2012

Studying Mate Choice and Species Extinction using EcoSim

Seyed Meisam Hosseini Sedehi
University of Windsor

Studying Mate Choice and Species Extinction using EcoSim

by

Seyed Meisam Hosseini Sedehi

A Thesis
Submitted to the Faculty of Graduate Studies
through Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2012

© 2012 Seyed Meisam Hosseini Sedehi

Studying Mate Choice and Species Extinction using EcoSim

by

Seyed Meisam Hosseini Sedehi

APPROVED BY:

_____
Dr. Stéphanie Doucet
Department of Biological Sciences

_____
Dr. Ziad Kobti
School of Computer Science

_____
Dr. Robin Gras, Advisor
School of Computer Science

_____
Dr. Scott Goodwin, Chair of Defense
School of Computer Science

May 10, 2012

DECLARATION OF CO-AUTHORSHIP/PREVIOUS PUBLICATION

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

In this thesis, we tackle the two biological questions of Mate Choice and Species Extinction using a predator-prey ecosystem simulation named EcoSim. We conducted two experiments to studies these questions. In the first experiment, the impact of three mate choice models, random, similar, and exaggerated ornamental mate, on the behaviour and characteristic of individuals is studied. Experimental results showed very interesting genetic and behavioural differences between each model. In the second experiment, the effect of demographic, environmental, and genetic factors on the species extinction was investigated. Experimental results illustrated that some combinations of these factors can lead to predict the near future extinction with some interesting insights about the effect of these factors.

DEDICATION

This thesis is dedicated to my parents who are always in my heart and to my beloved sister for her endless support and encouragement.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Dr. Robin Gras for his continuous support, guidance and patience throughout this study. Gratitude is also expressed to my committee: Dr. Stéphanie Doucet for her detailed and constructive comments and Dr. Ziad Kobti for his support and guidance. I am also grateful to my colleagues at EcoSim Lab for their consultation and help in all the time of research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Biological and ecological processes are complex systems formed by the complex interactions between biotic and abiotic components. Studies of these processes are difficult and in some cases infeasible due to these intricate relations. Moreover, data collection, due to a huge number of individuals dispersed on large landscapes, takes a long time for most of these processes. Furthermore, the manipulation of natural parameters for empirical research, to study their effects on the process and the analysis of some long term processes such as evolution, is impossible in a reasonable time. Laboratory approaches allow running tests by providing simplified and tractable systems where interactions of factors can be eliminated or restricted, and tests can be under control. These approaches are complementary and help to expand the knowledge about complex processes. In this thesis, two biological phenomena related to individuals' interactions, Mate Choice and Species Extinction, are studied.

In biology, the pairing of individuals for copulation including the raising of offspring is known as mating. Mate choice is a mechanism in which two individuals choose each other to mate. This mechanism relies on the behavioural and morphological traits that attract and stimulate mates, such as offering of nutrition, territories, nest sites or other resources needed by the mate for breeding [Andersson and Iwasa 1996].

The theory of sexual selection introduced by Charles Darwin, has stated why extravagant male traits have not been eliminated by natural selection, although these traits make an individual more conspicuous to predators and definitely reduce its survival

[van Dijk et al. 2010]. Moreover, growing and preserving ornamented traits requires an energetic cost. The interaction between the female preferences and the male traits plays an interesting role in the evolutionary process, especially for their impact on the different characteristics and behaviors of the individuals who form various species [Bateson 1983].

The evolutionary process has an undeniable effect on the sexual selection mechanisms. The studies of gradual long-running evolution are typically a combination of empirical and laboratory researches. Lancaster et al. [2009] made an empirical laboratory experiment on common side-blotched lizards (*Uta stansburiana*) with alternative reproductive strategies based on assessing the throat color and the dorsal patterning of males to investigate the effect of multiple traits on the offspring survival. They concluded that females choosing males with adaptive combinations of these traits, "would exhibit high fitness due to a high number of their progeny inheriting the appropriate trait combinations that increase their survival". Robinson et al. [2011] investigated the female preference on the evolution of an exaggerated male ornament using the dorsal fin of *Xiphophorus birchmanni* male fish as a courtship signal. The authors state that the female preference selects enlarged dorsal fins in male. Moreover, they conclude that males who court more raise their dorsal fins more, and smaller males with smaller absolute dorsal fins court more. They mentioned according to dependency between male size and dorsal fin size, it is difficult to clarify which one is more important in this case. In addition, they explained that the female preference was involved in the evolution of the enlarged dorsal fins in males for this species.

In addition to mate choice which is an interaction between mates, species extinction which explains the survival ability of a species is studied in this research. The

conservation of endangered species and expansion of their longevity have always encouraged scientists to be in search for the fundamental reasons of species' extinction. Each species can combine one or more distinct populations with similar ecological niche. Populations' extinction which is a milestone of biology and ecology has applications in conservation biology, biological control, epidemiology and genetics [Griffen and Drake 2008], [Drake et al. 2011].

Many laboratory tests have been done on species extinction. For instance, Drake et al. [2011] have investigated experimentally the effect of population size on the population's extinction of the flea *Daphnia magna* in two phases: initial and quasi-stationarity. They concluded that the population size has less effect on populations with high resilience, but habitat size, and environmental variability have more impact. In another experiment, Drake and Griffen [2010] have used laboratory populations of *Daphnia magna* to test the population dynamics due to declining levels of food provision. They showed that extinction will be revealed by slowing down the growth rate.

In general, producing an appropriate condition and an exact repeatable experiment is difficult, in particular when more than one factor has an effect on a process. Simulation techniques can be a good alternative to consider several factors together. One such technique, called Individual-Based Model (IBM) [DeAngelis and Mooij 2005], has been used in simulation of ecological and evolutionary processes such as ecological speciation [Thibert-Plante and Hendry 2009], genomic complexity [Adami et al. 2000], gender change [Zhang et al. 2010] and population dynamics [Letcher et al. 1998].

Gras et al. [2009] introduced a predator-prey ecosystem simulation called EcoSim, which is the combination of an IBM with Fuzzy Cognitive Map as the behaviour

model for the agents. This model allows investigating different aspects of life by evolving individuals in a multi-level food chain simulation. The predators act as a pressure factor on a prey and can be seen as an environmental stress. The prey eats grass and the grass availability is based on a spatial diffusion model leading to a dynamically changing environment. Moreover, this model contains a speciation process which generates new species from existing ones by an evolutionary process.

In this study, EcoSim is used as a platform for the investigation of mate choice and species extinction. This work is only a preliminary study on these phenomena. In this study some natural characteristics such as gender are omitted to simplify complex interaction between various conditions and parameters. We tried to test some controversial hypotheses, and also make some new hypotheses using a simulation based study of biological processes.

The main contribution of the author in this study is for analysing the information gathered from Ecosim and integrating the extracted knowledge to expand the understanding of these complex processes. For this purpose, we used different attributes of Ecosim, such as demographic, spatial, and genetic factors, called features. These features were gathered at each time step from individuals currently existing in the world. 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 to assess the quality and the generality of the models obtained. Then we extract highly significant rules from these models and compared their meaning with real rules known from biological observations.

Mate choice uses the average information of the whole population of prey individuals and species extinction uses the average information of each prey species' individuals. After some feature selection based on machine learning techniques, we have applied a decision tree method on the selected features to extract some rules. By analyzing the obtained results, we were able to understand better the effect of these features on each process by extracting meaningful rules. The organization of this study is as follows:

Review of literature (2^d chapter) composed of three sections. The first section includes a study of mate choice mechanisms and a review of some papers on this topic using individual-based models. The second section contains a study of species extinction that looks at some papers applying individual-based model approaches. The last section reviews the characteristics of EcoSim briefly.

Design and Methodology (3^d chapter) includes four sections. The first and third sections review the models used in this study for mate choice and species extinction. The second section shows the features preparation done by calculating the average information about individuals for the mate choice process. The last section is the presentation of the features preparation for species extinction using the average information about individuals belonging to specific species.

Analysis of Results (4th chapter) encompasses two main sections. The first section consists of two experiments on mate choice: the study of the behaviours of the mate choice models, and the discrimination of models based on average features. The second section contains two experiments on species extinction including the prediction of

species extinction in the near futures based on average species features and categorized features.

      Conclusions and Recommendation (5th chapter) summarizes and discusses the results of the experimental studies of mate choice and species extinction. Moreover, it suggests some modification to improve the analysis in the future work.

CHAPTER II

REVIEW OF LITERATURE

The first section takes a look at some mechanisms of sexual selection that present the interactions between a choosy partner, in the form of the female preference, and a chosen partner, in the form of the male trait. As the individual in our simulation are asexual, we focus our review on mate choice mechanisms, which implies different effects of a mate preference for another mate trait on the mate choice from a direct or indirect point of view. The second section studies the different reasons among demographic, environment, and genetic factors that affect the survival or disappearance of a species. Finally, the last section takes a look at the EcoSim as the platform of this study.

2.1 Mate Choice

Reproduction is the natural process that is the outcome of mating occurring in the natural habitat of a given species in which new offspring is produced from its parents. The main grouping methods of reproduction are sexual and asexual [Alters 2000]. In asexual reproduction, an individual reproduces from a single individual whereas sexual reproduction requires the involvement of two individuals that can be of opposite sexes or hermaphroditic. In sexual reproduction, a new organism is created by combining the genetic material of two organisms.

Choosing an individual as a mate in the matting process is governed by sexual selection. "When the males and females of any animal have the same general habits of life, but differ in structure, color, or ornament, such differences have been mainly caused by sexual selection" (chap. 4, p. 89) [Darwin 1859]. Darwin's idea has opened a new

horizon to the concept of evolution and has had a highlighted role in the evolutionary biology. Darwin found that intricate ornaments and weapons in many animals seemed to have different roles in the animal survivorship and the sexual competition for mating or attracting the opposite sex. Various mechanisms of sexual selection have been observed in which contest and mate choice take noteworthy attentions in the biological studies [Andersson and Iwasa 1996]. "Mate choice may be operationally defined as any pattern of behaviour, shown by members of one sex, that leads to their being more likely to mate with certain member of the opposite sex than with others" [Halliday 1983]. One major evolutionary consequences of mate choice is that "variations in the behavior of members of one sex are correlated with variations in their mating success". Choosing a certain mate between some potential mates can have different kinds of immediate and long-term benefits. Choosing a mate uses different criterions of mate quality such as:

- Choice for high fecundity or fertility

- Choice for immediate gains and parental abilities such as feeding

- Choice for resource and for high mate status

- Choice for mate complementarily such as genetic complementarily, degree of relatedness, and complementarily in the reproductive behaviour

- Choice for the most effective courtship display [Halliday 1983].

Mate choice includes different methods such as random mating, assortative mating and disassortative mating. In random mating, all individuals can be potential partners regardless of any physical, genetic, or behavioural preference. And in assortative mating, the mates have more resemblance to each other with regard to some traits, unlike disassortative mating [Bos and Caligari 1995].

As mentioned, mate choice concentrates on the behavioural and morphological traits attracting mates during the mating process. Physical appearance as morphological traits has a great impact on sexual attractiveness, especially in the beginning of a courtship process, including visual perception, audition, and olfaction [Andersson and Iwasa 1996]. "If males vary in the vigour of their sexual display, then, assuming that the effect of such display is to increase female sexual motivation, female will be more likely to mate with the most vigorous males" [Halliday 1983].

The females have special preferences for choosing and attracting males according to their traits. In the interaction between female preferences and male ornaments what still remain controversial is why females prefer ornamented males and why males evolve conspicuous traits that are costly and likely reduce viability [Andersson and Simmons 2006]. Based on empirical studies, different mechanisms of evolution of mate choice based on pre and post-copulatory sexual selection have been presented (Table 1). Each mechanism tries to explain the relation between female preferences and male ornaments due to the specific point of view of a direct or an indirect evolution.

Fisher [1915] hypothesized two answers to the interactions of preferences and traits with the concept of indirect genetic interaction: (1) females with a strong trait preference will have sons with higher mating success, which spread their genes more, (2) the male trait indicates high heritable viability that can be inherited by the offspring. He stated that the exaggerated traits are the outcome of a positive feedback loop between female preference and male display [van Dijk et al. 2010].

Table 1: Some mechanisms of mate choice evolution [Andersson and Simmons 2006]

Mechanism	Explanation
Direct phenotypic effects	"If the ornament reflects the ability of the male to provide material advantages, such as a high-quality territory, nutrition, parental care or protection. Female choice might also evolve as a result of resistance to direct costs imposed by males."
Sensory bias	"Female preference favoring a male ornament can initially evolve under natural selection for other reasons, for instance in the context of foraging or predator avoidance. Males evolving traits that exploit this bias then become favored by mate choice."
Fisherian mechanisms	"If there are genetic components to variance in female preference and male trait, a female choosing a male with a large trait bears daughters and sons that can both carry alleles for a large trait, and for the preference for it. This genetic coupling might lead to self-reinforcing co-evolutions between trait and preference."
Indicator mechanisms	('good genes' or 'handicap-mechanisms') suggest that "attractive male traits reflect broad genetic quality. Inherent in such mechanisms is the maintenance of genetic variation. In addition, other advantageous genes and relative freedom from deleterious mutations might lead to high male condition and expression of sex traits. Female preference for such traits can provide genetic benefits to those of her offspring that inherit favorable alleles from their father."

In addition, a direct mechanism of mate preference shows that a strong mate preference can evolve if it improves the total fitness of the other mate through direct phenotypic benefits, such as avoidance of disease transmission, inbreeding, protection

from other coercive males, or through choice of a mate who gives better parental care or who is more fecund.

2.1.1 Papers Review

Although the sexual selection still has many open questions, various aspects of it have been studied in many papers. Here we focus our literature review on the few Individual-based modeling approaches that have been used to examine the effect of the morphological mate choice. This section reviews some of these works: genetic linkage between preferences and traits, the effect of exaggerated traits on the reproduction ratio, and a consequence of attractive mates on the longevity of offspring.

Lorch et al. [2003] examined the effect of condition-dependent (non-mating fitness and display) sexual selection on the rate of adaptation. They present an individual-based genetic model in which a diploid additive genetic system evolves based on covariance between traits. "This model includes four basic traits: (1) a resource acquisition trait that determines condition, (2) a trait that specifies the extent to which display trait expression depends on male condition, (3) a trait representing the condition-independent part of male display, and (4) the intensity of female mating preference". Each female produces one son and one daughter to make next generation. As a result, the authors mentioned that the variance in male display trait stays at a higher level by evolving the condition dependence. Moreover, costly male display decreases when sexual selection is excluded.

van Dijk et al. [2010] introduced an individual-based model of sexual selection and mate choice in a quantitative genetic context to study the genetic linkage between preference and trait. They present the multi-locus genetic model of male display and

female preference and a cross-over procedure that recombines the parental genomes once at each generation. In addition, they apply viability selection according to natural selection and the mate choice in their model is a mathematical equation based on male display. The authors conducted experiments by combining natural selection with sexual selection. Based on the results, they concluded that "strong mate selection causes strong genetic linkage." Moreover, "genetic linkage coincides with run-away Fisherian sexual selection." And finally they mentioned that "strong natural selection decreases the strength of the Fisher process."

In nature, females of some species have an attraction to multiple male ornaments. The main question here is to what extent each of the male ornaments indicates the quality, or reflects different aspects, to the female while paying also attention to the imposed cost. van Doorn and Weissing [2004] studied this phenomenon by a mathematical analysis of an individual-based model platform using two kinds of ornaments for individuals. In this model, the ornaments show different quality of mate on the basis of overlap or independent information of multiple ornaments. The effect of dependency between ornaments on the evolution of female preferences takes into account male quality and viability and female's choice cost. They stated that female preferences for multiple indicators of quality may evolve when the choice cost is low and also when the ornaments provide independent cues of an individual's genetic quality. In other words, the ornament diversity can be determined by the amount of independent information provided by ornaments.

Fawcett et al. [2006] presented two individual-based models for exploring the link between male attractiveness and offspring sex ratios based on female preference. The

first model applies the multi-locus genetic model, and the second model uses the quantitative genetic model based on a continuously variable male trait. In the first step, they simulated the evolution of a conspicuous male trait and female preference due to sexual selection in the absence of variation in offspring sex ratios. Afterward, they investigated the possibility to determine the offspring's sex by females' preference. On the basis of the simulations, the authors state that females with attractive partners should produce more sons than those with unattractive partners because these sons will inherit their father's attractiveness and these females enjoy high mating success.

In other work, Fawcett et al. [2011] developed an individual-based evolutionary model for investigating the dynamic interplay between sexual selection and sex-ratio adjustment. They utilized two types of males, which differ in their ornamentation and one type of female which can adjust offspring sex-ratios in relation to its partner's ornamentation. The mutation can alter ornaments and exchange ornamentation of offspring. They explained that females who mate with exaggerated ornamented males have more sons in comparison to those who mate to less-ornamented males. Consequently, the biased sex allocation weakens sexual selection and leads to a gradual decline in male ornamentation and female preference. Furthermore, they mentioned that species with little or no control over offspring's sex could include the exaggerated sexual displays.

In few species, females choose males with rare or novel traits instead of classical phenotypes. This rare preference could have some beneficial such as rare male phenotypes might be more immune to diseases or less attract predators that search common phenotypes. Kokko et al. [2007] applied an individual-based model to study this

problem. In this model, a male phenotype is determined by a gene having k alleles in a haploid structure. The females have a choice gene which either have a preferred allele or mate randomly. In addition, males have temporal viability selection with some randomly selected genotypes that suffer an extra mortality risk or the most common genotype that is affected by the extra mortality risk. Each female samples n males to evaluate phenotypic frequencies and mates with the rarest male in its sample. The fitness of a phenotype is dependent on its frequency relative to other phenotypes. The authors stated when females choose a costly mating preference for rare males this preference will spread or reduce until stopped by frequency-dependent selection. Moreover, if females are usually prevented from expressing their preference, that preference can evolve into much higher frequencies and converge into fixation. They mentioned that sensible preferences bring about viability benefits evolve more easily than those that do not and the rarity mating preferences happen less often than others mating preferences.

2.2 Species Extinction

The second biological phenomenon that will be reviewed here is species extinction. There are many factors in extinction that can be classified into the three main realms of Demography, Genetics and Environment [Griffen and Drake 2008]. Demographic factors are impacted by population growth, reproduction rate and individual lifespan, and include: population variability, initial population size and migration [Ovaskainen and Meerson 2010]. A larger population is less susceptible to extinction. In addition, a small population far below carrying capacity, which is able to increase quickly, can avoid extinction from demographic stochasticity. Furthermore, migration can increase population size when the negative density-dependence is weak.

Genetic factors correspond to a shortage of genetic variation, which can be caused by a decrease in fitness due to inbreeding depression [Reed et al. 2003]. Diminishing genetic variation may increase extinction risk by limiting the adaptation ability to stressful environments. Moreover, migration can reduce extinction risk by decreasing the genetic consequences of small population size.

Finally, factors such as habitat quality, habitat fragmentation and environmental stressors have a major role in extinction as environmental factors [Patten et al. 2007]. For instance, the temporal variation in food availability can increase the extinction risk by decreasing the long-term growth rate of the population. The effects of most of these extinction factors depend on interactions with other factors and conditions. It is therefore important to carefully study the effect of each factor. In real life, it is difficult to identify or compute an exact effect of these factors separately; it is even harder to do it altogether.

2.2.1 Papers Review

As mentioned in previous discussions, different factors could influence the extinction of species. The vast numbers of these factors and the huge area affected by them have been studied in different papers.

Walters et al. [2002] utilized an individual-based model to explore the effect of demographic and environmental stochasticity on vulnerability of *Red-cockaded Woodpecker* populations. This model contains breeding territories and non-breeding space that the birds must cross to find the breeding territories. Each territory can be created or lost based on some conditions. In addition, males in this model can compete for two resources of breeding territories and empty territories, and females only move into territories containing a male. They stated that they observed two major results: 1) the

distribution and the density of territories impact on population behaviour the same as population size, especially within a certain range of population sizes, and 2) the environmental stochasticity has small effects on vulnerability and consequently, small populations could be stable.

Hovel and Regan [2007] assessed how habitat fragmentation and loss influenced predator–prey interactions and cohort size for a group of settling blue crabs using a spatial individual-based model. Cells of this model are divided into three habitat types: sea-grass patch interior, sea-grass patch edge and matrix. In addition, it includes three types of prey, meso-predators and top-level predators. Top level predators only hunt meso-predators and prey can be eaten by meso-predators. Each organism has a different level of habitat preference. Three types of movement are created: random movement, predation avoidance movement, and directed hunting. They showed that sea-grass habitat fragmentation and loss strongly influenced prey cohort size. Nevertheless, factors such as predator hunting strategy, prey movements and patterns of settled prey can alter the effect of habitat fragmentation and loss on prey. In addition, the presence of top-level predators, the ability of predators and meso-predators to detect and respond to prey, and prey movements have great effects on prey cohort size. Finally, they showed that increasing prey mobility can reverse the patterns of prey survival versus sea-grass fragmentation.

Uchman [2000] studied the influence of individual variability on population dynamics using an individual-based model. This model implements differences in individual assimilation rates due to intra-specific competition and variability of initial weights. The number of individuals in the population oscillates due to the availability of resources. Moreover, individuals are characterized by their weight that varies with the

assimilation of resources and the loss due to respiration. They concluded that a population consisting of identical individuals becomes extinct when resources are entirely consumed or are not sufficient. In contrast, including individual variability, the model produces longer population lifespan due to the oscillated behaviours of population size which allow the resources can increase when the number of individuals decreases. Furthermore, they explained that the average extinction time is not a monotonic function of the degree of individual variability.

Natural hybridization is a considerable menace that can cause species extinction. Wolf et al. [2001] investigated extinction risk of two hybridizing annual plant species using an individual-based model simulating the life cycle and including some ecological parameters. This model contained the habitat, including K patches in which each patch is occupied by only one flowering adult, and a sympatric population containing members of two species. Due to the interactions among the selfing rate, pollen production, pollen competition, seed production, and seedling competitive ability of each class of plants, three species can emerge in this model: native species, invading species, and hybrid species. The authors stated that different parameters such as population size, variance in pollen-tube growth rate and pre-zygotic reproductive barriers impact on extinction risk. Furthermore, the competitive ability, initial frequency, and selfing rate of the native taxon strongly affect extinction. They explained that when there is no habitat differentiation, one of the parental species or the hybrids replace the other two taxa. Finally, they claimed that hybridization is a significant genetic threat to both rare and abundant species, also it can lead to an imminent extinction.

