categories: demographic stochasticity, genetics and environmental factors (Griffen and Drake, 2008a), although admittedly there is overlap between these broad categories.

Random fluctuations in demographic factors such as birth rate and death rate can have dramatic effects on populations. The effect of demographic stochasticity is greater in smaller populations than in larger ones (Rai, 2003). In addition, there are factors relating to the transmission of genes from one generation to the next. Genes may be lost from a small population and the gene frequencies may be modified due to drift or inbreeding (Rai, 2003; Reed et al., 2003). Diminishing genetic variation may increase extinction risk by limiting the ability to adapt to stressful environments. Lastly, environmental factors such as natural catastrophes (including fires, floods, earthquakes, and volcanoes), temperature, availability of food, competitors, predators, and diseases influence the population by changing the demographic parameters. For example, Gregory and Courchamp (2010) advanced experimental

---

[*] Corresponding author. Tel.: +1 5192533000x3003.
E-mail addresses: mashaye@uwindsor.ca (M. Mashayekhi), macphe4@uwindsor.ca (B. MacPherson), rgras@uwindsor.ca (R. Gras).

http://dx.doi.org/10.1016/j.ecoinf.2014.02.001
1574-9541/© 2014 Elsevier B.V. All rights reserved.

evidence suggesting that predators can produce the so-called Allee effect (a reduction in population size making extinction more likely). Further, there is reason to believe that volcanic activity had a major role to play in five mass species extinctions (Palike, 2013).

Using mathematical modeling to study species extinction is prohibitively difficult, and consequently most results in this area are approximations at best, especially if a consideration of a mixture of relevant factors is desired (Ovaskainen and Meerson, 2010). Similarly, results obtained from laboratory experiments often conflict with field studies (Griffen and Drake, 2008a). Moreover, observing and studying species in nature in order to extract species information is a highly time consuming and complicated process given that populations exist within an interacting network of species, along with being distributed in a patchy manner over a heterogeneous space. A promising attempt to overcome these difficulties is the application of Individual-Based Modeling (IBM) as a bottom-up approach (DeAngelis, 2005; Grimm and Railsback, 2005). A bottom-up approach simulates the behavior of individuals and their interactions, creating a dynamic complex system which can lead to the emergence of new phenomena in the system. This approach has been used in the simulation of ecological and evolutionary processes such as ecological speciation (Thibert-Plante and Hendry, 2009), conservation applications (Lopex-Alfaro et al., 2012), and gender change (Zhang et al., 2010). Gras et al. (2009) have developed EcoSim, an IBM system including a behavioral model based on Fuzzy Cognitive Map (FCM) (Kosko, 1986). Several studies have validated some of the patterns observed in EcoSim. For example, Devaurs and Gras (2009) have analyzed the species abundance patterns observed in the communities generated, based on Fisher's log series. They demonstrated that their simulations produce results relating to species abundance patterns that cohere with patterns observed in real ecological systems. In another study, the chaotic behavior of the system with multi-fractal properties has been established, which agrees with what has been observed for real ecosystems (Golestani and Gras, 2010). Golestani et al. (2012) added small, randomly distributed physical obstacles into the simulations to investigate the influence of obstacles on the distribution of populations and species, the level of gene flow, as well as the mode and tempo of speciation. Hosseini Sedehi et al. (2012) were able to predict species extinction in EcoSim with high accuracy, and earlier work by Mashayekhi et al. (2013) prior to the current article involved being able to predict features important to both speciation and extinction in EcoSim. These studies demonstrate the potential of EcoSim simulations to approximate some important features of real ecosystems, although admittedly it does have its limitations such as the absence of abiotic factors (climate, fluctuations in temperature, precipitation, wind, soil changes, or geographic features such as mountains, valleys, rivers, lakes). However, many of these limitations exist in laboratory-based studies as well.

