

Mathematical Modeling of Biogeography-Based Optimization

1. Background

Biogeography is the science and study of patterns of species distribution, including migration, extinction, and speciation. The science of biogeography can be traced to numerous early observers of nature, including 19th century naturalists Alfred Russel Wallace and Charles Darwin. Until the 1960s, biogeography was mainly descriptive and historical. In the early 1960s, Robert MacArthur and E. O. Wilson began working on mathematical models of species assembly on islands, culminating with their classic book *Equilibrium Theory of Island Biogeography* [Mac63, Mac67]. Since then, biogeography has become a major area of research [Han97]. A recent search of Biological Abstracts, a biology research index, reveals that 21,460 papers related to the subject of biogeography were published in the year 2009.

A recent paper by this proposal's PI is the first that uses biogeography to motivate an optimization algorithm [Sim08a]. The application of biogeography to engineering is similar to what has occurred in the past few decades with genetic algorithms (GAs), neural networks, fuzzy logic, particle swarm optimization, and other areas of computer intelligence. Several decades ago, very few people could have foreseen the possibility that Darwin's theory of evolution, or models of neuronal activity, or the imprecise reasoning of common sense, or the swarming behavior of insects, could ever have an impact in engineering. Yet all these paradigms have had a large influence in the area of optimization.

In biological species distribution models, it is important to measure the quality of an island or habitat patch. For example, geographical areas that possess the necessary resources to support a species are said to have a high island suitability index (ISI) [Wes87]. Features that correlate with ISI are called suitability index variables (SIVs) and include factors such as rainfall, diversity of vegetation, topographic features, land area, and temperature. SIVs can be considered the independent variables of the island, and ISI can be considered the dependent variable. Islands with a high ISI tend to have a large number of species and large populations, while those with a low ISI have a small number of species and small populations.

Biogeography is analogous to problem solving. A good solution to an optimization problem is like an island with high ISI and high species diversity, and a poor solution is like an island with low ISI and low species diversity. Highly fit solutions resist change more than low diversity solutions. By the same token, highly fit solutions tend to share their features with low-fitness solutions. This does not mean that the features disappear from the high-fitness solution; the shared features remain in the high-fitness solutions while at the same time appearing as new features in the low-fitness solutions. This is similar to an individual of a species moving to a new island while other individuals remain in the original habitat. Poor solutions accept many features from good solutions. This addition of features to poor solutions tends to raise their quality. Poor solutions are therefore more dynamic than good solutions. This new approach to problem solving is called biogeography-based optimization (BBO).

Overview of Proposal

This proposal hypothesizes that incorporating biogeographical details into the BBO algorithm will improve its performance. This hypothesis is based on the premise that natural biogeography is an optimization process, as discussed in Section 2. It is also supported by preliminary results in Section 5.3, which show that when BBO migration curves take on the sigmoid shapes that are often found in nature, optimization performance improves relative to simpler linear migration curves. Section 3 gives an overview of BBO, which is the optimization algorithm that is motivated by natural biogeography, and

which is the focus of this proposal. Section 3 reviews some of the preliminary successes of BBO and provides some hypotheses about why it is effective as an optimization algorithm. Section 4 summarizes the objectives, the significance, and the relationship to current and previous work, of the proposed research.

Section 5 comprises the main portion of this proposal, and discusses each task of the proposed research in detail. We propose to develop analytical math models to quantify the performance of BBO. The use of mathematical analysis of the BBO computational approach is the cornerstone of this proposal. Section 6 gives a summary of the timeline of the proposed research. Section 7 discusses how the results will be disseminated and integrated into educational activities. Section 8 discusses how under-represented minorities will be involved in the research.

2. Biogeography is an Optimization Process

Island biogeography is a model that describes the distribution of species, and it has often been studied as a process that maintains species equilibrium on island patches. One reason that biogeography has been viewed from this perspective is that the equilibrium viewpoint was the first to place biogeography on a firm mathematical footing in the 1960s [Mac63, Mac67]. However, the equilibrium perspective has been augmented by biogeographers in recent years.