2.3 EcoSim

Gras et al. [2009] have presented EcoSim, an IBM including a behavioural model based on Fuzzy Cognitive Maps (FCM) [Kosko 1986]. EcoSim models biological concepts and processes such as genomes, evolution, speciation and interbreeding. Several studies have been done using EcoSim. Devaurs et al. [2010] have shown that the behaviour of this model is realistic by comparing the species abundance patterns observed in the simulation with real communities of species. Furthermore, the chaotic behaviour [Golestani and Gras 2010] and multi-fractal property [Golestani and Gras 2011] of the system observed in real ecosystems have been proven. Golestani et al. [2012] have also measured the effect of geographic barrier on the speciation in EcoSim. In [Majdabadi Farahani et al.], the diffusion of a disease in a population of prey and different healing policies have been studied in EcoSim.

In this model, two organism types, prey and predator, 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 some amount of grass, which is the primary producer, and meat. The availability of grass is calculated based on a spatial diffusion model that shaped the dynamic environment. Predators live on prey and prey live on grass. The predators act as a pressure factor on prey (environmental stress).

To observe the evolution of an individual behaviour and ultimately ecosystems over thousands of generations, several conditions need to be fulfilled: 1) every individual should possess genomic information; 2) this genetic material should affect the individual behavior and consequently its fitness; 3) the inheritance of the genetic material has to be done with the possibility of modification; 4) a sufficiently high number of individuals

should coexist at any time step and their behavioural model should allow for complex interactions and organizations to emerge; 5) a model for species identification, based on a measure of genomic similarity, has to be defined; 6) a large number of time steps need to be performed. These complex conditions pose computational challenges and require the use of a model which allies the compactness and easiness of computation with a high potential of complex representation.
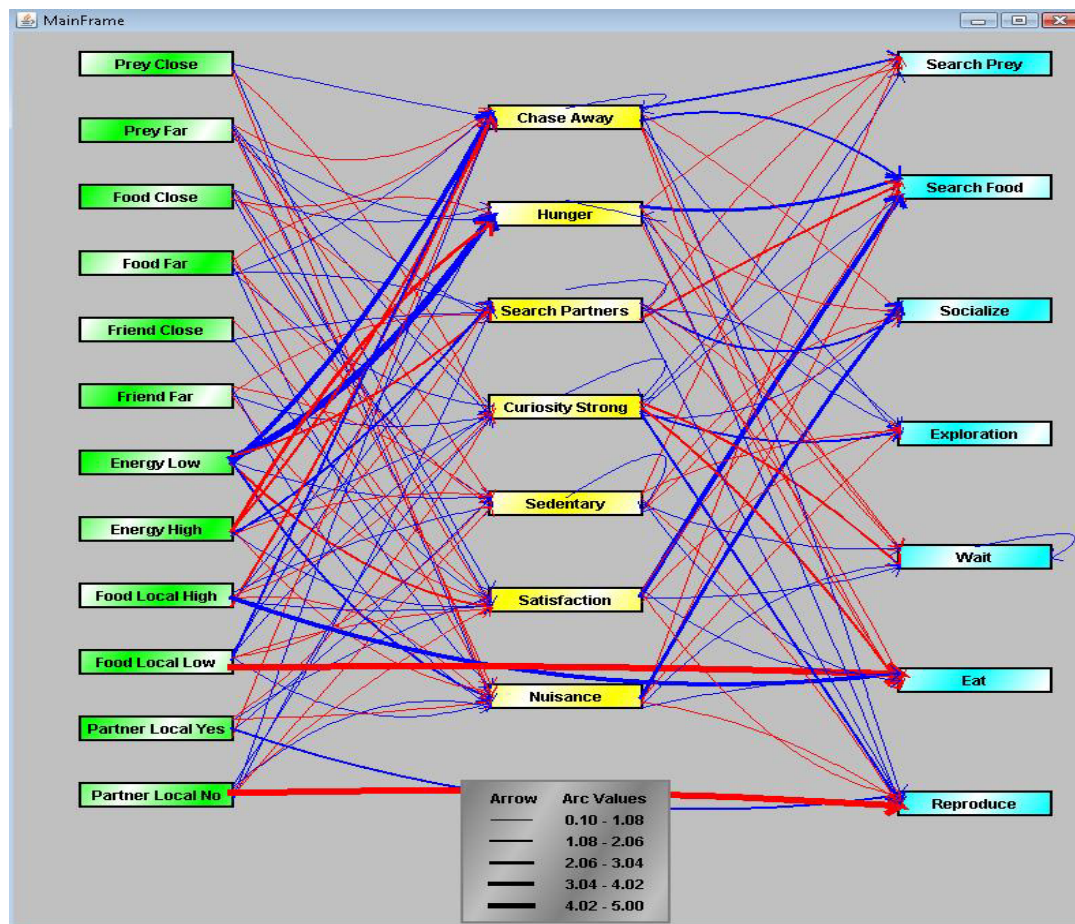
Figure 1: A sample of Predator's FCM including concepts and edges. The width of each edge shows the influence value of that edge. Color of an edge shows inhibitory (red) or excitatory (blue) effects [http://sites.google.com/site/ecosimgroup]

Every individual of EcoSim 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 and edges representing the influence of concepts on each other (Figure 1). When a new offspring is created, it is given a genome which is a combination of the genomes of its parents with some possible mutations.

In EcoSim, the FCM is not only the base for describing and computing the agent behaviours, but also the platform for modeling the evolutionary mechanism and the speciation events. Formally, a FCM is a graph which contains a set of nodes C, each node C_i being a concept, and a set of edges I, each edge I_{ij} representing the influence of the concept C_i on the concept C_j. A positive weight associated with the edge I_{ij} corresponds to an excitation of the concept C_j from the concept C_i, whereas a negative weight is related to an inhibition (a null value meaning that there is no influence of C_i on C_j). An activation level a_i (explained later) is associated to each concept. A FCM allows computing the new activation levels of the concepts of an agent, based on its perception and on the current activation levels of its concepts using normalized matrix product (look at Figure 2 as an example).
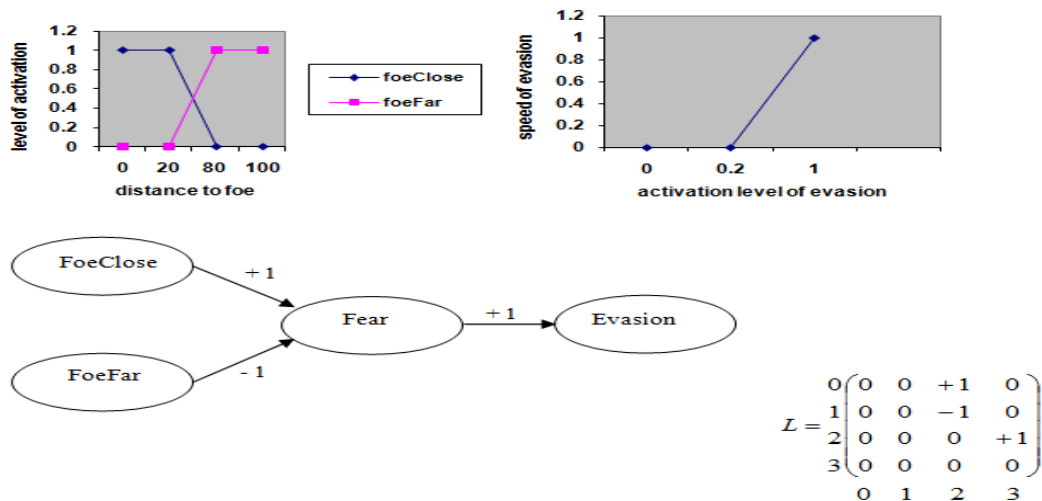
Figure 2: A FCM for detection of foe (predator) and decision to evade with its corresponding matrix L (0 for 'Foe close', 1 for 'Foe far', 2 for 'Fear' and 3 for 'Evasion') and the fuzzification and defuzzification functions [Gras et al. 2009]

There are three kinds of concepts: sensitive (such as distance to food and to sexual partner), internal (such as fear, hunger and satisfaction) and motor (such as escape, eat and reproduce). In addition, each concept has an activation level which depends on the current perceptions and its past internal state. The current activation level of a concept is computed based on the FCM and is used to choose the next action of the agent. The environmental fuzzified information of an agent is used to compute the activation level of a sensitive concept. Activation level of an internal concept is influenced by the sensitive concepts of the agent. And lastly, the action of agent is selected based on the defuzzified activation levels of the motor concepts that are affected by the sensitive and internal concepts. EcoSim iterates continuously, and each iteration, called time step, consists of the computation of the activation level of the concepts, the choice and application of an action for every individual. A time step also includes the update of the world: emergence and extinction of species and growth and diffusion of grass.

Each individual has an energy level that is affected by its actions. All actions decrease the energy level except the eating action. The energy is provided by the primary or secondary resources found in the world. For example, prey individuals gain 250 units of energy by eating one unit of grass and predators gain 500 units of energy by eating one prey. At each time step, each agent spends energy depending on its action and on the complexity of its behaviour model (number of existing edges in its FCM). On average, a movement action such as escape and exploration requires 50 units of energy, a reproduction action uses 110 units of energy and the choice of not action results in an small expenditure of 18 units of energy.

One of the actions performed by the individuals is reproduction. Several factors play roles in reproduction. For reproduction to be successful, the two parents need to be in the same cell, have enough energy, chose the reproduction action and be genetically similar. The organisms cannot determine their genetic similarity with their potential partner. They try to mate and if the partner is too dissimilar, the reproduction fails. The result of the reproduction action is a unique offspring with a genome which represents a combination of the parental genomes. The newborn receives an initial amount of energy equivalent to the energy that the two parents spend in reproduction. Moreover, all individuals can move in this model. The movement can be due to exploration, foraging or socialization. The movement of individuals can make different populations with various species. Figure 3 shows the spatial distribution of individuals in one specific time step of simulation's world.
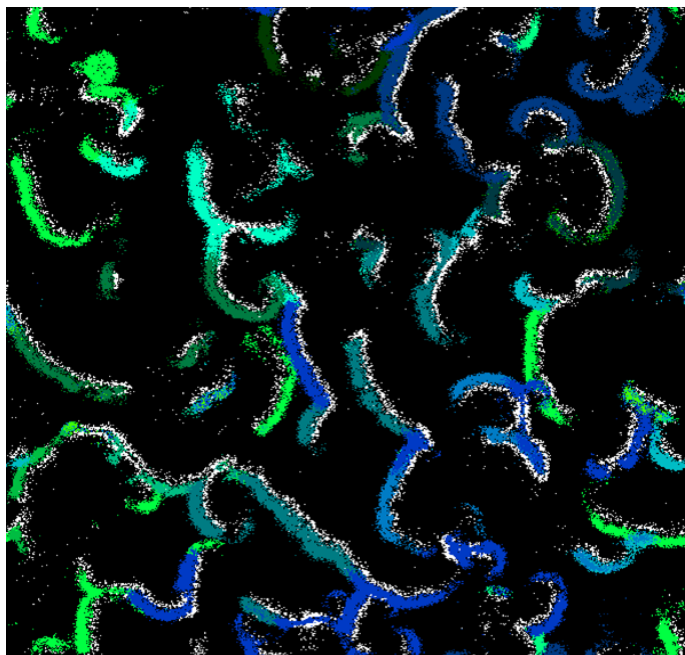
Figure 3: The status of the world in one specific time step, white color represents predator species and the other colors show different prey species

This model takes advantage of a speciation mechanism. A species is represented as a set of individuals having a similarity genetic characteristics [Mallet 1995]. The speciation mechanism implemented in EcoSim is based on the gradual divergence of populations which contain individuals that are more and more genetically different. This gradual divergence can lead to situations where some conspecific individuals cannot interbreed. To reflect the incipient process of speciation, a splitting of the species in two sister species is then performed. The splitting mechanism produces two clusters of individuals with high intra-cluster similarity and strong inter-cluster dissimilarity. 2-means clustering technique is implemented to allow for (1) the splitting of an existing species *S* into *S1* and *S2*, and (2) the clustering of individuals that initially belonged to *S* into one of either *S1* or *S2* [Aspinall and Gras 2010]. The speciation method begins by finding the individual in a species *S* with the greatest distance from the 'species genome' called the species center. If this distance is greater than a predefined threshold for speciation (which is two time greater than the threshold for reproduction), the 2-means clustering is performed. Otherwise, species *S* remains unchanged. If clustering is to be performed, two new species are created – one centered on a random individual, denoted *Ir*, and another centered on the individual which is the most genetically different from *Ir*, denoted *If*. Subsequently, all remaining individuals in *S* are added to one of the two new sister species – whichever species the individual is more genetically similar. After recalculating the new centers for the two new species, the process of clustering is repeated for convergence. After the 2-means clustering is completed, two new sister species (*S1* and *S2*) emerge. A single splitting events can only produce 2 sister species at a given time step. However, if one or both of the two resulting sister species are still

genetically heterogeneous after the splitting, other splitting events can occur on these new species at the next time step resulting in a final splitting of the initial species in more than two sister species in a very short period of time.

It is worth to notice that the speciation mechanism is only a labelling process. The information about species membership is not use for any purpose during the simulation but only for post-processing analysis of the results. It can be viewed as an online hierarchical clustering process. Since clustering is a difficult and time consuming task, it is impossible to apply it to the whole population of individuals' genomes (at some time steps there are more than 500,000 existing individuals) at every time step. We have therefore chosen a heuristic hierarchical approach in which the clustering is done through the whole process, in a given time step only a small subset of the whole population being clustered by our species splitting mechanism. However, we have observed that the quality of the clusters obtained by this heuristic approach is better, in term of both inter and intra cluster similarity, than the one obtained by applying a global clustering algorithm to the whole population. That can be explained by the fact that the loss in quality due to the hierarchical heuristic approach is less than the one due the high complexity of solving the whole clustering problem.

It has been observed that after an initialization period (between 500 and 1000 time steps), the individuals that are genetically similar (member of the same species) are also geographically close to each other in the world. Moreover, when a speciation event occurs, the two genomic clusters formed lead to two spatially separated populations. After splitting, the two sister species are still very similar leading to high number of hybridization events (Figure 4).
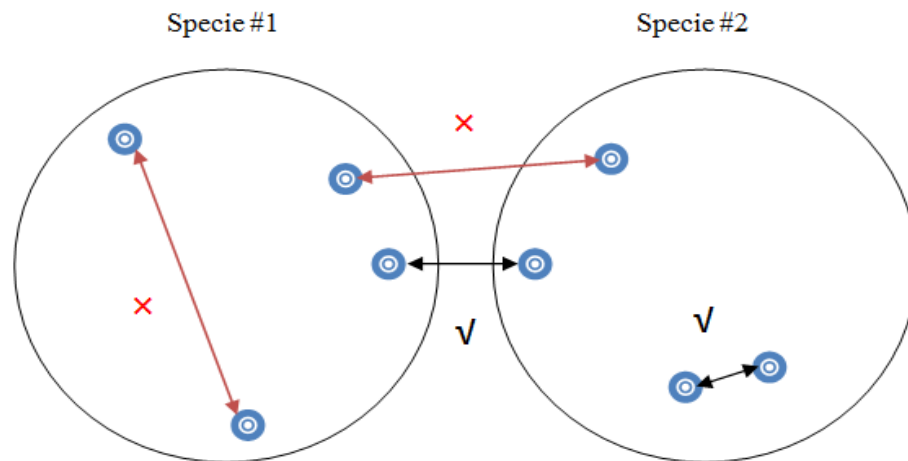
Figure 4: Reproduction between individuals in different situations. The red arrows indicate situations in which the genetic distance between individuals is greater than the threshold for reproduction and reproduction is stopped. Black arrows connect individuals with genetic distance smaller than the threshold for reproduction, indicating that these individuals can interbreed even if they belong to different species [Golestani et al. 2012]

As the two sister species continue to diverge, two completely isolated species emerge. The genetic distance between the two new species increases with time and rapidly becomes larger than the within cluster distance generating strongly isolated clusters in the genomic space. In this model, a new individual is a member of the species of one of its parents whom the most similar. Normally, the two parents are from the same species unless a hybridization event occurs.

CHAPTER III

DESIGN AND METHODOLOGY

3.1 Mate Choice Models

In EcoSim, the reproduction process is implemented according to the algorithm of Figure 5 in which an individual does not have specific gender. When an individual reaches to the proper age of mating with an enough amount of energy, it can mate with other individuals. In this algorithm, A and A′ present the first (choosy) and the second (chosen) partners who are trying to mate each other. D(.) shows the genetic distance between two individuals that should be less than T in which T has a predefined value. A choosy partner selects randomly a chosen partner for reproduction process based on meeting three conditions, including: 1) similar genetic distance between partners, 2) enough energy of chosen partner for this process and bring the new offspring 3) willing of chosen partner to mate right away with choosy partner.

```
if A.energyLevel > 0.125 · maxEnergyPrey then

    for all A′ of the same type in the same cell

        if A′.energyLevel > 0.125 × maxEnergyPrey and

            D(A,A′) < T and
            A′ has not acted at this time step yet and
            A″'s choice of action is also breeding

    then

            interbreeding(A,A′)
            A.sexualNeeds ← 0
            A′.sexualNeeds ← 0
            if A′ satisfies all the criteria, the loop is canceled
```

Figure 5: The algorithm of reproduction in EcoSim [Gras et al. 2009]

26

If all conditions meet, the reproduction process happens and consequently, a new offspring will be born. Otherwise the reproduction is failed. The random candidate for mating is selected from the five nearest individuals of choosy partner who are presented in the predefined distance. This list of the nearest individuals is called the local list.

Although in EcoSim, the mate choice occurs due to a random selection, in general, a sexual reproduction is a result of an interpersonal attraction. For studying the effect of a morphological attraction, as mate choice, on the sexual selection and on the population in the long run, the amount of energy as an indicator of a physical attraction is chosen. Energy represents the potential ability of an individual for mating and creating an offspring. Moreover, it shows the capacity of an individual to forage and consume food. As a consequence, energy can be assumed as a physical indicator of the strength or the ability to survive of an individual.

For evaluating energy as a morphological trait in mate choice, two models are presented. In the first model, called "*Max Energy*", each choosy individual prefers to mate with an individual candidate who has the highest value of energy in the local list. For this purpose, a priority list of the local list's individuals is generated based on energy. In this list, the first candidate has the highest amount of energy, and the last candidate has the lowest one. The choosy partner selects a chosen partner from this list. The first individual, who meets the three mating condition mentioned before, will be the chosen partner for reproduction process. This model demonstrates a mating with an exaggerated partner. In the second model, called "*Similar Energy*", each choosy individual selects a partner based on positive assortative mating. Each individual makes a priority list of local partners based on maximum similarity (minimum distance between the values of energy)

and mates with the more similar candidate who fulfills all other mating conditions. This model is used to implement the idea of genetic compatibility between mates. Clutton-Brock [2007] stated that strong "selection on females to maximize the growth and survival of their offspring may generate selection pressures for mating with genetically compatible partners".

To summarize, the original model, called "*Random*", doesn't consider the morphology of mate, the *Similar Energy* model selects a similar mate, and the *Max Energy* model chooses the most exaggerated mate based on the amount of energy as a morphological trait.

3.1.1 Feature Preparation

For investigating the mate choice models on the long-time evolution, 34 features are prepared by computing their average on all existent individuals at each time step without paying attention to the species identity. These features include information about demography, death, action, perception, and genomes of individuals and are described in Table 2 and Table 3. Demographic features include the characteristics of the population such as Species-Ratio, the ratio of all the current species to the entire population, Death-Ratio, the ratio of the number of dead individuals to the total population, Interbreeding-Ratio which is the ratio of new-born individuals with parents from two different species to the whole new-borns.

Furthermore, some action features, which show the percentage of the individuals choosing one action at a given time step like escape, forage and eat, are calculated. In addition to actions, some perception features are chosen, which depict the perception of an individual from its environment such as Distance-Food, Distance-Friends and

Distance-Predators. Only the Food-Availability feature is not directly measured from the individuals' information and is calculated by counting the total amount of food in the world to represent the availability of food in the entire world as an environmental factor. This feature can be affected by the population size and the individuals' dispersion.

Some few features need a little more explanations. The Genetic-Diversity measures how much diversity exists in the gene pool of the world. The entropy is commonly used as an index of diversity in ecology and increasingly used in genetics [Sherwin 2010]. In our case, the genetic diversity is the entropy of the genomes of all individuals, and represents the level of similarity between all the genomes of all individuals [Khater et al. 2011]. Genetic-Diversity, that shows the variety of alleles at all loci of the whole population, is calculated by the equation (.1.):

$$GD = -\sum_{i=1}^{loci} \sum_{j=1}^{allleles} p_{ij} \ln(p_{ij})$$
(.1.)

where p_{ij} expresses the frequency of alleles j at locus i in the entire population calculated by a discretization technique. The minimum and maximum value of each locus is calculated, and this range is discretized into several bins by fix step. MaxGenetic-Diversity is the number of bins having a frequency greater than zero. The next measure, Genetic-Complexity is an evaluation of the behavioural model complexity. As this complexity rely on the number of existing edges in the FCM, and the edges values are coded in the genome, Genetic-Complexity is computed by counting the number of active loci in the genome, that is the ones having a value greater than a small value ($ActiveLocus_j > \varepsilon$). For each individual i, the greater this measure is the more complicated the individual is considered. It is computed by the equation (.2.):

$$GC = \frac{\sum\limits_{i=1}^{indiv} \sum\limits_{j=1}^{nbLocus} ActiveLocus_{ij}}{Population\ Size} \qquad \textbf{(.2.)}$$

The Parental-Investment measures the energy cost that each parent has to pay for the breeding process and for the offspring care. This energy is partially transmitted to the offspring and defines its initial energy level. The last feature is Distance-Evolution which is the average genetic difference between the current population and the initial genome. It represents the accumulation of genetic variations in the population since the beginning of a run. It is computed by the equation (.3.):

$$DE = \frac{\sum\limits_{i=1}^{indiv} D(CR, C_i)}{Population\ Size} \qquad \textbf{(.3.)}$$

where CR and C_i refer to the reference genes and the individual genes respectively and D(.) is the Euclidian distance function.

