Università degli Studi di Pisa

Dipartimento di Informatica
Dottorato di Ricerca in Informatica

Ph.D. Thesis

The Emergence of Diversity and Stability: from Biological Systems to Machine Learning

Claudio Felicioli

Supervisor
Fabrizio Luccio

Abstract

The observation of emergent properties of biological systems has been the inspiration of successful technologies opening new fields of computer science like artificial neural nets, swarm intelligence algorithms, evolutive algorithms, etc. In this work we focus on the emergence of negative feedback cycles: self-regulatory mechanisms able to react to alterations of some environmental parameters (temperature, gas concentrations, solar light, etc.) in order to compensate, preserving the environment in a state suitable for life. We make the hypothesis that speciation events play a central role for feedback formation and, and in order to select the negative cycles, the arising species need to be strongly connected to the environment, therefore the speciation needs to be sympatric (a speciation mode where new species arise without geographical isolation). As an intermediate result, we propose a simulative model of sympatric speciation and apply it to the field of evolutive algorithms. We propose some variations of the standard island model, a model used in evolutive algorithms to evolve multiple populations, to obtain dynamics similar to the sympatric speciation model, enhancing the diversity and the stability of the evolutive system. Then we propose a technique to define a metric and calculate approximated distances on very complex genetic spaces (a recurring problem for several evolutionary algorithms approaches). Finally, we describe the more complex model of negative feedback cycles emergence and discuss the problems that, in the current model formulation, make it not applicable to real world problems but only to *ad hoc* defined resource spaces; conclusively we propose possible solutions and some applications.

Contents

Table 1: Table of symbols and variables - Sympatric Speciation Model

Symbol/Variable	Description
S	The geographical bi-dimensional space
l	The side of S
p	A point in S
$d_s(p_1, p_2)$	The geographical distance of two points in S
G	The genetic space
g	A point in G
$d_g(g_1, g_2)$	The genetical distance of two points in G
$f(g_i)$	The fitness function
R	The resource set
k_{res}	The density of resources
$e_{resource}$	The energy of a resource
A	The population of agents
a_i	An agent in A
SS_i	The static internal state of agent a_i
DS_i	The dynamic internal state of agent a_i
e_i	The energy of agent a_i
$range_{mating}$	The maximum distance between two mating agents
$range_{gathering}$	The maximum distance between a resource and an agent competing for it
$e_{metabolism}$	The energy level that every agent loses in the metabolism step of every iteration
e_{mating}	The minimum energy level that an agent needs in order to be able to mate
$step_{max}$	The maximum geographical distance of the position of an agent between two consecutive iterations
$e_{matingattempt}$	The energy lost by an agent attempting a mating
$e_{newborn}$	The starting energy level of a new agent
$maxmutation$	The maximum genetical distance between the genotype of a newborn agent and the mean parental genotype
$startingagents$	The number of agents in the initialization
$startinggenotype$	The genotype of the agents in the initialization
$startingenergy$	The energy of the agents in the initialization

Table 2: Table of symbols and variables - Simulacrum Genotype

Symbol/Variable	Description
l_{sg}	The size of the bit string
p_{sg}	The probability of a point mutation

Table 3: Table of symbols and variables - Modified Island Model

Symbol/Variable	Description
$n_{islands}$	The number of islands
A_i	The population of agents in the island i
pop_i	The size of A_i
m_{ij}	The agents migrating from the island i to the island j

Table 4: Table of symbols and variables - Feedback Emergence Model

Symbol/Variable	Description
$resTypes$	The number of resource types
$resEnergy_r$	The energy level of the resource type r
$R_i[r]$	The number of resources of type r in island i
d_{sr}	The rato of the spontaneous degradation sr
R_{input}	The resource with a constant incoming flux in the system
$radiation$	The number of the *input* resource added to each island in each iteration
$islandV$	The volume of the islands
$agentV_i$	The volume of the internal environment of agent i
t_{enz}	The transformation rate of the enzyme enz
ee_{enz}	The energy efficency of the enzyme enz
$cell_i$	A cell in the grid-like resource space
$maxN$	The maximum length of a spontaneous degradation trajectory

Chapter 1

Introduction

The observation of emergent properties of biological systems has been the inspiration of successful technologies [4] [63] [68] opening new fields of computer science like artificial neural nets, swarm intelligence algorithms, evolutionary algorithms, etc.

A very intriguing mechanism that we can observe in almost every complex biological system where different components (biotic and not biotic) interact are the negative feedbacks: they are self-regulatory mechanisms able to react to alterations of some environmental parameters (temperature [82] [96], gas concentrations [50] [51] [78] [91] [57], rainfall and solar light [11] [64], sea salinity [35], growing or shrinking of a species population size, etc.) in order to counterbalance the alteration, preserving the environment in a dynamical homeostatic equilibrium suitable for life.

More fascinating is the spontaneous emergence of negative feedback and the suppression of potentially catastrophic positive feedback, by the only means of a simple mechanism like natural selection.

In order to formulate an hypothesis about negative feedbacks emergence and then be able to reproduce and exploit them in artificial systems of computer science interest, several mechanisms and intermediate problems need to be considered.

The first mechanism is the sympatric speciation: a speciation event where new species arise without the need of geographical isolation. As an intermediate result we present a simulative model of sympatric speciation that, at our knowledge, differs from the models available in the literature.

Given the sympatric speciation model, we propose an application in the field of evolutionary algorithms; we describe some variations of the standard island model (a speciation model widely used in evolutionary algorithms to evolve multiple populations) obtaining useful emergent properties observed in the sympatric speciation model (enhanced biodiversity and stability of multi-species systems).

Working with evolutionary algorithms, we face the recurring problem of defining a metric and calculating distance for very complex search spaces [12] (like the space of computer programs source code). We propose a solution inspired by biological observations of correlation between genetic variability and reproduction compatibility. Adopting the proposed technique we can calculate approximated distances independently of the complexity of the genetic space.

In conclusion we describe a model of negative-feedback emergence and we explain the complications that, in the current model formulation, make it not applicable to real world problems; we propose possible solutions and some applications.

1.1 Structure of the Work

In the following sections of this chapter, first we expose the aims of the thesis, then we introduce some concepts of theoretical biology (the Gaia theory, the environmental negative feedback cycles, and some related problems like theoretical lacks and teleological accusations), then we introduce the agent-based approach to modeling, in the last section we introduce the evolutionary algorithms approach to machine learning, focalizing to the multi-modal problems.

In chapter 2 we cite previous works on topics related to the thesis.

In chapter 3 we give some definition for several terms we utilize in the thesis, underlining correlations between biological elements and computer simulation model elements.

In chapter 4 we propose a model for the sympatric speciation mode. Starting with some biological observations, we describe in the details the proposed simulative agent-based model and illustrate an example of execution of an implementation of the model. Then we discuss the features of the model and the observed emergent properties. After an analysis of the role of every parameter of the model, we finally discuss about the minimality of the model and the mechanisms tested in the model development phase but removed from the final version.

In chapter 5 we propose some contributions to the field of evolutionary algorithms. The first proposal (section 5.1) is a method for genetic distance measurements, not dependent to the search space complexity The second proposal (section 5.2) is a modified island model, suitable for multi-modal optimization problems: we expose the addressed problems (the diversity maintaining and the stability of the evolving system), then we report the experimental observations for the several different tested model variations, focalizing on the features of locality of selection and of variable fertility.

The third proposal (section 5.3) is a method, specific for the island model variation, that exploits the two levels of selection (local level and global level) in order to destabilize some locations of the system avoiding a whole system destabilization, obtaining an enhancement of the diversity emergence. We discuss the addressed problem, then we describe the method and illustrate an example of a run.

In chapter 6 we propose a model for feedback emergence. We start exposing our emergence hypothesis, then the proposed agent-based simulative model. In the model discussion we address several problems faced and the model minimality. In the observations we expose the experimental results and the problem of feedback emergence when we make use of random generated system definitions (in opposition to *ad hoc* system definitions).

In the conclusions (chapter 7) we give a short summary of the various proposed contributions described in the thesis.

1.2 Aims of the Work

Emergent properties observed in the natural environment have inspired the development of very successful technologies in several computer science fields [4] [63] [68]; for example: the neurophysiology studies of biological neural systems inspired the artificial neural networks and the connectionist approach to machine learning, the social behaviours of ants colonies and bird swarms inspired the Swarm intelligence systems, the darwinian evolutionary paradigm of random mutation followed by natural selection inspired the development of evolutionary approach algorithms.

In the case of the evolutionary paradigm, computer science has developed several techniques proved to be human competitive in several fields of technological development [46] [2] [60] [47] [88] [89] [5] [87] [48] [24] [54] [71].

The evolution mode of a population of organisms that, by the means of random mutation and selection, change itself to fit better the environment, is known as micro-evolution.

With the progress in the studies of the livings systems, the understanding of micro-evolution dynamics consolidated and the biological debate shifted from micro-evolution to environmental system level dynamics and macro-evolution related phenomena: the speciation events, the biotic/non biotic interactions, the alterations of the homeostatic environmental equilibrium caused by the metabolism of the living beings, the promotion or inhibition of environmental feedback cycles, the stabilizing effect of negative environmental feedback cycles, that preserve the environmental parameters in a range suitable for the living beings.

In literature several publications presented models based on macro-evolution (Tierra [76], Avida [69], and works focused on the compatibility between neo-darwinist paradigm and Gaia theory [17] [18] [19]).

A very intriguing and debated phenomenon is the emergence, in a system with strong interdependency between species and non biotic environment components, of altruistic mechanisms like the environmental regulation, evolved thanks to the typical selfish mechanism of natural selection.

The aim of this work is to study these system level emergent properties in order to model and exploit them in computer science fields.

We want to stress that the emergent properties analysis, the natural environment observations, hypothesis formulation and testing about the mechanisms involved in the emergence, have the only purpose of finding plausible hypothesis about the emergence dynamics: it's enough that we are able to successfully reproduce the emergence of the target property in a simulative model; in this work we don't try to verify the validity of our hypothesis in the natural environment, but only in the artificial models.

Our method consists in the developing of a simulation model where a target property is permitted to emerge, then we try to minimize the elements and parameters of the model, detecting and removing elements that turn out to be irrelevant or that are redundant, then, after the model minimization, we search possible applications in computer science.

The two main phenomena we consider in this work are the sympatric speciation and the negative feedback formation. The proposed applications are all in the evolutionary algorithms branch of machine learning.

1.3 Biological Systems

1.3.1 Gaia Ecological Paradigm

In the last century there has been a change in the interpretation of evolutionary dynamics: the paradigm of a static environment, seen as the stable and passive substrate that implicitly directs the evolution of life forms to better adapt to it, changed to a modern paradigm where the species not only are adapting to the environment, but their activities alter the environmental homeostatic equilibrium.

When the environmental homeostatic equilibrium is altered by the activities of the living beings that populate it, some new ecological niches can be created, thus new species can speciate from the existing ones in order to exploit the new formed niches; this environment mutation mechanism eventually keeps the evolvability of the system to a high level.

Not only every alteration can create new niches, but it can eventually (and almost surely in a system where components have developed a strong level interdependency, like the natural environment) alter the existing ones. In the worst cases an equilibrium alteration can destroy some niches, or it could create and activate a positive feedback cycle that would destroy the whole system's equilibrium, causing a mass extinction.

For several events of mass extinction happened in the past Earth's ages there are strong clues of a reason internal to the ecological system (in opposition to events caused by external events, like volcano eruptions, meteor impacts, etc.). For example, the extinction caused by the apparition of oxygen in Earth's atmosphere, starting from the Paleoproterozoic era: the oxygen was produced by photosynthetic archaea and bacteria and it was accumulated in the atmosphere reaching, about 2.4 billion years ago, a concentration that drove to extinction most of the anaerobic organisms (almost all the currently living organisms). The homeostatic chemical equilibrium was altered by the oxygen biotic production and caused mass extinction because it destroyed almost all of the ecological niches for anaerobic life; on the other hand, every alteration creates new niches, and as a matter of fact today the oxygen plays a main role for both photosynthesis and respiration.

The complexity of a biological system can be measured from the biodiversity level (the number of different species composing it), from the number of distinct inter-specific interactions and from the number of biotic-abiotic interactions (we need to consider non biotic components like the soil, the minerals, the oceans, the gases, etc. as reactive and essential components of the biological system).

The current biological system, the biosphere, has an incredibly high level of complexity and we could guess that a system with so many (apparently) chaotic interactions must be unstable; but instead of being so, in the natural environment we are able to observe a strong active resistance to almost any destabilization, stronger where complexity (biodiversity and interactions) is higher, thanks to the actions of lots of complex spontaneously evolved auto-conservative mechanisms.

The biosphere now is not seen anymore as a stable and passive substrate for life evolution, but as an energy and matter fluxes system in a (strongly) dynamical equilibrium state, actively maintained by the biotic-abiotic components interactions. This view is the one described by the Gaia theory (it be introduced in the next introduction section).

As an example of the role and the strength of the equilibrium maintaining activity of living beings, we know that the current Earth's chemical state is really far from the chemical equilibrium: without the action of living beings a lot of chemicals (like, as an example, the oxygen) quickly disappear from the atmosphere (in Figure 1.1 the surface fluxes of some gases measured on Earth are compared to the estimated fluxes in absence of the biological system).

The self-regulatory, equilibrium maintaining, mechanisms are unselfish (whole system level useful but potentially local level resource consuming) indirect behaviours of the living beings that have evolved thanks to the typically selfish dynamic of the natural selection.

The neo-darwinist point of view is in opposition to the fact that the evolution of a quality that is useful to the global system can be imputable to the exclusive effect of a selfish mechanism like the selection for resource competition. They state that a theory that expects the spontaneous emergence of some negative feedback cycles makes an implicit use of concepts like the group selection or teleological principles.

Another opposition to the self-regulatory biotic-abiotic systems is the not optimal choice of biotic active alteration and the (eventually resource consuming) stabilization of the environment, when the optimal selfish choice for every species is to simply adapt to the current environment.

All these critics made in opposition to the Gaia theory have been addressed by and completely or partially settled in several works.

1.3.2 Gaia Theory and Negative Feedbacks Cycles

Today there is abundance of well documented empirical evidences about several negative feedback cycles regulating critical parameters like the temperature [82] [96], the frequency of precipitations [11] [64], the sea salinity [35], the concentrations

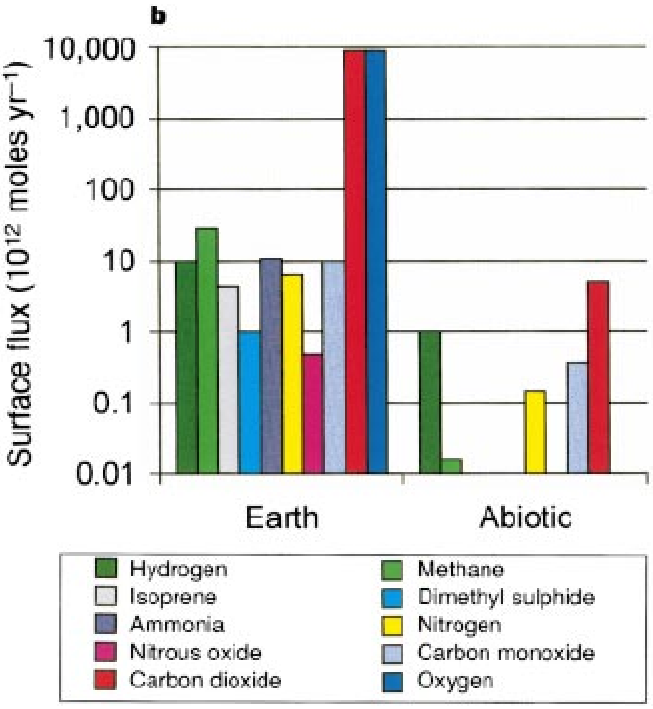


Figure 1.1: Comparison between the current surface fluxes of some gases (on the left) and the same fluxes in the chemical equilibrium estimated in absence of the biological system (on the right). This graph is presented in [51].

of gases in the sea [50] [51] [78] [91] and in the atmosphere [57], the number of individuals of some species, etc. The effect of these negative feedback cycles is the maintaining of the stabilized parameter in some boundaries life-permitting ranges.

The Gaia theory [55] [56] [57] [58] [59] predicts, and it's the first to analyzing in details, the negative feedbacks.

Initially the Gaia theory was proposed as an hypothesis; the informality of the hypothesis and several other lacks caused it to be embraced by some new-age movements, but to be rejected by the scientific community. The complains for teleological elements were dropped in 1988, after a conference at the American Geophysical Union, where the hypothesis was proposed again, reformulated as the Gaia theory [55]. Only in the last years the Gaia theory was accepted by the scientific community, but several disagreement issues are still alive [39]. Today there are a lot of academic publications based on Gaia theory [31].

The Gaia theory still has some lacks: even if it is based on spontaneously emergent auto-regulation systems, a model describing the emergence through classical consolidated evolutionary dynamics of such systems is still missing.

Prosecuting the main objective of proving the plausibility of the existence of unselfish system level self-regulatory mechanisms without any need of teleological implications [39] [40], keeping aside the consistency with the real ecological system, several researches have formulated some simulation models exhibiting dynamics expected by the Gaia theory [17] [18] [19] [91].

One of the first simulation model developed, the most simple and famous, is Daisy-world: it shows how a multi-species self-regulatory mechanism can work thanks only to the selfish selection.

In [51] Timothy M. Lenton described the standard daisy-world formulation:

> "Daisy-world is an imaginary grey world orbiting, at a similar distance to the Earth, a star, like our Sun, which gets warmer with time. The world is seeded with two types of life, black and white daisies. These share the same optimum temperature for growth, 22.5°C, and limits to growth of 5°C and 40°C. Initial conditions on the planet are so cold that daisy seeds cannot germinate. As solar forcing increases and the temperature reaches 5°C, the first seeds germinate. The paleness of the white daisies means that they are cooler than their surroundings, hindering their own growth. The black daisies, in contrast, warm their surroundings, enhancing their growth and reproduction. Hence black daisies come to dominate the initial community. As they spread, the black daisies begin to warm the planet. This increases the growth rate of all daisies, an environmental positive feedback that reinforces the spread of life. As the warmer, darker daisies are closer to the optimum temperature than the white daisies, they remain dominant. Soon the limited area of planet surface constrains the explosion of life. When daisies fill the world, the average temperature has risen close to the optimum for daisy growth. As the sun warms, the

temperature rises to the point at which white daisies begin to appear in the daisy community. As it warms further the white daisies gain the selective advantage over the black daisies and gradually take over. Eventually, only white daisies are left, and when the solar forcing gets too high, self-regulation collapses. The self-regulation of Daisy-world is impressive: although the solar input changes over a range equivalent to 45°C the surface of the planet is maintained within a few degrees of the optimum temperature for daisy growth." [51]

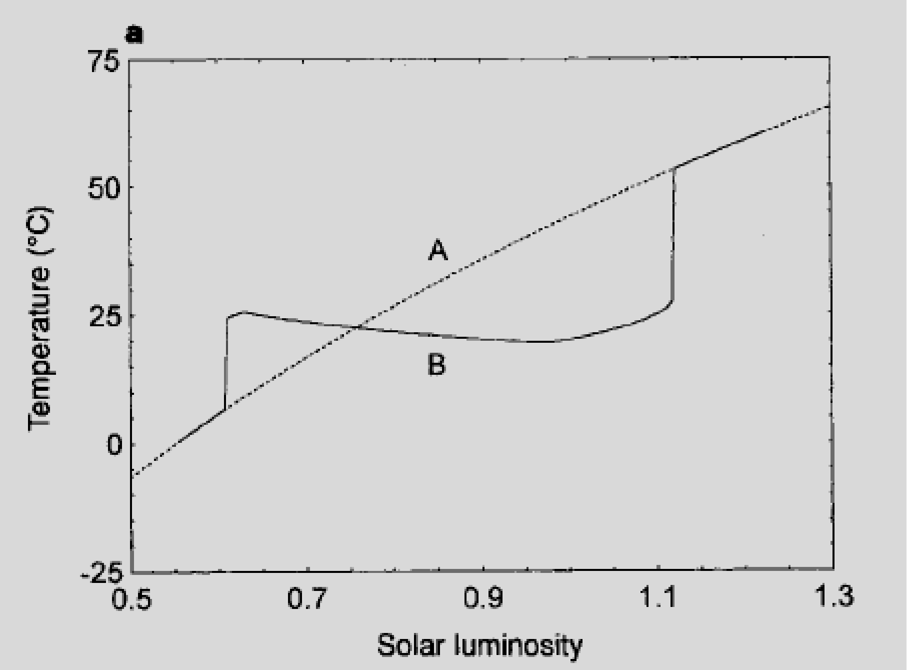


Figure 1.2: Temperature of the surface of the world at the variation of solar luminosity. The solid line represents daisy-world's temperature variation, the dotted line is the temperature of a control planet without flowers. This graph is presented in [51].

1.3.3 Feedback Cycles Formation and Inhibition

There are a lot of daisy-world model variations, but a model able to show not only that the self-regulation is possible, but also the dynamics of negative feedback cycles formation, starting from the speciation events, is still missing (and this lack is keeping alive a debate about the teleology of group selection).

A further interesting and essential feature of the biosphere is the fact that, while negative feedback cycles can emerge, on the other hand there is an inhibition (but the

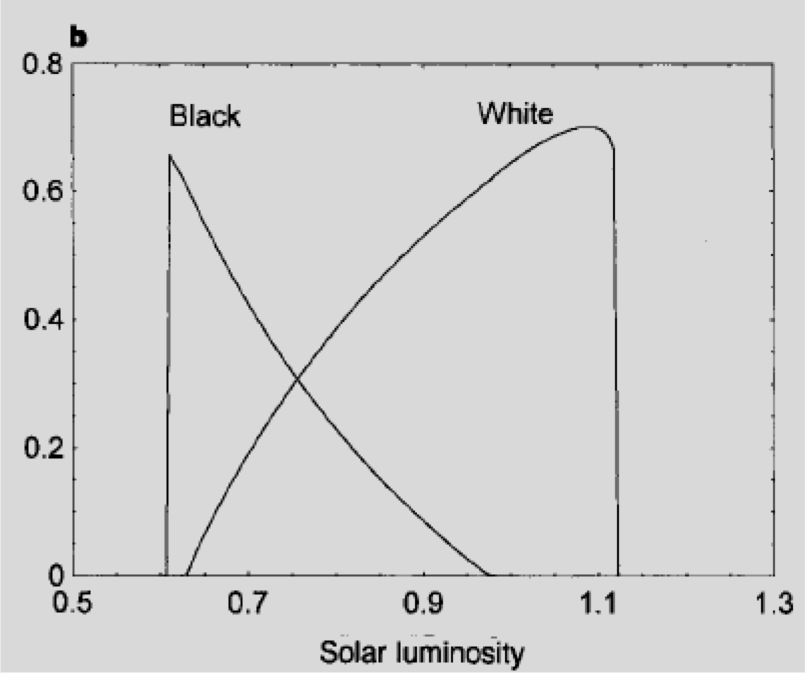


Figure 1.3: Distribution of the black daisies and white daisies populations on the surface of the world at the variation of solar luminosity. This graph is presented in [51].

geological history of the Earth shows that there can be some faults to the inhibition mechanism) of positive feedback cycle formation.

It is an obvious fact that, in the natural environment, the only possible existing positive feedback cycles are the ones quantitatively less influential than the negative feedback cycles that regulate the same parameters. This was true, at least, before the industrial modern era, and geological and fossil studies show that the inhibition system has made several faults in past ages, usually followed by mass extinctions.

In order to develop a model with emergence of feedback cycles, we need to include a model of speciation, as the species-species and species-environment interactions are the base elements that constitutes any observed feedback cycle.

An hypothesis we can formulate is that there is a selective pressure acting against or in favour of each speciation event, depending on the effect of the new capabilities on the existing system at two different levels: a local effect and the corresponding local selection, depending on simple competition with the other living beings, and a second non-local effect and the corresponding global selection, depending on perturbation of the global homeostatic equilibrium and on the eventuality of the formation of feedbacks cycles, positive or negative.

Obviously, in order to enable the hypothesis of a selection from the global level dynamics of the system (global selection) against or in favour the speciation phenomena, we need to assume that the speciation model through which the species arise is able to work from within the system, not from an environmentally isolated area (an island). This type of speciation is the sympatric mode speciation.

1.4 Agent Models

There are a lot of different interpretations and definitions for the term *agent*, because there are a lot of possible uses for the agents, that need different sets of features. A quite general definition that can describe the essential characterization of almost every agent implementation can be:

> An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives.

The typical agent approach model is composed by a population of agents, each of them able to act independently, and a geographical environment, where the agent lives and can move around.

The geographical environment can be an euclidean space with a given number of dimension (a typical setting is the bi-dimensional real space $\mathbb{R}^2$), a grid, a graph, a single point, etc.

The most important differences between the various approaches to the agents are about the type of individual capabilities of an agent, in terms of perception (of other agents and/or the environment), internal state complexity, capability of computation, environment alteration capabilities, communication with other agents, stochasticity of its actions, locomotion, capability to dynamically create new agents (usually by a fork-like procedure).

Another choice to make in every agent-based approach is the synchronization: every agent can act when it wants to act, randomly timed, continuously in the time, or all the agents act in synchronization, in a serial or a parallel way.

The agent model adopted in this works is the following:

- The agents are placed in a bi-dimensional limited finite geographical space in $\mathbb{R}^2$. In some models described in the following chapters the geographical space can be a graph of connected bi-dimensional limited finite surfaces of $\mathbb{R}^2$ (islands). In some model the islands can be only single point-shaped.

- The agents are synchronized (in some stages serial, in other parallel) and we call every action step an *iteration*.

- Each agent can interact with other agents within a fixed range of locality in the geographical environment. The possible interactions are limited to competition for resources and sexual reproduction.

- Each agent can percept the geographical environment and other agents within a fixed range of locality in the geographical environment. The perception of the environment is limited to an unordered list of reachable resources (described in the following chapters). The perception of the other agents is limited to a unordered list of agents enabled to mate and reachable. The two ranges (resource perception range and potential partners range) can be different.

- Each agent can alter the environment within a fixed range of locality in the geographical environment. The environment alteration is limited to resource collecting, removing them from the geographical environment. Only in the most advanced models the agents are able to put resources (the resources they have previously collected or some different resources they can produce by metabolizing of absorbed resources) in the environment.

- The agents have no control of their movement, that is stochastic Brownian-like diffusive motion. In the models where the agents are placed in a graph of connected islands, the Brownian-like motion is applied to the surface of the island where the moving agent is placed, and if the agent reaches the island geographical limits it will migrate to a connected island.

- Each agent has an internal state, divided in a read-only and in a read-write parts.

- The computations of an agent can alter only his internal writable state.

- The agents can execute stochastic computations.

In addition to these quite standard settings, for the particular needs of our work, the agents have two additional particular features (they are not new but quite common for agent-based approaches applied to the field of Evolutionary Algorithms):

- The agents can die, being removed from the agent population.

- The agents can create new agents to add to the current population in two possible ways (dependently on the particular model). The first one is the asexual reproduction: a new agent is added to the simulation placed in a point of the geographical space near to the position of the father; the new agent read-only internal state is a mutated copy of the read-only internal state of its parent. The second possible reproduction method is the sexual reproduction: it needs that two agents are mutually in the mating range and that both want to reproduce, a new agent is added to the simulation placed in a point of the geographical space, near to the halfway point of its parents geographical positions, the new agent read-only internal state is a mutated copy of a combination of the read-only internal states of its parents.

For the only purpose of simplicity of some system dynamics descriptions, we sometime quit from the agent paradigm and we describe some computations of the model, attributable to individual actions or synchronized interactions of groups of agents, that will alter the agent population or some agents internal states, as done by a non-agent mechanism. We want to assure that, independently of the paradigm used to describe some agent state alteration mechanism, we can explain it in a less simple way, but from within the agent paradigm.

1.5 Evolutionary Algorithms

The Evolutionary Algorithms [4] [45] are probabilistic search algorithms that maintain a population of candidates in the search space and evolve such a population by the application of reproduction, mutation and selection.

An iteration of the typical evolutionary algorithm consists of:

- Selection of some representative candidates in the current candidates population, through a selection method and a fitness function.
- Generation of a new candidates population from the selected representative, through a reproduction method.
- Random mutation of the new candidates population, through a mutation method.
- Replacement of the old candidates population.

Usually the population size is constant.

The fitness function has a main role in the standard formulation of an evolutionary algorithm: it's defined over the problem space and it's applied to a point in the search space in order to numerically evaluate its goodness as a potential solution to the target problem.

Usually the selection method needs the fitness values of all the search points of the current generation, or the rank order (based on the fitness values) of search points, to be able to select among them the population representatives.

The usual reproduction methods are cloning, sexual reproduction through a recombination method, or a combination of asexual and sexual reproductions methods.

There are a lot of different selection methods that make use of the fitness values, some of them are Linear Ranking Selection, Exponential Ranking Selection, Truncation Selection, Fitness Proportional Selection and Tournament Selection.