We often view stability and optimality as competing objectives; for example, a simple system is often more stable than a complex system, while an optimal system is typically complex and less stable than a simpler system. However, in biogeography, stability and optimality are two sides of the same coin. For example, biogeographical optimality involves diverse, complex communities which are highly adaptable to their environment, while biogeographical stability involves the persistence of existing populations. However, field observations show that complex communities are more adaptable and stable than simple communities [Har06, p. 82], and this idea has also been supported by simulation [Elt58, Mac55, McC00]. Observation indicates that species migration optimizes the habitability of islands [Har06, p. 82], and this idea has also been supported by biogeographical simulation [Elt58, Mac55, McC00]. The equilibrium/optimality issue in biogeography thus becomes a matter of semantics, because **equilibrium and optimality are simply two different perspectives on the same phenomenon in biogeography.**

The duality of equilibrium and optimality is commonly found in both natural and man-made systems. For example, in geomorphology (the study of landforms and the processes that form them), the observation of dynamic steady-state channel geometries of rivers has been derived as channel adjustments that minimize stream power, and the branching networks of river systems have a hydromechanical interpretation related to minimal total energy expenditure [Sel85]. As another example, microeconomic theory is based on price convergence to an efficient market-clearing equilibrium, but it is presumed to arrive there through individual utility-maximizing decisions of buyers and sellers [Bad01].

A dramatic example of the optimality of biogeography is Krakatoa, a volcanic island in the Indian Ocean which erupted in August 1883 [Win08]. The eruption was heard from thousands of miles away and resulted in the death of over 36,000 people, mostly from tidal waves whose remnants were recorded as far away as England. The eruption threw dust particles 30 miles high which remained aloft for months and were visible all around the world. Rogier Verbeek, a geologist and mining engineer, was the first visitor to Krakatoa, six weeks after the eruption, but the surface of the island was too hot to touch and showed no evidence of life. The island was completely sterilized [Whi93]. The first animal life (a spider) was discovered on Krakatoa in May 1884, nine months after the eruption. By 1887, dense fields of grass were discovered on the island. By 1906, plant and animal life was abundant. Although volcanic activity continues today on Krakatoa, by 1983 (one century after its desolation) there were 88 species of trees and

53 species of shrubs [Whi93], and the species count continues to increase linearly with time. Figure 1 shows the increase of bird and reptile species on Krakatoa since its eruption. Life immigrates to Krakatoa, and immigration can make the island more habitable, which in turn makes the island more friendly to additional immigration.

There are many other examples of biogeography as an optimization process, which is a type of mutually optimizing life/environment system. These examples include the Amazon rainforest [Har06], the temperature of Earth [Har06], the atmospheric composition of Earth [Len98], and the mineral content of the ocean [Lov90].

Many of the tasks proposed here involve the modification of the basic BBO algorithm to incorporate additional features from natural biogeography. The logical motivation for this approach can be summarized as follows.

Premise 1: Biogeography is an optimization process.

Premise 2: The biology community has quantified many details of biogeographical processes.

Hypothesis: Incorporation of biogeographical details in BBO will improve its performance.

The validity of our hypothesis will be tested using newly developed theoretical models (Markov models and dynamic system models) and simulation (standard optimization benchmark problems).

3. Biogeography-Based Optimization

In BBO, solutions are analogous to islands, and solution features are analogous to species. Note that in BBO, islands represent problem solutions. This is much different than island GAs, in which islands represent *populations* of solutions [Sko05].

BBO is not intended to be a simulation of natural biogeography. We are rather inspired by biogeography as the optimization metaphor that motivates the new algorithm that is called BBO. This is similar to other biologically-inspired evolutionary algorithms. An evolutionary algorithm is not usually a precise simulation of its motivating framework, but is only an approximate simulation, developed for the purpose of optimization. There are obvious disanalogies between natural biogeography and BBO, but we use natural biogeography as a motivating framework; we are not trying to simulate natural biogeography.

We use the notation $x(s)$ to denote a solution feature of island x . The number of islands is equal to the population size. The number of species on each island is equal to the dimension of the problem. The emigration rate μ of each solution is proportional to fitness, and is the likelihood of sharing solution features with other solutions. The immigration rate λ of each solution is inversely proportional to fitness, and is the likelihood of replacing current solution features with new features. Although there can be many varieties of BBO, the original algorithm [Sim08a] is outlined below, which depicts one BBO generation.

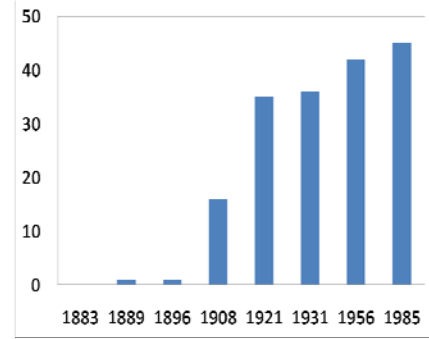


Figure 1 – Number of bird and reptile species on Krakatoa. Data is from [Lom10, p. 491].