The overall aim of this study is to use an Individual-Based Modeling approach to investigate a wide variety of important factors contributing to extinction, along with investigating their predictive potential using methods that circumvent the difficulties with empirical studies and mathematical modeling. To achieve this aim, we analyzed the information gathered from EcoSim followed by the integration of the extracted knowledge to verify species extinction realization in EcoSim under the three broad categories of genetic, environmental, and demographic in line with (Griffen and Drake, 2008a). We used individual-based computer simulations that take into account species interactions (including the effects of predation) and which are relatively inexpensive to run and which take a relatively short period of time to complete. We designed an approach based on a combination of feature filtering, focusing on the most informative features, and predictive model building. We evaluated the accuracy of the predictive model building to assess the quality and the generality of the models obtained, with an eye towards extinction prediction. In addition, this predictive model helped us to extract some effective prediction rules based on these filtered features. This approach increases the testability of ecological and biological mechanisms of species extinctions.

2. Material and methods

2.1. EcoSim

In EcoSim, prey species and predator species are simulated in a torus-like discrete world, which is a 1000 × 1000 matrix of cells. Besides prey and predators, every cell in this world may contain a limited amount of grass, which is the first level of a 3-tiered food chain (grass–prey–predator), and a limited amount of meat (originating from captured prey). The availability of grass is calculated based on a spatial diffusion model that shapes the dynamic environment, where predators live on prey and prey live on grass. The predators act as an environmental stressor on prey. In EcoSim, the fuzzy cognitive map (FCM) not only serves as the basis for describing and computing agents' behaviors, but also functions as the platform for modeling evolutionary mechanisms and speciation events. A species is represented as a set of genetically similar individuals (Mallet, 1995). The speciation mechanism implemented in EcoSim is based on the gradual divergence of populations containing individuals that are increasingly genetically different. Each individual acts according to its FCM which is coded in its genomes and assigned to it at birth time. The FCM is a directed graph containing nodes called “concepts” along with “edges” that represent the influence of concepts on one another. There are three kinds of concepts: sensitive (such as distance from food and sexual partners), internal (such as fear, hunger and satisfaction) and motor (such as escape, eating and reproduction). A positive weight associated with an edge in the agent's FCM corresponds to an excitation, whereas a negative weight is related to an inhibition. A null value means that there is no influence between the two concepts. The environmental information of an agent is used to compute the activation level of a sensitive concept. Further, the activation level of an internal concept is influenced by the sensitive concepts of the agent. Lastly, the action of the agent is selected based on the activation levels of the motor concepts that are affected by the sensitive and internal concepts. Fig. 1 shows an example of a simple FCM where there are three influence edges: closeness to a foe (foeClose), distance from a foe (foeFAR), and fear. Activation of the concepts foeClose and foeFar are computed by fuzzification of the real value of the distance from the foe. Further, the defuzzification of the activation of evasion provides information regarding the speed of the evasion.

Each iteration of EcoSim represents a time step, consisting of the computation of concepts' activation level, along with the choice and application of an action for each individual. A time step also includes an update of the world: the emergence and extinction of species, along with the growth of grass, and the diffusion of grass that involves the probability of grass spreading to the neighboring cells. The energy of a given individual, which is provided by the primary or secondary resources found in the world, decreases depending on its actions. One of the actions performed by individuals is reproduction, with several factors playing important roles in the reproductive process. To be successful, the two parents need to be in the same cell, have sufficient energy, choose the reproductive action, and be sufficiently genetically similar. The organisms cannot ascertain their genetic similarity with their potential partner. Although this is a limiting feature of the EcoSim simulation, it does not entail an absence of reproductive barriers, given that if individuals attempting to mate are sufficiently genetically dissimilar, the reproduction fails. The failure of reproduction in this case is effectively representative of a postzygotic reproductive barrier. A successful reproductive action eventuates in a unique offspring with a genome representing a combination of the parental genomes. The newborn individual receives an initial amount of energy proportional to the energy that the two parents spend in reproduction. (For more details about EcoSim, please refer to Appendix B.)

2.2. Data preparation

We extracted our data from nine different runs of EcoSim, each run involving 10,000 time steps, including all the applicable demographic,