We refer to the previously defined algorithm class, with the only reproduction method of sexual reproduction, as the classic generalized Evolutionary Algorithm model.

1.5.1 Diversity in Evolutionary Algorithms

In the evolutionary algorithms literature, the term diversity (or genetic diversity) is often used without any definition, but usually it is referred as the search space points (the population) diversity, opposed to the fitness values diversity (sometimes referred as phenotypical diversity).

With an opportune fitness function choice, if we plot the search points in the search space, after some classic generalized Evolutionary Algorithm iterations, usually we can see the formation of a cluster of points; the genetic diversity can be seen as such cluster's diameter.

Several works define the loss of diversity as the proportion of individuals of a population that are not selected during the selection phase [86] [14] [65].

In literature [12] [97] [98] [8] a lot of different measures of diversity are used:

- The number of different distinct genotypes (where a genotype is a point in the search space).
- The number of different distinct phenotypes (where a phenotype is a fitness value, calculated giving a search point as the input of the fitness function). In this case the phenotypical diversity is used in order to indirectly approximate the genetic diversity, this choice is usually done every time a metric for the search space is difficult to define and/or to compute.
- The edit distances between recurring structures observed in the search points.
- Entropy measures over the genotypes population.
- A problem domain specific measure.

It's a common assumption that Evolutionary Algorithms choices that obtain higher diversity levels are better [86] [85], but the real correlations between the maximum fitness value reached by the search points and a diversity measure has not been investigated for every single measure proposed in the literature [14] [98] [8].

A population with a high genetic diversity can maintain its exploration ability in the evolutionary search and can avoid premature convergence to a local optimum [65], therefore there are several publications that focus on methods aimed to maintain the diversity in the population (and usually for every different method, a different diversity measure is used) [26] [85] [86] [66] [8].

1.5.2 Multi-Population Systems

Genetic diversity, as it is commonly intended in Evolutionary Algorithm literature, has the limit to not considerate the biodiversity in the number of different species (the presence of more than a single cluster in the search space), but it's simply the biological counterpart of the diversity internal of a single species (the diameter of the only existing cluster).

Several publications point out that multi-population systems with different populations in an inter-specific competition or cooperation, can work better (faster search and/or better fitness levels reached) than the classical mono-population system, where there is only intra-specific competition [3] [67] [24] [23] [25] [72] [71].

Usually, in the case of a multi-population Evolutionary Algorithm, the number of different populations and the relationships between them (competition OR cooperation) is statically defined in the design phase, with a specific problem domain fashion [7] [72]. Several publications show that a statical choice of these search parameters can negatively affect the search process [37].

More important, in the case of multi-modal optimization problems we want to obtain not only a single solution but several alternatives, therefore the presence of multiple clusters in the search space is needed.

In this work we focus on the island model approach for multi-population evolutionary algorithms, proposing a solution for the problem of the statical number of species.

1.5.3 Metrics for the Search Space

The field of application of Evolutionary Algorithms is so wide that the search space can possibly be a very complex space.

The most general and powerful search space, widely used in the literature, is probably the space of imperative language program codes [45]. Other complex widely used search spaces are the space of electronic circuits [92] or some type of algebraic functions spaces.

In these cases a metric to use for search space distance measurement is really difficult to define.

Usually the search space metric is needed in order to evaluate the genetic diversity (and in these cases of complex spaces the genetic diversity is approximated by phenotypical diversity), but the genetic distance of two search points can have several other purposes [12]. As an instance, in this work the genetic distance of two mating agents is needed in order to obtain the fertility rate of the mating couple.

The usual approximation of making use of the phenotypical distance is a strong approximation in mono-population systems, but it's obviously a wrong choice in the more interesting and powerful multi-population systems.

We propose a new solution to this crucial problem, based on biological observations of links between genetic variability and reproduction compatibility.

Chapter 2

Related Works

2.1 Speciation and Diversity Maintaining in Evolutionary Algorithms

In order to use the evolutionary paradigm for the solution of multi-modal or multi-objective optimization problems, it is necessary to introduce speciation events in the evolutionary process of EAs.

Speciation is obtained in evolutionary algorithms using niching mechanisms and co-evolutionary models [99].

Several niching techniques have been proposed, which we can divide in three main classes: modification of the mechanism of selecting individuals for the formation of a new generation (*crowding models* [61]), modification of the parent selection mechanism (*fitness sharing* [33] [34] [36] or *sexual selection* [90]) or restricted application of selection and recombination mechanism (*grouping* individuals into partially isolated sub-populations [9]).

In co-evolutionary models the success of an individual depends not only on his fitness but on the interactions with the other individuals. Co-evolutionary models can be formulated without an explicit fitness definition, introducing open-ended evolution and facilitating evolution in dynamic environments. The co-evolutionary models can be classified in competitive [70] or cooperative [73].

Gavrilets in [27] reviewed several analytical results on speciation, considering the three basic geographic modes: allopatric, parapatric and sympatric. Finally Gavrilets showed that the sexual conflict mechanism (one sex evolves to attract the other to mating, while the latter evolves to keep the reproduction rate on optimal level) can lead to sympatric speciation.

Todd and Miller [90] presented a model where sexual selection and natural selection are coupled. Sexual selection allows species to create their own optima in fitness landscape, allowing populations to escape from local optima. The authors showed that sexual selection coupled with natural selection can lead to sympatric speciation.

Sánchez-Velazco and Bullinaria [80] proposed "gendered selection strategies for genetic algorithms". The gender selection strategies are based on a sexual selection mechanism, where females are selected not only on the basis of their fitness, but considering their potential to produce fit offspring with partners.

Ratford, Tuson and Thompson [74] proposed the seduction function. The seduction function gives low measure for couples of individuals with very low or high Hamming distance between their genotypes, and a high value for individuals with fairly similar genotypes. The authors applied their mechanism alone and in combination with crowding and spatial population model. Even if the use of the seduction function for the partner selection can successfully locate several local optima in multi-modal problems, the observed loss of diversity is very high, causing the extinction of all the local optima except one.

Laumanns, Rudolph and Schwefel [49] proposed a model, based on a predator-prey mechanism coupled with a graph topology, for multi-objective optimization problems. Deb [15] extended Laumanns' model with modified predator-prey model using different weighted vectors associated with each predator.

Li [52] proposed a predator-prey genetic algorithm for multi-objective optimization introducing a dynamic spatial structure to the predator-prey population, allowing dynamic changes of the prey population size.

The use of a spatial structure is effective in maintaining a better population diversity [9]: in [43] [42] and [53] the authors examined the effects of introducing dynamical ecological features altering the spatial structure.

Dreżewski and Kisiel-Dorohinicki in [21] proposed a co-evolutionary version of the evolutionary multi-agent system, showing an improved diversity maintaining suitable for multi-modal optimization problems.

Dreżewski and Cetnarowicz in [20] considered the sexual selection as a speciation mechanism for agent-based evolutionary algorithms. The proposed co-evolutionary multi-agent system with sexual selection is applied to multi-modal optimization problems and compared to classical evolutionary algorithms.

Dreżewski and Siwik in [20] proposed two techniques for maintaining population diversity in agent-based multi-objective co-evolutionary algorithms, one based on the flock operators and the other on the sexual selection.

Kondrashov and Shpak in [44] discussed the role of non-random mating as a prerequisite for sympatric speciation. The authors proposed a model with similarity-based non-random mating (assortative mating) and showed that speciation can take place even in absence of sexual selection.

Kirkpatrick in [41] showed that an assortative trait, evolving as a reproductive barrier between two hybridizing populations, can lead to isolation and speciation.

Dieckmann and Doebeli in [16] proposed a sympatric speciation model, showing how the assortative mating can lead to speciation even if the assortative trait is not linked to resource competition.

2.2 Artificial life

Artificial life studies the logic of living systems in artificial environments. The goal is to study the phenomena of living systems in order to come to an understanding of the complex information processing that defines such systems.

Ray [76] [77] developed the *Tierra* software, a computer simulation in which computer programs compete for central processing unit (CPU), time and access to main memory.

Ofria, Adami and Brown developed *Avida*, an artificial life simulation software inspired by Tierra [69] [1]. The main difference from Tierra is that in Avida every digital organism lives in its own protected region of memory, and is executed by its own virtual CPU. The speed at which a virtual CPU runs is determined by the tasks that the organism performs, giving more CPU time as a reward.

Chapter 3

Definitions

In this work we consider mainly two different theoretical fields, the biological field and the simulation models field.

There are some elements of the biological field represented in the simulation field; we refer to them as the simulation model counterpart of a biological element.

A simulation element that is the counterpart of a biological element is neither a comprehensive nor a coherent representation of the biological element because we need to have only a simplified version of a sub-set of the features that we can observe in the natural environment: we don't want to develop biologically realistic models, but biologically inspired ones.

3.1 Biological Systems

There are some really big problems about giving definition of several biological elements: the first more obvious problem is the limited understanding of biological dynamics, especially system level dynamics like the biosphere auto-regulation. The second problem is epistemological, about living beings definition and species definition: there is not in the zoological scientific community a strong agreement about these subjects and the debate is generations old [62].

In this work we do not engage in biological or epistemological debates and we adopt definitions arbitrarily.

Biosphere : a dynamical system composed by a biotic component (the species) and a not-biotic component (the inanimate portion of the environment).

Genotype : all the genetic information contained in a living being; it is almost stable during the whole life of the living being.

Phenotype : the expression in a living being of his genotype in his local environment (the portion of the biosphere that interact with the living being).

Living Being : the base element of the biotic component of the biosphere, it's characterized by a genotype (inherited from his parent/s) and a phenotype; some living beings can reproduce themselves (sexually and/or asexually) giving a partially random modified copy of the genotype (or a partially random modified copy of a combination of all the genotype of the partners in the case of sexual reproduction) to a new living being.

Resource : the component (biotic or not-biotic) of biosphere that a living being needs in order to live and to reproduce (the needed resources can differs depending on the living being considered).

Mutation : the random modification of a genotype.

Selection : the non uniform distribution of the resources needed in order to live and reproduce, the death of living beings not able to obtain resources and the proliferation of the others.

Evolution : the dynamic of the system in terms of variety and distribution of phenotypes, efficiency of living beings's resource collection, variation of system species and species interactions. Such a dynamics is determined by the combined effects of mutation and selection.

Species : is the characterization that fit a particular group of living beings that share some level of similarity, and where some component of such a group are eventually capable of interbreeding, producing fertile offspring of the same species.

Speciation : the process of formation of a new species in the biosphere.

Environmental Feedback Cycle : the activation and the effect of the activation of a subsystem of the biosphere system, in response to an alteration of some parameter of the biosphere system.

Negative Feedback Cycle : an environmental feedback cycle that react to an alteration of some environmental parameters (temperature, gas concentrations, solar light, growing or shrinking of the size of a species population, etc.) in order to counterbalance the alteration and maintain the environment in a dynamical equilibrium.

Positive Feedback Cycle : an environmental feedback cycle that reacts to an alteration of some environmental parameters in order to amplify the alteration.

3.2 Simulation Models

In the thesis we describe more than a single simulation model, therefore only elements common for every model are defined here.

We assume as defined (they are defined later for every different model) the environment, the agents, the agent internal state and the agent static internal state.

The environment is the simulation model counterpart of the biosphere, the agents are the simulation model counterparts of the living beings, the agent internal state is the simulation counterpart of the phenotype and the agent static internal state is the simulation model counterpart of the genotype.

Not every simulation model has a counterpart for resources, where represented they are defined.

We assume as defined some other elements that have not an explicit counterpart in the definition exposed for the biological field case: geographical space, geographical metric, genetic metric. All the models described are iterative, so the iteration is implicitly defined.

Population : the set of all the active agents in the simulation.

Genetic Space : the smallest space containing all the possible agent static internal states.

Mutation : the variation from a given point of the genetic space to another random target one, with a probability distribution on the targets inversely proportional to the distance (obtained from the genetic metric on the genetic space) of the target from the starting point.

Selection : the process that at each iteration, for each agent in the population, choose if it has to die (be removed from the population) or to live (based on fitness function or resource consumption that could be defined). Among all the survivals, this process can select some agents able to reproduce themselves (in the case of sexual reproduction it does not select single agents but couples based on their relative distance obtained from the geographical metric on the geographical space).

Evolution : the dynamic of the system in terms of variety and distribution of agent internal states, efficiency of agent's resource collection, variation of the system species and species interactions. Such a dynamic is determined by the combined effect of mutation and selection.

Species : the concept of species has no role in the run of the simulation but only in the data analysis process. A species is defined as every set of agents whose genotype form a cluster in the genetic space (such a definition allows an agent to be in more than one species and allows a hierarchical species structure).

Speciation : the events that, during the run of the simulation, lead to formation of a new species.

Feedback Cycle : during the simulation run, a system level change in the dynamics of the agents that can be explained by the means of another variation of agents dynamics or resource (where defined) dynamics.

Negative Feedback Cycle : a feedback cycle whose effect is to counterbalance the changes that have activated it.

Positive Feedback Cycle : a feedback cycle whose effect is to amplify the changes that have activated it.

We want to stress the fact that species, speciation, evolution and feedback cycles have not, by definition, any role in the run of the simulations, but they are used only in the data analysis process.

In the model description, when is clear that we are referring to the simulation models field and not to the biological field, we can use the biological term in order to refer to the simulation model counterpart.

Chapter 4

The Sympatric Speciation Model

Speciation is the evolutionary process by which new species arise.

The zoological scientific community defines different speciation dynamics believed to occur in the nature. Speciation rate is strongly debated but realistically it's too slow to directly test the speciation hypothesis; on the other hand a lot of indirect evidences have been gathered to support the assumptions.

The four most accepted modes of speciation are allopatric, parapatric, peripatric and sympatric speciation. The key feature that characterize the four modes is the role of geographical isolation (figure 4.1).

Isolation is a process that divides the starting population in different geographical zones, the islands; the interactions between living beings (competition, cooperation, mating) located in different islands is limited or absent, and limited or absent can be the migration of a living being from one island to another one.

In the biosphere the isolation can occur through several different ways: there could be real islands, but a river or a mountain range can isolate as well. But even considering different species sharing the same geographical area we can still consider them as partially isolated because different species need different types of resources, so the competition is limited to the shared needed resources. This extended notation of partial isolation usually is labeled as ecological niche.

The aim of this work is to study the dynamics of negative feedback formation; in chapter 6 we make the hypothesis that the speciation event plays a fundamental role for feedback formation and, in order to select the negative ones, the arising species need to be strongly connected to the environment. This does not mean that there is not the need of a new ecological niche to occupy for the arising species, but that if a species arises and evolves far from a given ecosystem then, when the speciation is complete, we cannot hope that, putting back all the new species in the same starting ecosystem, they would interact well. In the biosphere there are a lot of examples of this situation: when a species is transported from a continent to another, the consequences are difficult to predict, and often the new species weakened or compromised the whole local ecosystem. In particular, species evolving isolated cannot direct their genetic trajectory in order to inhibit system level positive

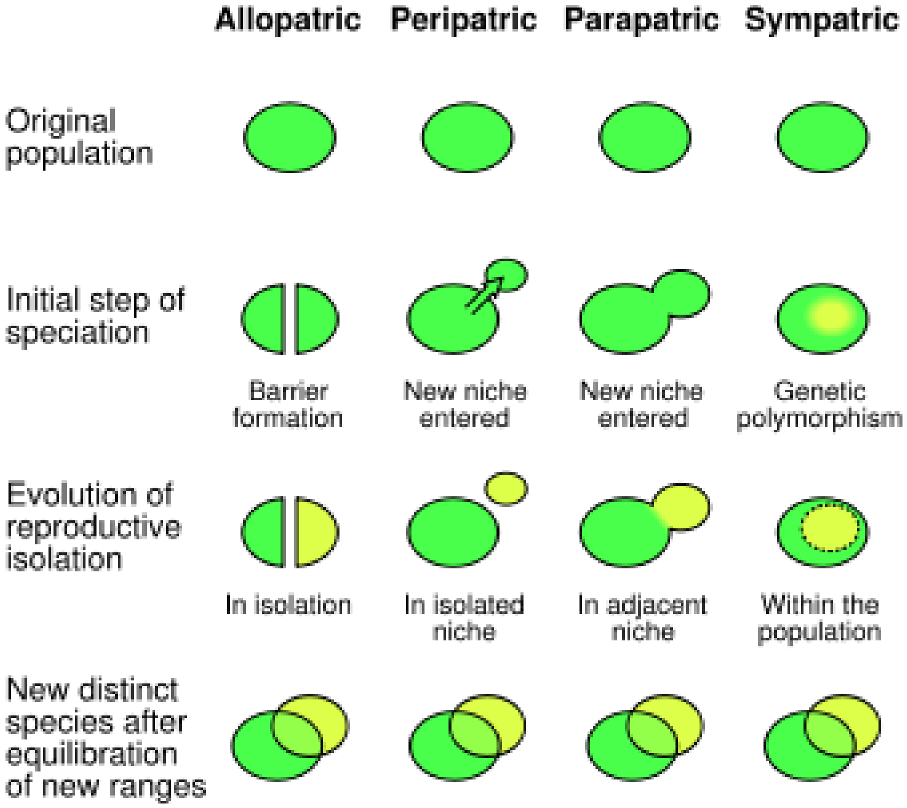


Figure 4.1: The schematized comparison of the differences in geographical isolation of allopatric, parapatric, peripatric and sympatric speciation modes. Figure drawn by Ilmari Karonen and published on wikipedia.

feedbacks or to promote negative feedbacks.

Therefore we need to focus on the more general sympatric speciation mode.

In the sympatric speciation mode, from a single starting species two will evolve and diverge, but without any geographical isolation. Even if the new diverged species share the geographical space but live in separate ecological niches, in the starting and critical stage of sympatric speciation the two potential diverging population are so similar to be considered in the same exact ecological niche (unless there is an *hopeful monster* case).

The main problem for a sympatric speciation manifests itself when a sub-population starts to differentiate: the original population and the sub-population are fully in contact an therefore they are in a situation in which there is competition, but the two population are so similar (in the hypothesis of slow mutation) that, in this situation, it's expected that one population quickly extincts.

4.1 Biological Observations

In 1966 Maynard Smith [84] proved the plausibility of the sympatric speciation hypothesis and today exist well documented empirical evidences of sympatric speciation.

Currently there is a lot of debate about the validity of this mechanism and, from the initial theoretical work of Smith, only in the recent years several works discussed evidences of sympatric speciation events.

A very interesting example of sympatric speciation in its critical starting stage (when the ecological niche is the same) is the case of *Microtus savii*'s speciation, a small rodent whose geographical distribution is limited to a large part of the italian peninsula and Sicily, studied in 2003 from a strong biological point of view by Galleni et al [13]. A chromosome investigation carried out in *M.savii* from central Italy and from Calabria showed different amounts of heterochromatin in the sex chromosomes; hybrids were formed from crosses but no offspring at all was obtained from the crosses between hybrids.

The most intriguing observation on the two different species of *Microtus savii* is the fact that even if they occupy the same niche the two population are well geographically separated, but not isolated: all the rodent living in the same zone are of the same species even if there is not any geographical obstacle to isolate the two populations and there is not any mating preference for the rodent of the same species.

The work explains the situation addressing to the hybrid population in the border zone between the two different species to act like a biological barrier of limited reproduction rate. They develop a simulation model where a mixed population of rodent of the two species are capable to randomly move in a bi-dimensional space, competing for limited resources that limit their population; when a pair of rodent of different sex are geographically near they can produce an offspring, generating an infertile hybrid if the two parents rodents are not of the same species. They show how even a starting situation with uniformly geographically distributed rodent of the two species quickly collapses to a situation with perfect isolation of the two species in two or more populations, separated by stable borders populated by hybrids. In absence of hybrids the system shows a fast extinction of one of the two species.

In this case the effect of the hybrids with limited reproductive capability is to stabilize the coexistence of the two similar populations in the same niche.

This case of study shows that there are possible mechanisms that permit the coexistence of similar populations in the same niche, a situation needed for the sympatric speciation mode. But it's not enough, we need a model that enables not only the maintaining of such diversity for the time necessary to form fully differentiated species, but that permits the initial sub-population formation.

4.2 The Simulative Model

The model of sympatric speciation is an iterative, dynamic, stochastic model based on agents.

The agents of this model are synchronized. Thanks to the geographical space abstraction they can be seen as moving entities on a planar space ($\mathbb{R}^2$).

The agents can perceive other agents in a fixed range ($range_{mating}$) in the geographical space and interact with them (for reproduction purpose).

The agents can perceive the resource components of the environment in a fixed range ($range_{gathering}$) in the geographical space and can compete for them with the near agents in order to increase their energy level.

The energy level that every agent loses in the metabolism step of every iteration is $e_{metabolism}$.

Another parameter of the model is the minimum energy level e_{mating} that an agent must have in order to be able to mate.

4.2.1 Components

The environment : is composed only by the *geographical space* (S), a *population of agents* (A) and the set of the resources (R).

The geographical space : is a bi-dimensional square of side l. As *geographical space metric* we adopt the Euclidean metric.

$$S = \{(x, y) \mid 0 \leq x, y \leq l,\ x, y \in \mathbb{R}\}$$

$$d_s(p_1, p_2) = \sqrt{(p_1[0] - p_2[0])^2 + (p_1[1] - p_2[1])^2} \qquad p_1, p_2 \in S$$

The genetic space : is a bi-dimensional space. As *genetic space metric* we adopt the Euclidean metric.

$$G : \mathbb{R}^2$$

$$d_g(g_1, g_2) = \sqrt{(g_1[0] - g_2[0])^2 + (g_1[1] - g_2[1])^2} \qquad g_1, g_2 \in G$$

The Fitness Function : defined on the genetic space.

$$f : G \to \mathbb{R}$$

The resources : each resource is characterized only by its *position* in the geographical space. The number of resources of the environment is a fixed number $(k_{res} \cdot l^2)$.

$$R = \{p_i \mid p_i \in S\}$$

$$|R| = k_{res} \cdot l^2$$

If an agent wins the competition for a resource, the energy gained is a fixed value $e_{resource}$.

The agents : an agent is characterized by its internal state, further divided in *static internal state* and *dynamic internal state.* The size of the agent population is not fixed nor directly limited.

$$a_i = SS_i \cup DS_i \qquad a_i \in A$$

The dynamic internal state : of an agent consists of its *energy level* value (an integer) and the *position* of the agent (a point of the geographical space, or geographical position).

$$DS_i = \{e_i, p_i\} \qquad e_i \in \mathbb{N},\ p_i \in S$$

The static internal state : of an agent consists of his *genotype* (a point of the genetic space, or genetic position).

$$SS_i = \{g_i\} \qquad g_i \in G$$

Thus an agent a_i can be represented as $\{e_i, p_i, g_i\},\ e_i \in \mathbb{N},\ p_i \in S,\ g_i \in G$.

As geographical space abstraction we mean a view of this model considering the position p_i in the geographical space of an agent as a parameter of the environment and not of the internal state of the agent (in this case an agent a_i is composed only by $\{e_i, g_i\}$).

4.2.2 The Iteration

The agents are synchronized. A synchronization cycle is defined as an iteration and each iteration is divided in several steps.

For the only purpose of simplicity of some system dynamical descriptions, sometime in the work we give up the agent paradigm and describe some computations in the model, attributable to individual action or synchronized interaction of groups of agents, that alter the agent population or some agents internal states, as done by a non-agent mechanism. We want to assure that, independently from the paradigm used to describe some agent state alteration mechanism, we can explain it in a less simple way but from within the agent paradigm. Of course dynamics like resources regeneration and resources stochastic distribution are outside the agent system.

An iteration schema is composed by several sequential steps:

1. Movement
2. Resources Regeneration
3. Competition
4. Metabolism
5. Reproduction

And now a detailed description of each step dynamics:

Movement Step : Each agent $a_i \in A$ must do a movement action: its current geographical position is stochastically changed to a new geographical position extracted with an uniform statistical distribution over the portion of S distant from its starting position at most $step_{max}$:

$$p_i \to p_i', \qquad p_i' \in S,\ d_s(p_i, p_i') \leq step_{max}$$

Resources Regeneration Step : For each resource $i < (k_{res} \cdot l^2)$, its new position in the geographical space $p_i \in S$ is randomly determined.

Competition Step : For each resource $i < (k_{res} \cdot l^2)$, all the agents capable of perceiving it form the set of opponents that compete in order to obtain it:

$$O_i = \{a_j \mid d_s(p_j, p_i) \leq range_{gathering},\ a_j \in A,\ p_i \in R\}$$

For each opponents set O_i, a winner is stochastically selected based on the opponents fitness values of each agent's genotype. The selection method used is Fitness Proportional:

- For each opponent its fitness value is calculated.

$$v_j = f(g_j) \qquad opp_j = \{e_j, p_j, g_j\} \in O_i$$

- The sum of fitness values of the opponents is calculated.

$$v_{total} = \sum_{k=0}^{|O_i|} v_k$$

- A stochastic value is extracted with uniform statistical distribution $U(0, v_{total})$.

$$v_{select} \in [0, v_{total}] \qquad v_{select} \in \mathbb{R}$$

- The winner is calculated.

$$winner \in \mathbb{N}, \qquad \sum_{k=0}^{winner-1} v_k \leq v_{select} \leq \sum_{k=0}^{winner+1} v_k$$

- The agent that win the competition gets its energy level increased.

$$opp_{winner} = a_x = \{e_x, p_x, g_x\}, \qquad a_x \in A$$

$$e_x = e_x + e_{resource}$$

Metabolism Step : Each agent $a_i \in A$ gets its energy level decreased by a constant value:

$$e_i = e_i - e_{metabolism}$$

If the remaining energy is $e_i \leq 0$ then the agent dies and it's removed from the agent population A.

Reproduction Step : All the agents whose energy level is equal or greater than the minimum mating energy e_{mating} compose the mating set:

$$MATE = \{a_i = \{e_i, p_i, g_i\} \mid e_i \geq e_{mating},\ a_i \in A\}$$

It is randomly selected (if existing) a couple of agents from the mating set that are near in the geographical space:

$$(a_x, a_y) \qquad a_x, a_y \in MATE,\ d_s(p_x, p_y) \leq range_{mating}$$

The selected copy mates (the process is described in section 4.2.3) and, independently to the result of their attempt, the two agents are removed from the mating set $MATE$.

4.2.3 Mating Mechanism

When a couple of agents a_x and a_y are selected for mating in the reproduction step, then they attempt to generate a new child agent to insert in the agent population A; such a child inherits their genotype.

The key element of the model is that the mating can fail to generate a child and that the fail probability is dependent to the genetic distance of the parents.

$$P_{mating}(a_x, a_y) = \frac{1}{1 + d_g(g_x, g_y)}$$

This means that the probability that a mating attempt successfully generates a child decreases proportionally with the genetic distance of the mating agents' genotypes.

The two agents get their energy level decreased by a fixed value independently to the success of mating.

$$e_x = e_x - e_{matingattempt}$$

$$e_y = e_y - e_{matingattempt}$$

If the mating is successful a new agent $a_{new} = \{e_{new}, p_{new}, g_{new}\}$ is inserted in the agent population.

- e_{new} is set to the constant starting level $e_{newborn}$ (the parameters of the model need to be such that this starting level is not greater than the sum of the energy paid by the parents in the mating process).

- p_{new} is set to the halfway point in the geographical space of its parents' positions:
$$p_{new} = \frac{p_x + p_y}{2}$$

- g_{new} is set near to the halfway point in the genetic space of its parents' genotypes:
$$g_{new} = mutation\left(\frac{g_x + g_y}{2}\right)$$

The function $mutation : G \rightarrow G$ selects two stochastic values in a uniform statistical distribution $U(-maxmutation, maxmutation)$ and uses them to the two coordinates of $\left(\frac{g_x+g_y}{2}\right)$.

In order to obtain a simulation working as intended, the value of $maxmutation$ needs to be small, such that the mutation function application to a point of the genetic space can produce only a minor variation of its corresponding fitness function value.

4.2.4 Initial Settings

The starting agent population is created of size *startingagents*, each agent in the starting population has the same genotype (*startingenotype*), a starting energy level value (*startingenergy*), and they are randomly placed within the geographical space.

The Fitness Function f is chosen such that only two different paths in the genetic space (sharing a common starting path) are fitness increasing in the genetic space; the paths starting point is the common starting genotype (*startingenotype*).

4.2.5 Termination Condition

The termination condition in our experiments is the number of iterations we want to run, but it can be set to any other condition, as an example it can based on the fitness level reached by the agents in the agent population.

4.2.6 The Result of a Simulation

In a classical search algorithm based on fitness function directed evolutionary dynamics, the result of the search process is considered to be the point of the search space with higher fitness that was found during the simulation.

In our case we consider the results of the simulation not simply a single agent, but the clusters that the genotypes of the various living agents constitute in the genetic space.

By adopting the view of the simulation interpreted as a search algorithm over the genetic space, directed by the defined fitness function, its final objective is not finding the best solution anymore, but finding several alternative solutions.

4.3 An Example of Execution of the Model

In order to implement the simulation model in a computer program and study its emerging properties we need to assign a value to every parameter of the theoretical model.