Table 2: The features of the mate choice based on information of all individuals in the entire population at each time step

#	Feature	Definition
01	Population-Size	The total number of individuals
02	Species-Ratio	The ratio of total species to the whole population
03	Speciation-Ratio	The ratio of new species to the whole species
04	Extinction-Ratio	The ratio of extinct species to the whole species
05	Food-Availability	The amount of grass in the entire world (all cells)
06	Speed	The average speed
07	Energy	The average energy
08	Age	The average age
09	Death-Ratio	The total number of deaths to the whole population
10	Death-Age	The average age at the time of death
11	DeathAge-Ratio	The ratio of deaths due to oldness to the whole deaths

Table 3: The features of the mate choice based on information of all individuals in the entire

population at each time step (continue)

#	Feature	Definition
12	DeathEnergy-Ratio	The ratio of deaths due to lack of energy to the whole deaths
13	Killed-Ratio	The ratio of killed individuals by predators to the whole deaths
14	Escape-Ratio	The ratio of escape from predators to the whole population (action)
15	Foraging-Ratio	The ratio of searching for food to the whole population (action)
16	Socialize-Ratio	The ratio of socialization among preys to the whole population (action)
17	Explore-Ratio	The ratio of world exploration to the whole population (action)
18	Eat-Ratio	The ratio of food consumption to the whole population (action)
19	Sedentary-Ratio	The ratio of immobile individuals to the whole population (action)
20	Reproduce-Ratio	The ratio of reproduction to the whole population (action)
21	ReproduceFail-Ratio	The ratio of failed reproduction to the whole population
22	Genetic-Diversity	The diversity of alleles for all loci based on the entropy calculation (refer to the equation (.1.))
23	MaxGenetic-Diversity	The max diversity of alleles for all loci
24	Genetic-Complexity	The number of loci having active alleles (refer to the equation (.2))
25	Parental-Investment	The ratio of energy which transfers to a new individual at the birth time and decreases the parents' energy as a cost of the offspring care
26	Distance-Evolution	The genetic difference between the reference genes (origin) and the current genes (refer to the equation (.3.))
27	Distance-Mating	The genetic difference between mates
28	Distance-Predators	The perception of the predators' distance
29	Distance-Food	The perception of the food's distance
30	Distance-Friends	The perception of the friends' distance
31	Inner-Energy	The perception of the amount of individual's energy
32	Quantity-LocalFood	The perception of the quantity of food in the vicinity
33	Quantity-LocalPartner	The perception of the quantity of partners in the vicinity
34	Compactness	The average number of individuals in a specific area (a cell) also called the population density

3.2 Species Extinction Model

According to the preceding discussion about EcoSim, the evolutionary mechanisms include interbreeding, mutation, and speciation. These mechanisms produce the evolution process over a long time in which a new species can appear or a species can disappear. Each species generates a massive raw data in its lifespan based on information of its individuals. Manipulation and analysis of these data needs suitable information retrieval tools. A dedicated class was added to the EcoSim to gather the information about all species which is used for species extinction. Furthermore, this class saves the information for each species separately per time step for more investigations.

3.2.1 Feature Preparation of Extinction

The study of species extinction includes 48 features, which are used in the prediction of the future extinction. These features are extracted from information about each species separately by taking the average values of all individuals' properties belonging to the specific species at every time step. Twenty-eight of these features are common with the features in the mate choice part, including: Speed, Energy, Age, Death-Ratio, Death-Age, DeathAge-Ratio, DeathEnergy-Ratio, Killed-Ratio, Escape-Ratio, Foraging-Ratio, Socialize-Ratio, Explore-Ratio, Eat-Ratio, Sedentary-Ratio, Reproduce-Ratio, ReproduceFail-Ratio, Genetic-Diversity, Parental-Investment, Genetic-Complexity, Distance-Evolution, Distance-Mating, Distance-Predators, Distance-Food, Distance-Friends, Inner-Energy, Quantity-LocalFood, Quantity-LocalPartner and Compactness. The descriptions of these 28 features are given in Table 2 and Table 3, although their calculations differ slightly as they are computed per species for the species extinction problem. The rest of features are described in Table 4. Twelve features of this

table, from number 09 to 20, represent spatial metrics characterizing the complex spatial dynamics of the world. These measures are calculated according to the position and the dispersion of the individuals belonging to each species separately.
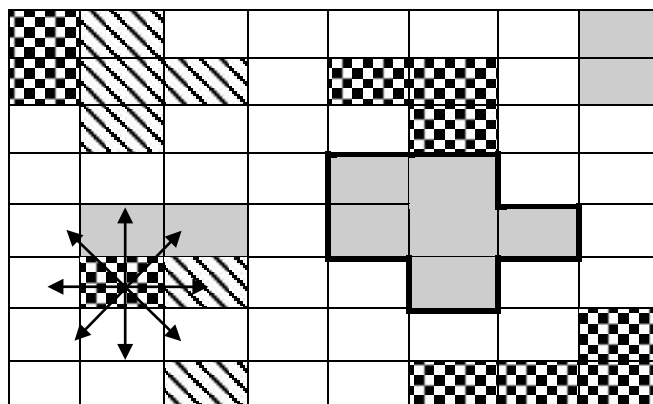
Figure 6: A sample part of the world. There are 3 patch types which show 3 species. The cells surrounded by thick solid lines shape illustrate one patch and also show the surface area of that patch. The cells' neighbours are showed by the arrows.

An ecological mosaic is a raster-based grid of categorical values to depict the distribution of a specific ecological measure [Parrott et al. 2008]. Most of these measures have been designed to describe the composition and configuration of patches, contiguous cells containing the same category value, in a landscape based on a grid (see Figure 6). In EcoSim, the individuals of each species are located in a couple of cells of the world and the adjacent cells containing individuals of the same species are considered as a patch. Thereby each species can be consisted of a number of patches called 'patch type'. The defined metrics in [Li and Reynolds 1993] and [Parrott et al. 2008] are applied to a two-dimensional world and adapted to EcoSim concepts. For example, Patch-Volume is the number of cells occupied by a species and Surface-Area is the number of cells in a patch, which are not shared by the adjacent cells of the same patch type. The Moore neighbourhood, comprising the eight cells surrounding the central cell, is used to define

the adjacency. Another measure is Contagion which is calculated by (.4.) and (.5.) to measure a dispersion of a patch type. A lower value of contagion shows many small patches and a higher value indicates few large ones.

$$RC = 1 - \frac{EE}{EE_{max}}$$

(.4.)

$$EE = -\sum_{i=1}^{pt}\sum_{j=1}^{pt} p_{ij}\ln(p_{ij})$$

(.5.)

$$p_{ij} = \frac{n_{ij}}{n_i} \qquad , n_i = \sum_{i=1}^{pt} n_{ij}$$

where, RC is contagion, $EE_{max} = pt \times \ln(pt)$ and pt is the number of patch types (in EcoSim the number of species). The n_{ij} is the number of adjacencies between cells of patch type j and cells of patch type i.

Spatial-Complexity (.6.) is a measure to describe how one patch type occupies a space, and it is calculated by considering the contents of successively square windows with size i ϵ [0 - n^2] (n is used to compute the maximum size of a window used to covers the space and its value is 5). SC value is lower for uniform or ordered patch shapes and has higher value for complex shapes.

$$SC = \frac{-\sum_{i=0}^{n^2} p_i\ln(p_i)}{\ln(n^2+1)} \qquad 0 < SC < 1$$

(.6.)

where p_i is the frequencies of the different possible occupation levels. The occupation level i shows how many of windows with size i contain i individuals for each possible placement of them.

Fractal-Dimension is used to quantitatively describe how one object occupies its volume [Foroutan-pour et al. 1999]. The count boxing method is applied to calculate the fractional dimension for each species. For this purpose, the world is covered with a 2-dimension filling box and the number of boxes (N(r)) required to cover the part of the whole world containing at least one cell related to the given species is recorded. This procedure is repeated with different size r (=2, 5, 10, 20 and 25) of the box and a graph of ln(N(r)) versus ln(1/r) is generated. Afterward, the fractal dimension is obtained by calculating the slope of the best-fit line through the points.

The Spatial-Diversity measures the distribution of individuals of a species based on the locations of all its individuals and is computed in two steps. The first step is the computation of the spatial centre of the species. EcoSim's world is a torus in which the opposite borders of the world (grid) are adjacent. Therefore, the circular statistics [Jammalamadaka and Sengupta 2001] [Md Ibne and Gras 2010] are applied to compute the centre of the spatial species distribution. The second step is the calculation of the average distance of all individuals to the center and the spatial standard deviation. More formally, if C is the center of one species, then:

$$TotalDisance = \sum_{i=1}^{n} dist(C, indv_i) \qquad (.7.)$$

$$SpatialSTD = \sqrt{\sum_i \frac{(dist(C, indv_i))^2}{n-1}}$$

The last measure is History-Movement that is calculated by considering the total movement of the center species in k (=100) time steps for each species. This value shows the trend of movement in long time.

Table 4: The features of species extinction based on the individuals' information for each species at each time step

#	Feature	Definition
01	Individual-Ratio	The ratio of species population to the whole population
02	Species-Number	The total number of species
03	Death-Energy	The average energy at the time of death
04	Parent1-MatingAge	The age of choosy partner at the time of mating
05	Parent1-MatingEnergy	The energy of choosy partner at the time of mating
06	Parent2-MatingAge	The age of chosen partner at the time of mating
07	Parent2-MatingEnergy	The energy of chosen partner at the time of mating
08	Interbreeding-Ratio	The ratio of births due to interbreeding to the total births
09	Spatial-Diversity	The dispersal place of a species' members based on the species center
10	Volume-Ratio	The ratio of the volume of a species patch (the number of cells that a species occupies) to the species population
11	Surface-Area	The number of outer cells in the species patch
12	Shape-Complexity	The ratio of the volume of a species patch to the volume of bounding box (smallest box that covers the volume)
13	Spatial-Complexity	The ratio of the volume of a species patch to the volume box
14	Space-Ratio	The measure shows how the species patches occupies the world
15	Patch-Number	The number of patches of one species
16	Patch-Size	The average size patches of one species
17	MultiSpecies-Cell	The number of cells that are shared between more than one species
18	Contagion	The measure shows the tradeoff between the size and the number of patches in the world
19	Fractal-Dimension	The measure describes how one species occupies its volume
20	History-Movement	The amount of a species movement during the specific time steps

CHAPTER IV

ANALYSIS OF THE RESULTS

This chapter presents the analysis of EcoSim's results on the two topics of mate choice and species extinction. For each topic, two experiments using a prediction process, based on the prepared features presented in the previous chapter, are conducted to distinguish between different cases of each topic. The machine learning techniques such as a Decision-Tree classifier (C4.5 / J48) and a Bayesian-Network (with K2: a hill climbing search algorithm), are applied to the results of prediction to evaluate the accuracy of data sets and to extract the prediction rules using WEKA [V3.6.4]. Moreover, five measures, including True-Positive Rate (TP Rate: sensitivity), False-Positive Rate (FP Rate), area under ROC curve (TP Rate versus the FP Rate) and F-measure (the weighted harmonic mean of precision and recall) as well as the overall accuracy, based on the confusion matrix, are used to evaluate the performance of the prediction process. The later discussions are based on the overall accuracy and the other measures bring to show the quality of the experiments.

4.1 Mate Choice

For assessing the sexual selection models, two experiments are conducted. In the first experiment, the major characteristics of three mating models are compared together by analysing the variations of features in time. This experiment focuses on the long-term trend of similarities and differences between models. The second experiment investigates the prediction of each model based on the proposed features. This experiment aims to find the general rules that can predict and explain the three different mating policies.

Both experiments use information from 15,000 time steps coming from time steps 10,000 to 25,000 of 3 independent runs for each model. We do not consider the first 10,000 time steps to give time to the system to stabilize and for the evolutionary process to affect the system.

4.1.1 Model Comparison

The chosen (second) partner for reproduction in the *Random* mating model is selected randomly from the local (same cell) partner list. This model implies a mating without any barrier that just needs to satisfy three main primary conditions. These conditions are: both partners having sufficient energy, the two partners being genetically similar enough, and both of them having a willingness to mate together with choosing the reproduction action. In the *Similar Energy* mating model, in addition to the three previous conditions, an individual prefers to mate with a candidate from the local partner list with a trait similar to its trait that presents the genetic compatibility between mates. Finally, in the *Max Energy* mating model, the preferred mate is a partner with an exaggerated trait. As mentioned in the previous chapter, the energy is used as an indicator of the morphological trait for the mate choice. For the *Max Energy* model, an individual with the higher amount of energy is stronger than the other individuals in the local partner list. It supposed that this individual has a higher fitness due to its ability of gaining and preserving that energy level.

The results are provided by taking the average of three separate runs for each model and illustrating the comparison between *Random*, *Similar Energy* and *Max Energy* mating models after the first 10,000 time steps. In the following pictures, weighted smoothing technique is applied to improve the quality of them. Two points should be

considered in analysing the results: 1) there is no gender implemented in this simulation and 2) the energy as a morphological trait has no direct effect on the viability of individuals but can be considered as an indicator of their potential fitness. The first characteristic of this comparison is the population size. For this measure, the *Max Energy* model is almost greater than the *Random* and the *Similar Energy* models (Figure 7).
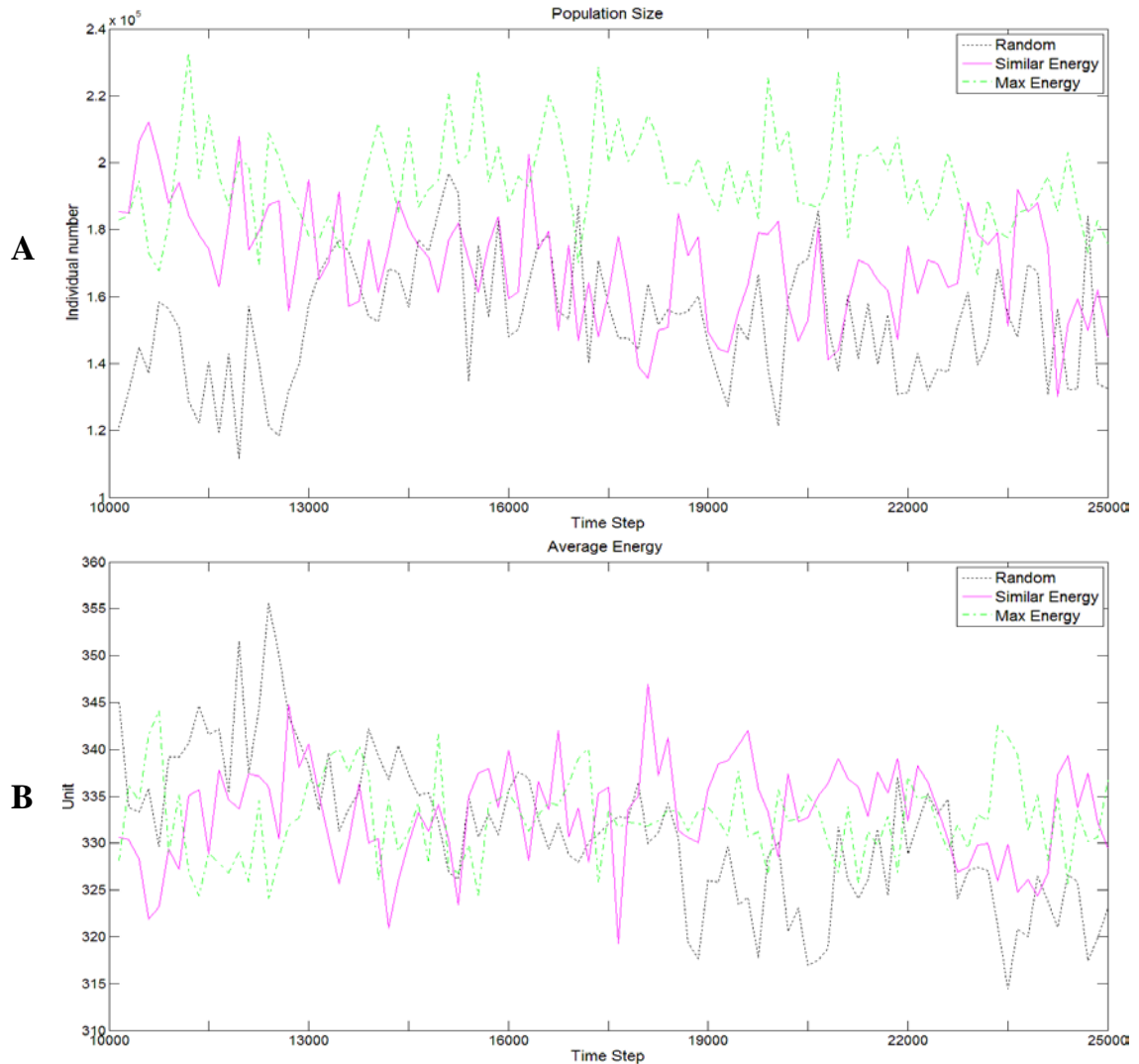
Figure 7: A) the population size, B) the average of energy for the time steps 10000 to 25000

As seen in Figure 7, the average energy of all individuals is approximately the same in all models. In EcoSim, doing each action consumes energy depending on the speed and Genetic-Complexity of an individual. Moreover, the reproduction process

takes an extra energy that decreases the parents' energy in the birth process by Parental-Investment as a cost to pay for mating and offspring care. On the contrary, only the action of eating can increase the level of energy of an individual. Consequently, the similarity between the energy trends of the mating models implies that the ratio of the eating action, with the amount of Parental-Investment, the speed and the amount of Genetic-Complexity are in balance with energy, and the rise of the one component compensates the fall of the other components. According to the results, see Appendix (A) Figure 30, the average speed all models are approximately in same range but the ratio of food consumption for *Max Energy* model is lower, then the energy difference returns to the Parental-Investment and the Genetic-Complexity that both are genetic concepts, and also food consumption.

Different figures show that the major discrimination is between the *Max Energy* and the *Random* models, and the *Similar Energy* model has an intermediate behaviour. Therefore, the comparisons will concentrate on the *Random* and the *Max Energy* mating models. The Parental-Investment of the *Max Energy* model has higher values in comparison to the other models and, its Genetic-Complexity has lower values (Figure 8). These two features explain a trade-off that have an effect on the energy level and explain the similar trend of the average energy observed for the different models. The higher amount of Parental-Investment means a higher mating cost. This cost, for *Max Energy* model in which individuals prefer a partner with the exaggerated trait, is higher than in the *Random* model in which individuals have not any mating preference.
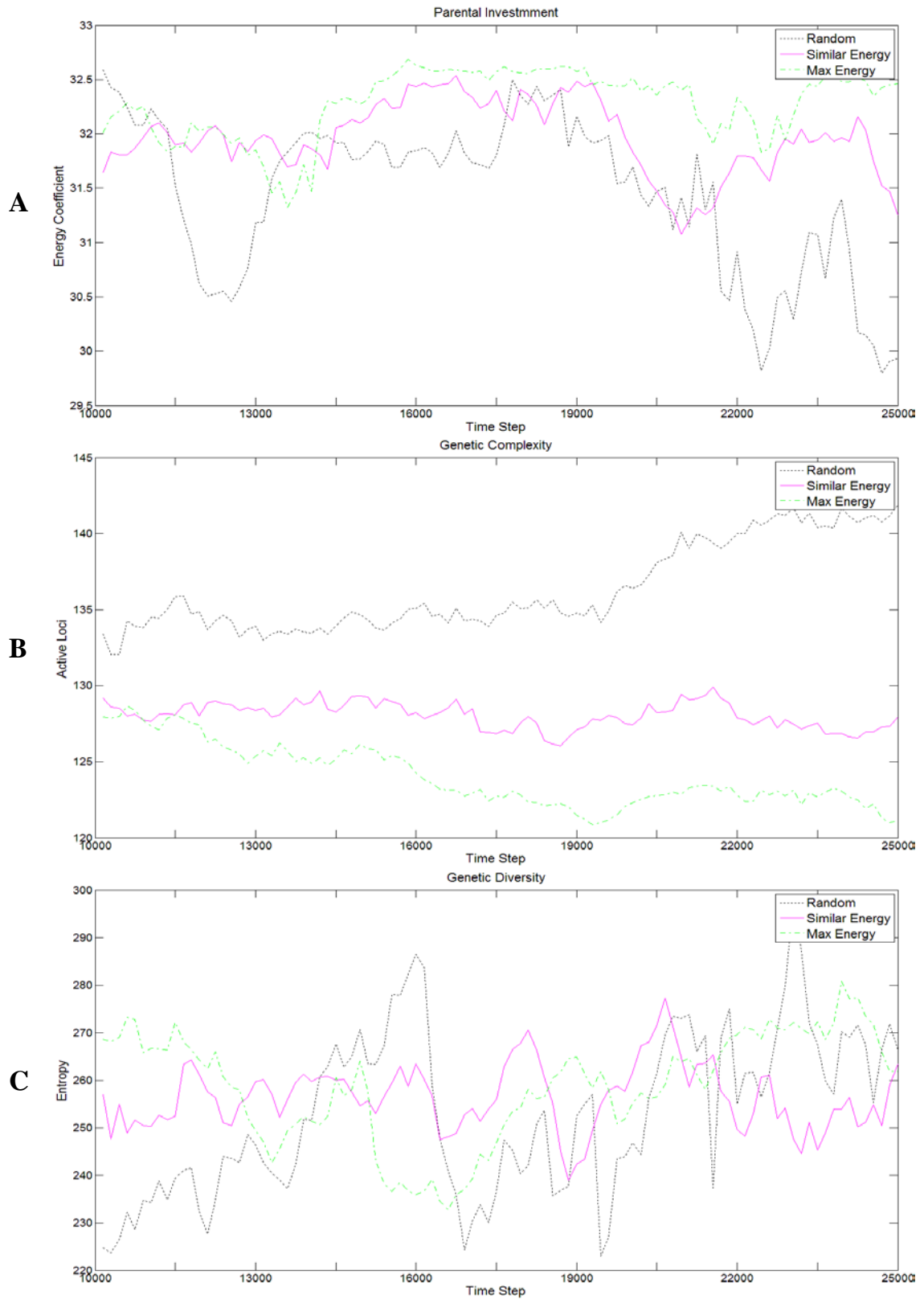
Figure 8: A) the average of Parental-Investment, B) the average of Genetic-Complexity C) the

average of Genetic-Diversity for the time steps 10000 to 25000

On the other hand, the Genetic-Complexity, showing the average summation of the active loci of individuals, in the *Max Energy* model is lower than the one of the *Random* model. That means that the evolved behavioural models (FCM) in the *Max Energy* model are simpler than those of the *Random* model. Nonetheless, the Genetic-Diversity of the three models is very similar. Based on these observations it can be inferred that the individuals of *Max Energy* model have a simpler behavioural model and more parental costs, although have the same genetic diversity as the other models. In addition to these observations, the Species-Ratio is similar for all models (see Appendix (A) Figure 30); it means that there is almost the same number of species per population size for the three models. This similarity reveals that a high genetic complexity does not necessarily imply a higher species ratio.