BBO Algorithm (One Generation)

Set each individual's emigration rate μ proportional to its fitness
Set each individual's immigration rate λ inversely proportional to its fitness
For each individual x_i
 For each individual feature s_k
 Use λ_i to probabilistically decide whether to immigrate s_k to x_i from another individual
 If immigrating then
 Use μ to probabilistically select the emigrating individual x_j
 $x_i(s_k) \leftarrow x_j(s_k)$
 End if
 Next individual feature
 Probabilistically decide whether to mutate individual x_i
Next individual

Immigration rate λ is a real number between 0 and 1, and represents the probability per solution feature that a solution feature will be replaced. The probabilistic selection of the emigrating island using μ is typically done using standard fitness-proportional (roulette wheel) selection.

The BBO algorithm above is a basic outline. There are many implementation issues that we do not address in this proposal. For example, initialization, solution representation, and mutation, can be implemented as in any other evolutionary algorithm. Although these issues are important, we do not discuss them in this proposal because they are not distinguishing characteristics of BBO. It is rather the migration mechanism (i.e., the sharing of solution features) that is its distinguishing characteristic. BBO implementation details are given in many of the references that accompany this proposal [Sim08a, Sim10b, Sim10c].

BBO has features in common with other evolutionary algorithms, but there are also fundamental differences.

- Although BBO is a population-based optimization algorithm, it does not involve reproduction.
- BBO has the unique feature that each solution uses its own fitness to decide whether or not to accept features from other solutions. **This is the primary, critical distinctive of BBO, and it can make a significant difference in optimization performance, as seen in Section 7.1 below.**

These points distinguish BBO from algorithms such as GAs and evolutionary strategies. BBO has the most in common with strategies such as particle swarm optimization (PSO) and differential evolution (DE) which maintain their solutions from one generation to the next, but which allow each solution to learn from its neighbors. However, BBO can be contrasted with PSO and DE in that BBO solutions are changed directly via immigration from other solutions (islands).

In view of its unique motivating framework (island biogeography), it is not surprising that BBO has differences from other evolutionary algorithms. Some open research questions include: How do these differences make the performance of BBO differ from other evolutionary algorithms? What do these differences say about the types of problems that are most appropriate for BBO? This proposal will be directed in part to answering these questions.

3.1 Initial BBO Results

The first tests of BBO were comparisons with seven other population-based optimizers on 14 benchmark problems, each with 20 dimensions [Sim08a]. BBO and the stud GA were the two best algorithms in terms of average performance. See [Ma10a] for more complete comparisons. BBO was also tested on a real-world sensor selection problem for aircraft engine health estimation, a problem of combinatorial

complexity. This was a good test for the applicability of BBO to practical engineering problems. BBO again provided the best performance, as shown in Table 1.

	ACO	BBO	DE	ES	GA	PBIL	PSO	SGA
Average performance	8.22	8.01	8.06	8.15	8.04	8.18	8.14	8.02
Best performance	8.12	7.19	7.60	8.05	8.02	8.08	8.06	8.02

Table 1 – Optimization results for an aircraft engine sensor selection problem [Sim08a]. The numbers show the average solutions over 100 Monte Carlo simulations, and the best solutions found during those 100 simulations.

Acronyms are defined as follows: ACO = ant colony optimization, BBO = biogeography-based optimization, DE = differential evolution, ES = evolutionary strategy, GA = genetic algorithm, PBIL = population-based incremental learning, PSO = particle swarm optimization, SGA = stud genetic algorithm.

Another recent practical test was a comparison of BBO and GA on a 30-bus electrical power optimization problem. It was seen that BBO performed an average of **50% better than a comparable GA** [Rar09]. Other practical applications of BBO have been published for power distribution problems [Bha10], antenna design [Sin10], image recognition [Joh10], cardiac disease diagnosis [Ovr10], gear train optimization [Sav09], and robot controller tuning [Loz11].

3.2 Why Does BBO Work?

The question “Why does BBO work?” can be addressed from several perspectives.

- (1) One approach is to say that we do not care why it works, we simply observe that it does work and so we take advantage of its capabilities. This is a practical engineering-oriented viewpoint but not satisfying to most researchers.
- (2) Another approach is to say that natural biogeography is an optimization process (see Section 2 above), so an algorithmic implementation of biogeographical processes should be able to optimize engineering systems. This explanation seems to be getting closer to the truth, but is still superficial.
- (3) Another possible explanation is that BBO has much in common with GAs in terms of information-sharing between solutions, and so all of the literature that discusses why GAs work also applies to BBO. This explanation begins to touch on some fundamental truths. First, we see that natural optimization processes that are as disparate as genetics and species migration have much more in common than initial appearances would indicate. Second, we consider the fact that biogeography generally involves populations that are smaller than those involved in natural genetics. That is, the number of islands in an archipelago is smaller than the number of individuals in a species. This indicates that BBO should be more effective than GAs for smaller populations. This is what we see in some of our preliminary results. We compared BBO and GA on 14 standard benchmarks with various dimensions and population sizes.
 - Population size 200: BBO performs better than GAs **57%** of the time
 - Population size 50: BBO performs better than GAs **79%** of the time
 - Population size 20: BBO performs better than GAs **86%** of the time

The indication from this data is that as population size *decreases*, relative BBO performance *improves*. This implication is important for practical problems, because fitness function evaluation for real-world problems can require simulations or experiments that take hours or days [Seb09].