All the parameters that are needed in order to implement the model will be summarized and their isolated effects to the system dynamics will be described in a following section.

We want to stress that the dynamics of the agent population in the implemented model depends on the choice of the parameters: even if is not difficult to find a parameter choice that grants speciation, there are choices that do not make the model work as intended; these cases give us hints of the different role that every parameter has in the emergence of speciation and are discussed in the following.

Another fundamental choice to do in order to obtain a speciation event is the fitness function definition. In the following example we make use of a fitness function with a well defined increasing gradient path in the genetic space. The gradient increasing path starts from *starting_genotype* and leads to a point of the genetic space from where there is a bifurcation: two separate paths with increasing gradient leading to different directions.

The implementation of the model with the described fitness function shows the emergence of speciation capabilities without the need of any geographical isolation.

The speciation event is dependent to parameter choice, but it shows a strong resilience to parameters variation and we didn't experience any difficulty to quickly set parameters that enable the speciation.

The dynamics of the agent population in the course of speciation, in both geographical space and genetic space, is shown, for an example of model execution, in the figures 4.2, 4.3, 4.4, 4.5 and 4.6.

The left-sided square in the figures represents the geographical space and the colored dots are the agent positioned; each dot in the geographical space is colored dependently on the agent genotype position in the genetic space; this enables a mapping from the two spaces and gives us the very interesting information that agents near in the geographical space tend to be near in the genetic space too.

The right-sided square in the figures represents the genetic space and the colored area represents the points of the space occupied by genotypes present in the agent population. The colors represent the density of representatives in the population for the particular genotype.

The small colored square on the left visualizes the values of the fitness function mapped on the genetic space: the yellow-red colors have a higher fitness value. The main starting gradient increasing path can be visualized on the diagonal; the bifurcation point is placed in the middle of the square and the two alternative paths are positioned one vertical and the other horizontal.

Figure 4.2 is a screenshot of the situation before the starting of the speciation: the only cluster of genotypes in genetic space, following the main starting gradient increasing path of the fitness function, has reached the bifurcation point.

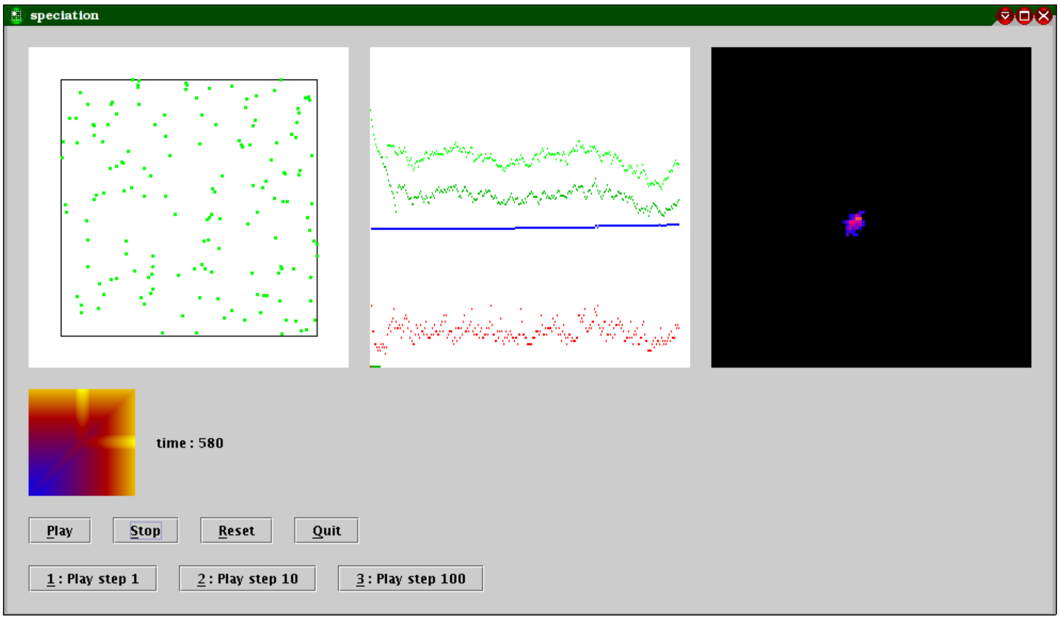


Figure 4.2: An example of execution of an implementation of the sympatric speciation model. A speciation is occurring. The left-sided square in the figure represents the geographical space with dots representing the agent positions, the right-sided one represents the genetic space with dots representing genotype density.

In the Figure 4.3 is shown the starting phase of the speciation, when the genotypes organize in weakly defined clusters and are still on the bifurcation point.

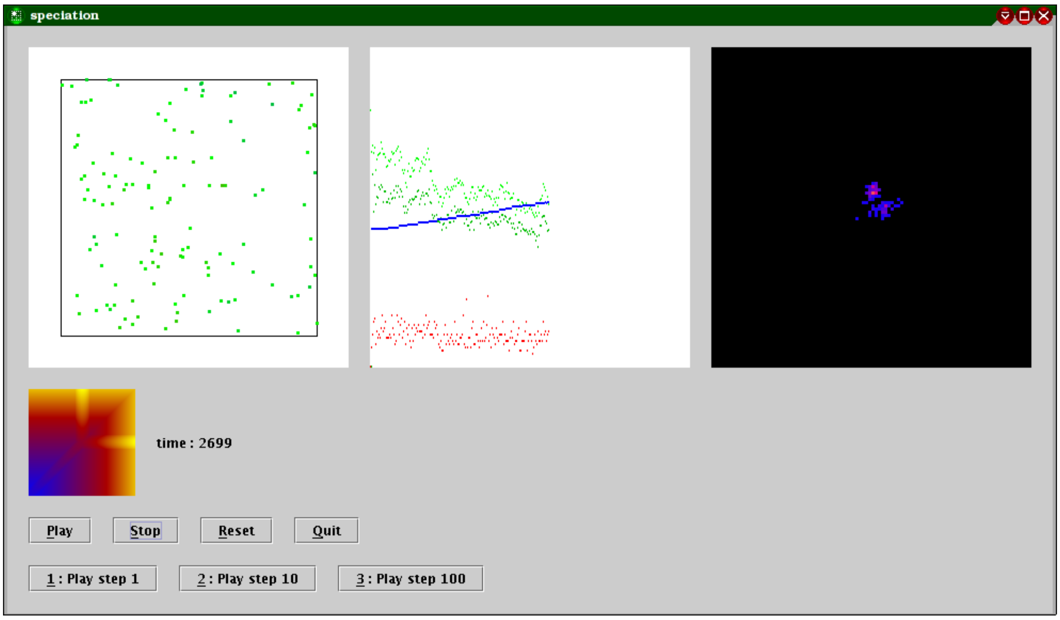


Figure 4.3: The same run of Figure 4.2: the genotypes start to organize in weakly defined clusters and are still on the bifurcation point.

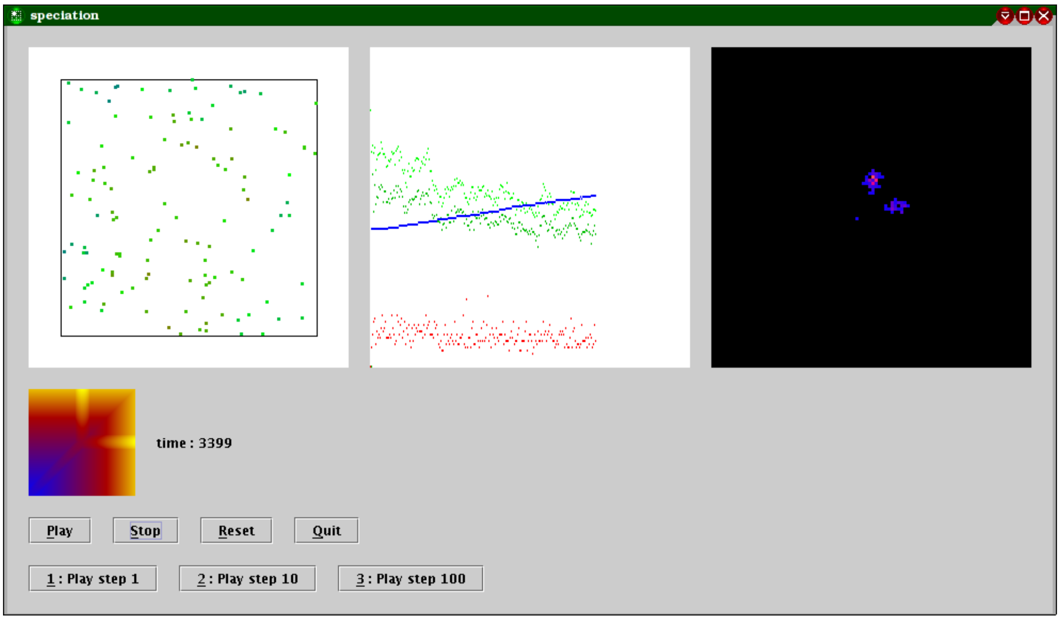
In the Figure 4.4 the two clusters in the genetic space are formed, but their closeness still allows the birth of hybrid agents.

In the Figure 4.5 the two clusters are more separated in the genetic space and it's shown an interesting emerging property of this dynamic of sympatric speciation: even if there is not any obstacle in the geographical space the agents tend to form isolated groups composed of representative of only one of the genetic clusters. Agents near in the geographical space tend to be near in the genetic space too. We can still see some hybrid agents in the boundary line: they are the green dots in the geographical space.

The situation in Figure 4.6 shows fully separated clusters in the genetic space and the absence of hybrid formation.

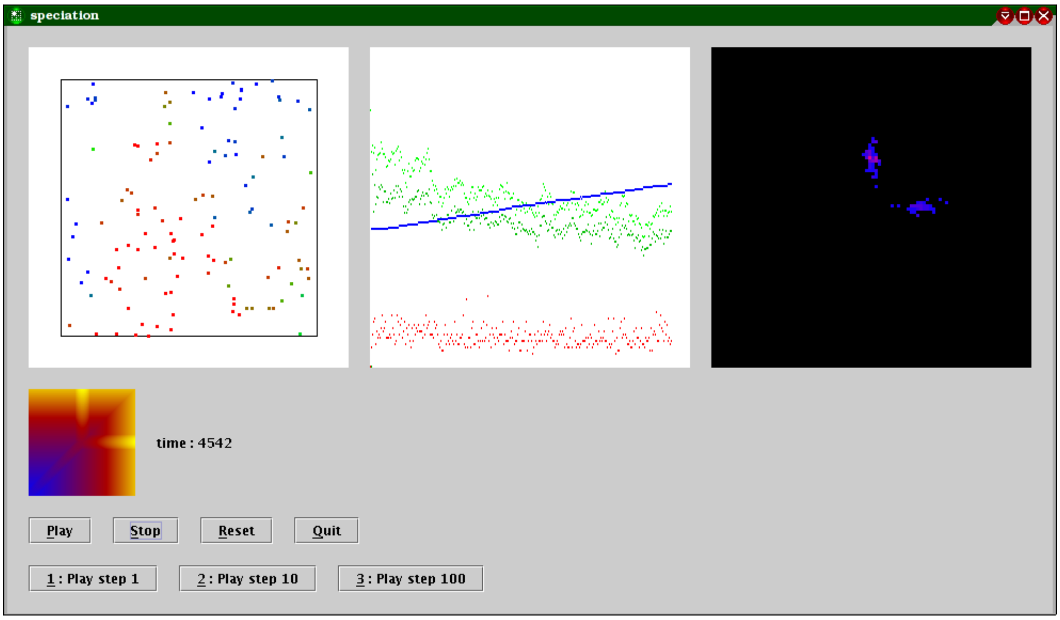
4.4 Model Discussion

The settings and the features choice adopted in the proposed model of sympatric speciation are standard in regard to the modeling approaches found in computer-science literature, but with one notable innovation, a simple mechanism fundamental for the emergence of speciation without any need of geographical discontinuities: the negative effect on the fertility of the genetic distance between couples of mating agents.



time : 3399

Figure 4.4: The same run of Figure 4.2: the two clusters in the genetic space are formed, but their closeness still allows the birth of hybrid agents.



time : 4542

Figure 4.5: The same run of Figure 4.2: the agents tend to form isolated groups in the geographical space composed of representative of only one of the genetic clusters.

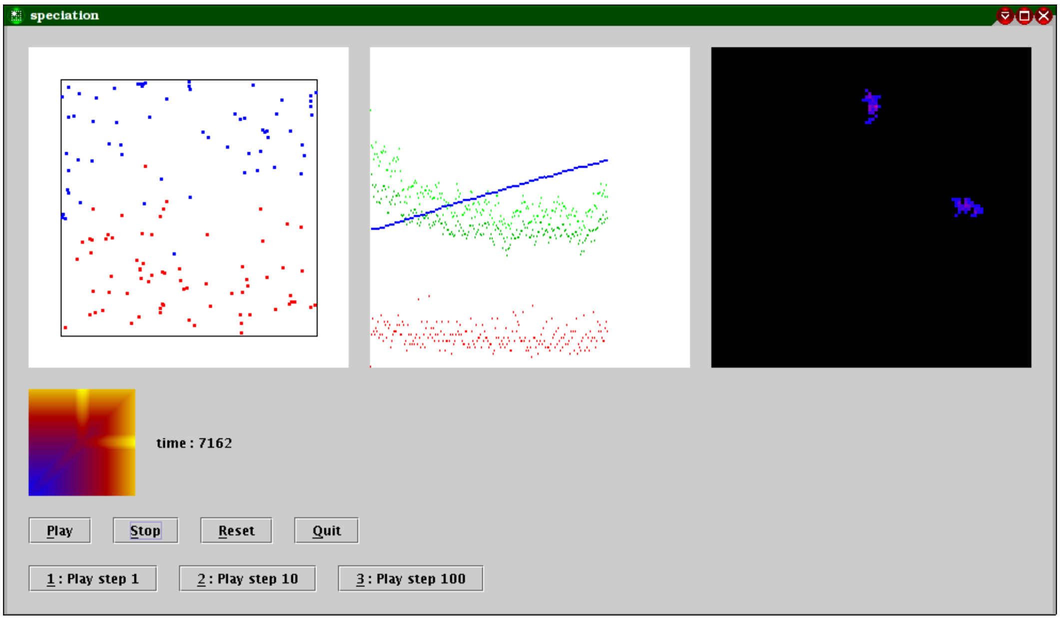


Figure 4.6: The same run of Figure 4.2: the two clusers are now fully separated in the genetic space.

The simulation dynamics of an implementation of the model of sympatric speciation can exhibit, dependently on the parameters choice, not only the expected emergence of the speciation, but another unexpected additional emergent property: the spontaneous formation in the geographical space of several zones where the agents populating it have their genotypes positioned in the same cluster of the genetic space. Another additional emergent property of the model is the cohesion of genetic clusters.

The most interesting features of the model, the observed emergent properties and the role of each parameter in the whole system dynamics, are discussed in this section.

4.4.1 Discussion About the Features of the Model

The Separation of the Geographical Space from the Genetic Space

The feature of separating the search space (the genetic space in our modeling approach) from a second distinct space (in literature it's usually addressed as geographical, like in our case) independent from the first one is present in several works in the field of evolutionary algorithms literature.

In absence of any geographical space, the selection phase of the evolutionary algorithm doesn't usually make use of the distance of pairs of sexually-reproducing agents.

In the island model approach, the islands' structure can be seen as a particular geographical space composed of only a very limited number of points (usually a limited migration capability from an island to a connected another one is granted to the agents).

Our choice for the use of a geographical space is explained by the necessity to have some type of locality over the selection process: the environment that defines the selective pressure over a particular agent (in our case the competition for the resources) is not the whole system environment but only a local fragment of it, centered on the agent's geographical position.

The locality of the selection is essential because it is the selective pressure that directs the genetic trajectory followed by a population subjected to the effects of genetic mutation.

Different selective pressures can move the genotypes of the agents toward different distinct directions in the genetic space, but if the selection pressure is the same for all the agents in the system then the genetic trajectory is the same too: no splitting of the genotype cluster can be expected and thus we will not obtain the occurrence of any speciation event. If we have not a locality of the selection pressure we cannot obtain speciation events without the necessity of isolation of the population in several sub-populations.

Making use of a geographical space not limited as the one we can see in the case of the island model, saves us from the additional parameter of the number of islands to put in the model, that is a problem in the field of multi-population evolutionary algorithms, as this number is defined at design time and usually is dependent to the problem that we want to solve.

Variable Fertility Discussion

For variable fertility we mean that the mating of two agents can produce a newborn agent with the probability:

$$P_{mating}(a_x, a_y) = \frac{1}{1 + d_g(g_x, g_y)}$$

This is a feature that, at our knowledge, in literature never appeared before in an evolution computer model; it is standard the assumption that a pair of selected agents (or a pair of search points) produces a new child agent with probability one.

There is a biological inspiration for this simple feature: the epistemological debate about a good definition of species stressed the existence of inter-species breeding, sometimes generating hybrids capable of making children, as a motivation to keep out the fertility from the qualities that define the species boundaries. From these cases we can assume the possibility that the fertility or infertility between two living organisms not only is determined by the genetic closeness (and this is almost obvious) but that it's a fuzzy property, decreasing when the distance increase. This is a naive and oversimplification assumption to do about the biological real world, where the high complexity of living beings prevents us to make this type of generalizations, but it worked out well in our simplified biologically inspired models.

When the adopted type of variability of the fertility coexists with the geographical space and the locality of mating (two agents can mate only if they are geographically near), we can expect several different effects:

- If in a zone of the geographical space the agents are genetically similar, then there is not any effect, because the probability is near to one.
- If in a zone of the geographical space two groups of genetically different agent populations are shuffled in *equal* proportions, then the effect is a reduction in reproduction rates for every agent because the energy cost of mating is paid independently to the success of mating.
- If in a zone of the geographical space two groups of genetically different agent populations are shuffled in *unequal* proportions, then the effect is a reduction in reproduction rates. The reduction is stronger in the outnumbered population and smaller for the main population.
- In the zones that constitute the borderline of two areas occupied by genetically different mono-population of agents, the effect not only is the reduction of the reproduction rates, but, due to the higher rates in the mono-population areas, any invasion attempted by a population will fail in the beginning, because the first invading agents will suffer a huge reduction of reproduction (remember that the reduction is stronger in the outnumbered population) thus the selective pressure imbalance will prevent any invasion.

We are able to appreciate the last described effect for the case of borderline zones when we are making comparisons with the work of Galleni [13]. It turns out that the effect of the adopted variable fertility has strong similarity with the effect of the hybrid zone of *M.savii*: it's the effect of diversity maintaining for different species living in the same niche.

The borderlines stability obtained by the variable fertility prevents us from the need to adopt further diversity maintaining mechanics.

Another quite obvious and expected effect is the last phase of speciation event: when two agent populations that constitute two different clusters in the genetic space are far from each other, then the interbreed disappears.

4.4.2 Observed Emergent Properties

Speciation

By observing a single genetic cluster dynamics in the genetic space when the population of agents is following a genetic mutation trajectory, we can note that the integrity of the cluster is not statical: several groups of genotypes tend to go away from the center of the cluster in every direction (more prominent in gradient increasing genetic directions).

Mapping these groups of differentiated genotypes on the geographical space we can observe that the agent composing them are positioned in the same local geographical areas.

It is the local selection that enables the separation from the genetic cluster, acting before the delayed global selection that, in asecond moment, can suppress or select the speciation.

The delay of global selection is caused by the fact that an event far in the geographical space takes some time (dependently to the velocity of diffusion of the agents) before being influential locally.

It's this delay from the local to the global selections that enables that, from a cluster in the genetic space centered in a point from which several directions are fitness increasing, different genetic clusters represented by groups of agents positioned in different geographical areas, can emerge without suppressing each other, and independently starting to constitute new populations, eventually will split the starting genetic cluster when the global level selection occurs.

We are able to observe the expected emergent property dependently to the parameter choice.

A species is a cluster in the genetic space, so it is possible to obtain not only speciation of different clusters, but eventually, dependently to the particular fitness function, the emergence of a hierarchical cluster structure.

Formation of Mono-population Areas

This emergent property of the simulation model was initially unexpected.

We are able to observe that, when a speciation is occurring in the system, the two genetically different populations tend to occupy separated areas of the geographical space, forming mono-population areas.

But, more in general, even if there is not any speciation event, if we map the geographical positions of agents with the genetic position of theirs genotypes, we note that agents that are geographically close usually are genetically close too.

Our interpretation of this emergent property of agent distribution is the contemporaneous effect of local selection and variable fertility. The local selection encourages variability among geographically far areas but, when considering the selection locality range, the selective pressure encourages the mono-population. Adding to the particularities of the local selection the previously discussed diversity maintaining effect of variable fertility, we can understand and expect the formation of the mono-population areas.

Cohesion of Genetic Clusters

For genetic clusters cohesion we mean the experimentally observed fact that a genetic cluster positioned in a zone of the genetic space, where the corresponding defined fitness function is a constant value, maintains a limited diameter rather than expanding indefinitely.

The genetic cohesion can be easily described as one of the direct results of the variable fertility: the genetic distance that two agent of the same species can reach is regulated by the fact that the greater it is, the less energetically efficient is their mating process.

Therefore, a genetic cluster with a big diameter denotes a species whose mating efficiency is low, and in this case the selective pressure tends to select a version of the species with a decreased diameter. But, on the other hand, the genetic cluster diameter cannot be indefinitely decreased (even if the fitness function is not a constant) because of the locality of selection and the effect of random genetic mutation of the newborns.

In the case of the constant value fitness function we see a cohesive cluster, with occasional speciation of alternative clusters. In comparison, evolutionary models where there is not any variable fertility, in absence of additional features, have no form of cohesion.

4.4.3 Parameter Role Discussion

Here we summarize all the parameters used for the description of the theoretical model. Any implementation of the theoretical model is a particular choice of parameters and, even if some choices prevent the emergence of the target properties of the model, it is not difficult to find a parameter choice that grants the emergence of speciation.

The singular effects to the system dynamics of every parameter are described. All the analysis are based on experimental evidences.

$range_{mating}$ **:** the geographical range for mating.

$range_{gathering}$ **:** the geographical range for resource competition.

We have not found any valid motivation for setting $range_{mating}$ and $range_{gathering}$ to different values. These two parameters define the range of the locality of selection. Their effects are linked to the parameters that define the geographical space size and the maximum distance that an agent can move in the movement phase. An increase of the locality range can be completely balanced by an appropriate increase of the geographical space size and of the maximum movement step size. When not balanced, the increase of the locality of selection causes an increase in the thickness of boundaries areas of limited reproduction rate that will form dividing different mono-population areas, and a second observed effect is the decrement of local genetic variability (actually, the variability is the same, but the locality range increases, so it's the same thing).

l : the size of the geographical space side.

$step_{max}$: the maximum variation of the position that an agent can do in the movement phase.

As previously stated, appropriate variations of $step_{max}$ and l can completely balance the variation of the locality range, therefore they, linked, have the same effect. But changing only one at the time, keeping constant all the other system parameters, we observe different effects.

If l is set over a minimum value experimentally found, any further increments do not have any big effect on the dynamics of the model; all the effects that we can obtain is to have an increased number of mono-population areas but their dimension and stability is the same for every value of l. If the l is too low, to say, about the dimension of a typical mono-population area that we can obtain with a very high value of l, then the lower this parameter is set, the more unstable the mono-population system becomes. If the mono-population system is destabilized then we cannot observe any speciation event.

The variation of $step_{max}$, keeping unchanged all the other parameters of the model, causes a variation of the delay of the global dynamics effect over the local effect. Increasing the $step_{max}$, even if the locality range is the same, the effect is transported around the geographical space with an increased diffusion velocity. A too big value of $step_{max}$ therefore destabilizes the system, on the other hand, a too small value slows down the system dynamics and, in presence of bifurcations in the fitness function, prevents speciation (this second effect can be mitigated by an increase of l).

k_{res} : the density of resources, as the number of resources to put in the geographical space for each unit of space.

$e_{resource}$: the energy gained by winning a competition for a resource.

The product $(k_{res} \cdot e_{resource})$ is proportional to the number of agents that we see in the simulation. Increasing or decreasing the density of population we obtain effects similar to the variation of the locality range of selection.

Keeping constant the product $(k_{res} \cdot e_{resource})$ and altering the $\frac{k_{res}}{e_{resource}}$ ratio we observe some very interesting dynamics.

When we increase k_{res} and proportionally decrease $e_{resource}$, the actual dynamics of the selection mode (fitness proportional) slowly changes, becoming similar to a truncated ranking selection mode. This has a really bad effect on the genetic variability of the genotypes and make the search process more prone to be trapped in a local optimum.

On the other hand, decreasing k_{res} and proportionally increasing $e_{resource}$, the actual dynamics of the selection mode (fitness proportional) slowly changes, becoming similar to a random selection mode (the selected agents are completely randomly chosen, regardless to their fitness values). This is a even worst scenario than the one of the previous case and all the dynamics of the system fail.

maxmutation **:** the maximum genetic mutation applied to each axis of the genetic space to the genotype of a newborn agent.

A really independent parameter, the variation of *maxmutation* alters the velocity of the search process in the genetic space for optima of the fitness function. The optimal value of *maxmutation* depends on the particularity of the defined fitness function. An increase in *maxmutation* promotes speciation events but, on the other hand, it causes a decrease of the genetic cluster cohesion and a generalized decreased rate of successful mating. If set to a value too high it destabilizes the system.

$e_{metabolism}$ **:** the energy decrease in metabolism.

e_{mating} **:** the minimum energy level that an agent must have in order to be able to mate.

$e_{matingattempt}$ **:** the energy spent by an agent for a mating attempt.

$e_{newborn}$ **:** the starting level of energy of a newborn agent.

This set of parameters regulate the life cycle of the agents. An increase in $e_{metabolism}$ not only prolongs the expected life (in term of number of iterations) of an agent, but, if not counterbalanced by a variation in the energy received by winning a competition for a resource, increases the number of living agents. The $e_{newborn}$ must be set to be $e_{newborn} \leq (2 \cdot e_{matingattempt})$ otherwise every successful mating adds energy to the system and it can eventually diverge. The lower $e_{newborn}$ is compared to $e_{metabolism}$, the higher is the infant mortality of agents. Increasing e_{mating} not only increases the time, in terms of number of iterations, before a new generation starts to have children, but simulates an increase in the selective pressure. If, for instance, we set $e_{mating} = 0$ (or, better, equal to $e_{matingattempt}$) then every agent tries to have children; the effect is that even the agent with the lower fitness level becomes able to generate an offspring (if they obtain any resource the energy is invested in children) and the selection focuses more on who starves than on who does not generate offspring. An increase in $e_{matingattempt}$ makes stronger the effect of the variable fertility, because every failed mate attempt costs more. A $e_{matingattempt}$ decrease on the other hand nullifies the effects of variable fertility: the mono-population areas disappear and no speciation occurs.

startingagents **:** the size of the starting agent population.

startingenotype **:** the genotype common to every agent of the starting agent population.

startingenergy **:** the energy level common to every agent of the starting agent population.

The *startingagents* is of no relevance because the size of the agent population of the system changes to a level depending to other parameters (the maximum average number of agents will be $\frac{k_{res} \cdot e_{resource}}{e_{metabolism}}$ but the effect of variable fertility puts the habitual number of agents under such a limit). *startingenergy* is of no relevance too. *startingenotype* depends to the fitness function definition.

4.4.4 Model Minimality and Mechanisms Dropped From the Minimal Model

The process of model designing was oriented to develop not only a model that enables the emergence of sympatric speciation, but a minimal model with such an emergent property.

Calling it minimal, we mean that the model is intended to be the simplest necessary, in terms of number of constitutive base mechanisms, in order to obtain the emergence of sympatric speciation.

We are not able to prove that a simpler but still operational model does not exists; anyway, our design process, considering the analysis of several different more complex model we initially developed, consists in tests of alternative choices of pruning of the base mechanisms of these models.

We can assume by experimental observations that, with the limited consideration of only the mechanisms or properties we analyzed, the proposed model is the simplest in terms of constitutive base mechanisms.

Essentiality of the Mechanisms Adopted in the Model

We can make a differentiation between the base mechanisms and the properties that emerge from the interaction of the mechanisms composition (system level emergent properties).

The emergent property whose presence define if the model is operational is the speciation in absence of any direct isolation of the agent population.

The essential emergent properties we must observe in the model in order to indirectly divide the agent population are any form of geographical locality in the selection and the spontaneous formation of mono-population areas in the the geographical space.

An alternative, with a different essential property and without the need to have a geographical space, is a form of genetic locality in the selection. This setting can eventually enable the formation of clusters of agents in the genetic space, but when the clusters become distant then any form of interaction between the two cluster disappear (on the other hand in our model if two agents are near in the geographical space they not only compete but there is the possibility of interbreeding). In the case of a residual interaction, then even if the speciation is possible, there is a lack of diversity maintinmnet. This is a bad feature if we consider the model to be implemented in an evolutionary algorithm (it removes the main advantage of evolutionary search over the standard search approach), but even if we are only interested in the theoretical biology value of the model, then it is not coherent to biological observations.