In spite of the mate chose's condition of *Max Energy* and *Similar Energy* models, the reproduction ratio of these models is similar to the *Random* model. This similarity means that the mate choice restriction does not have a deleterious effect on the number of reproduction. However, if the average energy of all models is similar, then why the population size of the *Max Energy* model is higher? This phenomenon might be due to the individuals' lifespan in this model, although the reproduction ratio, the reproduction failed ratio and the death ratio of this model, see Appendix (A) Figure 31, have not a significant difference compared to the *Random* model. The higher *Max Energy*'s population size can be caused by the more average age and the average death age of the individuals in this model. The individuals in the *Max Energy* model can live longer, as it can be seen in Figure 9, which can be explained by more chance of survival. The similar reproduction ratio and higher age of individuals leads to a higher number of offspring that

each individual can produce during its life and subsequently, a higher value of its fitness. The greater fitness of the *Max Energy* model can reveal that 1) mating with a 'strong' partner has a positive effect on the fitness of that individual and that 2) energy can be a good indicator for the morphological choice.
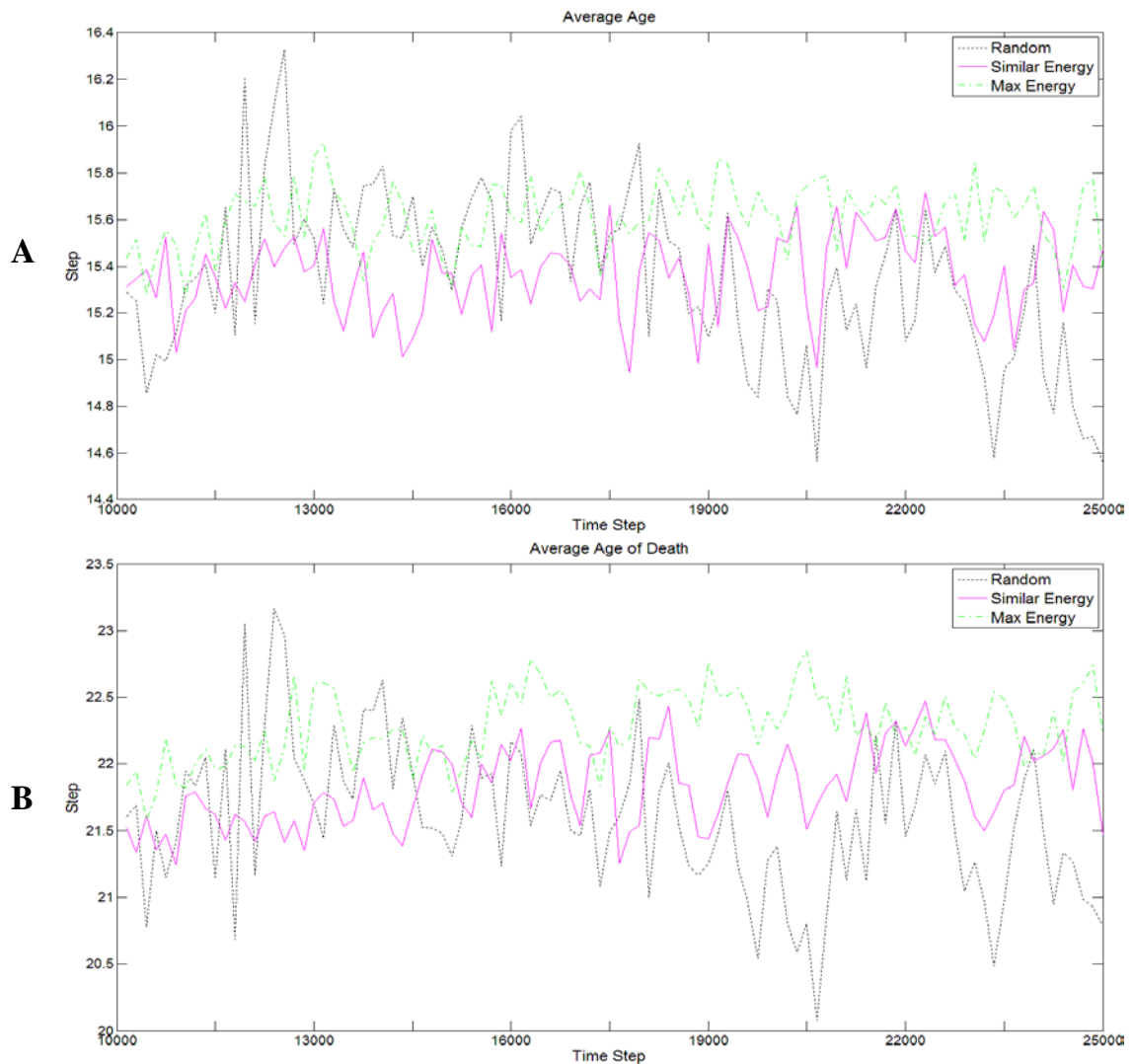
Figure 9: A) the average age, B) the average age of death for the time steps 10000 to 25000

Choosing a partner based on the energy as a morphological trait has an effect on the average of genetic distance between the two partners. The genetic mating distance in the *Max Energy* model is almost higher than the *Random* model (Figure 10). This means that the mating based on the exaggerated trait increases the average genetic distance of

partners (more dissimilarity between mates). We could expect that an increase of genetic distance between parents will also increase the genetic diversity of the population, but this is not the case. Perhaps this increase is smoothed by the natural selection pressure leading to the observed similar genetic diversity in all the models. But this phenomenon is still not well understood and needs more investigation.
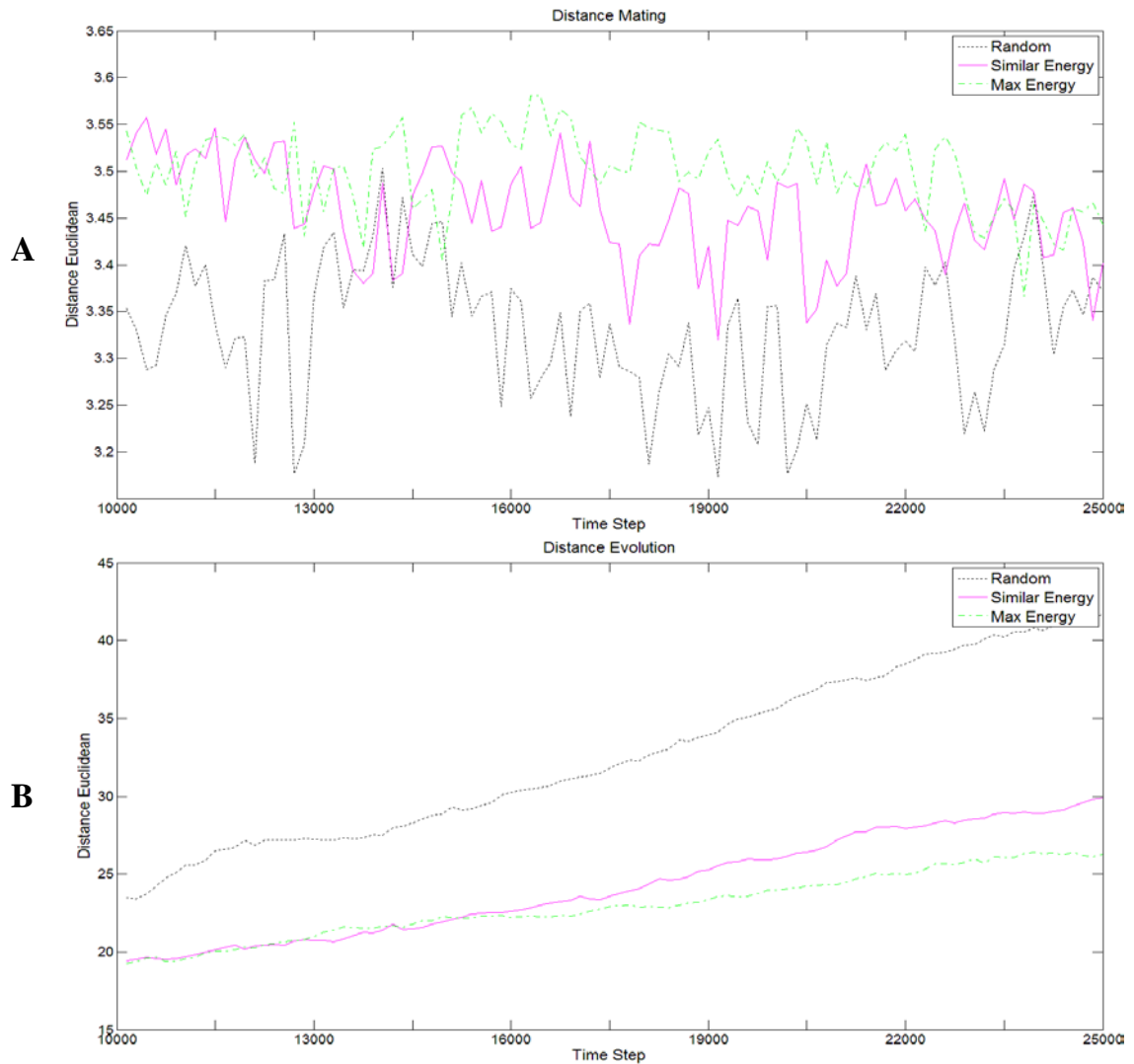
Figure 10: A) the distance of mating, B) average evolution of origin for the time steps 10000 to 25000

In addition to the average genetic distance of mating, the average evolutionary distance, showing the increase in genetic distance from the origin due to the evolutionary process, is highly lower in the mating models (Figure 10). The slope of the Origin model

is the sharpest, and the *Max Energy* model has the lowest one. This graph shows that in the *Random* model the accumulation of genetic variation from the origin of the run is much faster than for the models with mating policies. Although the *Random* model has higher evolutionary distance, this variety doesn't mean more various gene pools, as illustrated by the genetic diversity measure (Figure 8). This figure shows therefore a restriction in mating partner, even if it increases the genetic distance between mates, can slow down the evolutionary process. However, it does not affect the genetic diversity necessarily.
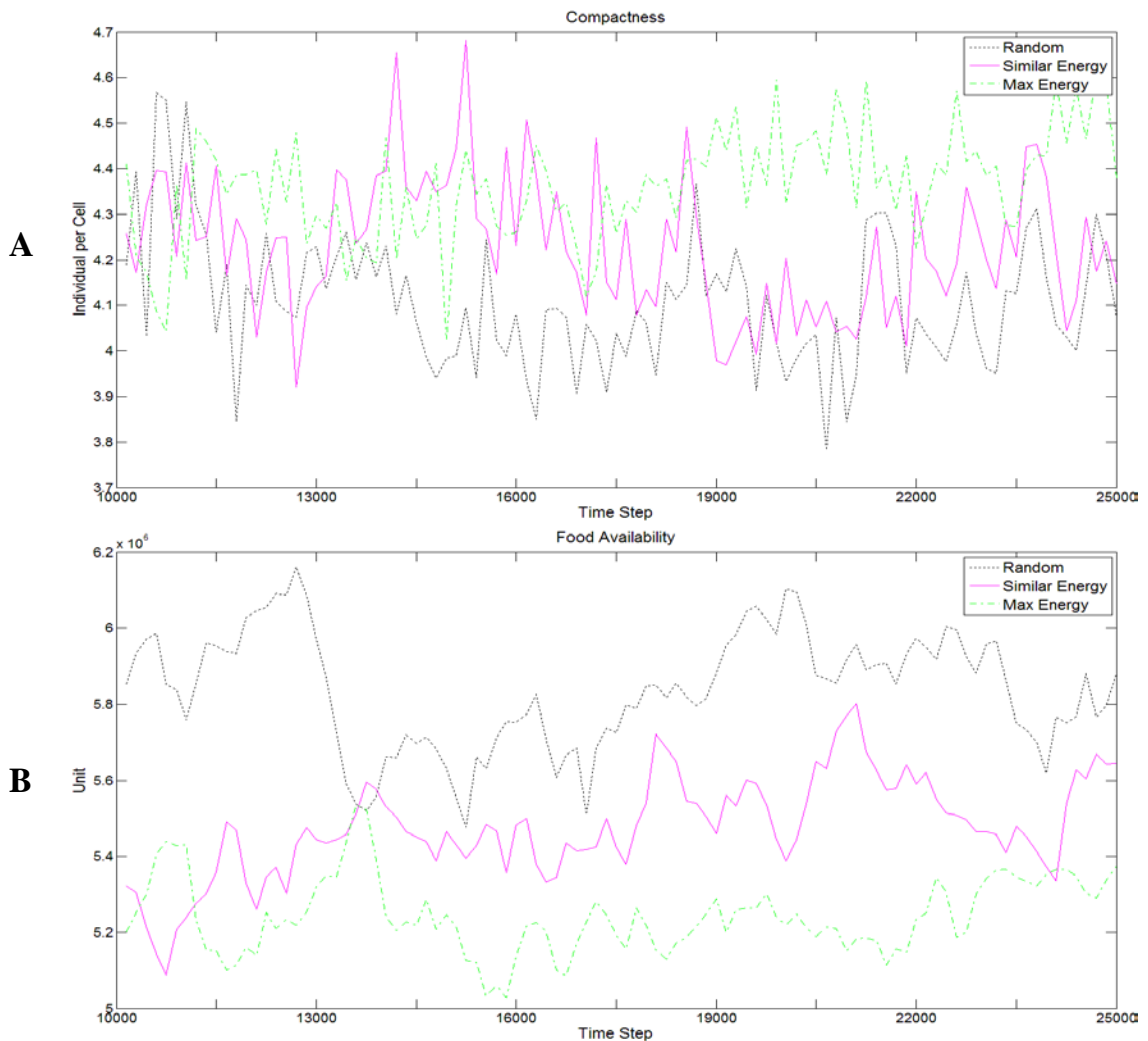
Figure 11: A) the compactness of individuals, B) the food availability for the time steps 10000 to 25000

The last difference when using the energy as a mate choice is on the dispersion of population. On the average, individuals of the *Max Energy* model make more compact groups as illustrated in Figure 11. This compactness has not had an impact on the reproduction ratio and the species ratio, although it affects the ratio of socialization, foraging and sedentary, see Appendix (A) Figure 32. Moreover, a higher compactness means a higher death ratio due to the lack of energy because the dense group of individuals cannot find enough food to eat. Consequently, the availability of food for the *Random* model is higher than for the other models which are caused by the larger dispersion of individuals who allow the grass to grow more and faster.

In this experiment, the comparison between different mating models has been done mainly by comparing different graphs. However, for a more formal evaluation the Table 5 shows the global statistic of population size, distance of mating and distance of evolution for each model. Based on this table we can conclude that *Random* and *Max Energy* models have a clearly separated behaviors for these three chrematistics (average separated by several standard deviations). The global statistics of other characteristics do not present as good separations as these three characteristics.

Table 5: The global statistic of some characteristics (mean & standard deviation)

Characteristic	Statistic	Random	Similar Energy	Max Energy
Population	**Mean**	1.51 e5	1.70 e5	1.93 e5
	STD	1.62 e4	1.24 e4	1.09 e4
Distance-Mating	**Mean**	3.32	3.45	3.50
	STD	0.06	0.03	0.02
Distance-	**Mean**	32.55	24.19	22.93
Evolution	**STD**	0.90	0.41	0.26

4.1.2 Prediction of the mate choice models

To analyse in detail the different mating models, we apply machine learning techniques to find the features and the rules that can differentiate and then explain each model. Three experiments are conducted based on three different sets, each of them being a combination of three runs one for *Max Energy*, one for *Similar Energy* and one for *Random* models. Each set contains 45000 samples (15000 samples for each run) corresponding to the time steps between 15000 and 25000. Two sets are merged together to prepare the training set and the test set. And the third independent set makes the validation set. As mentioned in the Feature Preparation's section, 34 features are gathered to study the mate choice in three steps: prediction of mate choice models, features reduction and rules extraction.

In the first experiment, all 34 features are used to discriminate between the mating models. The overall accuracy and other measures show that the predictions for the training and the test set are highly accurate (Table 6). The tree learned using j48 technique contains 79 leaves, each of them being a rule. This tree is composed of only 20 features out of the 34 used features.

Table 6: The evaluation measures of all features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.998	0.001	0.998	0.999	99.79%
Test	0.998	0.001	0.998	0.999	99.83%
Validation	0.323	0.339	0.283	0.629	32.28%

The accuracy of validation set is 32% which means that the model learned on the training set is too specific and can not make accurate prediction when the conditions are different. This low accuracy could be explained by the over-fitting problem. Furthermore,

the TP and FP rate of the validation show that only a few positive samples are classified correctly. Although the F-measure is low, the ROC measure shows that even the obtained classifier is weak, it is not random.

Decreasing the number of leaves can help to find more general rules. For reducing the number of rules and also bounding the effect of the over-fitting problem, the number of samples per leaf, which is a technique for pruning a tree -called Pruning Sample-, is increased to 1500. This pruning technique decreases the overall accuracy of the training set about 5%, but it increases the accuracy of the validation set about 16% (Table 7). Consequently, the new rules tree has nine leaves which is a strong reduction compared to the 79 rules we had before.

Table 7: The evaluation measures of all features using the J48 technique with the pruning samples equal to 1500 for distinguishing the mating models

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.942	0.029	0.942	0.982	94.18%
Test	0.945	0.027	0.945	0.983	94.49%
Validation	0.483	0.258	0.434	0.663	48.31%

In the next step, the most important features are selected based on feature selection techniques, such as Greedy-Stepwise, Linear-Forward-Selection, Best-First and Ranker (with merit greater than 0.2), which are implemented in WEKA [V3.6.4] and using default parameters, to find more general rules. Each technique provides a subset of the features (Table 8). The last column of this table shows the 10 selected features based on a voting mechanism.

Table 8: The selected features using the machine learning techniques for distinguishing the mating models. The last column presents the common features by a voting mechanism

#	Feature	Ranker	Best-First, Greedy-Stepwise, Linear-Forward-Selection	J48	Selected Features
01	Population-Size	✓	✓	✓	✓
02	Food-Availability	✓	✓	✓	✓
03	Species-Ratio		✓	✓	
04	Age	✓	✓	✓	✓
05	Distance-Predators			✓	
06	Quantity-LocalFood	✓	✓	✓	✓
07	Socialize-Ratio	✓	✓	✓	✓
08	Explore-Ratio			✓	
09	Foraging-Ratio	✓	✓	✓	✓
10	Sedentary-Ratio	✓	✓	✓	✓
11	Reproduce-Ratio			✓	
12	ReproduceFail-Ratio			✓	
13	Parental-Investment			✓	
14	Genetic-Complexity	✓	✓	✓	✓
15	Distance-Evolution	✓	✓	✓	✓
16	Distance-Mating	✓	✓	✓	✓
17	Genetic-Diversity	✓		✓	
18	MaxGenetic-Diversity			✓	
19	Killed-Ratio			✓	
20	DeathAge-Ratio	✓		✓	

The accuracy of the validation set with the selected features and the pruning samples equal to 1500 increases about 2% (Table 9), although the accuracy of the training set and the test set decrease about 6%. The tree only includes six features, Distance-Evolution, Foraging-Ratio, Sedentary-Ratio, Age, Genetic-Diversity and Food-

Availability that formed nine rules for the prediction of the mating models shown in Figure 12.

Table 9: The evaluation measures of the selected features using the J48 technique with the pruning samples equal to 1500 for distinguishing the mating models

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.887	0.056	0.887	0.972	88.68%
Test	0.887	0.056	0.887	0.971	88.74%
Validation	0.505	0.252	0.495	0.650	51.52%

This tree is similar to the tree which is generated using all features and the pruning samples equal to 1500 (Table 7). This similarity and the change of the accuracies imply that some features have a negative effect on the learning of the mating models, and also these machine learning techniques could not find the proper features set for this experiment. The rules tree illustrated in Figure 12 shows that each mating model is covered by three rules.

Three main features of this tree include the average age of all individuals, the evolutionary distance and the ratio of foraging that cover about 55% of samples' space of the training set containing the rules 1, 4 and 7 in Table 10. This table summarizes the rules generated by the selected features. According to this table, to distinguish a mating model based on the average features of all individuals, at most five features (rule 8) need to be evaluated. Moreover, these rules reveal that the main differences between these three mating models are linked to a combination of genetic, spatial and behavioural characteristic of the models.