4. Motivation for the Proposed Research

In this section we summarize the objectives, the significance, and the relationship to current and previous work, of the proposed research.

4.1 Research Objectives

There are three primary objectives of this work.

- (1) **The first objective** is to establish mathematical tools for the analytical investigation of BBO. These tools include Markov models, dynamic system models, and statistical mechanics models. Most evolutionary algorithm claims are made on the basis of simulation. One of the great needs in the evolutionary algorithm community is the development of theoretical results to more rigorously justify the algorithms that are being used, and to provide direction for future work and modifications. This project proposes to establish those results for BBO.
- (2) **The second objective** is to place BBO on a firm scientific foundation. BBO fundamentals have been established, but a stronger connection between BBO and biogeography needs to be made in order to take advantage of biogeography results. The areas that will be studied as part of this work include migration rate modeling, species populations, species interactions, and migration time correlation. These are areas of biogeography that are well known and that can therefore be exploited for better BBO performance.
- (3) **The third objective** is the implementation of the results in a software package that will be available on the Internet for general purpose optimization. The web-based software will be available on a dedicated web site at the PI's university. Source code will also be freely available for further in-depth study by other researchers and students. This will be one of the primary (but not the only) means for the dissemination of this research.

4.2 Expected Significance of Research

This work will be significant to the engineering community and to society in several ways.

- (1) **Tools** – This project will develop general mathematical tools for the use of BBO, including Markov models, dynamic system models, and statistical mechanics models. These tools have already been developed for GAs, but this work will extend those tools to BBO, which has a more sophisticated mechanism for sharing solution features. Preliminary results have already been obtained in this area as discussed in Section 5.1 below. One of the pressing needs in the evolutionary algorithm research community is theoretical results, and these tools can help provide that need.
- (2) **Optimization** – Just as other evolutionary algorithms have had a major impact on many diverse areas of engineering, the establishment of BBO will allow it to have an impact on any area of science or engineering that involves optimization, which is essentially every area of science and engineering. Two areas that have already benefited from BBO are sensor selection [Simo08a] and power system optimization [Rar09]. Other research that is currently being conducted with BBO includes medical imaging and diagnosis, automotive design, and image recognition.
- (3) **Software** – This project will involve the development of BBO software in a user-friendly web-based software package. Source code will also be freely available on the BBO web site. This will make it convenient for others to begin research in this new area.

4.3 Relation to Current Work

This project is closely related to the primary research interests of both PIs and thus naturally ties into their research plans. It is also related to current research priorities in the evolutionary computation community and thus is well connected with a national research program.

- (1) **Dr. Simon's research** focuses on biologically motivated optimization algorithms. He is the inventor of BBO [Sim08a], so this research program extends his present work. He is also a leader in the area of optimal filtering [Sim06], which has a strong potential to apply to BBO for uncertain fitness landscapes [Str01]. His relevant publications are listed in his biographical sketch. Dr. Simon is presently involved in a BBO-related NSF GOALI grant with the CMMI Division. However, that grant is focused on creating a multi-objective optimization framework for BBO and applying it to

automotive design. This present proposal is focused on the development of tools for BBO analysis, the influence of established biogeography theory on the BBO paradigm, and communication system applications.

- (2) **Dr. Smith's research** focuses on the application of island biogeography theory to mainland habitat patches, especially as it relates to conservation in fragmented systems. Much of this focus has been in the Great Plains ecoregion of North America where black-tailed prairie dog towns serve as model islands. Another area of interest is the influence of surrounding habitat types on local species. This is a modification of classic biogeography theory which assumes that intervening habitat between islands is inhospitable. In mainland systems, this is not always true and thus is a modification of the equilibrium theory that BBO is well-suited to investigate. Relevant publications are listed in his biographical sketch.
- (3) **BBO research** is being performed all over the world. Since its original publication in December 2008, BBO has been the subject of over 50 publications by over 60 authors [Sim10]. This research has included theoretical work such as extensions to constrained problems [Ma10b], hybridization with other algorithms [Joh10], and extensions to combinatorial problems [Mo10]. It has also included many applications, such as those discussed above in Table 1 and the following text. The rapidity with which BBO is spreading indicates that it is a competitive optimization algorithm that merits further research into its fundamental properties and theoretical basis.
- (4) **There are many open research areas** that have been identified in the area of evolutionary algorithms, and many of these are applicable to BBO. For example, the development of analytical tools is an important need, and this research will address that need for BBO. Characterizing the types of problems for which various evolutionary algorithms excel is an active area of research [Bau01] which will be studied in this proposal. Evolutionary algorithm convergence [Lau02, Sem03] is an open problem, and this research will obtain BBO convergence results. The maintenance of diversity in evolutionary algorithm populations [Tof03] is a current issue which this research will address.