We have already discussed in the previous model description sections the mechanism of geographical locality in the selection, how it works, its effects and the influence of the various parameters.

We have already discussed the spontaneous formation of mono-population areas in the the geographical space too, and we have found that a single feature can not only enable the formation of these areas, but, when they are formed, it has the main and essential role in the preservation of such areas. This base mechanism is the variable fertility and in the previous sections we have already described how it works, its effects and the influence of the various parameters.

Every component or elemental (non emergent) mechanism present in the described model, if removed or destabilized (destabilization that can be obtained by parameter choice, already discussed in details), suppresses the locality of selection or the formation of mono-population areas, or both; the resulting system is not operational.

Sexual Diversity

In the first developed non minimized model the agents internal static state was composed by one additional value: the gender.

The gender value could eventually be set to several different values, but, for a simple analysis, in the implemented models it can have only two separate values: male or female.

The gender of an agent was stochastically determined when an agent is producted by a successful mating.

The only role of the gender is the fact that every selected pair of mating agents needs to be composed by a male agent and a female agent.

This feature of the model allowed the exact implementation of a gender-dependent variability of the fertility, as in the case study of the *M.savii.*

We did not observe any qualitative variation in the system dynamics when we removed this feature.

Our hypothesis is that, in order to exploit the mechanism of sexual diversity, we need to have a diploid-like structure of genotypes.

We observe that sexual diversity is not an essential feature in order to obtain any mode of speciation in our simulation models.

Gestation Period

One of the various features tested in our models is the gestation period feature.

The gestation period is the time, expressed as a number of iterations, that it takes to a pregnant agent to release the newborn agent in the geographical space.

When a mating event is successful then the newborn agent is suspended from the simulation for the gestation period, and one of the agents of the mating pair becomes pregnant. A pregnant agent cannot set itself as enabled to further mating.

At the end of the gestation period the newborn is positioned in the geographical space in the same position of the pregnant "mother" agent, that returns to be non-pregnant.

An additional parameter is needed and is the length, expressed as the number of iterations, of the gestation.

We did not observe any interesting change the in system dynamics attributable to the effect of the gestation period. The only (not interesting) observed effects are a general slow down of the model, a decreased reproduction rate (if a pregnant agent dies before the end of the gestation period then the child agents is deleted too) and stronger fluctuations in the number of the living agents.

Diversity Maintaining: Decreased Fertility for Hybrid Agents

In the case study of *M.savii* the role of diversity maintaining when several mono-population areas are formed is ascribed to the presence, in the boundary zones, of a population of hybrid agents with decreased fertility [13] (they can still mate with every agent but the probability to produce any offspring is decreased even if the mating pair is composed to genetically identical hybrid agents).

In order to give this mechanism to the model we need to make explicit and functionally make use of the notion of species. We need to identify the different genetic clusters in the genetic space if we want to calculate if a newborn is to be considered an hybrid with decreased fertility or not.

This problem is really hard because if we make use of the cluster notation and identification in order to regulate the system dynamics, any possible cluster formation or split can be considered as attributable to this feature and not a emergent property.

Luckily, it turned out that the variable fertility essential mechanism is able to maintain the diversity, without any need of species notation. The decreased fertility for hybrid agents does not add anything to the model and can be dropped.

Aging

The aging feature consists of the addition of the age value in the internal dynamical state of agents.

In every iteration the ages of all the agents in the model are increased by one and the age of a new agent is set to zero.

The age of an agent has a role in the metabolism simulation phase: the metabolic energy paid by an agent is increased by the age multiplied by the aging parameter.

The effect of the aging in the model is to increase the selective pressure for an agent at the increasing of its age. The observed mutation of dynamics is the promotion of the generational replacement of agents.

This feature has very interesting effects and implications, but it is not essential to the model.

Sexual Maturity

Like for the aging mechanism, for the sexual maturity mechanism we need the age value for every agent, initially set to zero and increasing at each iteration of the simulation. But there is not any increase of the metabolic energy.

An additional parameter is needed and it's the minimal age, expressed as the number of iterations, that an agent needs to have in order to become able to participate to the mating phase.

This mechanism turned out to be completely useless, causing only a slow-down of the system dynamics.

Asexual Reproduction

In the described model we make use only of the sexual reproduction. This does not mean that we make use of any gender notation, but that, in order to generate a new agent, there is the necessity to have two agents and that the genotype of the newborn is a mutated composition of its parents genotypes.

We refer to asexual reproduction as the capability of a single agent to generate a new clone agent (we call it clone but it has a mutated copy of its parent genotype).

We can use the asexual reproduction as the only possible reproduction mode, or we can use both sexual and asexual (a choice that is quite popular in the evolutionary algorithms literature).

With only the asexual reproduction the effect of variable fertility is removed: the system diverges to a single population and, even if a speciation event occurs, the resulting multi-population system is so strongly unstable that it converges back to single population.

Mixing the sexual and the asexual reproduction modes, we obtain the only effect of the weakening of the diversity maintaining effect of the boundary zones that divide the mono-population areas.

Chapter 5

Applications of the Sympatric Speciation Model to Evolutionary Algorithms

The model for sympatric speciation, described and discussed in the previous chapter, can have some relevance in the field of theoretical biology, but the final use we want to make of it is not in the biological field: it's intended to be applied to evolutionary algorithms.

Moreover, such a modeling was done only because the sympatric speciation mode is a crucial component of our hypothesis about negative feedbacks emergence. The negative feedback emergence hypothesis will be discussed in chapter 6.

We have studied several variations of the sympatric speciation model and several variations of single mechanisms (for instance, the genetic distance measure) in order to show the usefulness of the application of our partial work to the evolutionary algorithms.

The first result we propose in this chapter is an approximation of the distance measure for the search space. The peculiarity of the biologically inspired approximation we propose is its applicability to every possible search space, with a sort of approximation that does not invalidate the measured distances for our particular purposes. This result is important because a distance measure for the search space is frequently used by the works found in evolutionary algorithms literature and because the search space can be extremely complex.

A second result we propose is a variation of the island model in order to solve the problem of the necessity to statically define the number of species in the design phase of the system.

The third and last result we propose is a technique that promotes alternative local optima search, through the destabilization of mono-population groups of island that emerges from the previously described model, coupled with a system global level stabilization that prevents the search from stopping.

5.1 An Approximated General Method for Genetic Distance Measurements Independent of the Search Space

The notion of distance in the search space is recurrent in evolutionary algorithms literature.

The measure adopted in a particular work is usually a problem dependent choice, a choice that sometimes, when the search space is complex [45] [92], brings out to be a very hard problem.

The needs for a definition of a distance measure metric can be various: it can be needed in order to evaluate the genetic diversity [14], but there can be some base mechanisms of the algorithm that depend on it [12].

5.1.1 The Problem

The most general and powerful search space widely used in literature is probably the space of imperative programming language codes [45]. Other widely used complex search spaces are the space of electronic circuits [92] and some type of algebraic functions spaces.

In these cases of complex spaces, a metric to use for the measurement of distances between two search points is really difficult to define.

When the problem is too much hard or if the effects of an arbitrary choice are not clear in terms of search outcomes, it's a common practise to approximate the genetic distance measure replacing it with a phenotypical diversity measure (the difference between the fitness values of the two searching points).

The usual approximation of using the difference between the fitness values can result to be an acceptable approximation in the case of mono-population systems, but it's obviously a wrong choice in the more interesting and powerful case of the multi-population systems.

5.1.2 Biological Observations

In the previous discussion about the biological observations that inspired the introduction of the variability of the fertility mechanisms, we cited some debates where the relationship between the genetic isolation and the fertility are discussed.

The outcome of the debates is that the reproductive fertility is a fuzzy property whose value depends on the genetic distance between the mating organisms.

It is a result of works in the field of genetic that, in order to evaluate the genetic reproductive compatibility of two organisms, for some particular cases of study it turns out that only a partial fraction of the genotypes is found to play a role: the concordance of the longer part of the genotypes is less important in order to quantify the reproductive compatibility.

5.1.3 The Proposed Solution

In order to obtain an approximated genetic distance measure, independent of a particular search space, we propose the addition, to every agent static internal state or to every search point, of a sequence of bits of a fixed length l_{sg}, that will be referred to as *simulacrum genotype*.

Every agent in the starting population or every search point in the starting search set has the same simulacrum genotype, for example a string of l_{sg} zeroes will do the work.

When a sexual reproduction successfully generates a new agent or a new search point, the new simulacrum genotype is obtained from the recombination of its parental simulacrum genotypes, then a point mutation (bit switch) can occur with a probability p_{sg}.

The recombination z of two simulacrum genotypes x and y is carried out in the simplest way possible: for each bit $z[i]$ of the string we randomly pick the $x[i]$ or the $y[i]$ bit, with an equal probability.

The proposed approximated distance between two agents or search points is the edit distance of the two simulacrum genotypes.

5.1.4 Discussion

The proposed solution can seem a very naive approach, anyway the simple mechanics of the simulacrum genotype are able to mimic very well the accumulation of mutation that occurs in the search process over the search space.

The edit distance of two simulacrum genotypes approximates the distance, in terms of generations, from their nearest common ancestor.

The two parameters l_{sg} and p_{sg} can be tuned for every particular need. For example a very long bit string and a p_{sg} near to 1 decrease the approximation of the nearest common ancestor distance.

We observed that, when adopting this approximated distance in replacement of the exact genetic distance used in the previously described sympatric speciation model, all the system properties appeared unchanged.

We want to stress that the adoption of the described approximated distance not only saves us from the problem of a complex search space, but, in comparison to a less approximated *ad hoc* solution over complex spaces, it is probably more time complexity cost friendly, and, compared to the habitual approximation of making use of the difference between the fitness values, it works well also in the case of multi-population systems.

5.2 A Modified Island Model

The *island model* [94] is the most popular model used for the multi-population approach of evolutionary algorithms.

We want to propose a variation of the island model adopting techniques taken from the sympatric speciation simulation model described in chapter 4.

The typical evolutionary algorithm is a probabilistic search algorithm that maintains a population of points in a search space, evaluates the goodness of such search points thanks to a fitness function, and iteratively does:

1. Selection of a representative set of points in the population, through a selection method that makes use of the fitness values of the points.
2. Generation of a new population from the selected representatives, through a reproduction method.
3. Mutation of each point of the new population.
4. Replacement of the old population.

The iteration is repeated until a stop condition is reached.

When an evolutionary algorithm makes use of the island model there is an additional element: the island graph. An island graph is a graph where the vertexes are the islands and the edges are the possible ways of migration from an island to another. Additionally, in the island model case there are several candidates populations, one for each island.

In order to better describe the system, we label the number of islands $n_{islands}$ and give an index to each island, each cadidate population is labeled A_i where i is the index of the corresponding island. The migration rates defined for each pair of islands are labeled m_{ij} and they specify the number of search point that migrate from island i to island j in the migration phase. Even if the value of m_{ij} can be different from m_{ji}, usually the total number of points migrating toward an island is equal to the number of points that leave the island. The size of each A_i is a constant value $|A_i| = pop_i$.

An iteration of the probabilistic search algorithm becomes:

1. For each island i:
    (a) Selection of a representative set of points in the population A_i, through a selection method that makes use of the fitness values of only the points in A_i (in other words the selection is local to the single island).
    (b) Generation of a new population from the selected representatives through a reproduction method.
    (c) Mutation of each point of the new population.
    (d) Replacement of A_i with the new population.
2. For each island pair (i, j), m_{ij} randomly selected search points are removed from A_i and added to A_j (the random selection does not consider points that have already migrated in the same iteration).

Figures 5.1 and 5.2 show a run of our implementation of the previously described island model. In this example the search space is a segment of $\mathbb{R}$.

On the left part of the figures are represented the islands and the search populations. Each small group of 40 coloured points represents a particular island population (there are 400 islands in this example) and each coloured point represents the position in the search space of a particular search point.

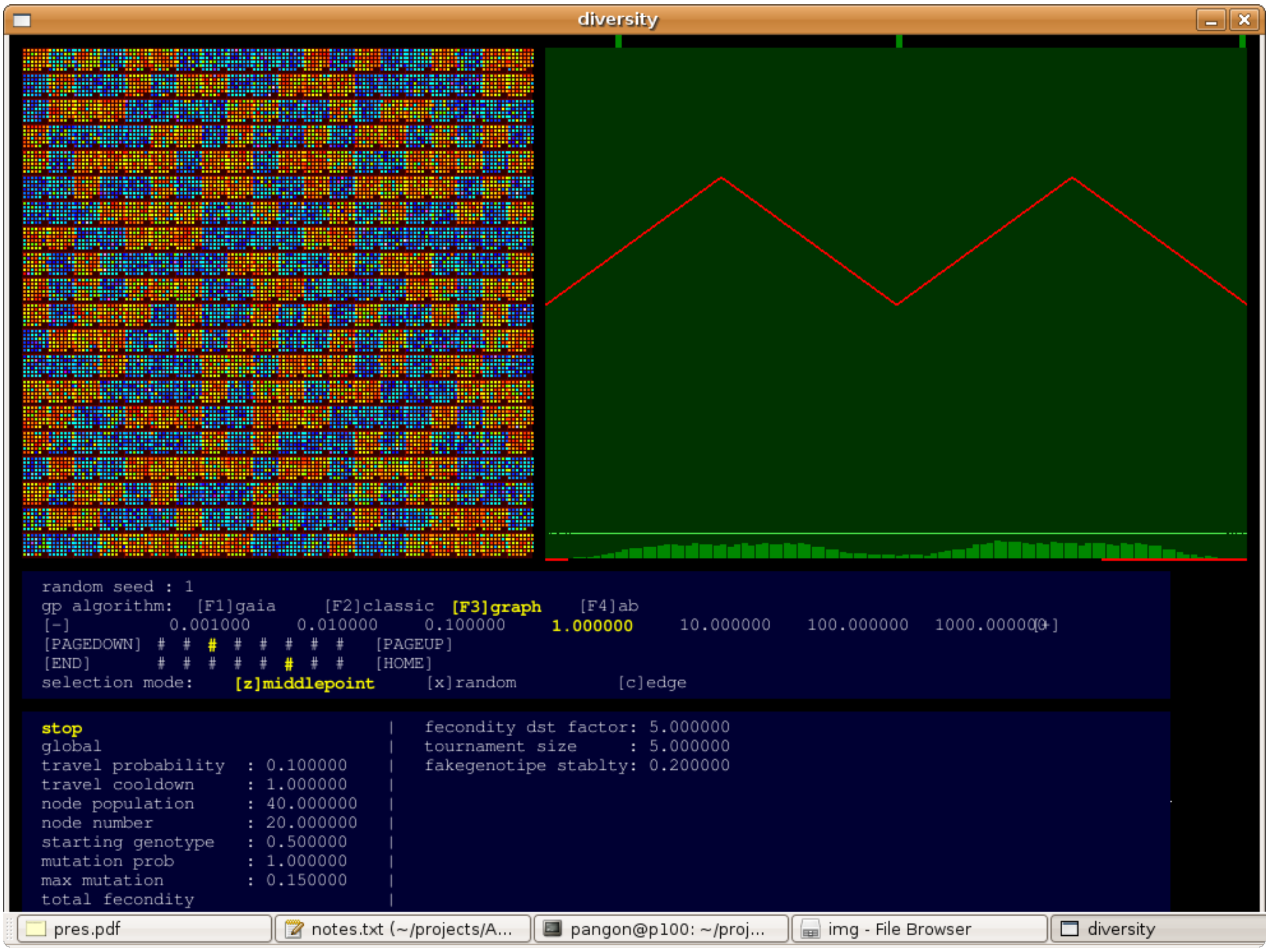
The right part of the figures represents the distribution in the genetic space of the genotypes. The green columns represent the number of genotypes and the red line is a plot of the fitness function.

Figure 5.1 shows that, if the starting population is set with a genotype placed in a point of the search space where two different directions are fitness increasing, then in the island model some islands can genetically go to one direction and some in the other one.

In figure 5.2 we can see that, even if initially a speciation can occur, in absence of any diversity maintaining mechanism the system is unstable and one of the two species quickly disappears.

5.2.1 The Problems

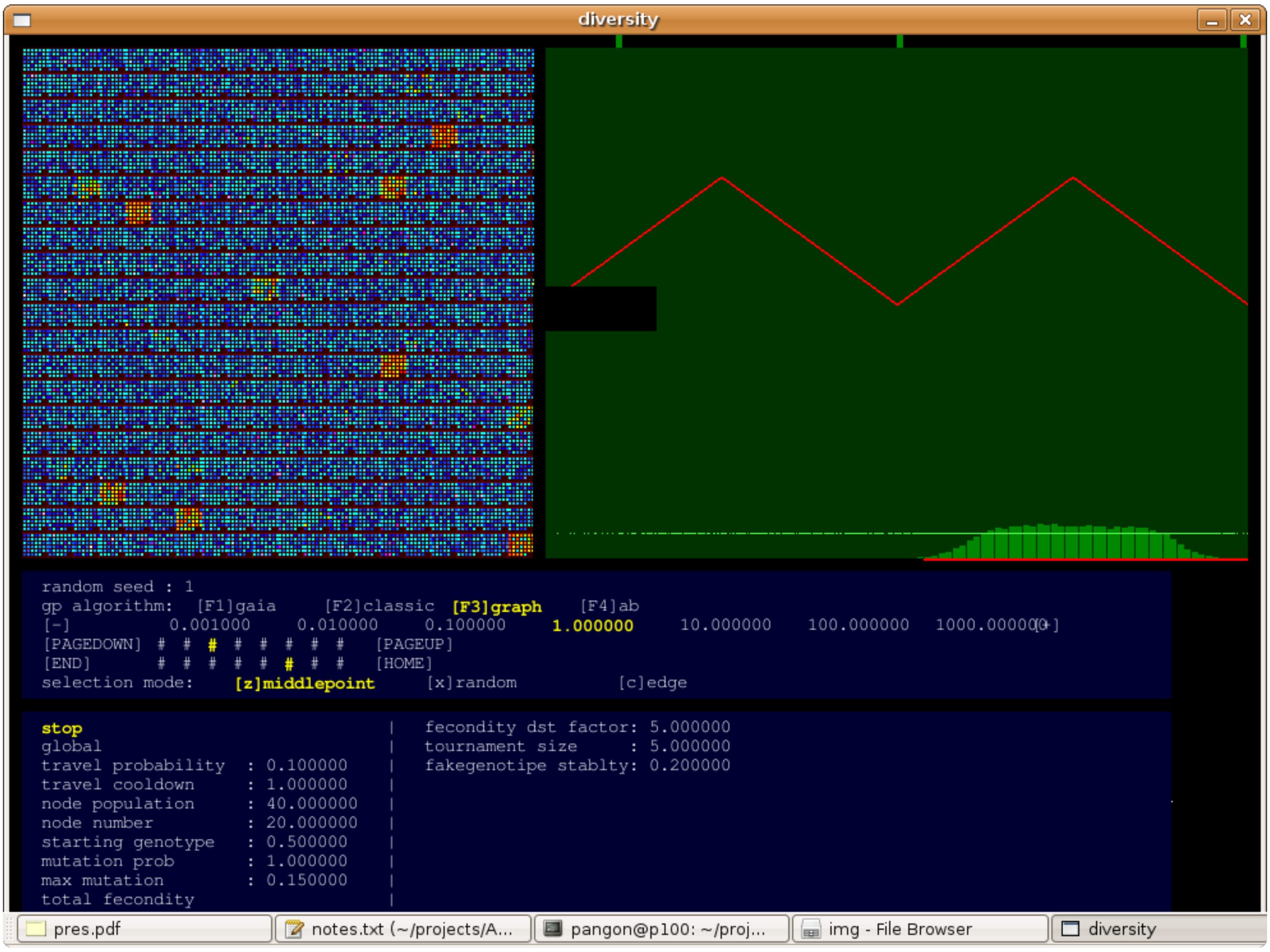
Even if our main and starting motivation of the adaptation of our sympatric speciation model to the island model is the emergent property of diversity maintaining, we realized that some other recurrent problems of the island model are solved by our approach.



```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph    [F4]ab
[-]            0.001000     0.010000     0.100000     1.000000     10.000000     100.000000    1000.000000[+]
[PAGEDOWN]  #  #  #  #  #  #  #  #   [PAGEUP]
[END]       #  #  #  #  #  #  #  #   [HOME]
selection mode:    [z]middlepoint     [x]random          [c]edge

stop                                |   fecondity dst factor: 5.000000
global                              |   tournament size     : 5.000000
travel probability  : 0.100000      |   fakegenotipe stablty: 0.200000
travel cooldown     : 1.000000      |
node population     : 40.000000     |
node number         : 20.000000     |
starting genotype   : 0.500000      |
mutation prob       : 1.000000      |
max mutation        : 0.150000      |
total fecondity                     |
```

Figure 5.1: An example of execution of a variation of the sympatric speciation model. The left-sided square in the figure represents the islands and the search points, while the right-sided one represents the genetic space genotype distribution.



```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph     [F4]ab
[-]           0.001000     0.010000     0.100000     1.000000     10.000000     100.000000    1000.000000[+]
[PAGEDOWN]  #  #  #  #  #  #  #  #   [PAGEUP]
[END]       #  #  #  #  #  #  #  #   [HOME]
selection mode:    [z]middlepoint      [x]random           [c]edge

stop                                  |   fecondity dst factor: 5.000000
global                                |   tournament size     : 5.000000
travel probability   : 0.100000       |   fakegenotipe stablty: 0.200000
travel cooldown      : 1.000000       |
node population      : 40.000000      |
node number          : 20.000000      |
starting genotype    : 0.500000       |
mutation prob        : 1.000000       |
max mutation         : 0.150000       |
total fecondity                       |
```

Figure 5.2: The same run of the Figure 5.1: even if initially a speciation can occur the system is unstable and one of the two species disappears.

Diversity Maintaining

The main problem we want to resolve in the application of the multi-population evolutionary algorithms to problems with alternative solutions, is the fact that, even if speciation is possible, in absence of any diversity maintaining mechanism the system is unstable and one of the two species quickly disappear.

The problem is well visualized by the example run of figures 5.1 and 5.2.

The Number of Islands

A recurrent problem, widely discussed in the field of multi-population evolutionary algorithms, is the definition of the $n_{islands}$ parameter.

The standard approach is to define it in the design phase of the model.

When the only purpose of a multi-population approach is to enhance the search by an additional competition element (if the migrants have a very high fitness they can destroy the native population of an island, invading the niche with their offspring), then a priori defined number of islands can be an acceptable solution.

But if the role of the islands is to support the divergence of several alternative clusters in the search space (in this case the role of the migration is to transfer solution structures from the other clusters, not to enhance competition, so invasion is usually avoided) the decision on the number of islands needs to be related to the expected number of alternative solutions of the problem, an evaluation usually impossible to make.

In the literature there are several solutions able to dynamicize the number of islands.

Islands Connections and Migration Rates

Other recurrent problems are the choices of the connectivity of islands and of the migration rates [6] [83].

When the number of islands is high the structure of the island graph and the rates of migration for each edge can be, depending to the problem, a crucial problem.

Some works adopt a full connectivity, some others set the islands in a chain, some others propose an island graph structure specifically designed for the particular problem to solve.

5.2.2 The Approach

We propose a solution to the described problems through the adaptation of the sympatric speciation model to the island model.

Adapting our sympatric speciation model to the island model we mean to enable the properties emergence by the reproduction of the features of the speciation model. The features or components that do not have a counterpart in the island model will be added to the modified speciation model, but for the other ones that have a direct counterpart, we need to found a solution that enables not only the emergence of speciation but also the search capabilities of the agent model.

From now we use again the agent paradigm and the genetic space is the search space. An agent is not a solution: its genotype is the solution. In the examples the genetic space is $\mathbb{R}$ (not $\mathbb{R}^2$ as in the basic sympatric speciation model description).

In order to measure the distances between two genotypes we adopt the Approximated General Genetic Distance previously described in section 5.1.

We compare the various alternative variations of the sympatric speciation model with the standard island model approach.

5.2.3 Locality

As previously stated, the island structures can be viewed as a very simple geographical space, where the agent movement is restricted to a migration from an island to another one. This type of geographical space enables a locality of selection: in the selection step of the island model a solution competes only against a local portion of the totality of the solutions.

In our model there are three basic activities of agents that have a locality component:

- Competition;
- Mating;
- Movement.

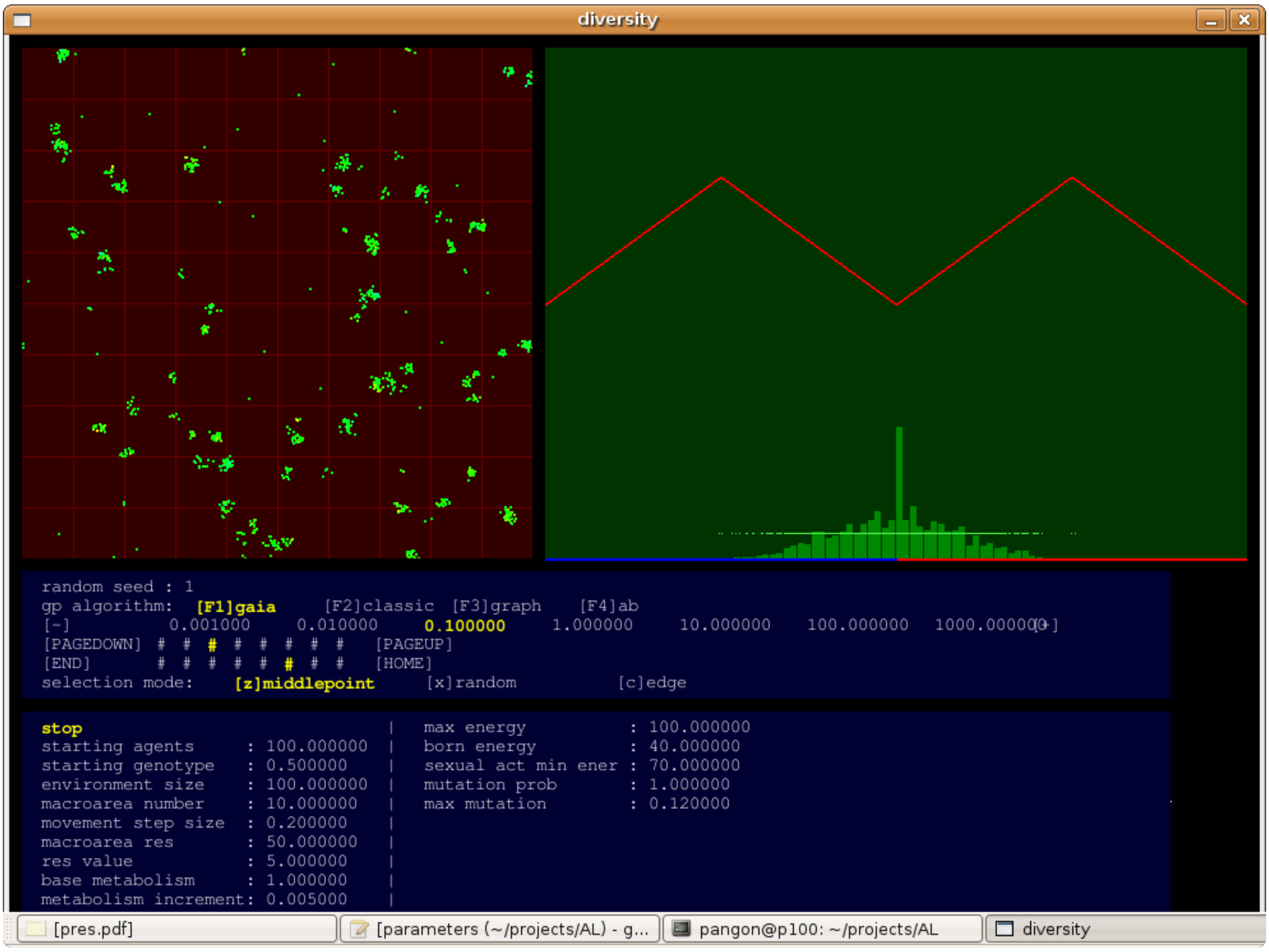
We have already observed experimentally that there is not a notable variation of the emerging properties of the model if the locality of the competition is different from the locality of mating, therefore we reduce to only two types of locality.

We tested several alternative variations from the sympatric speciation model of the geographical space and of the two described types of locality.

The first variation we tested is:

- Geographical Space: keeping the geographical space as a toroidal $\mathbb{R}^2$ surface (we described the sympatric speciation model with square surface in $\mathbb{R}^2$, but the dynamics are almost the same in the toroidal unlimited surface) but with a grid-like segmentation in macro-areas of fixed size.
- The locality of movement: keeping the movement in the geographical space as described in the sympatric speciation model.
- The resources are not placed on a point in $\mathbb{R}^2$, but assigned to macro-areas. Every macro-area gets the same fixed amount of resources at each iteration.

- The locality of competition: the agents populating a particular macro-area compete only for the resources locally assigned.
- The locality of mating: all the agents populating a particular macro-area are considered to be in the geographical mating range that enables couple random selection, but if two agents are in two different macro-areas, even if they are geographically very near, they are considered out of the geographical mating range.