The Distance-Evolution, as it appears in Figure 10, is very different in the *Random* model compared to the two models with mating policies. The *Max Energy* model has a higher average age (Figure 9), therefore this feature strongly affects the rule

of 7 but appear also on most of the rules. Although the graph of the Genetic-Diversity (Figure 8) has a lot of fluctuations, this feature could distinguish about 10% of all samples with the rules 2 and 5.
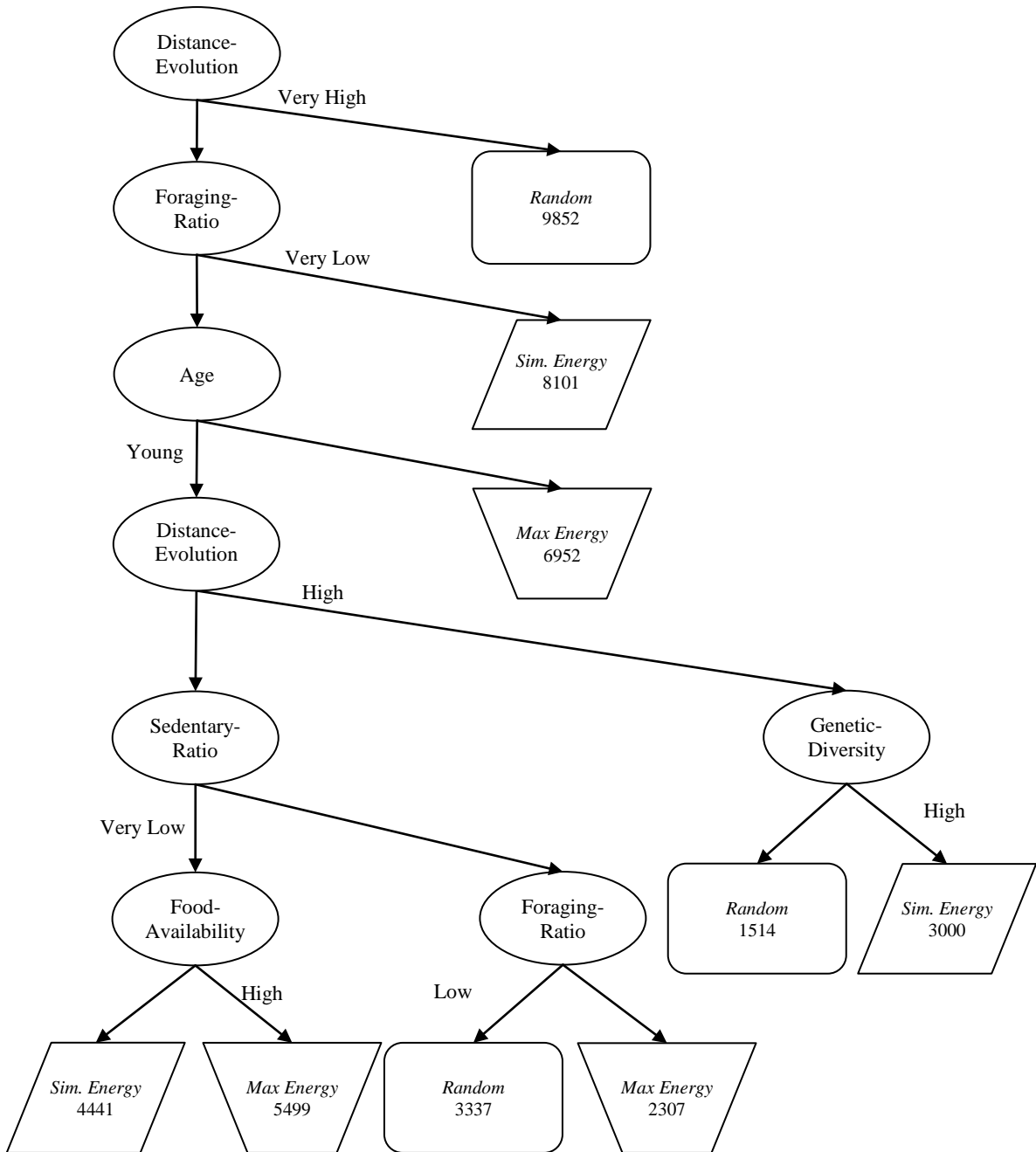
Figure 12: The rules tree of the selected features with the pruning samples equal to 1500 for distinguishing the mating models. The number in each leaf is the number of matched samples with the corresponding rule.

Table 10: The extracted rules of the selected features using the J48 technique for distinguishing the

mating models

#	IF	THEN
01	Distance-Evolution is very high	*Random*
02	Distance-Evolution is high AND Foraging-Ratio is not very low AND Age is young AND Genetic-Diversity is not high	*Random*
03	Distance-Evolution is not high AND Foraging-Ratio is low Age is young AND Sedentary-Ratio is very low AND	*Random*
04	Distance-Evolution is not very high AND Foraging-Ratio is very low	*Similar Energy*
05	Distance-Evolution is high AND Foraging-Ratio is not very low AND Age is young AND Genetic-Diversity is high	*Similar Energy*
06	Distance-Evolution is not high AND Foraging-Ratio is not very low AND Age is young AND Sedentary-Ratio is very low AND Food-Availability is not high	*Similar Energy*
07	Distance-Evolution is not very high AND Foraging-Ratio is not very low AND Age is not young	*Max Energy*
08	Distance-Evolution is not high AND Foraging-Ratio is not very low AND Age is young AND Sedentary-Ratio is very low AND Food-Availability is high	*Max Energy*
09	Distance-Evolution is not high AND Foraging-Ratio is not low AND Age is young AND Sedentary-Ratio is very low AND	*Max Energy*

As mentioned previously, some features can decrease the accuracy of prediction, but the techniques used could not detect and remove them properly. In a last step, each feature is evaluated separately for finding an appropriate features set for distinguishing

the mating models. Based on this experiment, some of the features have not an effect on the predicting and some of them have a small positive effect on it. Two features, including Foraging-Ratio and ReproduceFail-Ratio have a negative effect on the prediction which means that removing them increases the accuracy by about 15%. The interesting point is that all feature selection techniques, and the J48 chose them as the important features for mating models. By evaluating all 34 features and removing the unimportant features, a subset containing six features is selected. These features, including Sedentary-Ratio, Age, Genetic-Complexity, Distance-Evolution, Distance-Mating and Quantity-LocalFood improve the accuracy of the validation set about 17% in comparison to the previously selected features (see Table 11). However, the accuracy on the training set dropped down only by about 2%. Some features such as Genetic-Diversity, Genetic-Complexity and DeathEnergy-Ratio can increase the accuracy on the training set by about 6%, but only improve the accuracy of the validation set by about 1%. (71.13%), and therefore, these features are removed as well.

Table 11: The evaluation measures of the second set of selected features using the J48 technique and the pruning samples equal to 1500 for distinguishing the mating models

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.870	0.065	0. 869	0.958	86.99%
Test	0.866	0.067	0.867	0.956	86.63%
Validation	0.699	0.15	0.697	0.802	69.94 %

The tree based on these six selected features, shown in Figure 13, contains nine rules with a maximum depth of seven. The two first rules, containing two features of Distance-Evolution and Distance-Mating, can cover about 80% of the *Random* samples (12140 out of 15000). Quantity-FoodLocal is a part of individual's perception that shows the amount of food resources in a vicinity of individual. The amount of this feature gets

impacted by the food availability and the number of individuals in the vicinity whom are consuming the food.
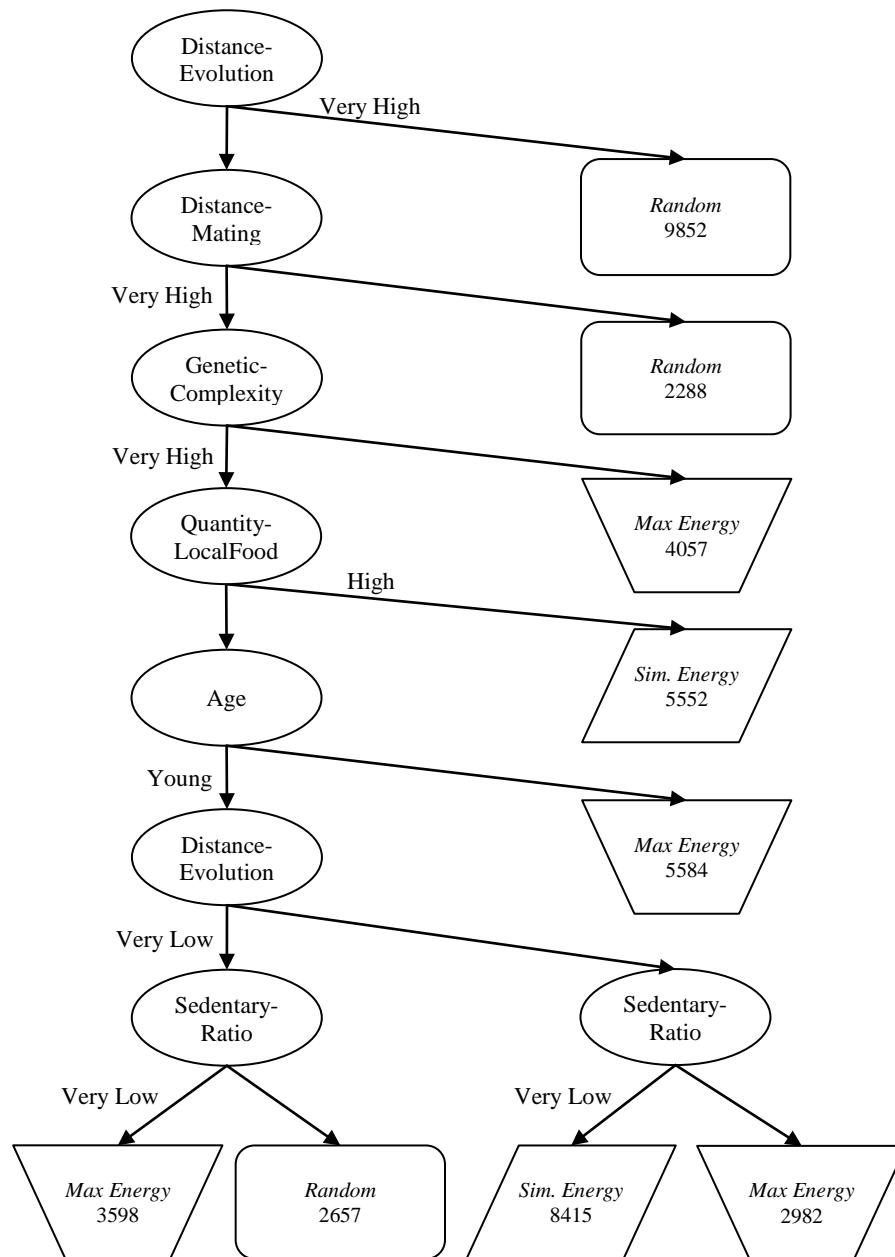
Figure 13: The rules tree of the second set of selected features with the pruning samples equal to 1500 for distinguishing the mating models. The number in each leaf is the number of matched samples with the corresponding rule.

Furthermore, the features such as population size and compactness have an indirect effect on this feature, for instance more individuals or more compact groups can decrease the amount of food in the vicinity of an individual. The interesting results of this analysis is that these rules show that adding a mating choice based on the energy trait can affect different aspects of the individuals such as their genetic characteristics. Moreover, these differences are specific of the mating models giving the possibility to accurately classify them.

4.2 Species Extinction

To study the extinction of species, two experiment sets are considered that aim to predict the species extinction in the next 100 time steps in EcoSim. Both sets, a combination of two runs using the time steps from 10000 to 20000, are used to build a training set and a test set. About 15% (70,000) of the combined samples are chosen randomly to form the training set and the remainder, composed of 410,000 samples, forms the test set. In addition, for both sets, an independent run is prepared to construct a validation set, containing 300,000 samples, to evaluate the overall accuracy of the learned model. Each run is formed by information about 300 species approximately.

The goal of the first experiment is to investigate the different features' effect on extinction. In this experiment, the dependencies and relations between features are also studied. In addition, some general rules for predicting species extinction are extracted by the machine learning techniques presented in beginning of this chapter. In the second experiment, the impact on the species extinction of five categories of features, demographic, age-energy, spatial, mating and genetic features, are analyzed in detail.

This experiment examines the relations between similar features and extracts the proper rules of each category.

4.2.1 First Experiment

As mentioned in the Feature Preparation section, 48 features have been computed to study species extinction. This part contains four steps: prediction of extinction, features selection, features dependencies and rules extraction. For these experiments, the J48 technique and a Bayesian-Network are applied to extract the rules and examine the features dependency respectively.

In the first step, all features are used to evaluate their respective importance on the prediction of extinction. The achieved accuracy for the test set is about 94% which indicate a good quality of the learned model. Moreover, the high accuracy of the validation set shows the generality of the learned model. The tree obtained from the J48 technique contains 1183 leaves composed by 18 features of 48 features. This large number of leaves imply a large number of rules but also an over-fitting problem. For decreasing the effect of these problems, the tree is pruned by increasing the number of samples per a leaf to 1000. The accuracy is almost unchanged, but the number of leaves is reduced dramatically to seven leaves. The results, obtained after applying the pruning technique (see Table 12), show that all features together have a good capacity for predicting the species extinction. All measures used for the evaluation confirm the high accuracy of the prediction on the training, test and validation sets. In addition, they prove that the rules created by the J48 technique are general enough to predict the extinction of a species in the near future even in the different situations from the ones used to learn the model.

Table 12: The evaluation measures of all features using the J48 technique and a pruning samples equal to 1000

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.942	0.075	0.941	0.971	94.16%
Test	0.942	0.073	0.942	0.971	94.20%
Validation	0.935	0.089	0.930	0.965	93.35%

Figure 14 present the tree obtained with the pruning samples equal to 1000, which contains seven rules made by six features, including Reproduce-Ratio, ReproduceFail-Ratio, Explore-Ratio, Age, Parent2-MatingAge and Spatial-Diversity. The number mentioned under each leaf gives the number of samples that are matched by this rule. Two rules of this tree with 20650 and 43643 samples have more impact on prediction of extinction and cover about 90% of the samples of the training set.
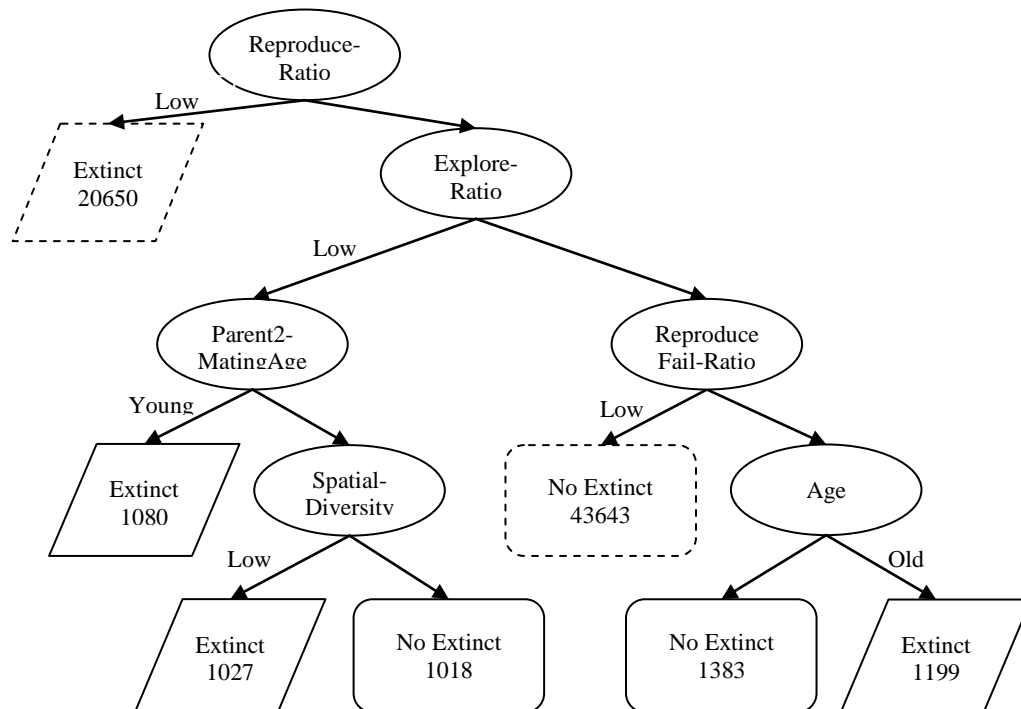
Figure 14: The rules tree of all features with a pruning samples equal to 1000. The number in each leaf is the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of matched samples

The rules extracted from Figure 14 are given in Table 13. For instance, based on rule number two, the species is more at risk of extinction when the exploration ratio is low, even when the reproduction ratio is adequate. Exploration can affect the gene flow in long-time evolution indirectly due to the level of interaction between individuals. Thus, the low exploration ratio can implies the lower gene flew. The interesting part of this rule is the role of a chosen partner in the extinction of a species. This rule implies that a species with few interaction and young mate is extremely vulnerable. One hypothesis is that the offspring survival of a young partner cannot be guaranteed because the ability of that partner for surviving and fitting to the environment is still doubtful.

Table 13: The extracted rules using the J48 technique

#	IF	THEN	Explanation
01	Reproduce-Ratio is low	Extinction	The species growth ratio is negative.
02	Reproduce-Ratio is not low AND Explore-Ratio is low AND Parent2-MatingAge is young	Extinction	The few interaction between individuals and the young chosen partners leads to become extinct.
03	Reproduce-Ratio is not low AND Explore-Ratio is low AND Parent2-MatingAge is not young AND Spatial-Diversity is low	Extinction	The few interaction of individuals with low spatial distribution declines the genetic diversity.
04	Reproduce-Ratio is not low AND Explore-Ratio is low AND Parent2-MatingAge is not young AND Spatial-Diversity is not low	No Extinction	
05	Reproduce-Ratio is not low AND Explore-Ratio is not low AND ReproduceFail-Ratio is low	No Extinction	
06	Reproduce-Ratio is not low AND Explore-Ratio is not low AND ReproduceFail-Ratio is not low AND Age is old	Extinction	The high mortality ratio in the old species steers toward extinction.
07	Reproduce-Ratio is not low AND Explorae-Ratio is not low AND ReproduceFail-Ratio is not low AND Age is not old	No Extinction	

The number of pruning samples has an impact on the combination of selected features. For example, if the pruning samples decrease to 500 samples, Individual-Ratio, Genetic-Diversity and Interbreeding-Ration replace Age. In addition, Parent2-MatingAge and Spatial-Diversity get exchanged with Parent1-MatingAge, Age and Spatial-Diversity. These relations express the dependency of some features with each other. For finding the important features and tracking the dependency between features, the feature space should be reduced. In the second step, the effective features are selected based on feature selection techniques.

Table 14: The selected features using the feature selection techniques. Last column presents the common features based on a voting mechanism

#	Feature	Ranker	Best-First, Greedy-Stepwise, Linear-Forward-Selection	J48	Selected Features
01	Individual-Ratio	✓	✓	✓	✓
02	Interbreeding-Ratio	✓	✓	✓	✓
03	Parental-Investment	✓	✓	✓	✓
04	Distance-Friends	✓	✓		
05	Quantity-LocalFood	✓			
06	Quantity-LocalPartner	✓		✓	
07	Escape-Ratio	✓			
08	Foraging-Ratio	✓	✓		
09	Socialize-Ratio	✓	✓		
10	Exploration-Ratio	✓		✓	
11	Eat-Ratio	✓		✓	
12	Sedentary-Ratio		✓		
13	Reproduce-Ratio	✓	✓	✓	✓
14	ReproduceFail-Ratio	✓	✓	✓	✓
15	Parent1-MatingAge	✓	✓	✓	✓
16	Parent1-MatingEnergy	✓	✓	✓	✓

Table 15: The selected features using the feature selection techniques. Last column presents the common features based on a voting mechanism (continue)

#	Feature	Ranker	Best-First, Greedy-Stepwise, Linear-Forward-Selection	J48	Selected Features
17	Parent2-MatingAge	✓	✓	✓	✓
18	Parent2-MatingEnergy	✓	✓	✓	✓
19	Distance-Mating	✓	✓		
20	Distance-Evolution			✓	
21	Genetic-Diversity	✓		✓	
22	Killed-Ratio	✓	✓	✓	✓
23	Death-Energy	✓	✓		
24	Death-Ratio	✓			
25	Age	✓	✓	✓	✓
26	Death-Age	✓	✓		
27	Spatial-Diversity			✓	
28	Contagion			✓	
29	Volume-Ratio	✓			
30	Speed	✓			

The last column of Table 14 and Table 15 show the selected features based on a voting mechanism. The achieved accuracy of the training set with these eleven features is about 93% close to the accuracy with of all features (Table 16). The training set, and validation set confirm that this feature selection doesn't decline the prediction of extinction's quality.

Table 16: The evaluation measures using the selected features and the pruning samples equal to 1000

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.935	0.086	0.935	0.959	93.48%
Test	0.935	0.088	0.934	0. 960	93.45%
Validation	0.951	0.074	0.951	0. 950	95.07%

In the third step, the dependency between the selected features is studied by a Bayesian-Network in which each node can have at most three parents. As shown in Figure 15, the leaves of this graph are made by three features: ReproduceFail-Ratio, Parental-Investment, and Parent2-MatingEnergy. Moreover, Individual-Ratio depends on most of the features except Partner1-MatingEnergy and Partner2-MatingEnergy. This dependency of Individual-Ratio expresses that the quality of the population size of a species relies on some other factors such as the average age and the birth ratio.
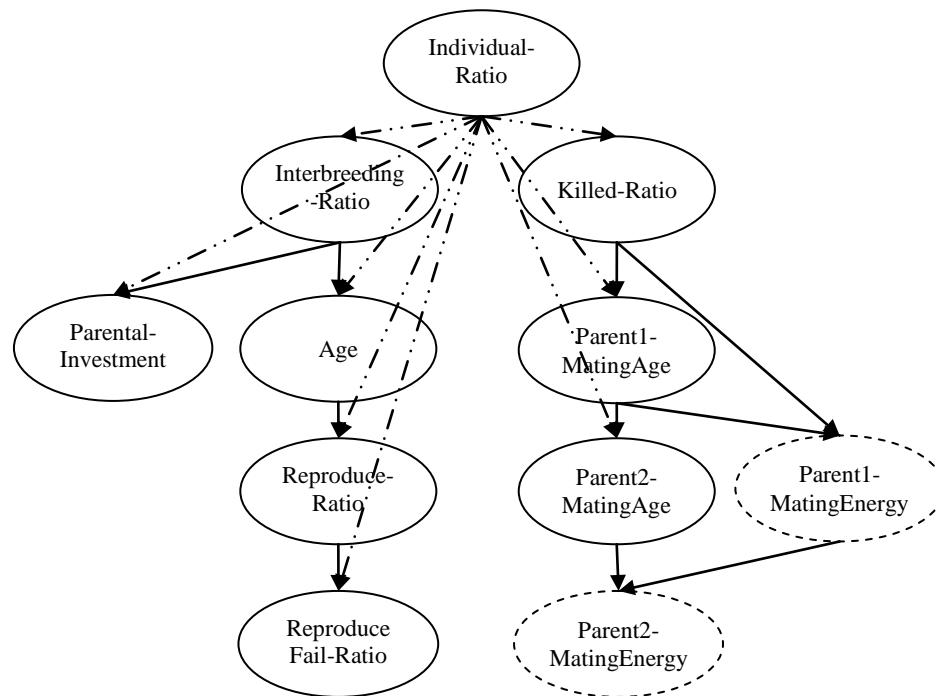
Figure 15: The dependencies of the selected features. Individual-Ratio has not a connection with dashed shape features

The Interbreeding-Ratio depends on Age and Parental-Investment which shows the effect of the average age of individuals and the parental cost of each individual. The effect of the average age is interesting as it is quite intuitive. However, neither nothing in our system decrease the fertility of the individuals with age nor increase the cost of reproduction for them. It should therefore be interesting to study in more detail why this

phenomenon appears. Finding the major features by this dependency graph is difficult, because the features can have some hidden dependencies. Nevertheless, it can be expected that the Individual-Ratio and Interbreeding-Ratio appear in the extinction rules due to their dependencies with other features.