5. Task Descriptions

This section discusses each task of the proposed research in detail. In summary, these tasks include:

- | | | |
|---|---|---------------------------------------|
| <ol style="list-style-type: none"> (1) Markov models (2) Dynamic system models (3) Statistical mechanics models (4) Migration rate modeling (5) Initial immigration (6) Species populations (7) Species interactions (8) Migration time correlation (9) Software development | } | <i>Mathematical modeling tools</i> |
| <ol style="list-style-type: none"> (6) Species populations (7) Species interactions (8) Migration time correlation (9) Software development | } | <i>Modifications to the BBO model</i> |

5.1 Markov Models

Markov models are tools that can be used to analyze the distributions involved in probabilistic algorithms. For a simple GA with roulette-wheel selection, the probability of transitioning from one population vector has been summarized in [Ree03]. This gives the transition matrix for the GA, and a wealth of Markov tools can then be used to analyze the behavior of the GA (time to convergence, limiting probability of population, etc.).

This analysis has recently been extended to BBO and GAs with global uniform recombination (GA/gur) [Sim10b, Sim10c]. The dimension of the transition matrix is equal to the number of possible populations, which is $(n+N-1)$ -choose- N , which grows very rapidly with both n (search space cardinality) and N (population size). Therefore, only problems of very small search space can be explored with this

method. However, some preliminary results showing the superiority of BBO are in Table 2.

Problem Type	Probability that the population converges to the optimum		Probability that the population does not contain any optima	
	GA/gur	BBO	GA/gur	BBO
Unimodal	40%	76%	57%	17%
Multimodal	57%	85%	41%	10%
Deceptive	62%	91%	36%	7%

Table 2 – The probability of convergence to the global optimum, and the probability of convergence to a population that does not contain any optima. Results are for a three-bit problem ($n=8$) with a population size of four ($N=4$) and a mutation rate of 0.1% per bit. BBO far outperforms GA/gur in every case. The table shows analytical results which have also been confirmed with simulations [Sim10b, Sim10c].

A careful look at the BBO algorithm on page 3 shows that BBO is very similar to GA/gur. But GA/gur does not have the selection pressure that comes from BBO’s immigration rate. This additional selection pressure makes a huge difference in performance, as seen in Table 2, at least for small populations, and indicates that BBO research has a high potential payoff.

The primary limitation of Markov models is that they can be applied only to problems of small dimension because of the rapid growth of the transition matrix with search space cardinality and population size. However, for many evolutionary algorithms one desires to prevent duplicates in the population in order to maintain diversity and improve performance [Ron98]. This assumes that $n > N$ which easily holds true for all practical problems. This reduces the number of possible populations to n -choose- N . The number of possible populations still increases rapidly with the problem dimension, but for relatively small n the reduction in transition matrix dimension could allow calculations that would otherwise not be possible. Combining this idea with the grouping of states to reduce the Markov dimension [Spe97, Ree03] will make otherwise intractable problems computable.

Another task that we propose is to extend BBO Markov analysis to other flavors of BBO. Standard BBO uses λ to decide whether to immigrate and μ as an emigration selection parameter. However, other versions of BBO can also be implemented. For example, μ could be used to probabilistically decide whether or not to emigrate and λ could be used as an immigration selection parameter. Another possibility is to immigrate or emigrate a randomly chosen solution feature, rather than considering the possibility for every feature. Yet another possibility is to always emigrate from the best island, or immigrate to the nearest island. Each of these variations will result in a different transition matrix and different analytical results for representative problems. The extension of Markov analysis to these various approaches will provide guidance as to the best approach to use for various types of problems.

5.2 Dynamic System Models

Dynamic system models can be used to analyze the properties of an evolutionary algorithm in the limit as the population size tends to infinity. Vector $p(t)$ is defined as a vector whose elements contain the proportion of each possible individual at the t -th generation, so $p(t)$ is an n -element vector, where n is the cardinality of the search space. Define f as the n -element fitness vector, and U as the $n \times n$ mutation matrix. Define M_k as the matrix whose i, j -th element is the probability that x_i and x_j cross over to produce x_k . It is shown in [Vos99, Ree03] that:

$$p_k(t+1) = \frac{p(t)^T \text{diag}(f) U^T M_k U \text{diag}(f) p(t)}{(f^T p(t))^2}.$$

This equation can be used to analyze the infinite-population GA, including trajectories and fixed points for various types of problems. One of the tasks in this proposal is to extend this method to BBO so that alternative BBO algorithms based on biogeography theory can be explored analytically.