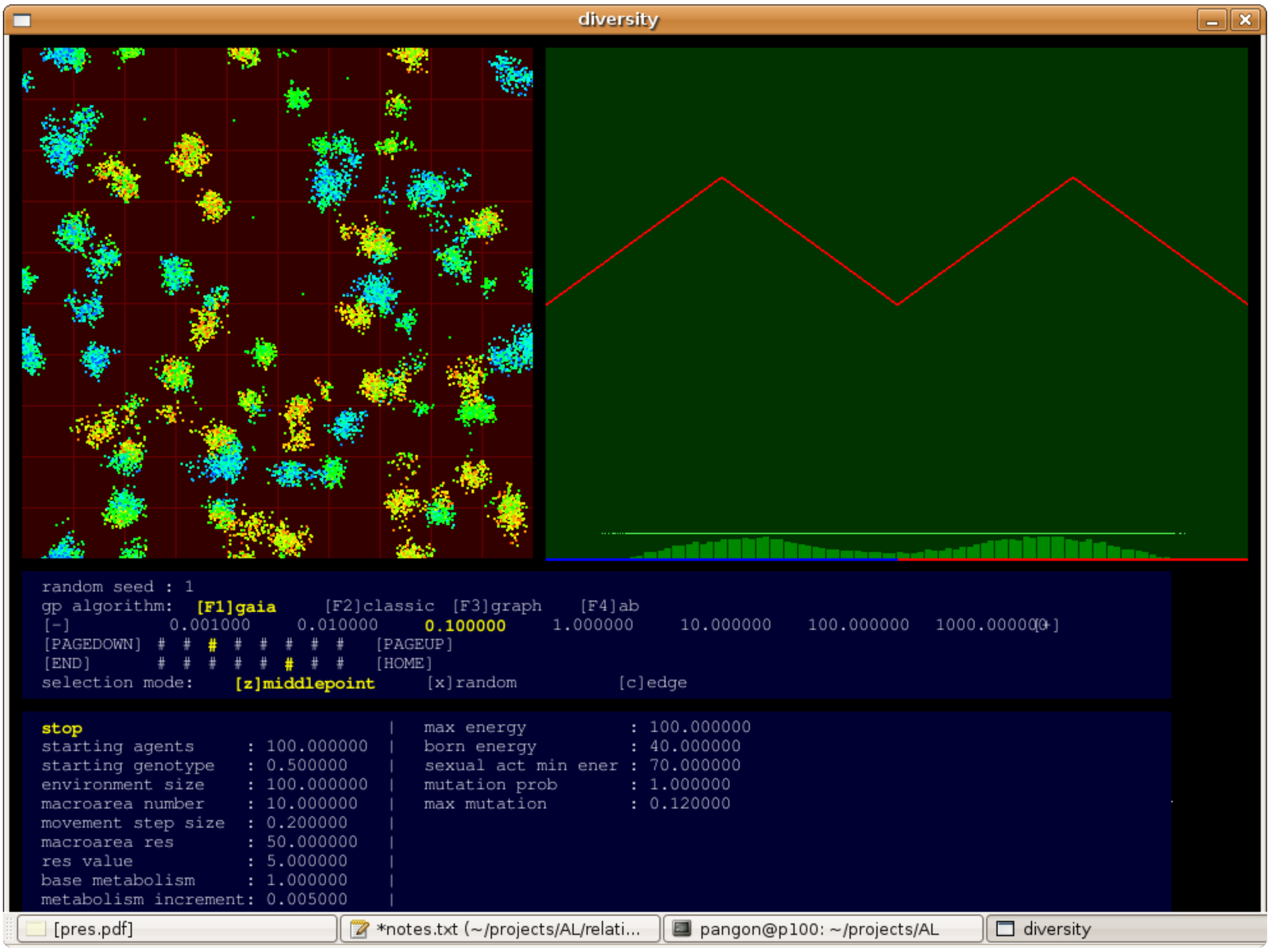


```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph    [F4]ab
[-]          0.001000     0.010000     0.100000     1.000000     10.000000     100.000000    1000.000000[+]
[PAGEDOWN]  #  #  #  #  #  #  #  #    [PAGEUP]
[END]       #  #  #  #  #  #  #  #    [HOME]
selection mode:     [z]middlepoint      [x]random           [c]edge

stop                               |  max energy          : 100.000000
starting agents      : 100.000000  |  born energy         : 40.000000
starting genotype    : 0.500000    |  sexual act min ener : 70.000000
environment size     : 100.000000  |  mutation prob       : 1.000000
macroarea number     : 10.000000   |  max mutation        : 0.120000
movement step size   : 0.200000    |
macroarea res        : 50.000000   |
res value            : 5.000000    |
base metabolism      : 1.000000    |
metabolism increment: 0.005000     |
```

Figure 5.3: An example of execution of a variation of the sympatric speciation model. The left-sided square in the figure represents the geographical space with dots representing the agent positions, the right-sided one represents the genetic space genotype distribution.

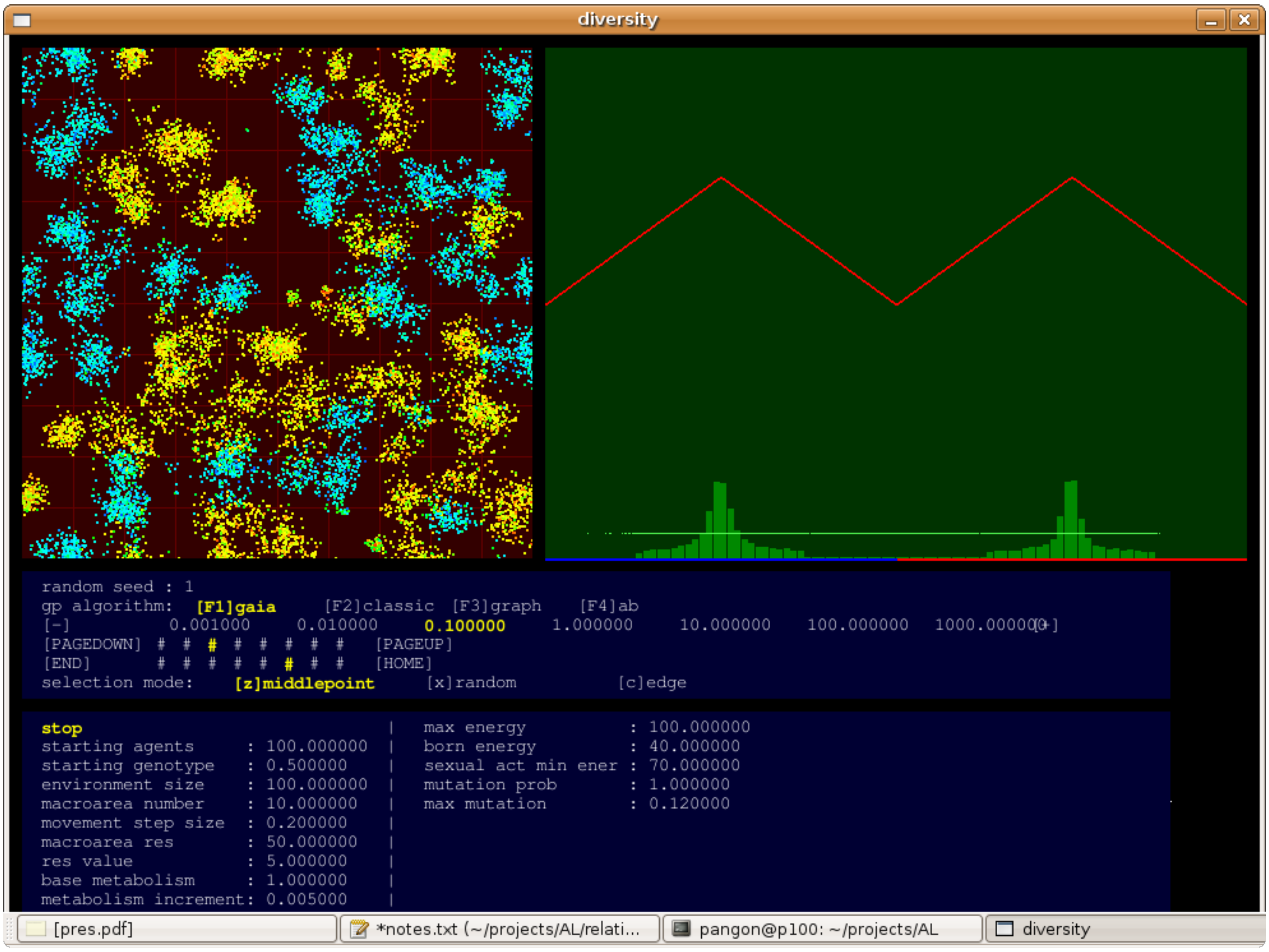
Figures 5.3, 5.4, 5.5 and 5.6 show the dynamics of this variation of the speciation model.



```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph    [F4]ab
[-]           0.001000     0.010000     0.100000     1.000000     10.000000     100.000000    1000.000000[+]
[PAGEDOWN]  #  #  #  #  #  #  #  #    [PAGEUP]
[END]       #  #  #  #  #  #  #  #    [HOME]
selection mode:     [z]middlepoint      [x]random            [c]edge
```

```
stop                                 |   max energy            : 100.000000
starting agents      : 100.000000    |   born energy           : 40.000000
starting genotype    : 0.500000      |   sexual act min ener : 70.000000
environment size     : 100.000000    |   mutation prob         : 1.000000
macroarea number     : 10.000000     |   max mutation          : 0.120000
movement step size   : 0.200000      |
macroarea res        : 50.000000     |
res value            : 5.000000      |
base metabolism      : 1.000000      |
metabolism increment: 0.005000       |
```

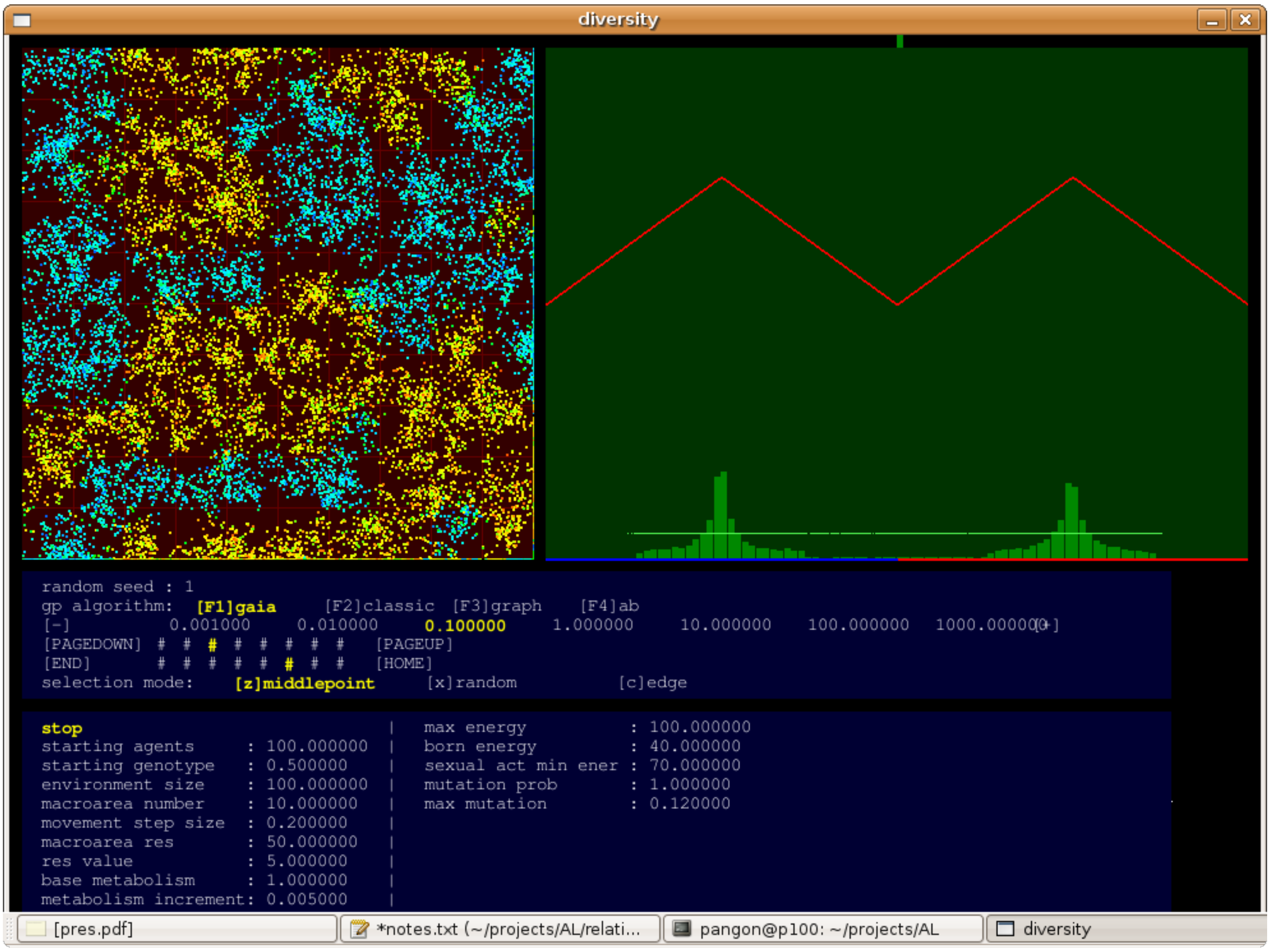
Figure 5.4: The same run of Figure 5.3: the explosive invasion of free islands.



```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph    [F4]ab
[-]           0.001000     0.010000     0.100000     1.000000     10.000000     100.000000    1000.00000[0+]
[PAGEDOWN]  #  #  #  #  #  #  #  #    [PAGEUP]
[END]       #  #  #  #  #  #  #  #    [HOME]
selection mode:     [z]middlepoint     [x]random           [c]edge
```

```
stop                                 |   max energy            : 100.000000
starting agents      : 100.000000  |   born energy           : 40.000000
starting genotype    : 0.500000    |   sexual act min ener : 70.000000
environment size     : 100.000000  |   mutation prob         : 1.000000
macroarea number     : 10.000000   |   max mutation          : 0.120000
movement step size   : 0.200000    |
macroarea res        : 50.000000   |
res value            : 5.000000    |
base metabolism      : 1.000000    |
metabolism increment: 0.005000    |
```

Figure 5.5: The same run of Figure 5.3: the hybrid agents are uniformly distributed in the macro-area, not on the boundaries.



```
random seed : 1
gp algorithm:  [F1]gaia     [F2]classic  [F3]graph    [F4]ab
[-]            0.001000     0.010000     0.100000     1.000000     10.000000    100.000000   1000.000000[+]
[PAGEDOWN]  #  #  #  #  #  #  #  #   [PAGEUP]
[END]       #  #  #  #  #  #  #  #   [HOME]
selection mode:     [z]middlepoint      [x]random          [c]edge

stop                              |  max energy           : 100.000000
starting agents      : 100.000000 |  born energy          : 40.000000
starting genotype    : 0.500000   |  sexual act min ener : 70.000000
environment size     : 100.000000 |  mutation prob        : 1.000000
macroarea number     : 10.000000  |  max mutation         : 0.120000
movement step size   : 0.200000   |
macroarea res        : 50.000000  |
res value            : 5.000000   |
base metabolism      : 1.000000   |
metabolism increment: 0.005000    |
```

Figure 5.6: The same run of Figure 5.3: the mono-population areas emerge.

The left-sided square in the figures represents the geographical toroidal space and the colored dots are the agent geographically positioned; each dot in the geographical space is colored in dependence of the agent genotype position in the genetic space, this enable a mapping from the two spaces. The red grid shows the geographical space segmentation in macro-areas.

The right-sided square in the figures represents the distribution in the genetic space of the genotypes. The green columns represent the number of genotypes, the red line is a plot of the fitness function.

In the Figure 5.3 is showed the starting phase of the simulation: the dynamics of the model variation in this phase are the same of the sympatric speciation model.

In the Figure 5.4 is showed a phase of the simulation where the first observed particularity of this model variation starts to show its nature: when the first two agents of the population move in a previously free macro-area, the discontinuity in the selective pressure, due to the discontinuity in resource distribution in the geographical space, increases their mating rate to a very high level compared to the level observed in the macro-area they come from. This, in addition to the fact that the newborns are placed almost on the borderline of the two macro-areas, causes an unfair advantage to the particular genotype of the first couple of invading agents in the competition for the first macro-area.

In the Figure 5.5 we can observe that with the macro-area variation no hybrid boundaries emerge in the simulation: when two macro-areas with genetically different populations start sending migrants from one macro-area to the other one, the hybrid agents are uniformly distributed in the macro-area, not on the boundaries. The effect is a decrement of the mating rate in the whole macro-area.

Figure 5.6 shows a quite stable situation. Even if there are not any hybrid boundaries, the mono-population areas emerge, but the boundaries are forced to be on the macro-areas grid lines.

5.2.4 The Adoption of the Islands with Fixed Population Size

Thanks to the previous first variation of the model we are able to see that we can adopt a simplification of the resource system needed in the sympatric speciation model.

The other variations we discuss adopt exactly the same concept of island utilized by the island model:

- There is not any geographical space: two agents in the same island are near but if they are in different islands they cannot interact.
- There is a graph structure of islands and the adopted connectivity of the graph is the more widely used: the total connectivity.
- The movement phase is replaced by the migration phase.

- There is not any energy level in the agent internal state: on each iteration all the agents are replaced by the next generation, composed by the offspring of some agents selected from the original population. The population size has a fixed value for every island.

5.2.5 Variable Fertility and Total Connectivity

We have tested how the base island model changes its dynamics by the only addition of the variable fertility mechanism.

The variable fertility mechanism is the same described in the base sympatric speciation model. In order to measure the distances between two genotypes we implemented a system with the Approximated General Genetic Distance and one with the exact genetic distance.

Figures 5.7, 5.8 and 5.9 show an example of the island populations dynamics: we have almost the same identical dynamics observed in the basic island mode case but with a prolongation of the number of iterations with the two coexisting solutions.

It's hard to isolate the only effect of the variable fertility in this model variation. Using the approximated general genetic distance measure we are able to better understand it.

It can be summarized by a late occurring effect: after the first iterations of the system, that have exactly the same dynamics of the basic island model, when the accumulation of mutations in the simulacrum genotypes (a genetic split typical and distinct for every different island, independently to the position in the search space) in the agent populations has reached a high value, then the effect of invading agents changes.

Before the genetic clusters separation, when a migrant agent arrives in an island, through the mating with native population it can eventually produce several hybrid agents. After the genetic clusters separation, a single agent has some troubles reproducing itself, because the success of a mating with the natives is low and no hybrid agents are born; therefore there is the necessity of at least two distinct migrants of the same genetic cluster and the rate between the number of components of the two different populations tends to increase giving an advantage to the main population.

With the only effect of variable fertility we are not able to reach the level of variability maintaining we have previously obtained in the sympatric speciation model.

5.2.6 Total Fertility and Local Connectivity

One of the basic observations done in the analysis of the sympatric speciation model is that the limited movement capabilities of the agents generate a delay of the global dynamics effects compared to the local effects.

Increasing the maximum movement step that an agent can do in the geographical space, the system suffers a destabilization.

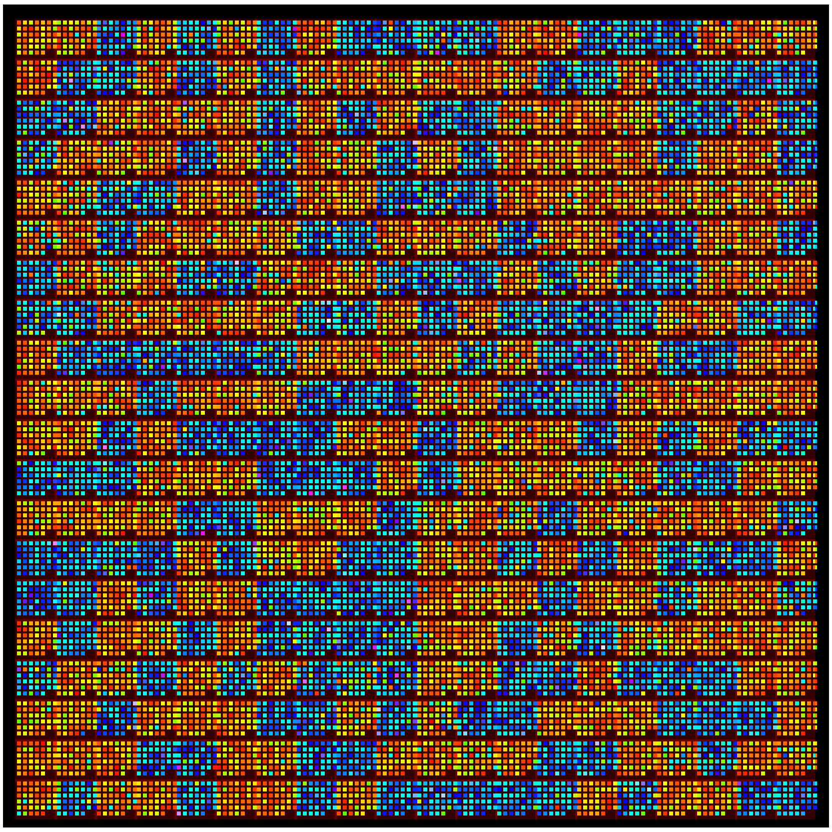


Figure 5.7: An example of a run of the variation *Variable Fertility and Total Connectivity* of the sympatric speciation model. Each group of 40 coloured points represents an island agent population (there are 400 islands). The colour of a represented agent indicates the position of the agent's genotype in the search space.

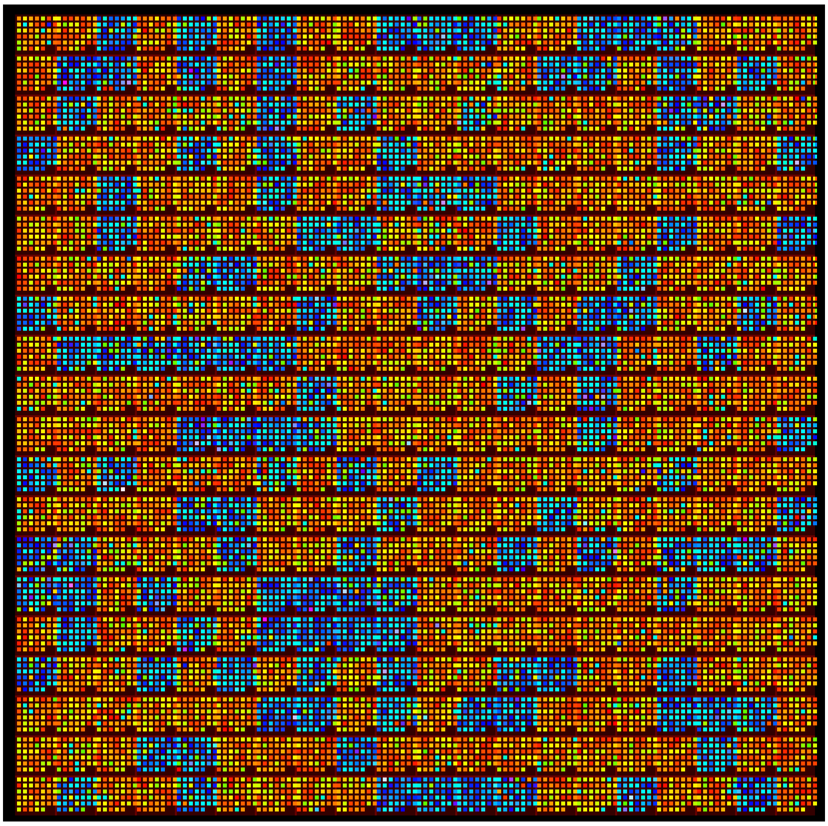


Figure 5.8: The same run of Figure 5.7: the red agents start to dominate.

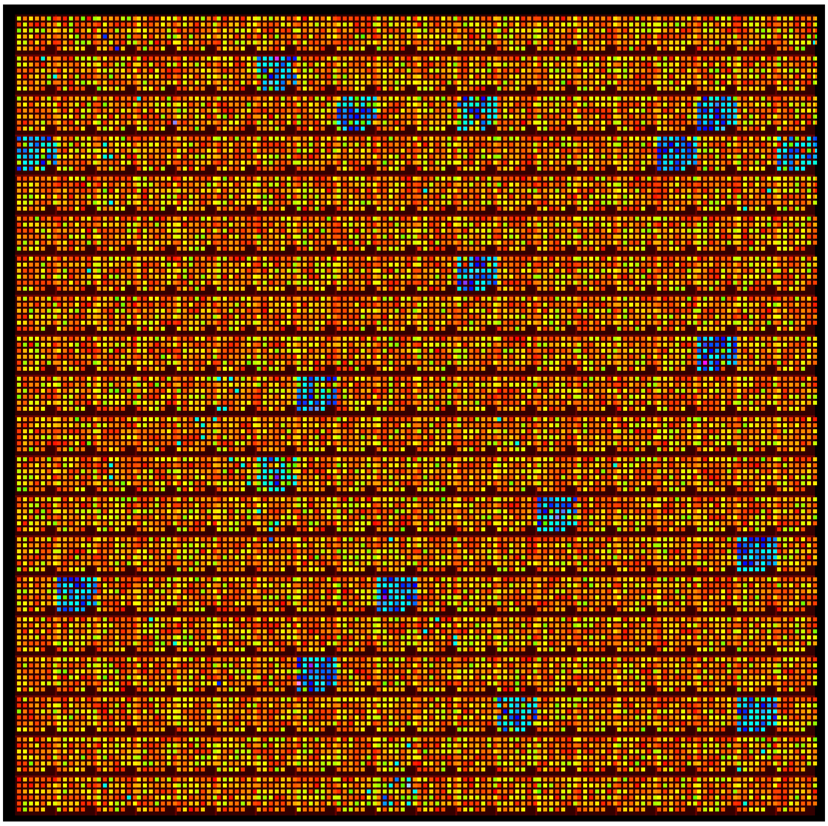


Figure 5.9: The same run of Figure 5.7: the blue agents will disappear.

In order to adopt a form of locality in the movement even if we have dropped the geographical space, we can change the connectivity of the islands: we want that a single migration or a small sequence of migrations of an agent cannot make it move to any possible island in the system, but only to a limited set of local islands (Figure 5.10 is an example).

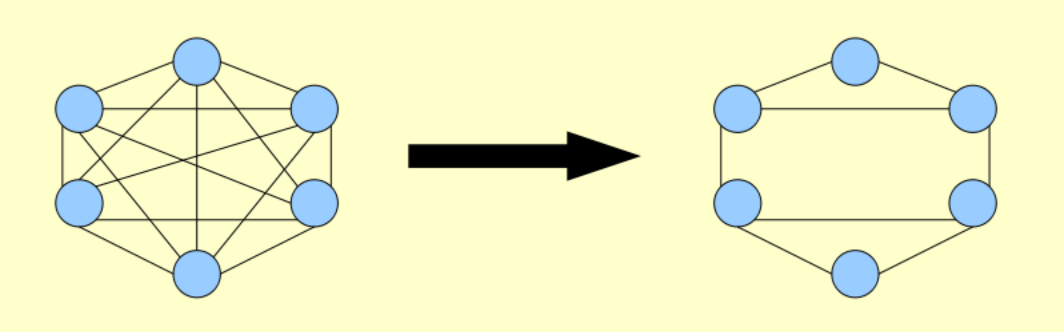


Figure 5.10: In order to obtain a form of locality in the movement, distant islands must exists.

In this particular variation of the model, we adopt a grid-like connectivity of the islands and a travel cool-down (an agent cannot migrate if it has already migrated within a fixed number of iterations; his cool-down is inherited by his offspring).

Figures 5.11, 5.12 and 5.13 show an example of the local connectivity of islands dynamics: we observed the emergence of mono-population areas like in the sympatric speciation model.

The boundaries of the mono-population areas are not defined by hybrid populated zones but are forced to be border islands groups. Unluckily, the border islands do not have the stability showed by the hybrid zones.

Even if mono-population areas emerged, there is not any diversity maintaining mechanism to stabilize the border islands groups, and, after a variable amount of iterations, the system converges to a single type of population in every island.

5.2.7 Variable Fertility and Local Connectivity

In this final variation of the model we adopt both variable fertility and local connectivity of the islands.

We adopt the same grid-like connectivity of the islands previously discussed in the *Total Fertility and Local Connectivity* model variation. We also adopt the travel cool-down parameter (an agent cannot migrate if it has already migrated within a fixed number of iterations, his cool-down is inherited by his offspring) in order to enhance the locality of migrations.

We adopt the same variable fertility mechanism used in the *Variable Fertility and Total Connectivity* model variation.