Figure 16 presents the tree obtained using the features selected by pruning 1000 samples per leaf. This tree is formed by six features, Individual-ratio, Parental-Investment, Reproduce-Ratio, Interbreeding-Ratio and Parent2-MatingAge, and contains six rules. Although the dependency graph shows the relation between the features, the selected features by the J48's tree do not correspond to these dependencies.
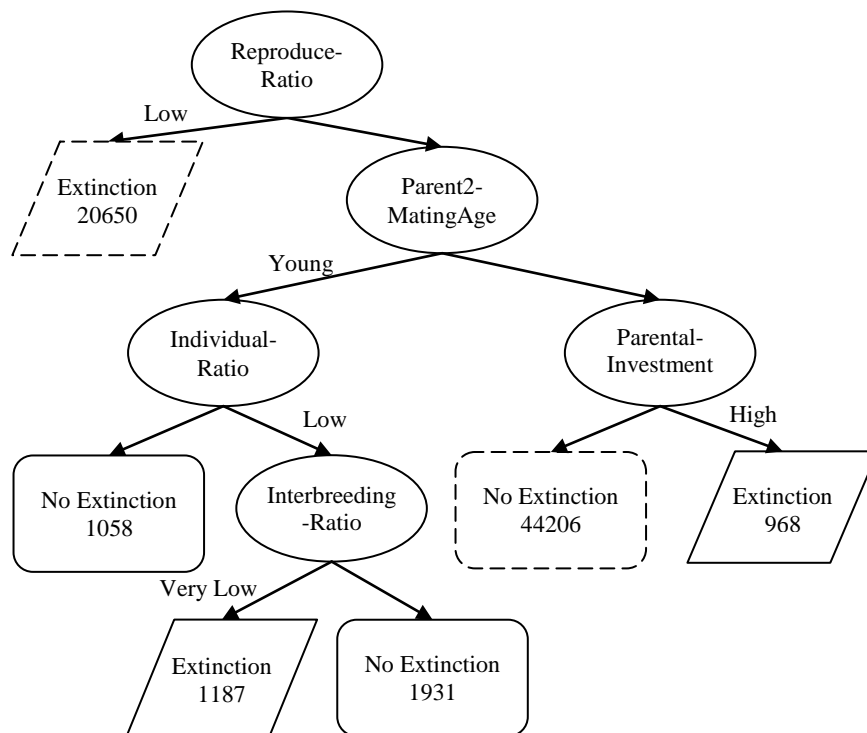
Figure 16: The tree using the selected features with the pruning samples equal to 1000. The number in each leaf is the number of the matched samples with the corresponding rule. The dashed shape shows a rule with a high number of matched samples

Two main rules, highlighted by dashes, cover about 93% of the training set. These two rules are: 'if Reproduction-Ratio is low, then that species will become extinct' and 'if Reproduce-Ratio is not low, AND Parent2-MatingAge is not young, AND Parental-Investment is not high, then that species will not become extinct'. The second rule explains that a species with an adequate birth ratio and a low parental cost can survive in the near feature because the growth population rate is positive and also the individuals of that species do not have risky investment on the mating process.

Comparing tree of Figure 12 to the one of Figure 16, it is noticeable that Reproduction-Ratio is the root of both trees. The rule using this feature reveals that a durable species depends mostly on how the growth ratio is. In addition to Reproduce-Ratio, Parent2-MatingAge is common among both trees confirming the importance of a chosen partner's age on the survival of a species in EcoSim. Comparing Figure 14 and Figure 16 it appears that Explore-Ratio, ReproduceFail-Ratio and Spatial-diversity get exchanged for Individual-Ratio, Parental-Investment and Interbreeding-Ratio. This replacement states that the population ratio of a species, the parental cost, and the ratio of inter-species breeding can compensate for the exploration ratio, which is an element for the gene flow due to movement of individuals, the unsuccessful birth ratio, and the species dispersal. Clearly, the Individual-Ratio can be affected by Explore-Ratio, ReproduceFail-Ratio and Spatial-diversity due to the impact of the diversity of the gene pool on the population size and also the effect of the growth ratio on it. In a rough conclusion, the extinction of a species can be predictable by different combination of features, because most of them have strong or hidden dependencies on each other and a

change in the combination of features can lead to several different rules due to the repetitive behaviour of machine learning techniques used.

4.2.2 Second Experiment

In this experiment, the features are divided into six categories and studied separately. These categories include: demographic, spatial, age-energy, mating, genetic and multi-category features. Each category contains the features related to the title of that category. This experiment is conducted to examine the effect of each category on the species extinction. In addition, for each category we attempt to find the main features that are need to build the general rules predicting extinction using the J48 technique.

4.2.2.1 Demographic Features

The first category, including five features: Species-Number, Individual-Ratio, Reproduce-Ratio, Death-Ratio, and Killed-Ratio, represents the demographic features. It has been shown that demographic features (factors) have an impact on the extinction of populations [Ovaskainen and Meerson 2010]. Figure 15 presents the dependencies of the demographic features using a Bayesian-Network technique. It can be seen in this figure that Individual-Ratio is caused by Reproduce-Ratio and Death-Ratio confirming the dependence of the population size with the birth and death ratio. It is obvious that the number of individuals leads to an increase or a decrease of the number of births and deaths, consequently it can affect ratio. It seems reasonable to expect that Individual-Ratio can have a great effect on the species extinction.
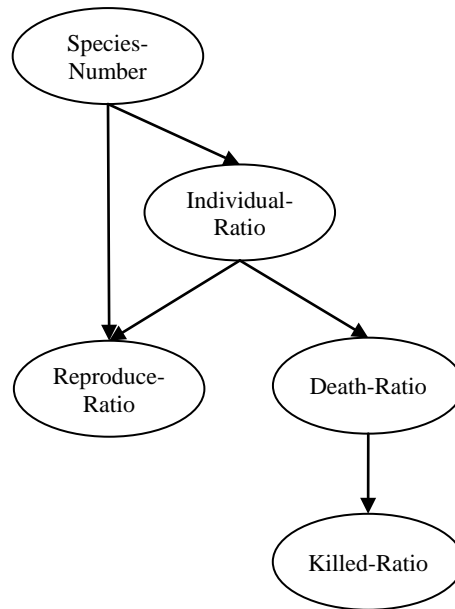
Figure 17: The dependencies of demographic features

The evaluation measures, showed in Table 17, prove that the demographic features are good enough for predicting the extinction in EcoSim. The accuracy is high even for the validation set. The tree based on the demographic features and with pruning 1000 samples per leaf is shown in Figure 18. This tree has a depth of five with five rules made by three features: Individual-Ratio, Reproduce-Ratio and Killed-Ratio.

Table 17: The evaluation measures of demographic features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.934	0.090	0.933	0.961	93.39%
Test	0.933	0.087	0.933	0.962	93.34%
Validation	0.948	0.074	0.948	0.975	94.83%

Species-Number and Death-Ratio do not appear in the tree. Three rules, highlighted by dash in Figure 18, cover most samples of the training set. The first two rules can predict about 78% of the training samples which demonstrate the importance of the individual ratio and the reproduction ratio in disappearing or in surviving of a species.
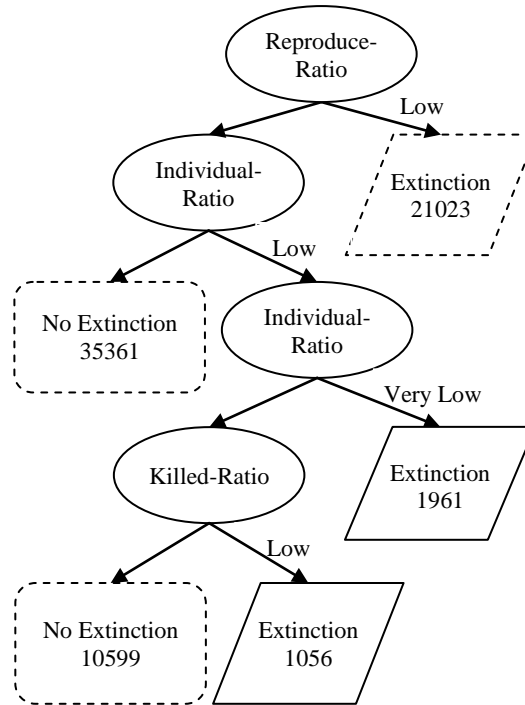
Figure 18: The tree of demographic features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

These three rules include: 'If Reproduce-Ratio is low, then that species will become extinct', 'If Reproduce-Ratio is not low, AND Individual-Ratio is not low, then that species will not go extinct' and 'If Reproduce-Ratio is not low, AND Individual-Ratio is not very low, AND Killed-Ratio is not low, then that species will not go extinct'.

4.2.2.2 Age-Energy Features

The second category that can affect the extinction combines seven features related to the average age and energy of species: Age, Energy, Death-Age, Death-Energy, DeathAge-Ratio, DeathEnergy-Ratio and Speed. For instance, the Age as a measure of the average age of individuals can express the fragility of a species toward extinction. Furthermore, the average energy of a species expresses the vulnerability or the resistance

of a species to hard situation. The dependencies between these features, given in Figure 19, show that Speed and Energy are caused by Age.
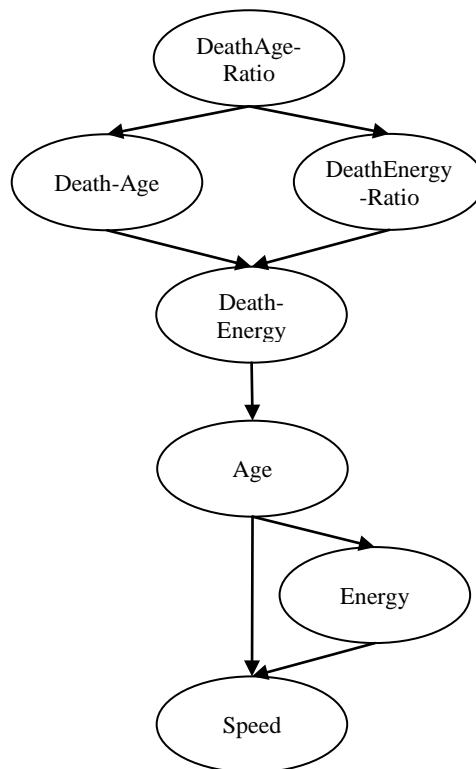
Figure 19: The dependencies of age-energy features

A species with an older average age cannot forage properly and consequently, its average energy decreases. Moreover, the amount of age can impact on the speed of a species. One more time, as there is no physical change for the individuals when they get older, these results are interesting but their reasons are not clear.

Table 18 shows the evaluation measures of the age-energy features. The overall accuracy in comparison to the selected features' accuracy drops about 3% which is negligible. The tree with a depth of five using the pruning 1000 samples per leaf is shown in Figure 20. This tree includes seven rules formed by three features: Age, Energy and Death-Energy.

Table 18: The evaluation measures of age-energy features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.907	0.121	0.907	0.926	90.73%
Test	0.908	0.113	0.908	0.926	90.79%
Validation	0.933	0.103	0.932	0.937	93.28%

Two rules cover about 85% of the training samples, shown by a dashed shape in Figure 20. These rules are 'if Death-Energy is low, then that species will become extinct' and 'if Death-Energy is not low, AND AGE is not old, AND Energy is not very high, then that species will not become extinct'.
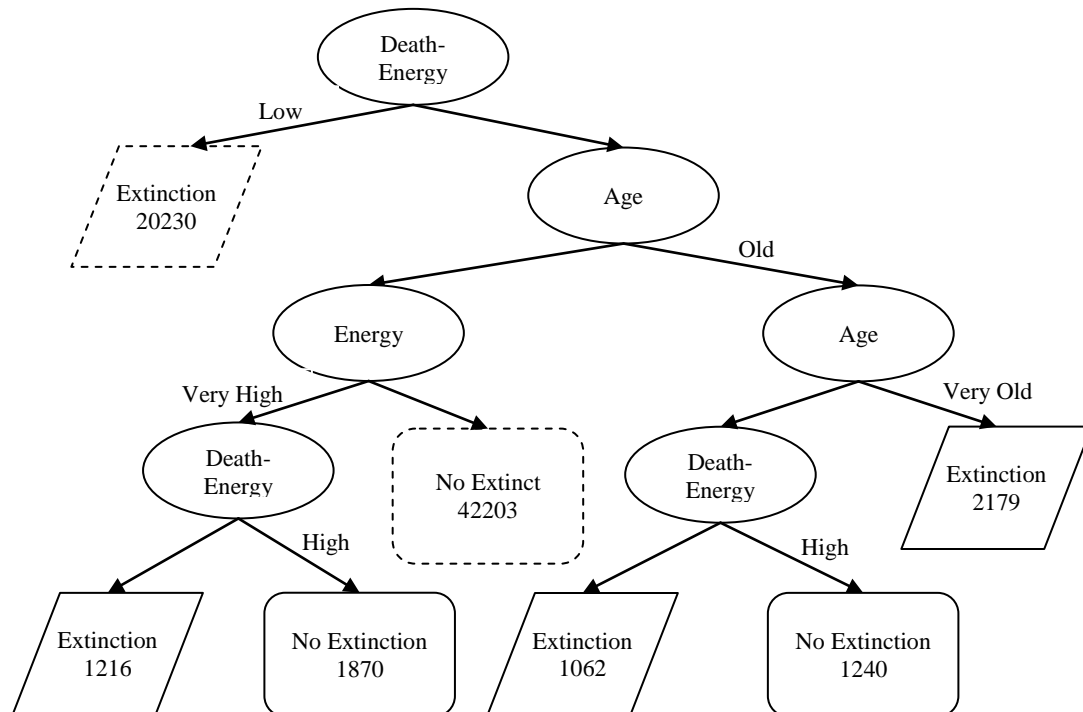
Figure 20: The tree of age-energy features with pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

The first rule states that the average energy at the death time has an impact on the extinction. This amount is a combination of two values: the energy of death due to an old age plus the energy of death due to the hunted preys by predators. The low amount of

Death-Energy possibly means that the species faces some severe conditions. The second rule expresses that a species can survive when its individuals are young with an acceptable amount of energy.

In a second attempt, four features: Death-Energy, DeathAge-Ratio, DeathEnergy-Ratio and Speed are removed to concentrate more on the rest of them. In Figure 21, the tree is given based on the three features: Age, Energy and Death-Energy. This tree is made by five rules with a depth of four. Three rules highlighted by dash can cover 95% of the training samples: 'if Age is too old, then that species will become extinct', 'if Age is not too old, AND Death-Age is old, then that species will not become extinct', and 'if Age is not too old, AND Death-Age is too young, then that species will become extinct'.

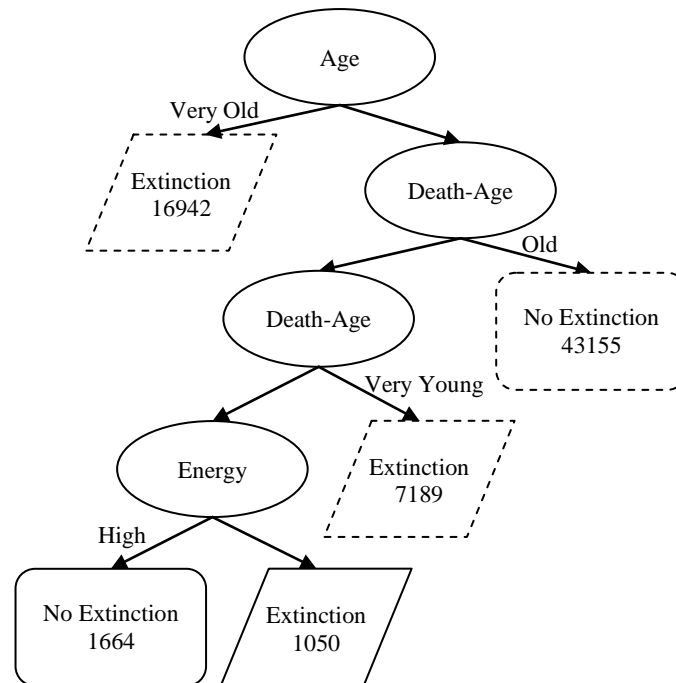**Figure 21: The rules tree for Age, Energy and Death-Age features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples**

The third rule means that a species contain individuals which are not old and in which the average age of death is very young will become extinct due to missing the younger individuals without replacement. When comparing trees of Figure 20 and Figure 21, it appears that Death-Age is replaced with Age and Death-Energy. The relation between these features can be observed in Figure 19.

4.2.2.3 Spatial Features

The third category works on the thirteen spatial features that encompass Spatial-Diversity, Volume-Ratio, Surface-Area, Shape-Complexity, Contagion, Spatial-Complexity, Fractal-Dimension, Space-Ratio, Compactness, Patch-Number, Patch-Size, MultSpecies-Cell and History-Movement. The spatial features are part of the environmental factors, as the habitat fragmentation, that have a major role in extinction [Patten et al. 2007], [Drake and Lodge 2004].
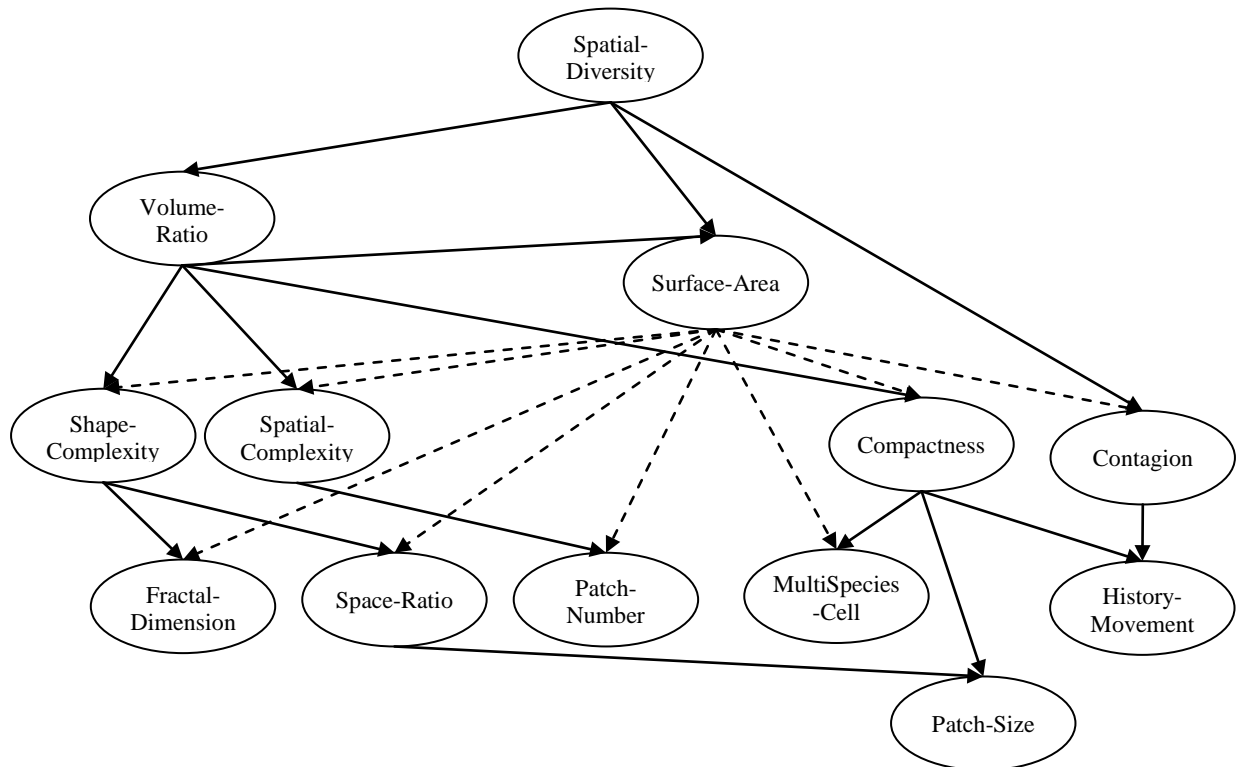
Figure 22: The dependencies of spatial features

According to Figure 22 that shows the graph of features dependencies, Volume-Ratio and Surface-Area have more connections with other features. As mentioned before, Volume-Ratio is the division between the population size of a species and the number of occupied area (cells). This feature shows how the species is distributed in the world and can be viewed as a measure for the access of the species to the natural resources. The second feature, Surface-Area, presents the boundary of dispersion of a species that shows the extent of species' territory. These features can explain the vulnerability of a species according to the occupied area, the availability and dispersal of food and partners in that area.

The results of spatial features using the J48 technique show that they are good enough to predict the extinction without losing the overall accuracy (Table 19). The tree obtained with pruning 1500 samples per leaf, given in Figure 23, contains six rules. Only Volume-Ratio and Surface-Area form this tree which can be explain by the effect they have on the other features as it can be seen in the dependency graph. The samples of training set are covered approximately equally with all leaves of this tree.

Table 19: The evaluation measures of spatial features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.919	0.106	0.918	0.949	91.86%
Test	0.919	0.098	0.919	0.942	91.88%
Validation	0.937	0.080	0.937	0.957	91.67%

The six rules are: 'if Volume-Ratio is very high, then that species will become extinct', 'if Volume-Ratio is not very high, AND Surface-Area is very high, then that species will become extinct else will not', 'if Volume-Ratio is very low, AND Surface-Area is high, then that species will not become extinct', and 'if Volume-Ratio is very

low, AND Surface-Area is not high, then that species will not become extinct else will become extinct'.
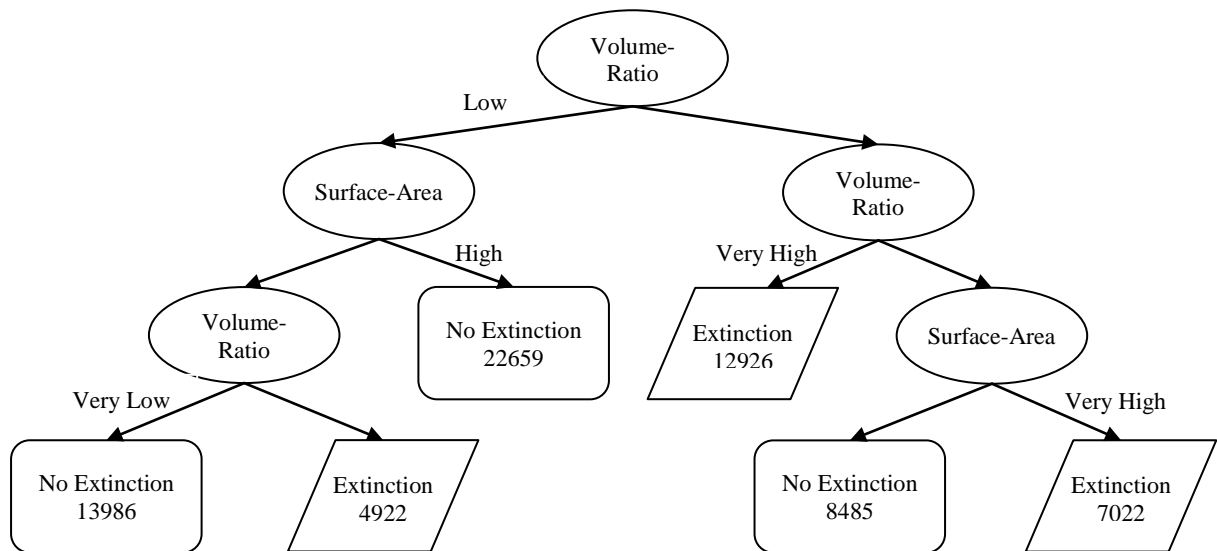

Figure 23: The rules tree of spatial features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

For the second rule, when Volume-Ratio is low the number of the occupied cells can be high or the number of individuals can be low. However, this rule also mentions that the Surface-Area is high. Therefore, based on the first part of rule, the population is distributed in the vast area, and consequently the species can survive because it has enough resources.