Dynamic systems analysis differs from Markov analysis in several ways. First, a Markov system has $(n+N-1)$ -choose- N states, where n is the size of the search space and N is the population size. But a dynamic system has only n states. This makes dynamic systems analysis more tractable for reasonably-sized problems. Second, Markov analysis is used to analyze the probability distribution of a finite population. But dynamic systems analysis is used to analyze the proportionality of each individual in an infinite population. Both tools are useful for gaining a qualitative understanding of evolutionary algorithms, each having its own particular advantages and limitations.

Dynamic systems models have not yet been developed for BBO. But just as Markov analysis has begun to be extended to BBO [Sim10b, Sim10c], this proposal will extend dynamic systems analysis to BBO. This will provide insight into the behavior of BBO for different types of problems and will allow analytical comparisons between GAs and BBO for various types of problems.

5.3 Statistical Mechanics Models

In physics, we often encounter systems that include many interacting components. Theoretically we could model the behavior of each component to analyze the behavior of the entire system. Practically, this is not possible because of the large number of components. Therefore we instead derive a statistical approximation for the behavior of the components. We lose some information when we do this, but we gain computational tractability, and we can still capture the essential behavior of the system. This approach has been taken with GAs [Ree03, Chapter 7] and we propose to apply the same idea to the generation of BBO models. As in Section 5.2, $p(t)$ is defined as the n -element vector whose elements contain the proportion of each possible individual at the t -th generation, where n is the cardinality of the search space. Vector $p(t)$ approximates the probability distribution of the population over the search space, and we approximate $p(t)$ using Fourier series, wavelets, cumulant generation functions, or other similar approximators. Although we have not developed this approach for BBO to any great extent, we propose it as a new mathematical modeling tool for BBO. This will allow us to obtain approximate analytical results describing the behavior of BBO.

The following tasks move away from tool development, and move toward the biogeographical foundations of BBO. The modeling tools developed in the above tasks will be used to analyze the BBO modifications described below. These modifications add complexity to BBO, but they also allow for the exploitation of mechanisms that improve optimization performance, and they leverage a large body of biogeography research.

5.4 Migration Rate Modeling

The initial model proposed for immigration and emigration rates is linear based on rank. The proposed model was linear to facilitate the theoretical determination of an equilibrium population for a given island. This model gives a closed-form solution for the equilibrium species count for each island [Sim08a]. However, there is no reason to suppose that these rates are linear. Empirical data suggest that these rates are probably convex nonlinear functions of the number of species as shown in Figure 2 [Mac67, Whi98]. Some preliminary results in this direction have very recently become available in [Ma10a]. In that paper, a set of standard optimization benchmarks are tested with 50 Monte Carlo simulations. In a comparison of linear migration curves like those shown in the original BBO paper [Sim08a], and sigmoid migration curves like those found in natural biogeography [Whi98], the sigmoid migration curves result in better optimization performance in every benchmark (23 out of 23 benchmarks). The important albeit preliminary conclusion that we draw from these results is that **incorporating features of natural biogeography in BBO can result in better optimization performance.**

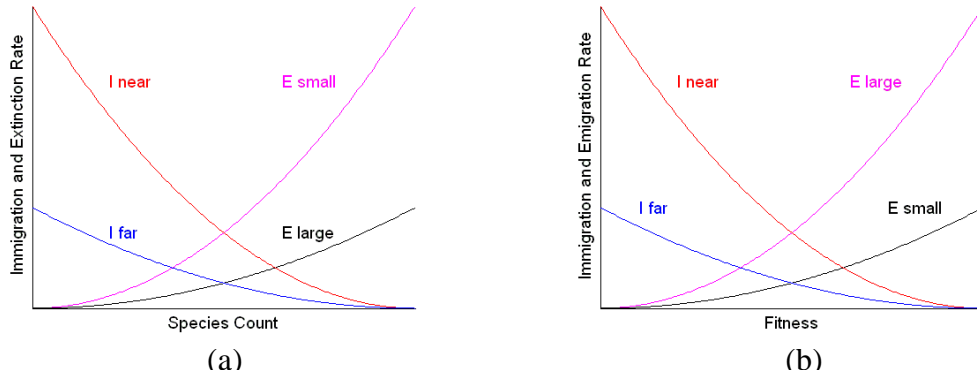


Figure 2(a) – Biogeography immigration curves for islands that are far from a mainland source or near a source, and extinction curves for islands that are large or small (adapted from [Whi98, p. 119]). (b) BBO migration curves. In BBO the E curves are used for emigration rather than extinction, so the “large” and “small” labels are reversed from Figure 2(a), and the horizontal axis is fitness rather than species count.