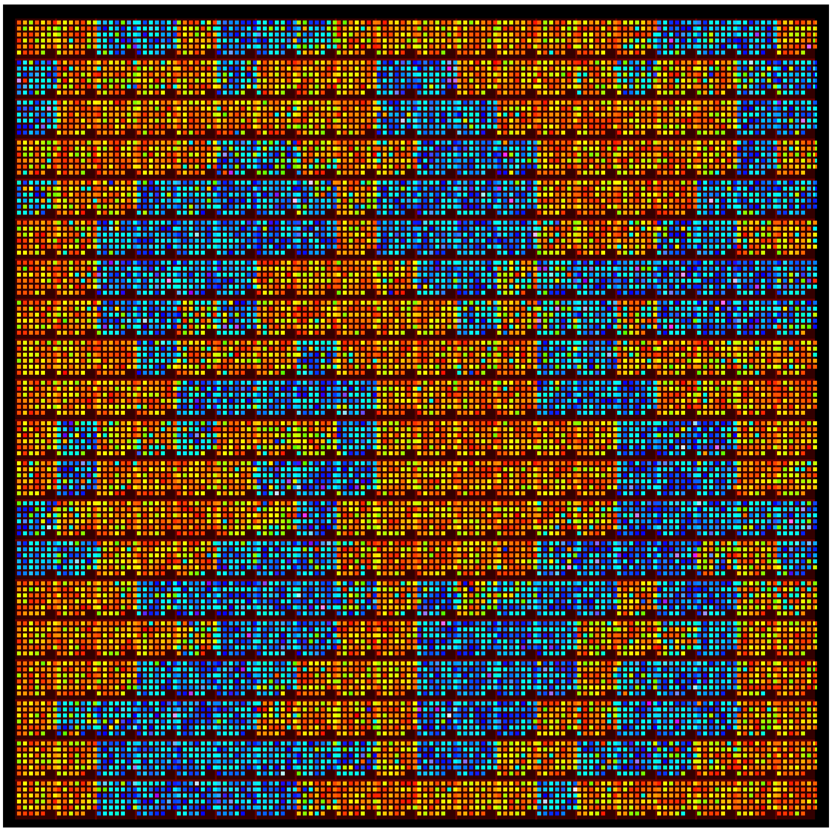


Figure 5.11: An example of execution of the variation *Total Fertility and Local Connectivity* of the sympatric speciation model. Each group of 40 coloured points represents an island agent population (there are 400 islands). The colour of a represented agent indicates the position of the agent's genotype in the search space.

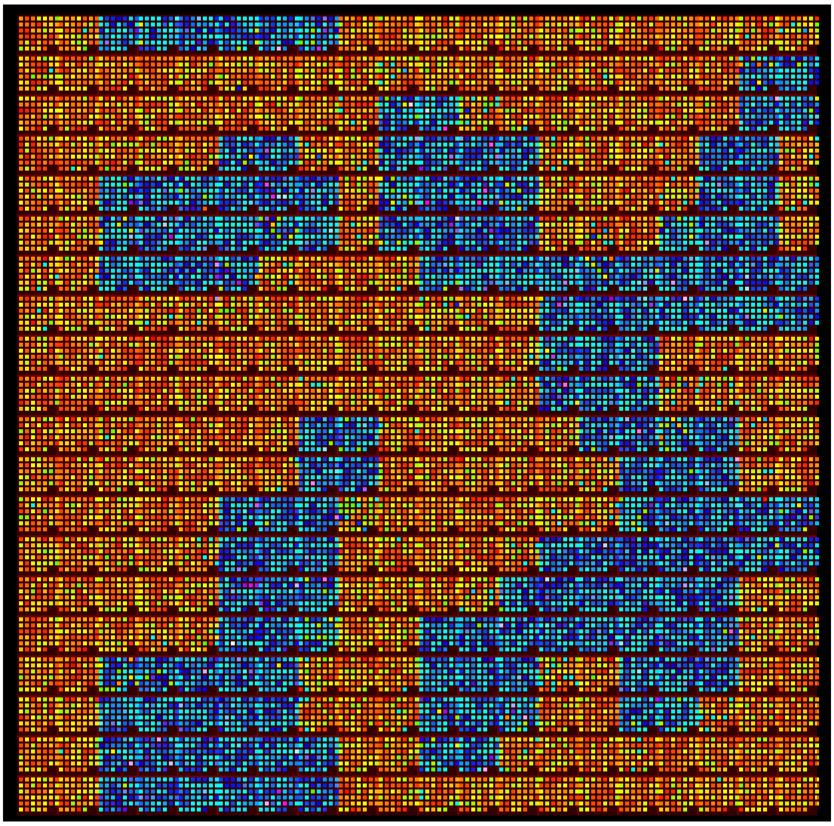


Figure 5.12: The same run of figure 5.11: the mono-population areas emerge.

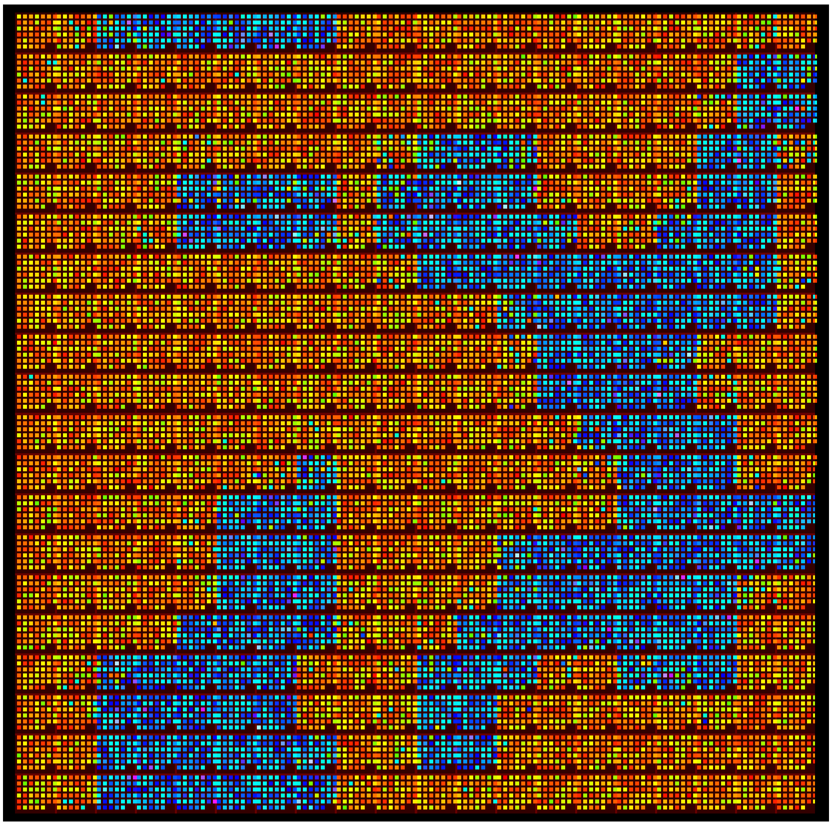


Figure 5.13: The same run of figure 5.11: there is not any diversity maintaining mechanism to stabilize the border islands groups, the system will converge to a single type of population in every island.

Figures 5.14 and 5.15 show an example of simulation run of this model variation: we observe the emergence of the mono-population areas like in the *Total Fertility and Local Connectivity* model variation, but in this case the mono-population areas are maintained by an emergent conservative mechanism.

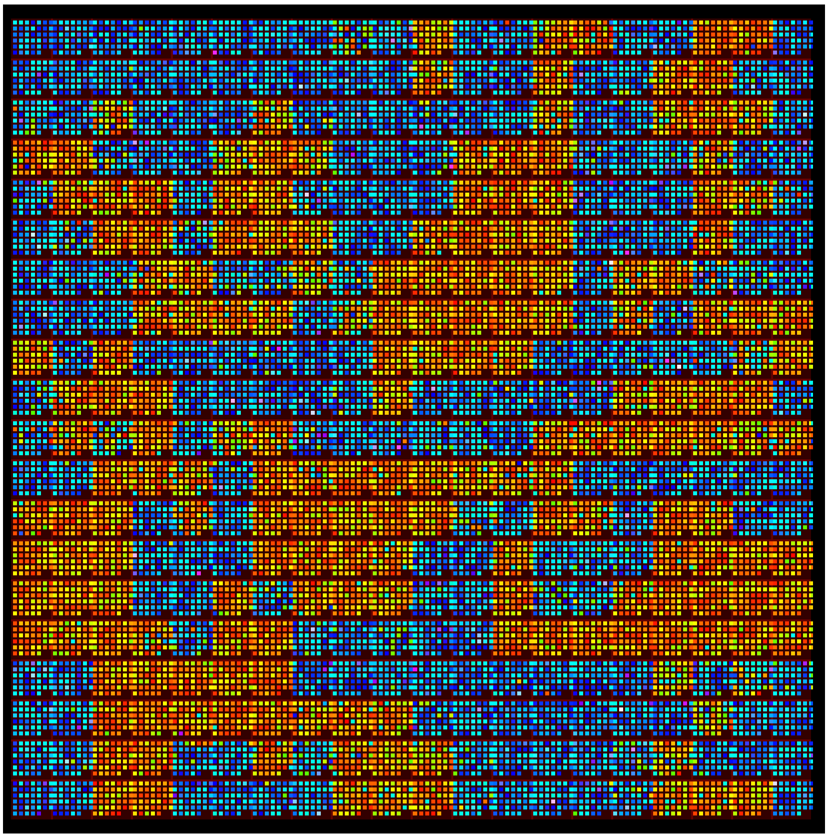


Figure 5.14: An example of execution of the variation *Variable Fertility and Local Connectivity* of the sympatric speciation model. Each group of 40 coloured points represents an island agent population (there are 400 islands). The colour of a represented agent indicates the position of the agent's genotype in the search space.

As in the previous model variation, the boundaries of the mono-population areas are not defined by hybrid populated zones but are forced to be border islands groups.

The effect of the variable fertility is a late occurring effect and in our tests (probably due to the excessive simplification of the search space) the adoption of the approximated general genetic distance measure makes the effect stronger. With the accumulation of mutations in the simulacrum genotypes in the different mono-population areas, the groups of border islands become more resistant to invasion, thanks to the decreased propagation of the invading genotypes.

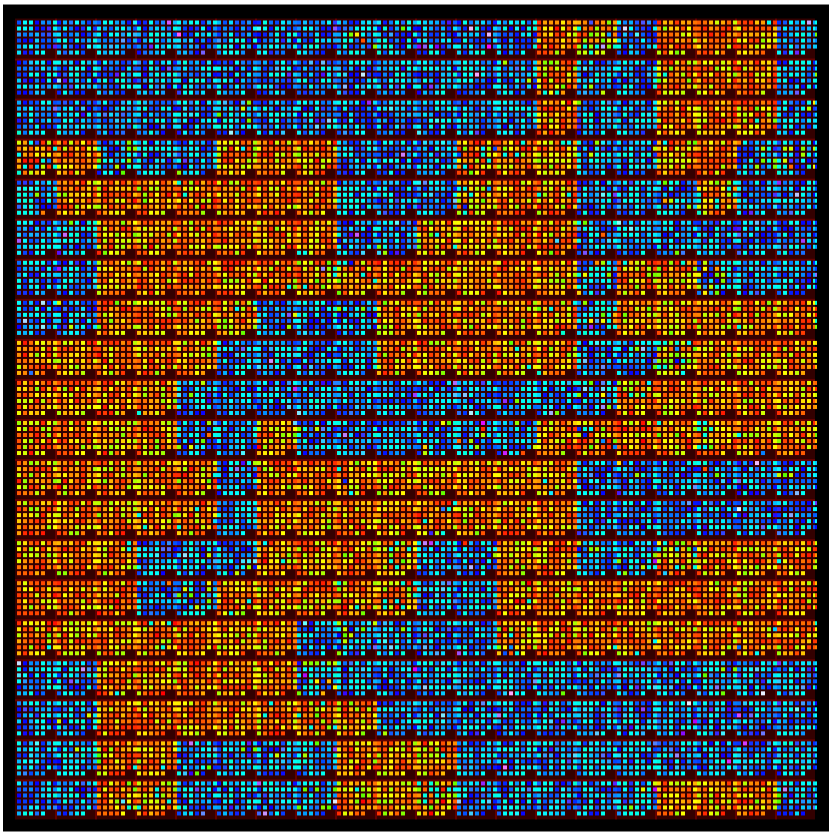


Figure 5.15: The same run of figure 5.11: the mono-population areas are maintained by a conservative mechanism.

5.2.8 The Proposed Solution

We have found a variation of the sympatric speciation model that, even if it drops the geographical space, adopting islands with a fixed population size (as in the typical island model), is able to show the emergence of mono-population areas in the island graph with boundaries stabilized by another emergent property.

The only two modifications of the default island model iteration are in the step of generation of the new populations from the representative sets and in the migration step.

The iteration dynamics becomes:

1. For each island i:
    (a) Selection of a representative set of agents in the population A_i, through a selection method that make use of the fitness values of only the agents in A_i.
    (b) Generation of a new population from the selected representatives, through a reproduction method. *The reproduction method, when it wants to select a pair of agents for sexual reproduction purposes, needs to abort such selection accordingly to the stochastic variable fertility method.*
    (c) Mutation of the genotype of each agent in the new population.
    (d) Replacement of A_i with the new population.
2. For each island pair (i, j), m_{ij} randomly selected agents are removed from A_i and added to A_j (the random selection does not consider agents that have already migrate in the same iteration, *or whose migration cool-down value is grater than zero*).

In addition to this simple variations of two steps of the iteration, the proposed model imposes that the island graph connectivity is a grid-like connectivity.

5.2.9 Discussion

We want to stress that the proposed model is compatible with the existing island model approaches. With some minor additions the evolutionary algorithms can improve the simulation process results:

- If the main role of the several isolated populations is only an enhancement of the competition mechanism, adopting our proposed modifications we enable an extension (in the number of iterations), or even a stabilization (if the fitness values are similar), of the state of multi-population, usually unstable.

- If the main role of the several isolated populations is the finding of alternative solutions, adopting our proposed modifications we resolve the problems of the choice of the number of islands and the island graph connectivity. Moreover, it improves the stability of the multi-population system.

If we have the possibility to make a radical change, we can obtain a stronger stabilization effect adopting a model more similar to the sympatric speciation model, but, in this case, the computational costs of the various locality mechanisms needed are higher than in the case of the approximation of the geographical space using the islands approach.

5.3 Island Destabilization Method

Unlike all the other proposals in the field of evolutionary algorithms, the island destabilization method we are going to describe does not have any biological inspiration.

5.3.1 The Problem

In the various runs of the different evolutionary search algorithms, we observed a very annoying recurrent situation that arise in the case of multi-modal fitness functions with multiple local optima.

When the search process, visualizable like a migration movement of a cluster in the search space, reachs a fitness bifurcation, we want it to find and maintain two different local optima of the fitness function, visualizable as isolated clusters in the search space (Figure 5.16 case A).

But if the cluster does not start from an area in the search space with a bifurcation in the fitness function value (a point where two different directions in the search space are fitness increasing), then one or more local optima of the fitness function will never be found (Figure 5.16 case B).

In order to improve the probability of reaching such a bifurcation point, we can increase the internal diversity of the cluster, i.e. increase the diameter of the cluster in the search space, or execute several runs from several different starting position in the search space.

Unluckily, if one or more populations have already reached a local optimum, even if the internal diversity of the cluster is so high that an alternative local optimum can be reached, a speciation, with formation of a mono-population isolated area, is very difficult to happen (Figure 5.16 case C).

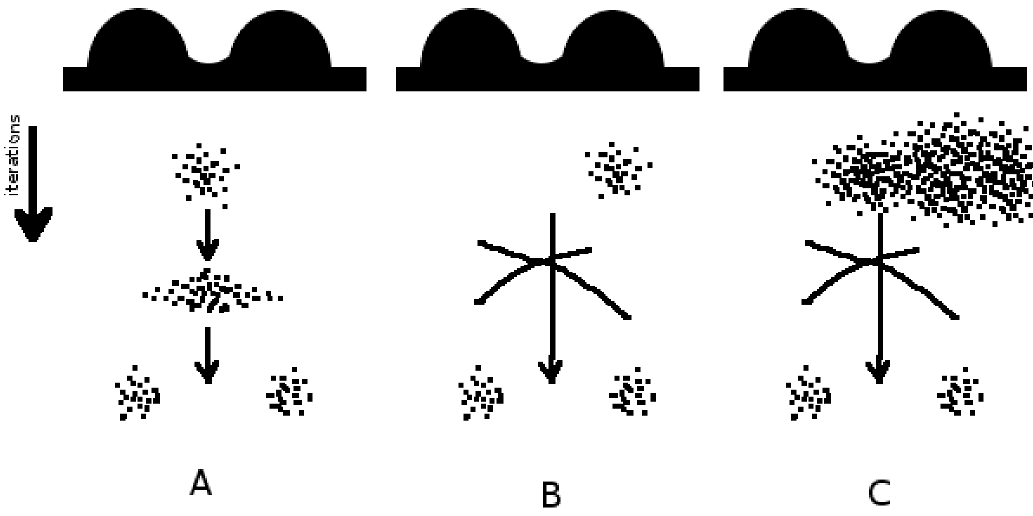


Figure 5.16:

In fact, if the diameter of the cluster is so wide that it touches another zone of the search space with a gradient of the fitness function pointing away from the center of the cluster, we see some agents with this genotype, but no a cluster splits from the original one in order to follow a potentially long fitness increasing path in the search space (Figure 5.17).

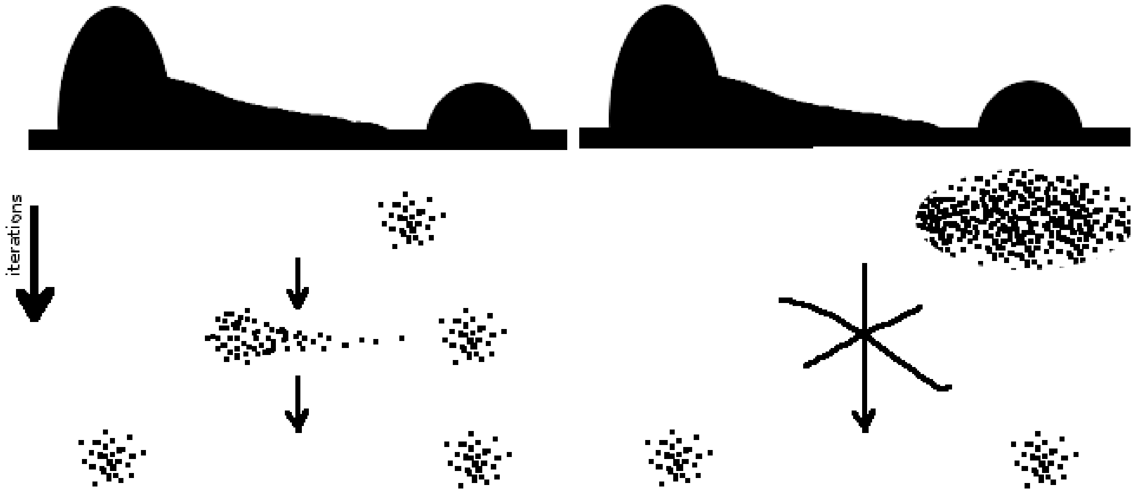


Figure 5.17:

5.3.2 The Proposed Method

We propose a method applicable to the Modified Island Model described in section 5.2 that enables speciation events in cases like the one visualized in Figure 5.16 C or, more interesting, the case visualized in the right side of Figure 5.17.

This method consists in a local destabilization of the search population.

We need to consider a single island whose population has reached a local optimum in the search space. What we want is to destabilize it, such that there is a limited probability that the current search cluster, characterized by the agents of the island population, disgregates and the population begin eventually to evolve in another direction.

If we are in the cases of Figure 5.16 case C or the right side of Figure 5.17, with a cluster diameter wider than the local optimum diameter, a disgregation of the island population can be equivalent to a situation of Figure 5.16 A or the left side of Figure 5.17.

If we consider only the single-island system, if we can obtain such a destabilization then we have lost the purpose of the search: every time the island population disgregates, a big chunk (or the whole) of the searching process information is irremediably lost.

What we want is a destabilization mechanism whose action is local to a single island or a small group of adjacent islands. The global system reacts with a delay repopulating the islands if the destabilization does not have found any fitness increasing path. But if the destabilized population starts to follow an increasing fitness path in the search space, then the delay of the global level reaction allows a speciation.

Analyzing the role of each parameter of the modified island model, we have found that we can obtain this mechanism with an opportune choice of some of these.

We can obtain such a local destabilization in the modified island model by contemporaneously:

Lowering the genetic stability of the newborns. Increasing the *maxmutation* parameter. If in the model the newborn genetic mutation is not applied every time (several works have this feature), but only with a fixed probability, this probability needs to be increased.

Lowering the selective pressure. This can be obtained modifying the selection method, by lowering the stochastical advantage of having a high fitness. The parameters to modify depend on the particular selection method (for example, in a tournament selection, this can be obtained decreasing the tournament participants).

5.3.3 Examples

In figure 5.18 is shown a particular situation where all the islands of the system are stabilized to an optimum, but the fitness function has a second local optimum. For this second local optimum there is not any cluster in the search space and it's fitness value is lower than the fitness of the first one. On the right side of the figure we can see the distribution of the genotypes in the search space. It's interesting that even if the cluster diameter is really wide no agent with genotype in the second optimum is born. In figure 5.19 is shown the effect of the only modification of the selective pressure.

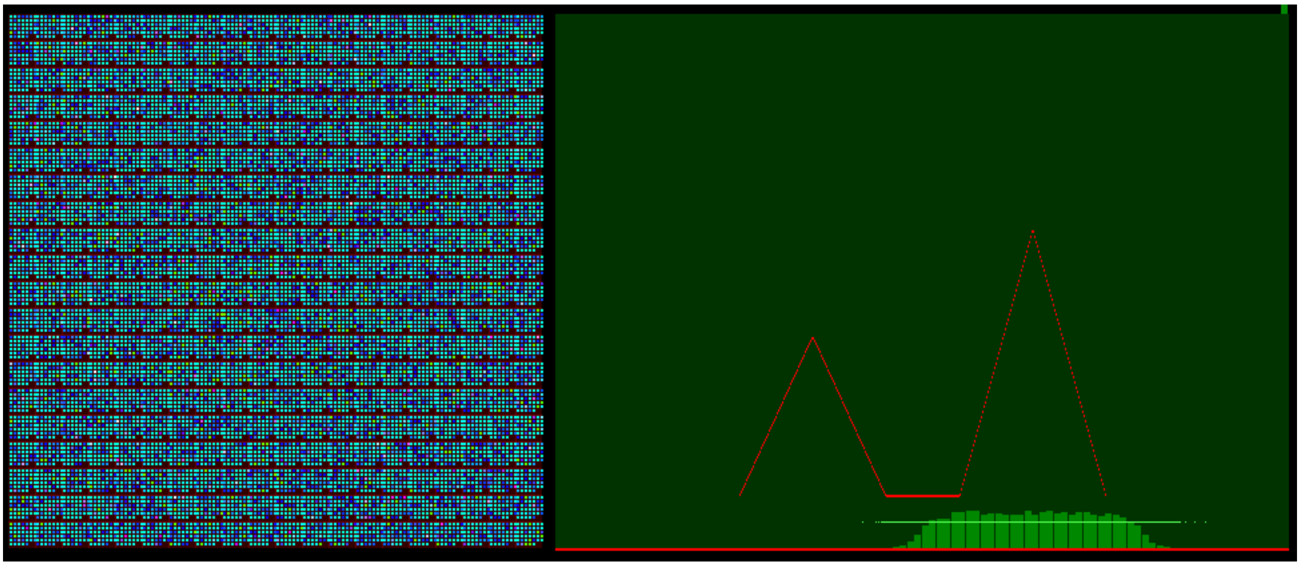


Figure 5.18:

Figures 5.20 and 5.21 show an example of the dynamic of the systems in the setting of fitness function with two local optima. In this case the starting genotype is not placed in the middle point between the two optima: all the starting agents are in the blue optimum. This is the case C of Figure 5.16. We can observe that some islands in the figure 5.20 have been destabilized to the red. Thanks to the decreased selective pressure, the delay in the global reaction gives to some destabilized islands (not all the destabilizing ones) the time to obtain a good fitness value. In figure 5.21 we can view this situation.

5.3.4 Discussion

The statistical nature of the single-island destabilization and the stabilizing global-effect of the migration from other islands of the system, prevents the global system to destabilize. In fact, we found that only single islands or small adjacent groups destabilize. This is true for a system with a high amount of islands.

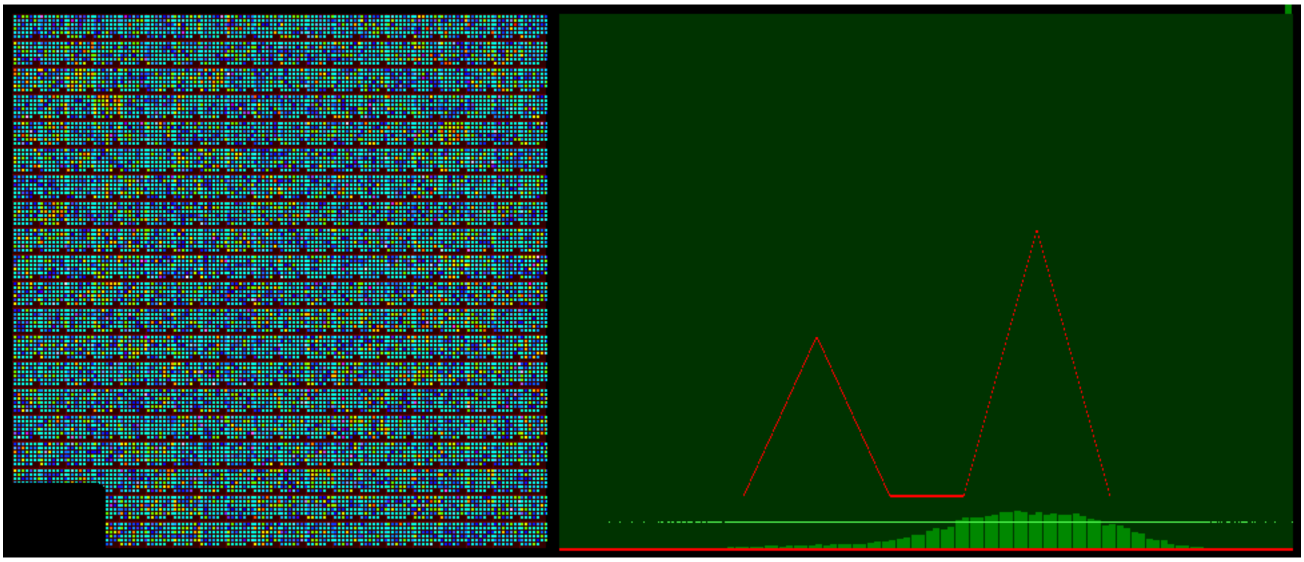


Figure 5.19:

In the absence of multiple local optima of the fitness function, the efficiency of the search results diminished only by a small factor (a deeper analysis is still needed), but in presence of multiple local optima the system gains an increased speciation capability.

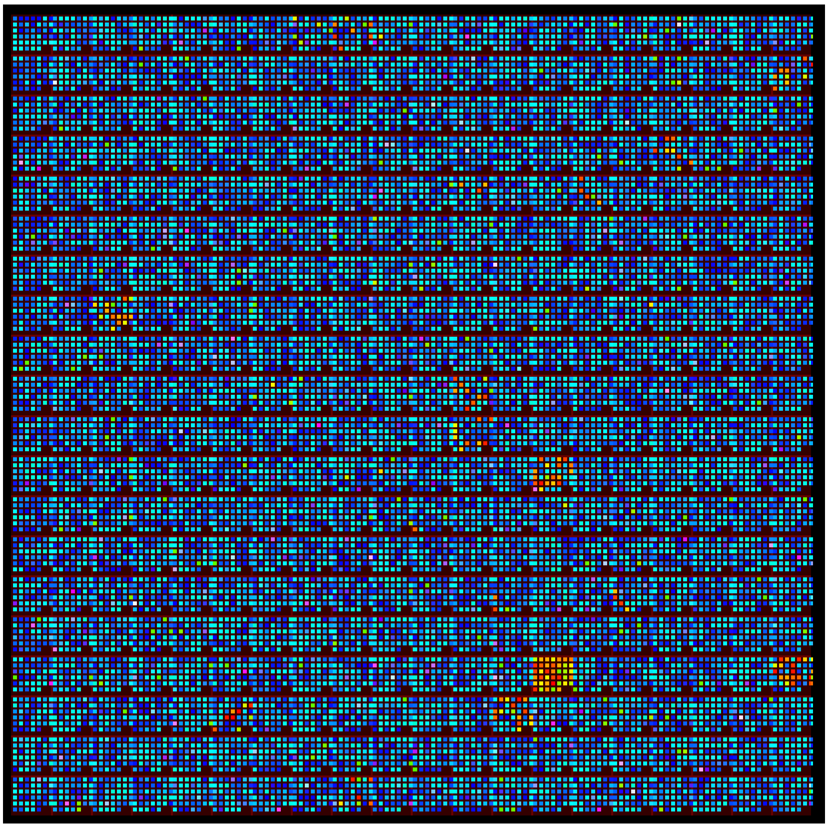


Figure 5.20: An example of execution of the *Island Destabilization Method.* Each group of 40 coloured points represents an island agent population (there are 400 islands). The colour of a represented agent indicates the position of the agent's genotype in the search space.