4.2.2.4 Mating Features

In the Forth category, a combination of six mating features is studied. These features are: Reproduce-Ratio, Interbreeding-Ratio, Parent1-MatingAge, Parent1-MatingEnergy, Parent2-MatingAge and Parent2-MatingEnergy. The Reproduce-Ratio and Parent1-MatingAge have more interaction with other features according to the dependency graph presented in Figure 24. The changes of Reproduce-Ratio, which is

affected by the abundance of recourses such as food and partner, can impact on the quantity and variety of individuals of a species. Having more individuals means faster consumption of the available food and makes severe situation for a species.
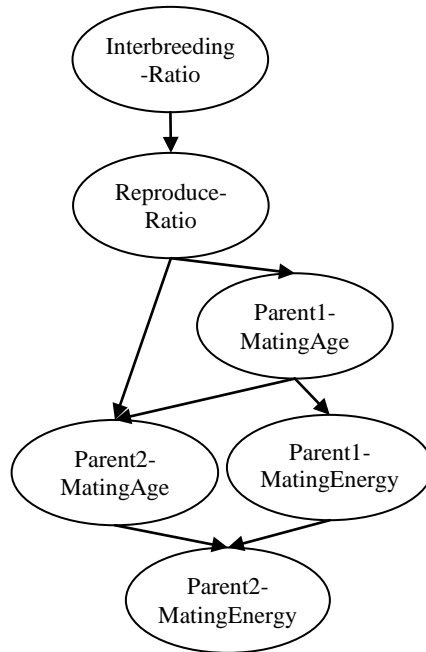
Figure 24: The dependencies of mating features

Table 20 shows that the evaluation measures with the mating features are as good as the measures with the selected features for the training, test and validation sets. These measures are achieved by pruning 1000 samples per leaf using the J48 technique. The tree of mating features contains five rules with two of them covering about 90% of the training samples.

Table 20: The evaluation measures of mating features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.933	0.092	0.933	0.941	93.34%
Test	0.933	0.091	0.933	0.943	93.32%
Validation	0.949	0.084	0.948	0.950	94.88%

This tree is made by four features: Reproduce-Ratio, Parent1-MatingAge, Parent2-MatingAge and Interbreeding-Ratio. The two main rules are: 'if Reproduce-Ratio is low, then that species will become extinct' and 'if Reproduce-Ratio is not low, AND (Parent1-MatingAge AND Parent2-MatingAge) are not young, then that species will not become extinct'. The second rules mention that the parental age has an impact on the survival or extinction of a species. A species that has a positive birth rate but has a young choosy partner in average can be at risk of extinction. This phenomenon is interesting, although it needs extra investigation. One hypothesis can be based on the amount of energy of younger individuals: If they have less energy in comparison to mature individuals, therefore, the cost of parental care can kill them and consequently, the number of older individuals will be increased in contrast to younger individuals.
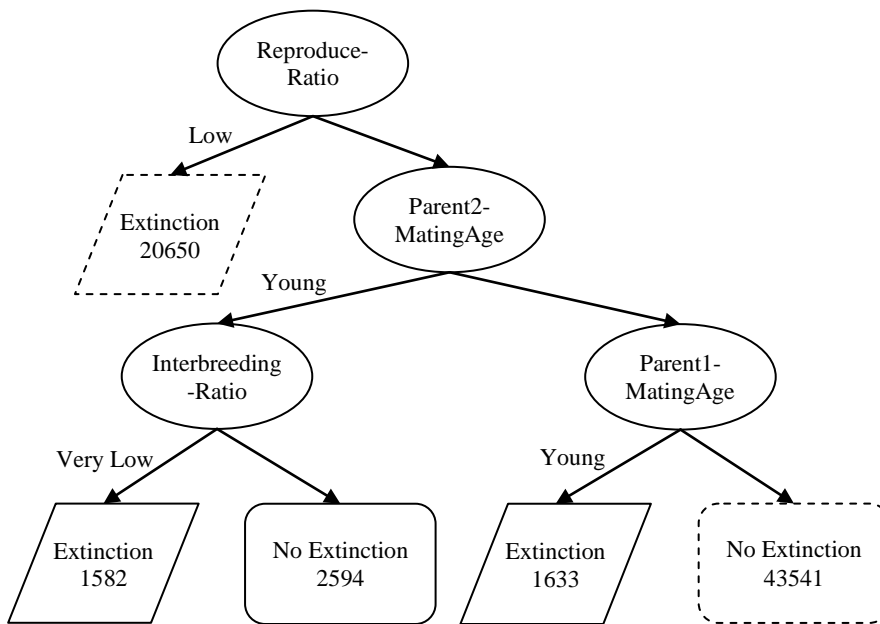
Figure 25: The rules tree of mating features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

4.2.2.5 Genetic Features

The fifth category examines the effect of five genetic features on the extinction. This experiment uses the features: Parental-Investment, Distance-Mating, Distance-Evolution, Genetic-Diversity and Genetic-Complexity. The dependency graph in the Figure 26 shows that Genetic-Diversity has more connection than the other features. This feature explains the diversity of the genes' pool for a species.
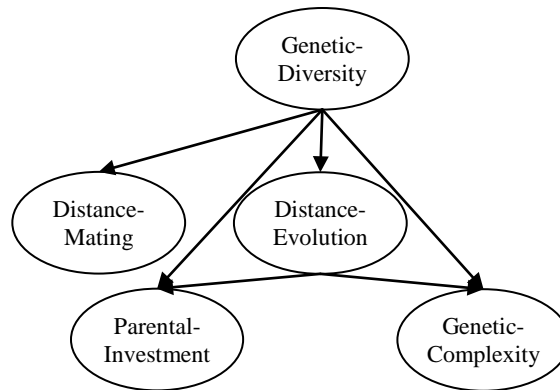
Figure 26: The dependencies of genetic features

Table 21 presents the quality of genetic features for the prediction of species extinction. The overall accuracy for the training, test and validation sets are similar to the ones obtained with the selected features. The tree contains four rules in which two of them, highlited by dash, cover about 95% samples of the training set. This tree is composed of three features: Parental-Investment, Distance-Mating and Genetic-Diversity. Although Genetic-Diversity has a lot of connections with other nodes in the dependency graph, the Distance-Mating is the root of the tree.

Table 21: The evaluation measures of genetic features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.933	0.097	0.932	0.927	93.28%
Test	0.930	0.099	0.930	0.926	93.02%
Validation	0.949	0.092	0.948	0.938	94.88%

The genetic features create four rules, including: 'if Distance-Mating is very low, then that species will become extinct', 'if Distance-Mating is not very low, AND Parental-Investment is not high, then that species will not become extinct' and 'if Distance-Mating is not very low, AND Parental-Investment is high, AND Genetic-Diversity is not high, then that species will become extinct else will not'. In all rules, the usage of Distance-Mating shows that the genetic similarity between partners has a key role on the extinction. One explanation might be that if the partners are too genetically similar, the diversity in a species declines and leads to extinction of that species. This interesting rule shows the impact of genetic distance between partners as a pre-zygotic barrier. Furthermore, the cost that should be paid by partners can put individuals at risk of death. This cost corresponds to the investment of each individual in the breeding process.
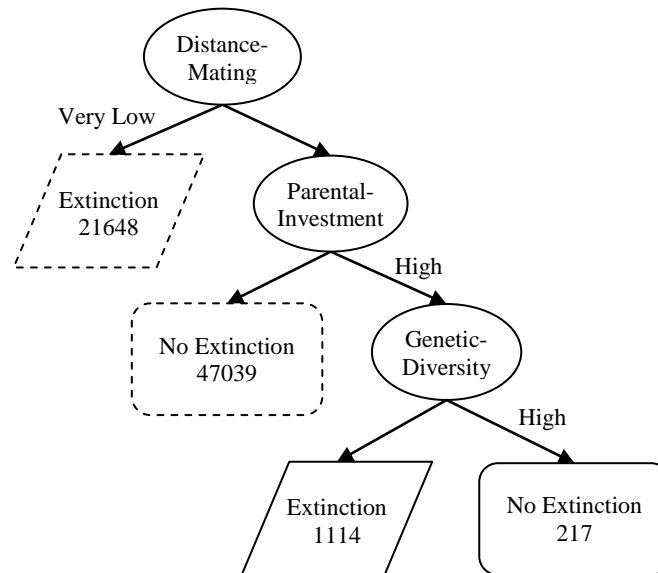
Figure 27: The rules tree of genetic features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

4.2.2.6 Multi-category Features

The last category includes six features form the previous categories. These features are chosen to study the effect of a combination of features on the extinction. The chosen features are the most important feature of each category. For this purpose, Species-Number from demographic, Energy from age-energy, Volume-Ratio from spatial, Interbreeding-Ratio from mating and Genetic-Diversity and Parental-Investment from genetic category are selected. The dependency graph between them is given in Figure 28. It shows that Interbreeding-Ratio is caused by Genetic-Diversity and Volume-Ratio. These dependencies can be due to the relation between interbreeding and an increase or a decrease of the genes' variety. Moreover, Genetic-Diversity is caused by Parental-Investment and Energy that it is acknowledgment validation of the parental cost as a part of genetic content in EcoSim.
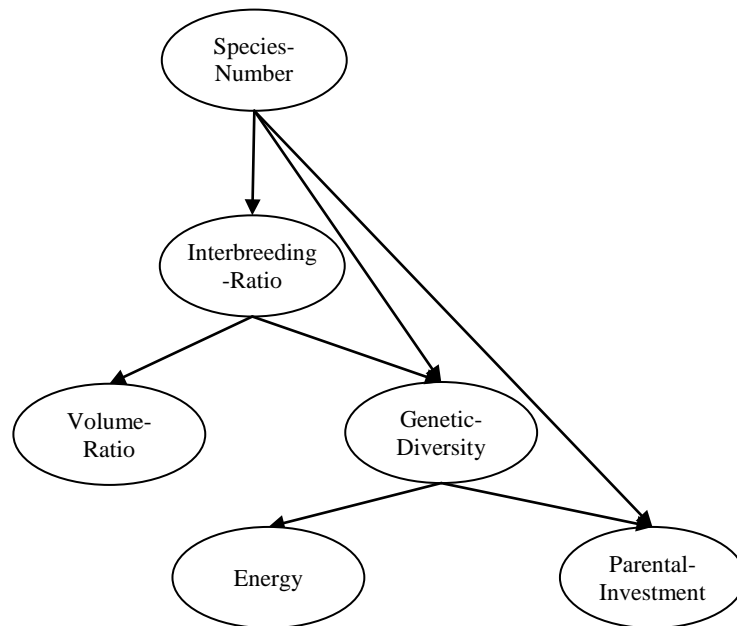
**Figure 28: The dependencies of multi-category features**

The accuracy measure of the multi-category features is high enough to predict extinction (Table 22), but a little less than the accuracy of the selected features. Three

features: Parental-Investment from genetic, Interbreeding-Ratio from mating and Volume-Ratio from spatial category, are used in the tree. This tree (Figure 29) is made by four rules. Each of them covers an acceptable number of the training samples which is interesting because in all other trees just some rules could cover most of the sample space.

Table 22: The evaluation measures of multi-category features using the J48 technique

Data Set	TP Rate	FP Rate	F-Measure	ROC	Accuracy
Training	0.904	0.128	0.903	0.917	90.40%
Test	0.903	0.129	0.902	0.917	90.32%
Validation	0.914	0.126	0.913	0.923	91.37%

These rules include: 'if Parental-Investment is very high, then that species will become extinct', 'if Parental-Investment is not very high, AND Interbreeding-Ratio is not very low, then that species will not become extinct' and 'if Parental-Investment is not very high, AND Interbreeding-Ratio is very low, AND Volume-Ratio is low, then that species will not become extinct else will become extinct'. The last rule implies that the individual dispersal can be affected by the ratio of interbreeding. The low interbreeding ratio shows that the species have a few gene exchanges with other species and consequently, the survival of this species is dependent to the compact spatial distribution of its individuals. Based on previous experiments, the survivorship of compact spatial distribution of individuals depends on the food availability. Therefore, this rule reveals the dependency of different features on the extinction of a species.
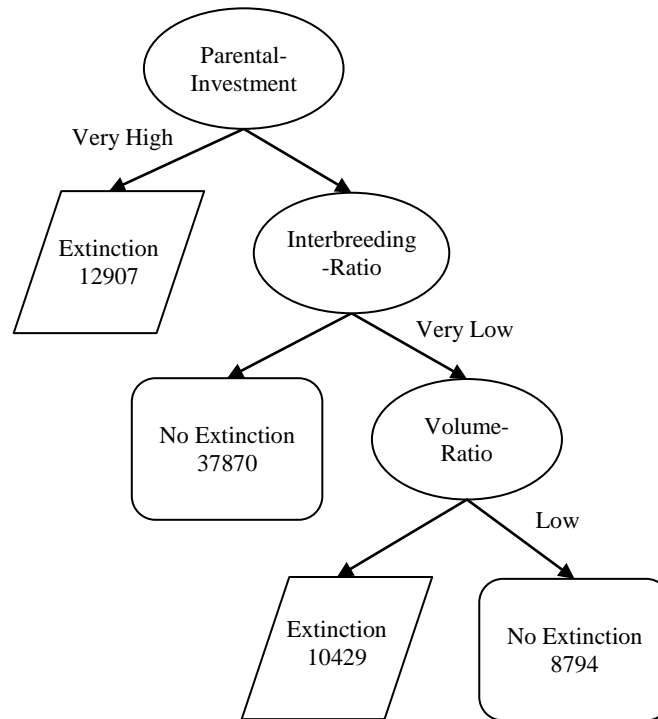
Figure 29: The rules tree of multi-category features with the pruning samples equal to 1000. The numbers in the leaves give the number of matched samples with the corresponding rule. The dashed shape shows a rule with a high number of the matched samples

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

In this study, we used EcoSim, an IBM platform with genetic traits. This model allows us to work on numerous features simultaneously. Based on this model, we conducted two experiments for investigating mate choice models and species extinction generating a huge amount of raw data. We apply a new strategy, based on a combination of feature selected techniques and machines learning tools to explore such a large data set. We have been able to obtain very interesting and promising results to find the effective features impacting on these phenomena. We have therefore validated our approach which can now be generalized and applied to many different data and many different situations.

5.1 The Mate Choice Models

Three models of morphological mate choice, based on energy as an indicator of strength, were examined. The results showed that energy as a physical signal for mating process can change some characteristic of species. Two experiments were conducted to evaluate the introduced mate choice models. In the first experiment, the general behaviour of *Similar Energy*, as a similar mate policy, *Max Energy*, as an exaggerated ornamented mate policy, and *Random*, as a random mate policy, were compared together. The results showed that the *Max Energy* and *Similar Energy* models have different behaviours in comparison to the *Random* model. However, the main differences occurred between *Max Energy* and *Random* models and *Similar Energy* model had intermediate behaviours between them.

Some features had almost opposite trends in *Random* and *Max Energy* models. For instance, Population-Size, Parental-Investment, Age, Age-Death, Distance-Mating, and Compactness, are higher in the *Max Energy* model, whereas Genetic-Complexity and Distance-Evolution, are lower in the *Max Energy* model. Some intuitive and interesting results have been extracted from these models:

- On average, individuals having a strong mate have a simpler behavioural model, more parental costs and form more compact groups. The good parent hypothesis predicts that individuals with the most exaggerated traits invest more in offspring [Hoelzer 1989]. Moreover, due to differential allocation hypothesis [Burley 1986], individuals paired to attractive partners provide more care. However these hypotheses has some conflicts with other empirical studies [Maguire and Safran 2010]. In our experiments, we considered an equal parental cost for both partners.

- A high genetic complexity does not necessarily imply a higher species ratio, but maybe genetic diversity has an effect on it.

- A restricted mate choice can slow down the evolutionary process.

- Contrary to our expectation, the increase of genetic distance between parents did not increase the genetic diversity of the population. However, the compatibility hypothesis implies that genetic dissimilarity between mates increases the genetic diversity [Mays and Hill 2004]. For instance, Ortego et al. [2008] observed that mating with genetically less-related partners in *Lesser Kestrels* avoid the detrimental consequences of reduced

genetic diversity. This phenomenon is not well understood in our simulation and it needs further analysis.

- The non-random mating policy can increase the average genetic distance of partners.

- Mating with a strong partner has a positive effect on the fitness of the offspring. It has been shown that "a strong mate preference can evolve if it improves the total fitness of the female through direct phenotypic benefits" [Andersson and Iwasa 1996].

In the second experiment, machine learning techniques were applied to find the proper rules for distinguishing between these three models based on selected features. The accuracy of the initial classifications on validation set was not good enough. Nevertheless, a sub set of features were extracted which had an acceptable overall accuracy on validation set. These six features included Sedentary-Ratio, Age, Genetic-Complexity, Distance-Evolution, Distance-Mating and Quantity-LocalFood. Based on this set, nine rules were extracted that could distinguish between mate choice models. The tree is a combination of genetic, perception and action features. However, the genetic features formed the main roots of tree.

Three extensions can improve the mate choice simulation in EcoSim: 1) adding sex to the individuals, 2) using genetic linkage between female preferences and male traits and 3) applying viability function to males based on conspicuous traits for escaping from predators. These modifications can make simulation more realistic. Moreover, different sexual selection hypotheses can be investigated by them.

5.2 The Prediction of Species Extinction

Different factors affect extinction of a species. These factures can be categorized into three areas of demographic, environmental and genetic. Two experiments were considered: using all features together and using several features' subsets. Different machine learning techniques were applied to find the most significant features for the extinction process, and the most accurate rules, which can be used to predict this phenomenon. In general, the overall accuracy of validation set (higher than 93%) for species extinction revealed that selected features are very good for predicting this biological phenomenon in EcoSim. However, we observed that varying the parameters of these techniques, such as the amount of pruning samples, can change the accuracy and the obtained rules.

Some of the features had dependencies together, and the Bayesian-Networks illustrated the relation between them. Moreover, we showed when a feature is removed, a number of other features could take its place and compensate its effect on the prediction of extinction. These dependencies could also exist in real ecosystems which are of course an even more complex system.

The experiments showed that 16 features have more impact on the species extinction and for many of them it confirms some observations in real ecosystems. These features include: Individual-Ratio [Lande et al. 2003], [Rai 2003], [Drake et al. 2011], Reproduce-Ratio [Ellison 1994], ReproduceFail-Ratio, Explore-Ratio, Age [Pearson 1995], [Doran et al. 2006], and [Finnegan et al. 2008], Energy [Evans et al. 2005], Death-Age, Death-Energy, Killed-Ratio, Parent1-MatingAge, Parent2-MatingAge, Parental-Investment, Distance-Mating, Interbreeding-Ratio [Rhymer and Simberloff 1996], [Wolf

et al. 2001] (although in our results having a hybridization avoids going extinct),Volume-Ratio and Surface-Area [Collins et al. 2009], [Hill et al. 1996]. Table 23 shows the main rules achieved by each experiment. One of the main rules that appeared in some feature sets is: a species with the low Reproduction-Ratio will become extinct in the near future.

Table 23: The major rules of different experiments for species extinction that lead to extinction

Features Set	IF
All, Selected, Demographic, Mating	Reproduce-Ratio is low
Age-Energy (1)	Death-Energy is low
Age-Energy (2)	Age is too old
	Age is not too old AND Death-Age is too young
Spatial	Volume-Ratio is very high
	Volume-Ratio is not very high AND Surface-Area is very high
	Volume-Ratio is very low AND Surface-Area is high
Genetic	Distance-Mating is very low
Multi-category	Parental-Investment is very high
	Parental-Investment is not very high AND Interbreeding-Ratio is very low AND Volume-Ratio is not low

When all features were used, only demographic factors, Reproduce-Ratio, ReproduceFail-Ratio and Explore-Ratio, appeared in rules. However, in selected features, Parental-Investment (a genetic factor) was added to the set of effective rules. The best balanced trees in which all rules could cover equally the training samples were related to the spatial and multi-category features. Among these two trees, the multi-category tree had less depth and rules that are more general. This tree is a combination of all three

factors: demographic, environmental and genetic with the features Interbreeding-Ratio, Volume-Ratio and Parental-Investment.

The individual ratio can be a good index to investigate the behaviour of species. Drake and Griffen [2010] have mentioned that the individual ratio can impact on the extinction due to lifespan of that species. Furthermore, the lifespan can be a measure to divide species into two groups of long and short lifespan' species. These groups can be analyzed separately to find major features that work on them, and to keep track the different effect of each factor on each group.