Modeling BBO immigration and emigration rates as nonlinear will allow more flexibility and open up a broader range of investigations of their effect on BBO behavior. The analytical tools discussed above will be extended to nonlinear migration so that the effects of nonlinearity can be studied not only empirically, but also theoretically. Another intriguing topic for research is the use of Markov analysis along with Monte Carlo simulations to optimize the rates for specific problem types.

5.5 Initial Immigration

Classical island biogeography theory indicates that immigration rate decreases as the number of species increases, as shown in Figure 2. In BBO this corresponds to a decrease in immigration rate as solution fitness increases. This means that as a solution becomes more fit, the probability of incorporating features from other solutions decreases. However, more recent advances in biogeography indicate that for some “pioneer” species (plants, for example), an initial increase in species count results in an initial increase in immigration rate [Wu95]. This is because these early immigrants modify the island to make it more hospitable to other species. That is, the positive effect of increased diversity due to initial immigration overcomes the negative effect of increased saturation (new immigrants are more likely to be already represented). In BBO this would correspond to an initial increase in immigration rate as a very poor problem solution initially improves its fitness. This can be viewed as a temporary positive feedback mechanism in BBO. A very poor solution accepts features from other solutions, increasing its fitness, which subsequently increases its likelihood of accepting even more features from other solutions. This is depicted in Figure 3. This idea can also be incorporated into other evolutionary algorithms, but its initial motivation comes from BBO.

5.6 Species Populations

In island biogeography, species with larger populations are more likely to emigrate [Han93]. This is because emigration is stochastic, so large populations have a higher probability of some individuals being carried by wind or flotsam to neighboring islands. A larger population of a BBO feature will make it less likely that the feature will become extinct. A larger population will also make it more likely that the feature will emigrate. This has the effect of giving feature-specific emigration curves.

Note that this is not the same as simply changing the migration rates for the whole island; it is rather adjusting the migration rates for each particular feature. This would result in each BBO island (solution) having separate emigration curves for each of its solution features.

Dispersal model – The ultimate goal of population modeling is to retain beneficial solution features in the population, while allowing a greater probability for the removal (extinction) of harmful solution features. Given some species s_k , its population p_{ki} on the i -th island can be modeled as follows [Adl94].

$$p_{ki}(t+1) = (1 - d_{ki})p_{ki}(t) + d \sum_{j=1}^N g_{kji} p_{kj},$$

where N is the number of islands, t is a time index (i.e., generation number of the BBO algorithm), d_{ki} is the fraction of species k that emigrates from the i -th island at each time step, and g_{kji} is the fraction of the population of species k that leaves the j -th island and arrives at the i -th island. Island proximity is used to obtain the number of individuals that leave island j and arrive at island i . Set $d_{ki} = d_k$ – that is, the fraction of species k that leaves island i is constant for all islands. Set $g_{kij} = h(i, j) d_k$ – that is, the fraction of species k that migrates from island j to island i is a function of the Hamming distance $h(i, j)$ between the two islands. All that remains is to define d_k . This can be proportional to the fitness contribution of the k -th feature s_k . That is, species that contribute to the habitability of an island are more likely to emigrate elsewhere. A species that positively contributes to habitability has its own “sub-fitness” and is thus fit enough to emigrate. This increases the habitability of destination islands. d_k can thus be obtained as the correlation between feature s_k and fitness. If a solution feature is present on an island, then the above equation describes the dynamics of the “population” of that feature.

Estimation of distribution algorithms use population-wide correlations between solution features and fitness [Lar02]. However, they do not maintain individual population members. This proposed BBO modification combines the maintenance of individuals in the population, and solution feature correlations with fitness.

Age – Another option for modeling species population is to use the age distribution of the individuals in the species [Rey01, p. 308]. Aging factors have been used for individuals in GAs [Gho98] but not for species in BBO. In GAs, old individuals have a higher likelihood of being purged from the population in order to prevent stagnation and to increase the GAs exploratory component. The idea in BBO is similar but not identical. In BBO, old individuals correspond to solution features rather than solutions. Old solution features are more likely to die and decrease the virtual population of that feature. The decrease in the population of a solution feature makes that feature less likely to be shared with other solutions. This gives BBO an opportunity to introduce new solution features to existing solutions.