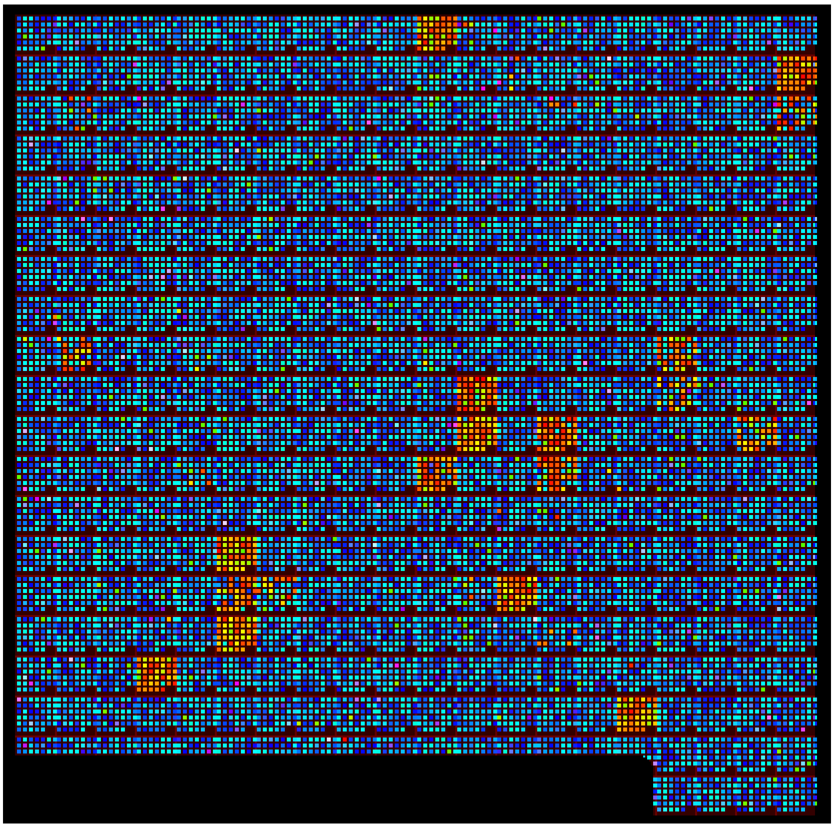


Figure 5.21: The same run of figure 5.20: the formation of the alternative solution islands.

Chapter 6

The Feedback Emergence Model

The final focus of the thesis is the development of a model of system auto-regulation through the emergence of negative feedback cycles.

The feedback cycles in the biological environment are constituted by interactions between biotic (species) and abiotic components of the biosphere and they target one or several parameters of the natural environment.

Negative feedback cycles are feedback cycles that became active when the value of the parameter they controls is shifting away from its homeostatic equilibrium value, and their effect is to compensate such alteration. Therefore, the negative feedback cycles role is to increase the stability of the homeostatic equilibrium of the system.

In order to explain the spontaneous emergence of auto-regulation, we unavoidably need to make use of group selection. There is group selection when a group-oriented altruistic behaviour is selected in a population, even if the altruistic individuals have lower fitness.

The debate about the group selection compatibility with the selfish base mechanism of natural selection started from Darwin's original theory of evolution and it continues today. As for the species definition problem, we do not want to enter in the group selection debate: we do not add any mechanism in the model whose purpose is to enable the group selection; groups selection emerges as an outcome of the previously modeled mechanisms of variable fertility and local selection.

6.1 Feedback Emergence or Inhibition Hypothesis

Now we are considering feedback cycles, but the same hypothesis can eventually be applied to other system-wide mechanisms, altruistic, like negative feedback cycles, or catastrophic, like positive feedback cycles.

A basic hypothesis is that the system evolved to a state where all the species and the abiotic elements of the system developed a strong level of interdependency through biotic active environment transformation: each species role on the system is not only the consumption of resources, but also the resources transformation.

The homeostatic equilibrium of the system depends on the resource transformation effect of the various species and the species have evolved in order to fit the transformed environment (the fitness changes every time the species in the system alter the homeostatic equilibrium).

We consider an island in the state of homeostatic equilibrium, where the species developed a high level of interdependency.

The fitness function is dynamical, because it is implicitly defined by the dynamical system environment (an evolutionary process can alter the environment and therefore the fitness).

There are input fluxes of external resources in the island (as, for the biological system, there is the solar radiations), but it is not constant, some fluctuations (periodical or chaotic) can occur.

6.1.1 Single Island Dynamics

We start considering only a single environment with several species that have developed a strong level of interdependency, where every species needs in order to survive some products of the other species. This single environment is isolated and there is not any limit to the velocity of internal diffusion of resources and organisms.

We assume that every time a speciation event successfully occurs, if the new species develops any new resource transforming capabilities that can alter the homeostatic equilibrium, then all the species in the system suffer a chaotic change (presumably a decrease) of their fitness.

Two cases can occurs:

- The homeostatic equilibrium variation enables the species to reach a **better** fitness than before the speciation.
- The homeostatic equilibrium variation enables the species to reach a fitness level **lower** than before the speciation.

We see a period of fast mutations where the existing species adapt to the altered environment. Additionally, any new resource transformation capabilities of the new species can create, combined to the capabilities of all the other species, a feedback cycle:

Positive feedback cycle : at the first notable fluctuation of the regulated parameter, the positive feedback cycle causes a potentially catastrophic alteration of the homeostatic equilibrium of the system with a corresponding great change of the implicit fitness function. The result of a catastrophic equilibrium alteration can eventually be a mass-extinction.

Negative feedback cycle : the fluctuations of the regulated parameter are dampened. With a less fluctuating environment, the species can adapt more efficiently to the more stable environment.

6.1.2 Multi Island Dynamics

Now we make the hypothesis that, if the system is composed by several islands where migrations can occur with a limited rate, then the dynamics described in the previous single island case can be different.

In particular, every time the fitness adaptation of the species of the island where the speciation occurs increases, the better adaptation causes an increase of the population and therefore in an increased outgoing migration flux. In this case the new species are encouraged to populate the other islands.

If, on the other hand, the new homeostatic equilibrium defines a new fitness function that enables lower adaptation for the current species, then less energy is obtained from the resources and the selective pressure increases. This causes not only a decrement of the outgoing migration flux, but also the island can be "captured" by the other island systems, because even if the incoming migration flux remains the same, it becomes proportionally more influential because of the decreased local population stressed by the increased selective pressure.

The two cases of feedback cycle formation are particular cases where the fitness adaptation levels are dropped almost to zero (positive feedback) or increased (negative feedback).

Our hypothesis is that speciation events in a strongly interdependent system can eventually cause the formation of feedback cycles, not only the biologically observed negative ones, but the positive ones too, and that, in a second delayed moment, the global level selection, in the case of negative feedback, supports the selection of the new modified environment, making it transfer the new species from the starting island to the others, or, in the case of positive feedback, prevents a fail of the global level system, resetting the local level failed system with the invasion of the species of the surrounding islands.

We want to stress that in this hypothesis the fluctuations of the incoming external resources flux are not essential to suppress the diffusion of the positive feedback cycles in all the islands of the system (assuming that minor stochastical fluctuations of the resource levels activate the feedbacks). On the other hand, we expect to see selection and widespread diffusion of positive feedback cycles even if there is not the additional advantage of the fluctuations suppression, but if we consider the fact that the negative feedback action is probably producted only by the development of fitness decreasing capabilities, then without the fluctuations the emergence is very unlikely to occur.

6.2 Simulative Model

The feedback emergence model is an iterative, dynamic, stochastic model based on agents.

This model is an extension of the modified multi-island sympatric speciation model (the *Variable Fertility and Local Connectivity* variation).

As in the sympatric speciation models, we need the variable fertility mechanism and the differentiation between local and global selections (in this case, as in the modified model, "local" is everything in the same island and "global" is every type of inter-island interaction).

6.2.1 Geographical Environment

The geographical environment is the same adopted in the *Variable Fertility and Local Connectivity* speciation model variation: a grid-like connected graph of islands.

The main difference is that the island population size is not fixed, but it depends on the efficiency of the resource transformation activities of the agents that populate the island: some more energy-efficient islands can be more populous than others.

Obviously, in this case the migration rates are not fixed, but (stochastically) proportional to the current population size.

The agents can interact (mating and resource competition) in the same island and they are considered all at the same distance; agents in different islands cannot interact.

6.2.2 Resources

A big change in this model is the way in which resources are represented.

There is not only a generic resource anymore, but several different types of resources are defined. Every resource unit has an energy potential value defined by the resource type: there are high-energy resource types and low-energy resource types.

As in the modified model, the resources are assigned to the island, for each island the number of resources for each type is stored in an array of values.

The unused resources does not disappear from the island: they accumulate for the next iteration.

Together with the resource types definition and the various energy potential value of every resource type, some spontaneous degradations and the various respective degradation rates are defined.

Resources Spontaneous Degradations

A spontaneous degradation from resource type A to resource type B with rate d_i ($0 < d_i < 1$, $d_i \in \mathbb{R}$) is the transformation of a fraction of the total number of resources of type A to type B, and it happens at each iteration of the simulation, in the resource spontaneous degradation step.

For all the defined spontaneous degradations, the energy potential level of resource type B is lower than the energy potential level of resource type A, in this way the total energy of the system can only decrease.

We define $R_i[r]$ as the number of resources of the type r in the island i. $resTypes$ is the number of resource types. $resEnergy_r$ is the energy potential level of the resource type indexed by r.

Resource spontaneous degradation step
For each island i:

1. Initialize to 0 the elements of an array $mods$ of size $resTypes$.
2. For each resource type a:
    (a) $apparent\_rate = 0$
    (b) $cumulative\_rate = 0$
    (c) For each spontaneous degradation sr that transforms the resource type a to the resource type b_{sr}:
        i. $cumulative\_rate += d_{sr}$
        ii. $apparent\_rate += (1 - apparent\_rate) \cdot d_{sr}$
    (d) $mods[a] -= R_i[a] \cdot apparent\_rate$
    (e) For each spontaneous degradation sr that transforms the resource type a to the resource type b_{sr}:
    $mods[b_{sr}] += (R_i[a] \cdot apparent\_rate) \cdot (d_{sr}/cumulative\_rate)$
3. For each resource type r: $R_i[r] += mods[r]$

For clarity purposes in the degradation step we have considered the values of the array $mods$ to be integer numbers and the degradation process to be deterministic, but currently we adopt a stochastic degradation method in order to obtain a chemically realistic resources dynamics. We can adopt a more elegant method: an optimized version of the Gillespie algorithm [29] [30] [75] [28].

Resources Homeostatic Equilibrium

There is an input flux of the most energetic resource type (indexed $input$), expressed by the number of such resource type units to add to every island ($radiation$) at each iteration step.

In order to be able to obtain an homeostatic equilibrium of the resources fluxes, we define an output resource type (indexed $NULL$) and we set the output resource to 0 at every iteration.

Resource input-output step
For each island i:

1. $R_i[input] += radiation$
2. $R_i[NULL] = 0$

If the spontaneous degradation system has not resources types from wich is impossible to find a degradation route to reach the $NULL$ resource, then, independently from the radiation rate and spontaneous degradation rates, after an initial transient number of iterations, a state where, for all resources, the sum of incoming fluxes is equal to the sum of outgoing fluxes will be unavoidably reached.

We refer to such a state as the resources homeostatic equilibrium.

6.2.3 The Agents

In this model each agent can be viewed as a portion of geographical space: in its internal state the agent has an array storing the quantity of every resource types that it contains.

We refer to the resources contained in the agent as its *internal environment* and to the resource contained in the island where the agent is placed as its *external environment*.

With an analogy to the cell mechanisms, every agent has *membrane pumps* and *enzymes*: at each iteration each agent interacts with the external environment in order to exchange resources in a diffusive-like fashion. The exchange results is influenced by the presence of membrane pumps that can pump inside or outside the internal environment some particular types of resource (in absence of pumps the internal environment and external environment concentration of resources are the same), then, after the resources exchange step, each enzyme of an agent transforms the resources of its internal environment.

Resources Exchange Step

The diffusion model we adopt needs to handle resource concentrations, therefore we need to define volumes for the islands ($islandV$) and for the internal environment of the agents ($agentV_i$). For each pump of each agent it's defined the resource type that it pumps, a value of its efficiency and the direction of pumping (inside or outside the membrane).

We do not describe the technical details of the exchange computation, but it's developed in order to obtain:

- In absence of any membrane pump, the internal and external environment resource concentrations are the same for every resource type.

- The inter-membrane resource flux caused by a membrane pump in a particular iteration is defined by the concentration of the resource type targeted by the pump (external environmental concentration for pumps that pump in, internal environmental concentration for pumps that pump out), multiplied by the pump efficiency.

- For every agent, the incoming resource flux of a particular resource type not directly caused by pumps has a value proportional to the external resource type concentration.

- For every agent, the outgoing resource flux of a particular resource type not directly caused by pumps has a value proportional to the internal resource type concentration.

In the resources exchange step of the iteration, we need to calculate the steady state where, for each agent and for each resource, the incoming and the outgoing fluxes are equal. Then from the concentrations of the steady-state and the volumes we can calculate the number of resources contained in each island and in each agent internal environment.

Resources Enzymatic Transformation Step

For each enzyme of each agent are defined, as for the case of spontaneous degradations, the starting resource type A and the target resource type B, and the transformation rate t ($0 < t < 1$, $t \in \mathbb{R}$). Additionally, for every enzyme, we need the energy-efficiency of the transformation: ee.

In the case that the resource type A is more energetic than the resource type B, then the energy an agent gains after an execution of such transformation is:

$$(resEnergy_A - resEnergy_B) \cdot ee$$

An enzymatic transformation, unlike the spontaneous degradations, can transform a low energetic resource type in a more energetic one. In that case there is not any energy gain but an energy consumption for the agent:

$$(resEnergy_B - resEnergy_A) \cdot (1 + ee)$$

Resource enzymatic transformation step

```
For each agent $i$:
1. Initialize to 0 the elements of an array $mods$ of size $resTypes$.
2. $energy\_variation = 0$
3. For each resource type $a$:
    (a) $apparent\_rate = 0$
    (b) $cumulative\_rate = 0$
    (c) For each enzymatic transformation $enz$ that transform the resource type $a$ to the resource type $b_{enz}$:
        i. $cumulative\_rate += t_{enz}$
        ii. $apparent\_rate += (1 - apparent\_rate) \cdot t_{enz}$
    (d) $mods[a] -= R_i[a] \cdot apparent\_rate$
    (e) For each enzymatic transformation $enz$ that transform the resource type $a$ to the resource type $b_{enz}$ with efficiency $ee_{enz}$:
        i. $prod = (R_i[a] \cdot apparent\_rate) \cdot (t_{enz}/cumulative\_rate)$
        ii. $mods[b_{enz}] += prod$
        iii. if $(resEnergy_B \leq resEnergy_A)$ then $energy\_variation += (resEnergy_A - resEnergy_B) \cdot ee_{enz} \cdot prod$
        iv. if $(resEnergy_B > resEnergy_A)$ then $energy\_variation -= (resEnergy_B - resEnergy_A) \cdot (1 + ee_{enz}) \cdot prod$
4. For each resource type $r$: $R_i[r] += mods[r]$
5. $e_i += energy\_variation$
```

Genotype

The genotype of the agents does not describe anymore a point in a genetic space with a defined fitness function.

There is not any explicit fitness function in this model, but the fitness of a particular agent can be indirectly considered as its ability to gain energy from the external environment.

The genotype now defines the several parameters of the agent: the internal environment volume $agentV_i$, the pumps (for each pump: the resource type, the direction of pumping and the efficiency) and the enzymes (for each enzyme: the starting resource type A, the target resource type B, the transformation rate t and the energy-efficiency of the transformation ee).

We need to define the genetic mutation of a genotype and the genetic recombination of two different genotypes. Mutation and recombination are related to the enzyme description and development problems; we will propose a possible implementation of the two mechanisms, but the only thing that cannot be changed is the fact that the various energy-efficiency parameters of the enzymes cannot be obtained by mutation or recombination, they need to be calculated from the enzyme descriptions.

Metabolism

In this model the metabolism energy value that every agent needs to pay at every iteration is not the same for all the agents, but it is calculated from the agent parameters.

The metabolism is the sum of various factors plus a basic minimum metabolism value.

The metabolism increases if there is: an increase of the agent internal environment volume, an increase in the number of the pumps and of the pumps efficiency parameters, an increase in the number of the enzymes and of the enzymes transformation rates, and, in dependence of the adopted genotype description, an increase in genotype description size.

6.2.4 Local Selection

The local selection feature is obtained thanks to the grid-like connections of the islands, as previously described in section 5.2.6 for the sympatric speciation model *Variable Fertility and Local Connectivity* variation.

6.2.5 Variable Fertility

The variable fertility feature is obtained adopting the same variable fertility mechanism described for the sympatric speciation model in section 4.2.3.

The genetic space now is very complex, therefore we need to add to the agents genotypes the previously described simulacrum genotype bit string (section 5.1), in order to calculate genetic distances.

6.2.6 The Iteration

A summary of the iteration computations is the following:

Summary of iteration

1. For each island i:
    - (a) $R_i[\mathit{input}] += \mathit{radiation}$
    - (b) Compute the resource spontaneous degradation step
    - (c) $R_i[\mathit{NULL}] = 0$
    - (d) Compute the agent/external environment resource exchange step
    - (e) Initialize the mating pool to an empty set
    - (f) For each agent j:
        - i. Compute the resource enzymatic transformation step
        - ii. $e_j -= \mathit{metabolism}(j)$
        - iii. If $e_j \leq 0$ then kill the agent (and release the internal environment resources to the island)
        - iv. If $e_j \geq e_{mating}$ then add the agent to the mating pool
    - (g) Until the mating pool size is ≤ 1, select 2 random agents a and b to mate:
        - i. Remove a and b from the mating pool
        - ii. $e_a -= e_{mating\_attempt}, e_b -= e_{mating\_attempt}$
        - iii. Compute the fertility level confronting their simulacrum genotypes
        - iv. If the fertility check does not fail, then create a new agent with a genotype equal to a mutated recombination of the genotype of the parent agents
    - (h) For each agent j:
        - i. Decrease the migration cool-down value
        - ii. If the migration cool-down is equal to 0, then migrate the agent to an adjacent island and reinitialize the cool-down

6.3 Model Discussion

6.3.1 The Choice of Resources Types, Degradations and Possible Enzymes

We have already explained that there is not any explicit fitness function in this model, but the fitness of a particular agent can be indirectly considered as its ability to gain energy from the external environment. In the previous models, in order to obtain a speciation event we had to define a fitness function with two or more local optima. Now, without an explicitly defined fitness function, we face the problem of finding a starting environment definition that enables speciation and, obviously, a negative feedback emergence.

The elements that have a role in the dynamical implicitly defined fitness function are:

- The resource type definitions, and their corresponding energy potential values.
- The spontaneous degradations defined for the resource types, and the degradation rates.
- The possible enzymatic transformations, and the maximum energetic efficiencies of the various transformations.
- The mutation mechanism of a genotype and the recombination mechanism of two different genotypes.

In the real case of the natural environment system, resource types, degradations and possible enzymatic transformations are bind to the low level physics of the real world.

But in the case of our artificial model it's pointless to calculate all the elements from a certain particular low level physical-like mechanism.

For this problem we adopted a solution based on an additional hypothesis: every random big resource type system (big in the number of resource types) with a coherent spontaneous degradation system (capable to reach an homeostatic equilibrium) and an energy coherent system of possible enzymatic transformations, has an acceptable probability to enable evolutionary dynamics of agents with speciation events and negative feedback cycles emergence.

We are aware that our hypothesis can be quite strong and can be view as opposed to anthropic reasoning.

Assuming this hypothesis, we adopted a method to randomly generate a resource type system and, from the generated resource types, calculate a coherent set of spontaneous degradation and a set of possible enzymatic transformations.

From the adopted method we can derive a "natural" way to describe enzymes in the genotype in the form of a string; the mutation and recombination then are string operations inspired by the real genetic mutation and recombination.

Resource Space

We define the resource space as a bi-dimensional grid with a Manhattan distance.

Then we define a particular cell of the grid (indexed *input*) as the maximum energy point with energy value $resEnergy_{input}$. The energy value of any other cell i is defined as:

$$resEnergy_i = resEnergy_{input} - Manhattan(input, i)$$

If the resulting energy $resEnergy_i$ is lower than 0 then we assume that $resEnergy_i = 0$.

From this cells space we randomly select $resTypes$ cells with energy value > 0. This cells, plus the *input* cell, are the resource types of the feedback emergence model. The $NULL$ resource can be viewed as any cell with energy value $= 0$.

Spontaneous Degradations

From a particular defined resource space, we can obtain all the spontaneous degradations and the corresponding degradation rates choosing only a maximum number of steps parameter ($maxN$).

For each resource type a corresponding cell in the resource space $cell_a$, we calculate all the possible trajectories composed by unitary steps with not-increasing energy value in the resource space, starting from $cell_a$, of length $maxN$ steps (the not-increasing energy value constrain needs to be valid for every unitary step in the trajectory).

For each possible resource type b, whose corresponding cell in the resource space $cell_b$ is the end point of at least one trajectory from $cell_a$, we define a spontaneous degradation from a to b with a degradation rate proportional to the number of such trajectories divided by the total number of possible trajectories. For the case of the degradation to $NULL$ we consider all the trajectories ending in a cell of the resource space with energy value equal to zero.

Enzymatic Transformations

All the possible enzymatic transformations are all the trajectories composed by unitary steps in the resource space, starting from a cell corresponding to a resource type a.

In the case of enzymatic transformation there is not the constrain that all the unitary steps of the trajectory need to be of not-increasing energy value, and the end cell of the transformation, indexed *end*, is not constrained to have a corresponding resource type.

The resource type product of an enzymatic transformation is the resource type b whose corresponding $cell_b$ is the cell with the greatest number of not-increasing energy trajectories of length $maxN$ that, starting from $cell_{end}$, ends in $cell_b$. If there is not such a resource type then the product resource type is a itself. For the case of the enzymatic transformation to $NULL$ we consider all the trajectories ending in a cell of the resource space with energy value equal to zero.

The enzymatic transformation rate t (if used by the model) is inversely proportional to the length of the corresponding trajectory.

The energy-efficiency of an enzymatic transformations is obtained from the number of energy increasing unitary steps of the corresponding trajectory.

Enzyme Description

In order to fully describe an enzyme we need to describe only the corresponding trajectory and the starting resource type a; the resource type b of the product, the transformation rate t and the energy-efficiency parameter ee can be calculated as previously described.

The more natural method to describe an enzyme is then a string that is the concatenation of the resource type index followed by a string on the alphabet of the four possible unitary steps on the resource space, sequentially describing each step of the trajectory.

Genotype Description

An agent genotype is composed by three parts:

1. All the simple numerical parameters, like the internal volume, or the efficiency of every possible pump (a value of 0 if the corresponding pump is not present, a negative value if the pump pumps out, positive if it pumps in).
2. The Simulacrum Genotype bit string.
3. A string describing the various enzymes, that we label *enzymatic string*.

The enzymatic string can be of fixed or variable length, so we need to add to the alphabet (currently composed only by the various resource type indexes, and by the 4 possible unitary step in the resource space) a separator character.

When the enzymatic string is parsed, every time a resource type index appear we consider the following character as the trajectory description, until a new resource type index or a separator character appears. Therefore we can describe several enzymes in the same enzymatic string.

Genotype Mutation

The mutation of the genotype numerical parameters part has the same simple dynamics described in the previous speciation models.

The mutation of the Simulacrum Genotype bit string is executed as described in section 5.1.

Every mutation event of the enzymatic string is of two possible types, inspired by observations of the real genetic sequence mutations:

Point mutation : a single character of the enzymatic string can be deleted, modified or added in a random position.

Segment mutation : a segment of the enzymatic string can be deleted or copied and added in a random position of the enzymatic string.

The segment mutation enables a partial re-utilization of already developed structures in the enzymatic string, allowing quick new useful enzyme development.

Genotype Recombination

The recombination of the numerical parameters of two genotypes is performed choosing, for each parameter, a random value included in the interval defined by the two parent parameter values.

The recombination of the Simulacrum Genotype bit strings is executed as previously described in section 5.1.

The recombination of the enzymatic strings is performed selecting some random recombination numbers ($rec[i]$) of incremental value, with values smaller than the size of the enzymatic strings. The recombined new enzymatic string is obtained copying the first $rec[0]$ character from one of the parent enzymatic string, then copying from the second parents enzymatic string the characters from $rec[0]$ to the $rec[1] - 1$, then again from the first parents enzymatic string the characters from $rec[1]$ to the $rec[2] - 1$, and so on, until the new enzymatic string is complete.

6.3.2 Avoid the Super Agent

Observing the biological systems we can see that a super organism fully independent from the other ecological components, able to do any possible enzymatic reaction, able to adapt to every possible environment, does not exist.

A first obvious hypothesis about the super organism absence comes from the observation that every additional capacity or increasing of any already existing capacity of a species, is paid in terms of additional resource consuming structures.

But there is another additional observation to do: in order to increase the energy-efficiency or to enable a new functionality, the internal system of every organism potentially needs to be modified.

A non specialized organism loses in the competition against a specialized one, then the fact that, in order to develop and optimize a new capability, it needs to alter its current internal state, every additional capability potentially causes a decrease of the theoretical maximal optimization for all the other existing capabilities.

An additional outcome is that even if we suppose to solve the problem of competition removing all the existing organisms and replacing them with a single species of super organisms, then we can expect a lower efficiency of the whole system caused by the limited optimization capability caused by the more constrained internal system.

Therefore we have decided to personalize the metabolism energy cost that every agent needs to pay at every iteration, calculating it from the internal environment volume, the number of the pumps and the pumps efficiency parameters, the number of the enzymes and the enzymes transformation rates, and, dependently on the adopted genotype description, the genotype description size.

About the second observation, the need to alter the internal environment in order to optimize the agent, we have already a solution from the dynamics of resource exchange: if an agents want to optimize the energy obtained from a particular enzymatic degradation of a resource type, then it needs to develop a pump for that resource type; then if the same agent wants to exploit a second type of resource, a second type of pump is needed, but the side effect of a second pump addition is an increased outgoing flux for every resource type, an increased outgoing for the first exploited resource type too, then there is an indirect side effect of decrease of optimality relative to the first resource type exploit.

6.3.3 Model Minimality

The dynamics in this model are quite more complex than the dynamics of the sympatric speciation model (the sympatric speciation is included in the feedback emergence model), therefore the process of minimization of the elements of the model is more difficult.

As for the case of the sympatric model, we are not able to prove that a simpler but still operational model does not exists. Our design process is based on the comparison of several models with alternative choices of the basic mechanisms.

We can assume by experimental observations that, with the limited consideration of the mechanisms or properties we analyzed, the proposed model is the simplest in terms of constitutive basic mechanisms.

One of the emergent property that are needed in our model in order to enable the emergence of negative feedbacks is the speciation, therefore all the minimization analysis done for the sympatric speciation model are still valid here. Therefore the following features are not essential:

- Sexual diversity (two genders);
- Gestation period;
- Hybrid agents with decreased fertility;
- Increasing of the metabolism with the aging;
- Sexual maturity minimal age;
- Asexual reproduction.

In the sympatric speciation model discussion all these features are analyzed in the details (see section 4.4.3).

In addition to the unnecessary features inherited from the speciation model, we have found several new others:

Enzymatic transformation rates : in our experiments, defining a constant transformation rate, the same for all the possible enzymes, does not alter the possible system dynamics. We have kept the transformation rates in the description of the model only for consistency with the spontaneous degradation rates (the spontaneous degradation rates, on the other hand, are essential in order to develop not trivial resource homeostatic equilibria).

Membrane advanced models : we call the pumps *membrane* pumps, but we don't have any membrane model in the resource exchange process. In the early models of feedback emergence, we tested several different biologically inspired models of membrane. Membrane with permeability to some resource types, models with limited inter-membrane fluxes, etc. Experimentally we found that not considering at all the membrane does not alter the possible system dynamics.

Complex enzymatic transformations : with more than one single product or with more than only one single reagent. We have not extensively experimented the system dynamics when the possible enzymatic reaction are in forms like $A + B \rightarrow C$ or $A \rightarrow 2B$, because of problems of enzyme description and resource types graph definition. We are able to obtain in some cases the negative feedback without this type of enzymatic transformations, therefore we have assumed that enzymatic reactions which form is more complex than $A \rightarrow B$ are unnecessary. Anyway, the model effectiveness is lower than the one we expected, this simplification of the transformation forms could have compromised the model, but more extended tests are needed.

Genotype String Mutation Methods

We have observed that, in the proposed enzymatic description string format, adopting only the point mutation method the agents take extremely long time before develop new enzymes, because there is the necessity of some point mutation accumulation. This problem disappears adopting the previously described method of segment mutation, thanks to the chance of a partial re-utilization of already developed structures. Even if the segment mutation is not strictly essential, we choosed to keep it.