APPENDICES

APPENDIX A

Mate Choice Model's Characteristics
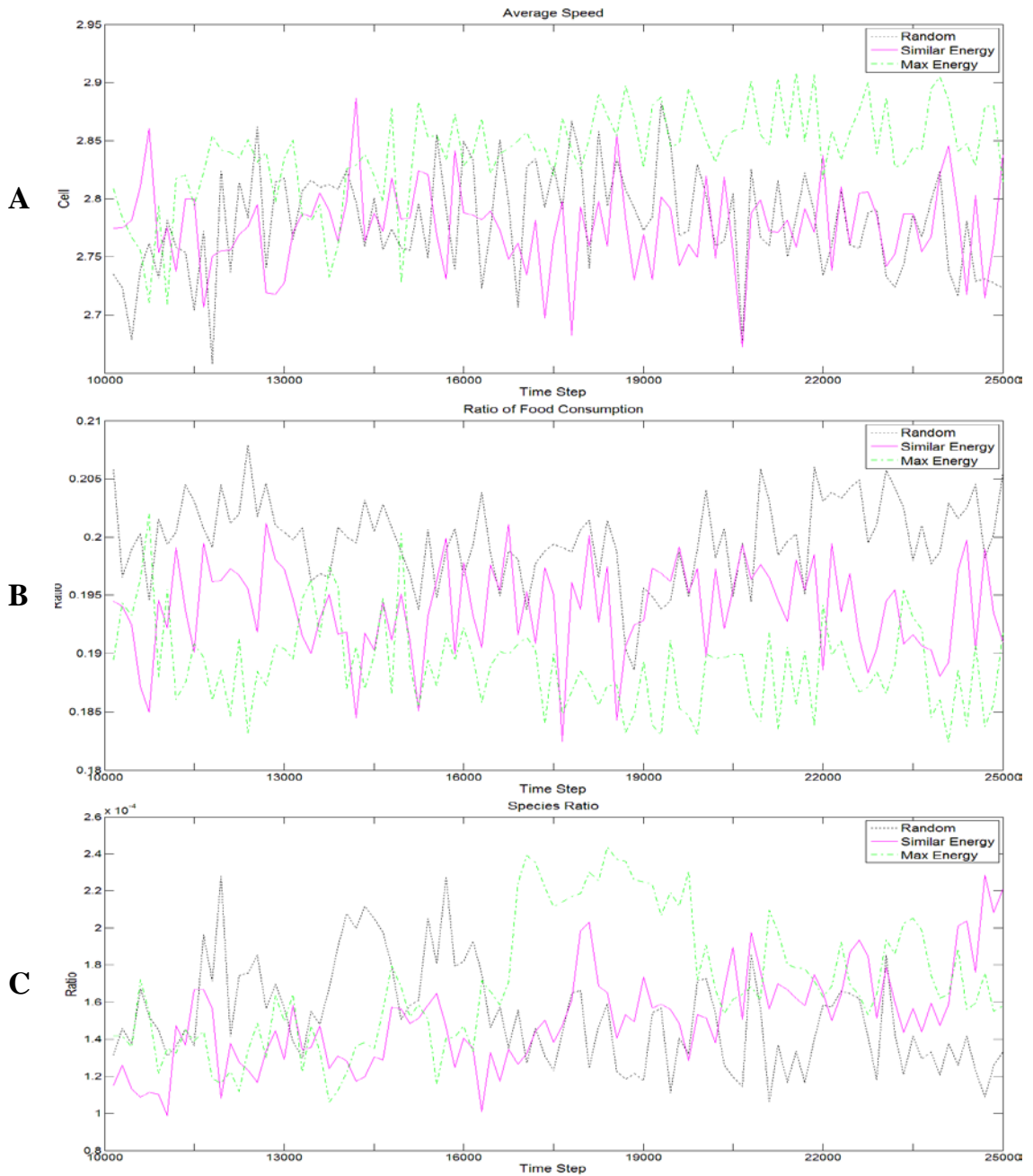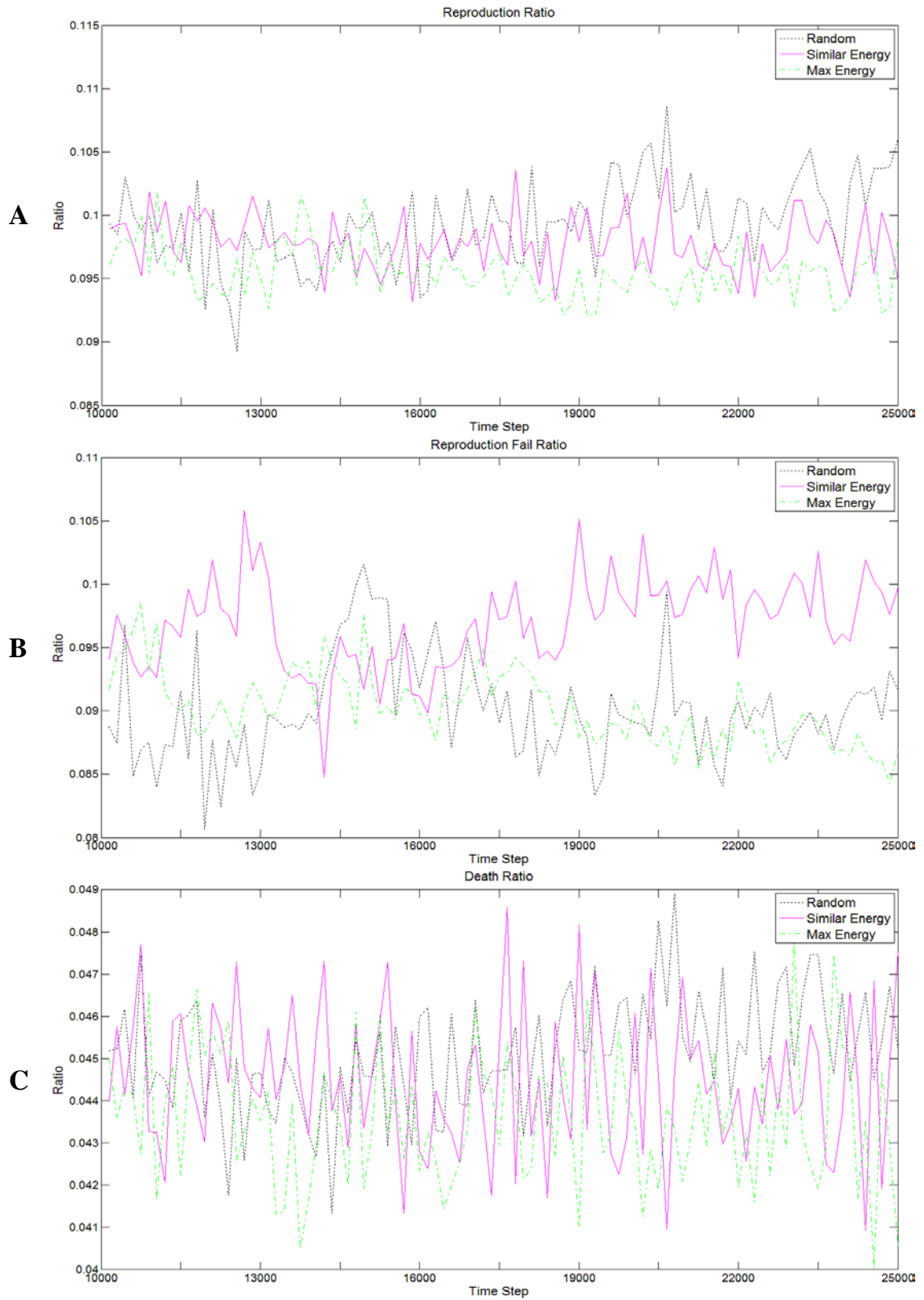

Figure 30: A) the average speed B) the ratio of food consumption C) the species ratio for the time steps 10000 to 25000

Figure 31: A) the reproduction ratio, B) the reproduction-fail ratio C) the death ratio for the time steps 10000 to 25000
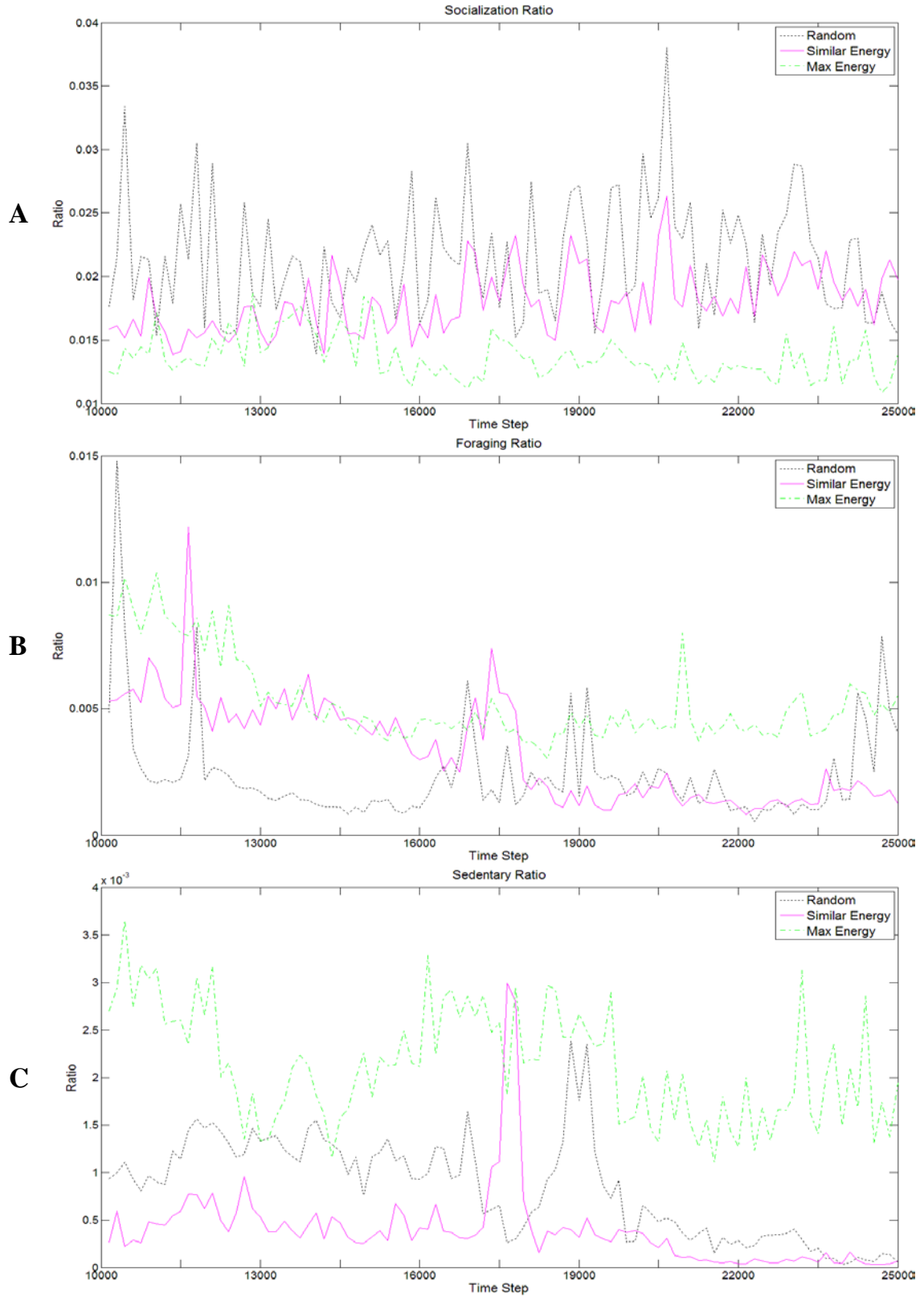
Figure 32: A) the socialization ratio, B) the foraging ratio C) the sedentary ratio for the time steps

10000 to 25000

REFERENCES

ADAMI, C., OFRIA, C., AND COLLIER, T.C. 2000. Evolution of biological complexity COMPUTER SCIENCES. *PNAS 2000*, 10.

ALTERS, S. 2000. *Biology: understanding life*. Jones & Bartlett Learning.

ANDERSSON, M. AND IWASA, Y. 1996. Sexual selection. *Trends in Ecology & Evolution 11*, 2, 53-58.

ANDERSSON, M. AND SIMMONS, L.W. 2006. Sexual selection and mate choice. *Trends in Ecology & Evolution 21*, 6, 296-302.

ASPINALL, A. AND GRAS, R. 2010. K-Means Clustering as a Speciation Mechanism within an Individual-Based Evolving Predator-Prey Ecosystem Simulation. *Active Media Technology LNCS6335*, 318-329.

BATESON, P.P.G. 1983. *Mate choice*. Cambridge Univ Pr.

BOS, I. AND CALIGARI, P.D.S. 1995. *Selection methods in plant breeding*. Springer.

BURLEY, N. 1986. Sexual selection for aesthetic traits in species with biparental care. *The American Naturalist 127*, 4, 415-445.

CLUTTON-BROCK, T. 2007. Sexual selection in males and females. *Science (New York, N.Y.) 318*, 5858, 1882-5.

COLLINS, C.D., HOLT, R.D., AND FOSTER, B.L. 2009. Patch size effects on plant species decline in an experimentally fragmented landscape. *Ecology 90*, 9, 2577-88.

DARWIN, C. 1859. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London: John Murra.

DEANGELIS, D.L. AND MOOIJ, W.M. 2005. Individual-Based Modeling of Ecological and Evolutionary Processes. *Annual Review of Ecology, Evolution, and Systematics 36*, 1, 147-168.

DEVAURS, D. AND GRAS, R. 2010. Species abundance patterns in an ecosystem simulation studied through Fisher's logseries. *Simulation Modelling Practice and Theory 18*, 1, 100-123.

VAN DIJK, D., SLOOT, P.M.A., TAY, J.C., AND SCHUT, M.C. 2010. Individual-based simulation of sexual selection: A quantitative genetic approach. *Procedia Computer Science 1*, 1, 2003-2011.

VAN DOORN, G.S. AND WEISSING, F.J. 2004. The evolution of female preferences for multiple indicators of quality. *The American Naturalist 164*, 2, 173-186.

DORAN, N.A., ARNOLD, A.J., PARKER, W.C., AND HUFFER, F.W. 2006. Is extinction age dependent? *Palaios 21*, 6, 571-579.

DRAKE, J.M. AND GRIFFEN, B.D. 2010. Early warning signals of extinction in deteriorating environments. *Nature 467*, 7314, 456-9.

DRAKE, J.M. AND LODGE, D.M. 2004. Effects of environmental variation on extinction and establishment. *Ecology Letters 7*, 1, 26-30.

DRAKE, J.M., SHAPIRO, J., AND GRIFFEN, B.D. 2011. Experimental demonstration of a two-phase population extinction hazard. *Journal of the Royal Society, Interface / the Royal Society 8*, 63, 1472-9.

ELLISON, P.T. 1994. Extinction and descent. *Human Nature 5*, 2, 155-165.

EVANS, K.L., GREENWOOD, J.J.D., AND GASTON, K.J. 2005. The roles of extinction and colonization in generating species – energy relationships. 498-507.

FAWCETT, T.W., KUIJPER, B., PEN, I., AND WEISSING, F.J. 2006. Should attractive males have more sons? *Behavioral Ecology 18*, 1, 71-80.

FAWCETT, T.W., KUIJPER, B., WEISSING, F.J., AND PEN, I. 2011. Sex-ratio control erodes sexual selection, revealing evolutionary feedback from adaptive plasticity. *Proceedings of the National Academy of Sciences 108*, 38, 15925-15930.

FINNEGAN, S., PAYNE, J.L., AND WANG, S.C. 2008. The Red Queen revisited reevaluating the age selectivity of Phanerozoic marine genus extinctions. *The Paleontological Society 34*, 3, 318-341.

FISHER, R.A. 1915. The evolution of sexual preference. *The Eugenics Review 7*, 3, 184.

FOROUTAN-POUR, K., DUTILLEUL, P., AND SMITH, D.. 1999. Advances in the implementation of the box-counting method of fractal dimension estimation. *Applied Mathematics and Computation 105*, 2-3, 195-210.

GOLESTANI, A. AND GRAS, R. 2010. Regularity analysis of an individual-based ecosystem simulation. *Chaos (Woodbury, N.Y.) 20*, 4, 043120.

GOLESTANI, A. AND GRAS, R. 2011. Multifractal Phenomena in EcoSim, a large scale Individual-Based Ecosystem Simulation. *ICAI*, 991-999.

GOLESTANI, A., GRAS, R., AND CRISTESCU, M. 2012. Speciation with gene flow in a heterogeneous virtual world: can physical obstacles accelerate speciation? *Proceedings of the Royal Society B: Biological Sciences* April.

GRAS, R., DEVAURS, D., WOZNIAK, A., AND ASPINALL, A. 2009. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life 15*, 4, 423-63.

GRIFFEN, B.D. AND DRAKE, J.M. 2008. A review of extinction in experimental populations. *The Journal of animal ecology 77*, 6, 1274-87.

HALLIDAY, T.R. 1983. The study of mate choice. *Mate choice 1*, 462.

HILL, J.K., THOMAS, C.D., AND LEWIS, O.T. 1996. Effects of habitat patch size and isolation on dispersal by Hesperia comma butterflies: implications for metapopulation structure. *Journal of animal ecology*, 725-735.

HOELZER, G.A. 1989. The good parent process of sexual selection. *Animal Behaviour 38*, 6, 1067-1078.

HOVEL, K. A. AND REGAN, H.M. 2007. Using an individual-based model to examine the roles of habitat fragmentation and behavior on predator–prey relationships in seagrass landscapes. *Landscape Ecology 23*, Sep1, 75-89.

JAMMALAMADAKA, S.R. AND SENGUPTA, A. 2001. *Topics in circular statistics*. World Scientific Pub Co Inc.

KHATER, M., SALEHI, E., AND GRAS, R. 2011. Correlation between Genetic Diversity and Fitness in a Predator-Prey Ecosystem Simulation. *AI 2011: Advances in Artificial Intelligence*, 422-431.

KOKKO, H., JENNIONS, M.D., AND HOUDE, A. 2007. Evolution of frequency-dependent mate choice: keeping up with fashion trends. *Proceedings. Biological Sciences / The Royal Society 274*, 1615, 1317-1324.

KOSKO, B. 1986. Fuzzy cognitive maps. *International Journal of Man-Machine Studies 24*, 1, 65-75.

LANCASTER, L.T., HIPSLEY, C. A., AND SINERVO, B. 2009. Female choice for optimal combinations of multiple male display traits increases offspring survival. *Behavioral Ecology 20*, 5, 993-999.

LANDE, R., ENGEN, S., AND SÆTHER, B.E. 2003. *Stochastic population dynamics in ecology and conservation*. Oxford University Press, USA.

LETCHER, B.H., PRIDDY, J.A., WALTERS, J.R., AND CROWDER, L.B. 1998. An individual-based , spatially-explicit simulation model of the population dynamics of the endangered red-cockaded woodpecker , Picoides borealis. *Biological Conservation 86*, 1-14.

LI, H. AND REYNOLDS, J.F. 1993. A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology 8*, 3, 155-162.

LORCH, P.D., PROULX, S., ROWE, L., AND DAY, T. 2003. Condition-dependent sexual selection can accelerate adaptation. *Evolutionary Ecology Research 5*, 6, 867-881.

MAGUIRE, S.E. AND SAFRAN, R.J. 2010. Morphological and genetic predictors of parental care in the North American barn swallow Hirundo rustica erythrogaster. *Journal of Avian Biology 41*, 1, 74-82.

MAJDABADI FARAHANI, Y., KHATER, M., AND GRAS, R. EcoDemics: Modeling Epidemic Spread in a Simulated Predator-Prey Evolutionary Ecosystem. *Journal of Artificial Life (submitted)*.

MALLET, J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution 10*, 7, 294-299.

MAYS, H.L. AND HILL, G.E. 2004. Choosing mates: good genes versus genes that are a good fit. *Trends in ecology & evolution 19*, 10, 554-9.

MD IBNE, SINA; GRAS, R. 2011. Computation of Population Spatial Distribution in Individual-Based Ecosystem Simulation. *IEEE ALIFE*, 1-5.

ORTEGO, J., CALABUIG, G., APARICIO, J.M., AND CORDERO, P.J. 2008. Genetic consequences of natal dispersal in the colonial lesser kestrel. *Molecular Ecology 17*, 8, 2051-9.

OVASKAINEN, O. AND MEERSON, B. 2010. Stochastic models of population extinction. *Trends in ecology & evolution 25*, 11, 643-652.

PARROTT, L., PROULX, R., AND THIBERTPLANTE, X. 2008. Three-dimensional metrics for the analysis of spatiotemporal data in ecology. *Ecological Informatics 3*, 6, 343-353.

PATTEN, M.A., WOLFE, D.H., SHOCHAT, E., AND SHERROD, S.K. 2007. Habitat fragmentation, rapid evolution and population persistence. *Evolutionary Ecology 7*, 235-249.

PEARSON, P.N. 1995. Investigating age-dependency of species extinction rates using dynamic survivorship analysis. *Historical Biology 10*, 2, 119-136.

RAI, U.K. 2003. Minimum Sizes for Viable Population and Conservation Biology. 3-9.

REED, D.H., LOWE, E.H., BRISCOE, D.A., AND FRANKHAM, R. 2003. Inbreeding and extinction : Effects of rate of inbreeding. *Conservation Genetics 4*, 3, 405-410.

RHYMER, J.M. AND SIMBERLOFF, D. 1996. Extinction By Hybridization and Introgression. *Annual Review of Ecology and Systematics 27*, 1, 83-109.

ROBINSON, D.M., TUDOR, M.S., AND MORRIS, M.R. 2011. Female preference and the evolution of an exaggerated male ornament: the shape of the preference function matters. *Animal Behaviour 81*, 5, 1015-1021.

SHERWIN, W.B. 2010. Entropy and Information Approaches to Genetic Diversity and its Expression: Genomic Geography. *Entropy 12*, 7, 1765-1798.

THIBERT-PLANTE, X. AND HENDRY, A. P. 2009. Five questions on ecological speciation addressed with individual-based simulations. *Journal of Evolutionary Biology 22*, 1, 109-123.

UCHMAN, J. 2000. Individual variability and population regulation: an individual -based model. *Oikos* 1987, 539-548.

WEKA. V3.6.4. *http://www.cs.waikato.ac.nz/ml/weka/*.

WALTERS, J.R., CROWDER, L.B., AND PRIDDY, J. A. 2002. Population Viability Analysis for Red-Cockaded Woodpeckers Using an Individual-Based Model. *Ecological Applications 12*, 1, 249-260.

WOLF, D.E., TAKEBAYASHI, N., AND RIESEBERG, L.H. 2001. Predicting the Risk of Extinction through Hybridization. *Conservation Biology 15*, 4, 1039-1053.

ZHANG, J., KHASAWNEH, M.T., AND BOWLING, S.R. 2010. Gender change in certain species - an agent-based modeling study. *2010 IEEE Systems and Information Engineering Design Symposium*, 225-228.

VITA AUCTORIS

NAME	Seyed Meisam Hosseini Sedehi
PLACE OF BIRTH	Tehran, Iran
YEAR OF BIRTH	1980
EDUCATION	School of Computer Science
	University of Windsor
	Windsor, Ontario, Canada
	2010 – 2012 M.Sc.
	Computer Engineering Department
	Amir Kabir University
	Tehran, Iran
	2005 – 2008 M.Eng.
	Computer Engineering Department
	Azad University (Tehran Branch)
	Tehran, Iran
	1999 – 2003 B.Eng