6.4 Observations

6.4.1 Homeostatic Equilibrium Alteration

In order to reach the resource homeostatic equilibrium the simulation needs to be executed agent-less for several iterations.

We can observe that when the agents start to populate the system the resource homeostatic equilibrium change; the changed new equilibrium induces agent adaptation, therefore additional homeostatic variations, and so on.

It's common to observe various periods of almost maintained homeostatic equilibrium, followed by quick alterations in the environment *and, at the same identical time,* in the agents genotypes.

Figure 6.1 show a simple example of the various resources types concentration dynamics in time (the concentration values are plotted on a logarithmic scale): starting with no resources in the system, thanks to the radiation input and the spontaneous degradations, we can observe an initial fast accumulation of the resources until the output fluxes equate the input fluxes. The homeostatic equilibrium level is not the same for every resource type. In the middle of the plotted run, the agents are added to the system. The number of agents is plotted (not in logarithmic scale, for a better visualization) in white color. We can see that the resource homeostatic equilibrium is altered by the effect of biotic resource transformations.

In Figure 6.2 there is a plot of some parameters (the average values of the whole population) of the agents. We can observe that the non chaotic random variations in 6.2 reflect the resource homeostatic equilibrium alteration in 6.1.

On the right side of Figure 6.1 we can observe that a new resource homeostatic equilibrium is reached after a period of fast accumulation of mutation in the agents (right side of Figure 6.2); the corresponding increase of the agent population size shows that the mutation of the agents transformation capabilities caused a better energy utilization.

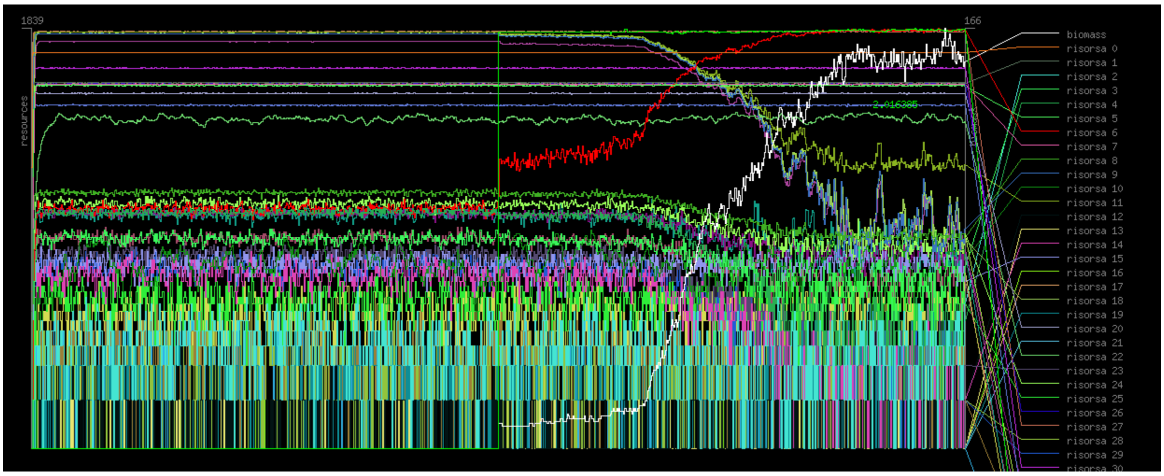


Figure 6.1: A plot of the resource type concentration levels during a simulation run, on a logarithmic scale. The white line represents the number of living agents, not in logarithmic scale.

Figures 6.3 and 6.4 are other two examples of biotic resource homeostatic equilibrium alteration.

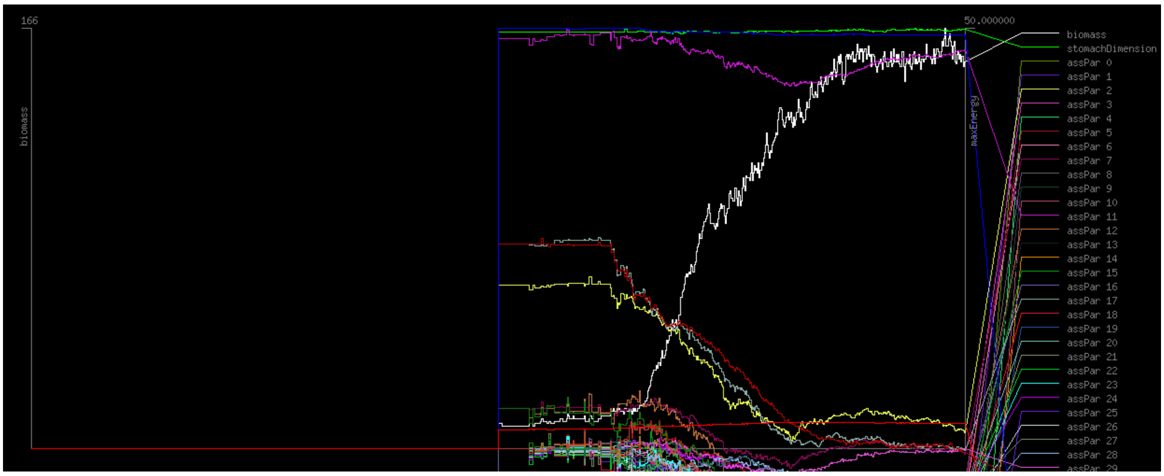


Figure 6.2: A plot of some agent parameter value (the average values of the whole population) variations during the same simulation run of Figure 6.1.

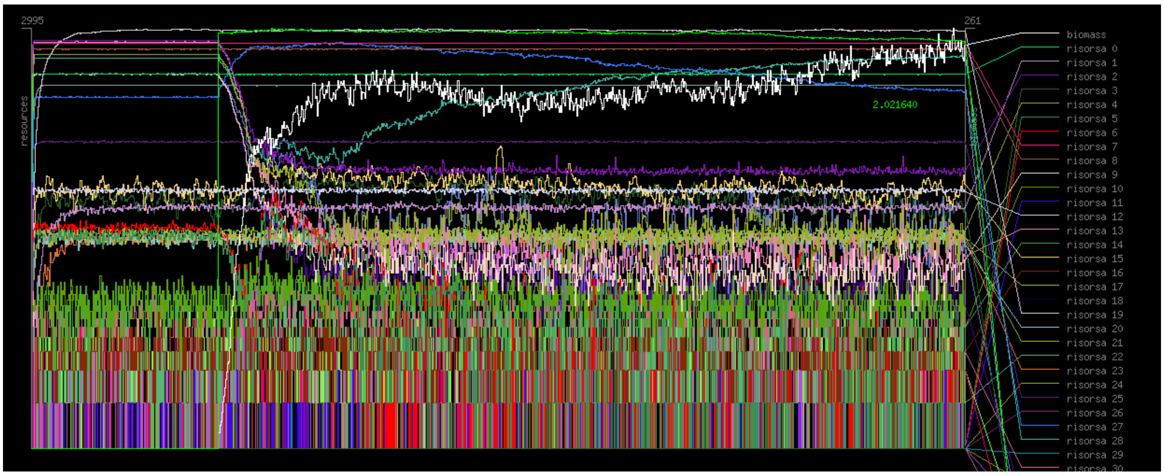


Figure 6.3: A plot of the resource type concentration levels during a simulation run, on a logarithmic scale. The white line represents the number of living agents, not in logarithmic scale.

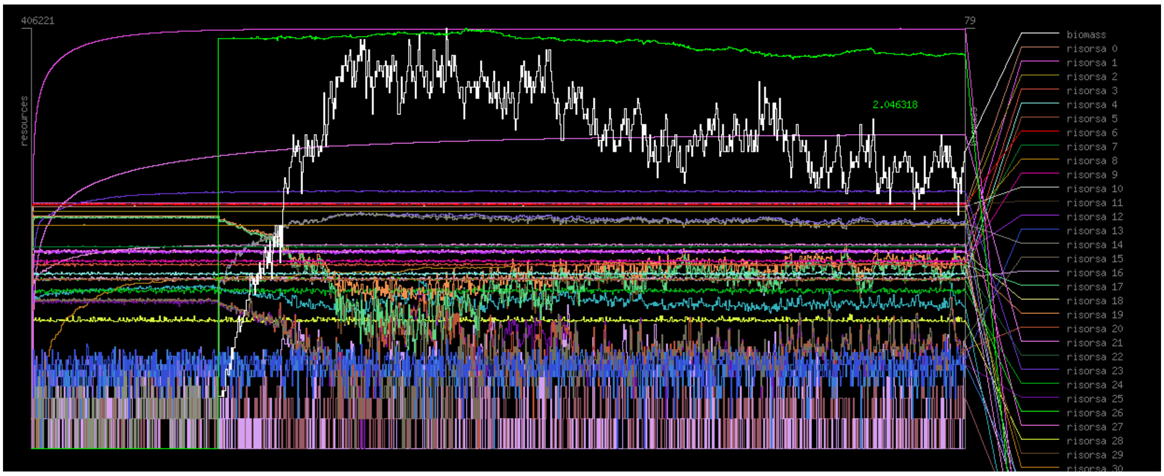


Figure 6.4: A plot of the resource type concentration levels during a simulation run, on a logarithmic scale. The white line represents the number of living agents, not in logarithmic scale.

6.4.2 Environmental Pollution

In our experiments we have observed pollution-like events that are able to compromise the system.

A pollution event happens when a species has an enzyme that transforms a resource type with quick regeneration rate (for example the irradiated resource type) to another resource type with a very low degradation rate. The effect that we observe is that the produced resource type accumulates, increasing its concentration in the external environment and therefore increasing its flux toward the agents internal environments.

If the presence of the new product become so high to be one of the main components of the external environment, then this produces a strong negative poison-like influence to the effectiveness of the membrane pumps: the agents need to use more energy in order to obtain the internal resource homeostatic equilibrium previous to the pollution.

The dynamics generated by polluting agents can be almost equal to the positive feedback dynamics: in absence of group selection there can be mass extinctions, but if there is a global selection and a high level of interdependency between the various biotic element of the system, then the pollution capability diffusion are inhibited.

6.4.3 Negative Feedback Emergence

We are able to observe the emergence of feedback cycles in *ad hoc* defined resource spaces, but the emergence of the feedback cycles in the random generated version of the resource space is actually a quite rare event.

The cause of this problem can be that the hypothesis opposed to the anthropic reasoning, previously formulated in section 6.3.1, is not valid. Alternatively it can be that we have not created a sufficiently interdependent biotic-abiotic interaction system, or that the developed systems are too much small in the number of components. It can be that the enzymatic transformation simplified form ($A \rightarrow B$) is too much weak, in this case we need to reformulate the random resource space generator.

In order to obtain an answer, more research about the random generated resource systems dynamics is needed.

With the limits of the current formulation of the model, where we can obtain negative feedback emergence only in *ad hoc* crafted resource system, differently from the case of sympatric speciation model, we are not able to find useful applications to the field of evolutionary algorithms.

Chapter 7

Conclusions

In this thesis we have presented and discussed several original contributions to the fields of theoretical biology and machine learning.

Sympatric Speciation Model

We have formulated a new biologically inspired simulative minimal model for sympatric speciation. The relevance of this model can be found in theoretical biology.

The proposed variable fertility mechanism has similarities with the assortative mating, but not only it does not require that agents have recognition capabilities, the fact that it is a post-mating barrier (the assortative is a pre-mating barrier) cause it to waste resources to the agents in multi-population areas (boundaries), making it a more efficient diversity maintaining mechanism.

Simulacrum Genotype

We proposed a biologically inspired solution to the problem of genetic space distance measurements. This solution is relevant in the field of evolutionary algorithms: can be applied not only to the proposed models, but to the data analysis process of pre-existing models.

Island Model Variations

We proposed several variations of the island model, a widely used model in the field of evolutionary algorithms, in order to develop an evolutionary system suitable for multi-modal optimization problems.

The variations are based on the introduction of features observed in the sympatric speciation model (variable fertility, local selection paired with a delayed global selection). The system is focused on maintaining a high population diversity.

Island Destabilization Method

We proposed a not biologically inspired method of diversity enhancement, the Island Destabilization Method, applicable to the previously described island model variation.

Feedback Emergence Model

We formulated a biologically inspired simulative model for negative feedback cycles emergence and positive feedback cycles inhibition. Even if the results of this model are limited because we are not able to observe the emergent properties for random generated system settings, in *ad hoc* settings we are able to observe the emergence. This model has a relevance for the Gaia theory.

Bibliography

[1] C. Adami. *Introduction to artificial life*. Springer (1998).

[2] D. Andre and A. Teller. *Evolving team Darwin United*. RoboCup-98: Robot Soccer World Cup II. Lecture Notes in Computer Science Vol.1604 (1999), pp. 346-352.

[3] P. J. Angeline and J. B. Pollack. *Competitive Environments Evolve Better Solutions for Complex Tasks*. Proceedings of the 5th International Conference on Genetic Algorithms (1993), pp. 264-270.

[4] W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela and R. E. Smith. *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*. (1999).

[5] H. Barnum, H. J. Bernstein and L. Spector. *Quantum Circuits for OR and AND of ORs*. Journal of Physics A: Mathematical and General 33(45) (2000), pp. 80478057.

[6] T. C. Belding. *The Distributed Genetic Algorithm Revisited*. Proceedings of the Sixth International Conference on Genetic Algorithms (1995).

[7] S. Bird and X. Li. *Adaptively choosing niching parameters in a PSO*. Genetic And Evolutionary Computation Conference: Proceedings of the 8th annual conference on Genetic and evolutionary computation (2006).

[8] E. K. Burke, S. Gustafson and G. Kendall. *Diversity in Genetic Programming: An Analysis of Measures and Correlation With Fitness*. IEEE Transactions on Evolutionary Computation (2004), pp. 47-62.

[9] E. Cantú-Paz. *A survey of parallel genetic algorithms*. Calculateurs Paralleles, Reseaux et Systems Repartis 10(2) (1998), pp. 141-171.

[10] W. Cedeño and V. R. Vemuri. *Analysis of Speciation and Niching in the Multi-Niche Crowding GA*. Theoretical Computer Science 229 (1-2) (1999), pp. 177-197.

[11] R. Charlson, J. Lovelock, M. Andreae and S. Warren. *Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate.* Nature 326 (1987), pp. 655-661.

[12] R. J. Collins and D. R. Jefferson. *Selection in Massively Parallel Genetic Algorithms.* Proceedings of the Fourth International Conference on Genetic Algorithms (1991).

[13] A. Cordelli, P. Cerrai and L. Galleni. *Artificial Life and Speciation, a Case Study: Heterochromatin and Speciation in the* Microtus savii *Group* (Rodentia-Arvicolinae). Biology Forum 96 (2003), pp. 87-104.

[14] J. M. Daida, D. J. Ward, A. M. Hilss, S. L. Long, M. R. Hodges and J. T. Kriesel. *Visualizing the Loss of Diversity in Genetic Programming.* Proceedings of the 2004 congress on evolutionary computation (2004), pp. 1225-1232 Vol.2.

[15] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, Chichester, UK (2001).

[16] U. Dieckmann and M. Doebeli. *On the origin of species by sympatric speciation.* Nature 400 (1999), pp. 354-357.

[17] K. Downing. *EUZONE: Simulating the Evolution of Aquatic Ecosystems.* Artificial Life 3(4) (1997), pp. 307-333.

[18] K. Downing and P. Zvirinsky. *The Simulated Evolution of Biochemical Guilds: Reconciling Gaia Theory and Natural Selection.* Artificial Life (1999).

[19] K. Downing. *Emergence of Emergence Distributions: Using Genetic Algorithms to Test Biological Theories.* Proceedings of the 7th International Conference on Genetic Algorithms (1997).

[20] R. Dreżewski and K. Cetnarowicz. *Sexual Selection Mechanism for Agent-Based Evolutionary Computation.* ICCS 2007, volume 4488 of Lecture Notes in Computer Science (2007), pp. 920-927.

[21] R. Dreżewski and M. Kisiel-Dorohinicki. *Maintaining Diversity in Agent-Based Evolutionary Computation.* ICCS 2006, volume 3993 of Lecture Notes in Computer Science (2006), pp. 908-911.

[22] R. Dreżewski and L. Siwik. *Techniques for Maintaining Population Diversity in Classical and Agent-Based Multi-objective Evolutionary Algorithms.* ICCS 2007, volume 4488 of Lecture Notes in Computer Science (2007), pp. 904-911.

[23] R. A. P. Durazo and M. V. Rendón. *Similarities Between Co-evolution and Learning Classifier Systems and Their Applications.* Lecture notes in computer science (2004).

[24] D. B. Fogel, T. J. Hays and D. R. Johnson. *A Platform for Evolving Characters in Competitive Games*. Proceedings of the Congress on Evolutionary Computation (2004), pp. 1420-1426.

[25] C. M. Fonseca and P. J. Fleming. *An Overview of Evolutionary Algorithms in Multiobjective Optimization.* Evolutionary Computation 3-1 (1995), pp. 1-16.

[26] S. Garcia, J. Levine and F. Gonzalez. *Multi Niche Parallel GP with a Junk-code Migration Model.* Genetic Programming, Proceedings of EuroGP 2003 (2003), pp. 327-334.

[27] S. Gavrilets. *PERSPECTIVE: MODELS OF SPECIATION: WHAT HAVE WE LEARNED IN 40 YEARS?* Evolution 57(10) (2003), pp. 2197-2215.

[28] M. A. Gibson and J. Bruck. *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels.* Journal of Physical Chemistry A 104 (2000), pp. 1876-1889.

[29] D. T. Gillespie. *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions.* Journal of Computational Physics 22(4) (1976), pp. 403-434.

[30] D. T. Gillespie. *Exact Stochastic Simulation of Coupled Chemical Reactions.* The Journal of Physical Chemistry 81(25) (1977), pp. 2340-2361.

[31] J. Gillon. *Feedback on Gaia.* Nature 406 (2000), pp. 685-686.

[32] M. Glickman and K. Sycara. *Evolvability and Static vs. Dynamic Fitness.* Workshop Proceedings of Artificial Life VII (2000).

[33] D.E. Goldberg and J. Richardson. *Genetic algorithms with sharing for multimodal function optimization.* Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms (1987), pp. 41-49.

[34] P. Hajela and C.Y. Lin. *Genetic search strategies in multicriterion optimal design.* Structural Optimization 4 (1992), pp 99-107.

[35] G. Hinkle. *Marine Salinity: Gaian Phenomenon? Gaia in Action.* Floris Books (1996), pp. 75-88.

[36] J. Horn and N. Nafpliotis. *Multiobjective optimization using niched pareto genetic algorithm.* IllGAL Report 93005, Illinois Genetic Algorithm Laboratory, University of Illinois, Urbana-Champaign (1993).

[37] C. Z. Janikow and C. J. Mann. *CGP Visits the Santa Fe Trail - Effects of Heuristics on GP.* Proceedings of the 2005 conference on Genetic and evolutionary computation (2005), pp. 1697-1704.

[38] S. Kauffman. *The Origins of Order*. Oxford University Press (1993).

[39] J. Kirchner. *The Gaia Hypothesis: Fact, Theory, and Wishful Thinking*. Climatic Change 52 (2002), pp. 391-408.

[40] J. Kirchner. *The Gaia Hypothesis: Can It Be Tested?* Reviews of Geophysics 27 (1989), pp. 223-235.

[41] M. Kirkpatrick. *Reinforcement and divergence under assortative mating*. Proceedings of the Royal Society B: Biological Sciences 267 (2000), pp. 1649-1655.

[42] M.A. Kirley. *A Cellular Genetic Algorithm with Disturbances: Optimisation Using Dynamic Spatial Interactions*. Journal of Heuristics 8(3) (2002), pp. 321-342.

[43] M. Kirley, X. Li and D. Green. *Investigation of a cellular genetic algorithm that mimics landscape ecology*. Simulated Evolution and Learning - SEAL98, volume 1585 Lecture Notes in Artificial Intelligence (1988), pp. 90-97.

[44] A.S. Kondrashov and M. Shpak. *On the origin of species by means of assortative mating*. Proceedings of the Royal Society B: Biological Sciences 265 (1998), pp. 2273-2278.

[45] J. R. Koza. *Genetic Programming*. Encyclopedia of Computer Science and Technology (1998), pp. 29-43.

[46] J. R. Koza. *Genetic Programming II: Automatic Discovery of Reusable Programs.* (1994)

[47] J. R. Koza, F. H. Bennett, D. Andre and M. A. Keane. *Genetic Programming III: Darwinian Invention and Problem Solving.* (1999)

[48] J. R. Koza, M. A. Keane, M. J. Streeter, W. Mydlowec, J. Yu and G. Lanza. *Genetic Programming IV: Routine Human-Competitive Machine Intelligence.* (2003).

[49] M. Laumanns, G. Rudolph and H.P. Schwefel. *A Spatial Predator-Prey Approach to Multi-objective Optimization: A Preliminary Study*. Lecture Notes in Computer Science (1998), pp. 241-249.

[50] T. Lenton. *Redfield Revisited: Regulation of Nitrate, Phosphate and Oxygen in the Ocean*. Global Biogeochemical Cycles (1999).

[51] T. Lenton. *Gaia and natural selection*. Nature 394 (1998), pp 439-447.

[52] X. Li. *A Real-Coded Predator-Prey Genetic Algorithm for Multiobjective Optimization*. Lecture Notes in Computer Science (2003), pp. 207-221.

[53] X. Li and M. Kirley. *The Effects of Varying Population Density in a Fine-grained Parallel Genetic Algorithm.* Proceedings of the 2002 Congress on Evolutionary Computation v.2 (2002), pp. 1709-1714.

[54] Y. Lin and B. Bhanu. *Evolutionary Feature Synthesis for Object Recognition.* IEEE transactions on systems, man and cybernetics. Part C, Applications and reviews (2005), pp. 156-171.

[55] J. Lovelock. *The Ages of Gaia: A Biography of our Living Earth.* Oxford University Press (1988).

[56] J. Lovelock. *Gaia: The Practical Science of Planetary Medicine.* Gaia Books Ltd. (1991).

[57] J. Lovelock and A. Watson. *The regulation of carbon dioxide and climate: Gaia or geochemistry.* Planetary Space Science 30(8) (1982), pp.795-802.

[58] J. Lovelock. *Gaia: A New Look at Life on Earth.* Oxford University Press (1979).

[59] J. Lovelock and L. Margulis. *Atmospheric homeostasis by and for the biosphere: The Gaia hypothesis.* Tellus 26 (1974), pp.2-10.

[60] S. Luke. *Genetic programming produced competitive soccer softbot teams for RoboCup97.* Genetic Programming 1998: Proceedings of the Third Annual Conference (1998), pp. 214-222.

[61] S.W. Mahfoud. *Crowding and Preselection Revisited.* Parallel Problem Solving from Nature PPSN-II, Elsevier (1992). Evolution 57(10) (2003), pp. 2197-2215.

[62] J. Mallet. *A species definition for the modern synthesis.* Trends in Ecology & Evolution (1995).

[63] P. Marrow. *Nature-Inspired Computing Technology and Applications.* BT Technology Journal (2000).

[64] R. Monastersky. *The Plankton-Climate Connection, In From Gaia to Selfish Genes: Selected Writings in the Life Sciences.* The MIT Press (1991).

[65] T. Motoki. *Calculating the Expected Loss of Diversity of Selection Schemes.* Evolutionary Computation (2002).

[66] H. Mühlenbein. *Evolution in Time and Space - The Parallel Genetic Algorithm.* Foundations of Genetic Algorithms 1 (1991), pp. 316-337.

[67] J. T. Ngo and J. Marks. *Spacetime Constraints Revisited.* Computer Graphics 27 (1993), pp. 343-350.

[68] D. Noble. *The rise of computational biology.* Nature Reviews Molecular Cell Biology (2002).

[69] C. Ofria and C.O. Wilke. *Avida: A Software Platform for Research in Computational Evolutionary Biology.* Journal of Artificial Life 10 (2004), pp. 191-229.

[70] J. Paredis. *Coevolutionary Computation.* Artificial Life 2(4) (1995), pp. 355-375.

[71] J. B. Pollak and A. D. Blair. *Co-Evolution in the Successful Learning of Backgammon Strategy.* Machine Learning 32 (1998), pp. 225-240.

[72] M. A. Potter. *The Design and Analysis of a Computational Model of Cooperative Coevolution.* (1997).

[73] M.A. Potter and K.A. De Jong. *Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents.* Evolutionary Computation 8(1) (2000), pp. 1-29.

[74] M. Ratford, A.L. Tuson and H. Thompson. *An Investigation of Sexual Selection as a Mechanism for Obtaining Multiple Distinct Solutions.* Technical Report 879, Department of Artificial Intelligence, University of Edinburgh (1997).

[75] M. Rathinam, L. R. Petzold, Y. Cao and D. T. Gillespie. *Stiffness in Stochastic Chemically Reacting Systems: The Implicit Tau-Leaping Method.* Journal of Chemical Physics 119(24) (2003), pp. 12784-12794.

[76] T.S. Ray. *An approach to the synthesis of life.* Proceedings of Artificial Life II (1992), pp. 371.

[77] T.S. Ray. *An evolutionary approach to synthetic biology: Zen and the art of creating life.* Artificial Life 1(1/2) (1994), pp. 195-226.

[78] A. Redfield. *The biological control of chemical factors in the environment.* American Scientist 46 (1958), pp. 205-221.

[79] D. Roberston and J. Robinson. *Darwinian Daisyworld.* Journal of Theoretical Biology 195(1) (1998), pp. 129-134.

[80] J. Sánchez-Velazco and J.A. Bullinaria. *Gendered Selection Strategies in Genetic Algorithms for Optimization.* Proceedings of the UK Workshop on Computational Intelligence UKCI (2003), pp. 217-223.

[81] Saunders. *Daisyworld and the Future of Gaia. Gaia in Action.* Floris Books (1996), pp. 75-88.

[82] D. Schwartzman and T. Volk. *Biotic enhancement of weathering and surface temperatures on Earth since the origin of life.* Palaeogeography, Palaeoclimatology, Palaeoecology (Global and Planetary Change Section) (1991), pp. 357-371.

[83] Z. Skolicki and K. Jong. *The Influence of Migration Sizes and Intervals on Island Models.* Proceedings of the 2005 conference on Genetic and evolutionary computation (2005), pp. 1295-1302.

[84] J.M. Smith. *Sympatric Speciation.* The American Naturalist 100 (1966), pp. 637-650.

[85] P. Snijders, E. D. Jong, B. Boer and F. Weissing. *Multi-Objective Diversity Maintenance.* Proceedings of the 8th annual conference on Genetic and evolutionary computation (2006), pp. 1429-1430.

[86] A. Sokolov and D. Whitley. *Unbiased Tournament Selection.* Proceedings of the 2005 conference on Genetic and evolutionary computation (2005), pp. 1131-1138.

[87] L. Spector, H. Barnum and H. J. Bernstein. *Genetic Programming for Quantum Computers.* Genetic Programming 1998: Proceedings of the Third Annual Conference (1998), pp. 365373.

[88] L. Spector and H. J. Bernstein. *Communication Capacities of Some Quantum Gates, Discovered in Part Through Genetic Programming.* Proceedings of the Sixth International Conference on Quantum Communication, Measurement, and Computing (2003), pp. 500-503.

[89] L. Spector, H. Barnum, H. J. Bernstein and N. Swamy. *Finding a Better-Than-Classical Quantum AND/OR Algorithm Using Genetic Programming.* Proceedings of 1999 Congress on Evolutionary Computation (1999), pp. 2239-2246.

[90] P.M. Todd and G.F. Miller. *Biodiversity through sexual selection.* Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems (1997), pp. 289-299.

[91] T. Volk. *Gaia's Body: Toward a Physiology of Earth.* Copernicus (1998).

[92] J. Wang, Z. Fan, J. P. Terpenny and E. D. Goodman. *Knowledge Interaction with Genetic Programming in Mechatronic Systems Design Using Bond Graphs.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (2005), pp. 172-182.

[93] A. Watson and J. Lovelock. *Biological homeostasis of the global environment: the parable of Daisyworld.* Tellus 35B (1983), pp. 284-289.

[94] D. Whitley, S. Rana and R. B. Heckendorn. *The Island Model Genetic Algorithm: On Separability, Population Size and Convergence.* Journal of computing and information technology (1999).

[95] G. Williams. *The Molecular Biology of Gaia.* Columbia University Press (1996).

[96] P. Williamson and J. Gribbin. *How Plankton Change the Climate.* New Scientist 16 (1991).

[97] M. Winederg and F. Oppacher. *A Linear Time Algorithm for Determining Population Diversity in Evolutionary Computation.* IASTED International Conference on Intelligent Systems and Control (2003), pp. 270-275.

[98] M. Winederg and F. Oppacher. *The underlying similarity of diversity measures used in evolutionary computation.* GECCO (2003).

[99] E. Zitzler and L. Thiele. *Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach.* IEEE Transactions on Evolutionary Computation 3(4) (1999), pp. 257-271.