In island biogeography, species age influences extinction rate, mobility, and reproductive ability [Gro06]. Just as individual mortality for some species is high at a young age, low at middle age, and high again at old age, species mortality follows the same inverted bell-shaped curve. Young species tend to be unstable and susceptible to extinction. Middle-aged species are well-established but still mobile. Old species are genetically and geographically stagnant and less adaptable. In BBO, species represent solution features, so solution features that have been recently introduced to a solution have a higher extinction rate and a low emigration rate. Middle-aged features have a lower extinction rate and higher emigration rate, and older features revert to the pattern of high extinction rate and low emigration rate.

Effective population size – In biology, not all population sizes are equal. In general, we can write the *effective* population size of a species as $N_e = k(kN - 1) / (V + k(k-1))$, where N is the population size, k is the mean number of surviving progeny, and V is the variance in the number of progeny [Gro06, p. 388]. A large variance in the number of surviving progeny indicates a poor

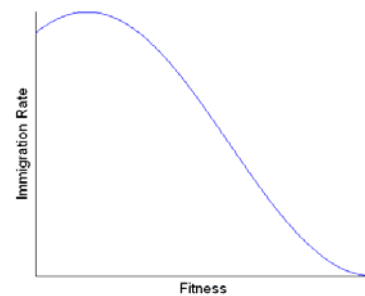


Figure 3 – This proposed model shows that immigration initially increases with fitness (contrast with Figure 2). This gives improving solutions momentum to continue improving. As a solution continues to become more fit, immigration eventually begins to decrease in order to give less fit solutions relatively greater opportunity.

distribution of genes in the next generation, resulting in a genetic bottleneck.

In BBO, effective population size is modeled by the abundance of a single species. This is different than the solution diversity discussed in Section 7.4 above. Here, the idea is to estimate the diversity of a solution *feature*. Population size N is related to the correlation between solution features and fitness. Mean progeny k is proportional to N , which assumes a constant reproductive capacity for all species. This gives $N_e \approx aN / (V + aN)$, where a is a constant. Variance V is inversely proportional to the correlation between the solution feature and island distance. (Island distance is related to differences in solution feature space.) This shows that solution features that exist in a wide variety of islands (where “wide variety” refers to solution feature space) are considered robust in terms of dispersion. Effective population size is a more specific parameter than raw population size, and is thus used to determine solution feature sharing and extinction rates.

5.7 Species Interactions

Predator/prey relationships – In biology, certain species have adversarial relationships. These relationships do not necessarily harm the prey species. For instance, prey may respond to predators by reducing the exploitation of their own resources, thus benefiting themselves in the long term [Han97]. However, another scenario is the one in which predators reduce prey to such an extent that one or both populations face extinction.

Predator/prey relationships can be inferred from a BBO population by examining solutions and noting which pairs of solution features have a low probability of coexisting. Those solution features are modeled as a predator/prey pair. Combining this information with the ISI contribution of each solution feature results in the predator being defined as the adversary that is positively correlated with ISI, and the prey defined as the adversary that is negatively correlated with ISI. The predator/prey populations might lead to a nonzero equilibrium population, or it might lead to the extinction of one or both populations [Got08, Han97]. This information is used throughout the population of islands to increase the likelihood of predator presence, and reduce the likelihood of prey presence. This should improve the mean fitness of the population of solutions. This idea can be combined with Task (1a) in order to determine how this affects BBO convergence and stability. Most predator/prey models are for two-species systems. These existing models can be used in BBO, but a more complete description would be obtained if multi-species predator/prey models are used.

Resource competition – In contrast to the above, similar species compete for similar resources in nature. Therefore, it is unlikely that many similar species occupy the same island, especially if they have large populations [Til94]. In BBO, this means that it is unlikely that solution features emigrate to islands that already have large similar populations. Alternatively, it could mean that emigration rate is not affected, but survival likelihood is lower. This also means that if two solution features have equal probability of extinction, then the feature most similar to other features in the solution will become extinct. This is a different type of interaction than that described in the preceding paragraph. However, both models are plausible, and competition is generally viewed in biology as a more significant driver of community composition than predator/prey interactions.

We do not discuss here how similarity will be measured, because there are many different ways that it could be measured (e.g., Hamming distance). We are primarily interested here in what to do with the similarity after it is measured. Also, we assume that similarity measures are restricted to solution features that represent similar features of a solution. The choice of which solution features can be compared in a meaningful way is problem-dependent.

The two perspectives discussed in paragraphs (a) and (b) above have an analogy in GAs. Sometimes niching is used in GAs to encourage similar solutions to crossover [Sta02], and sometimes fitness sharing is used in GAs to artificially increase the fitness of unique solutions and give them a better chance of

crossing with diverse population members [Yu05].

5.8 Migration Time Correlation

In natural island biogeography, if a species moves in a given direction, it is more likely to continue in that direction to the next island. This is because emigration is influenced by prevailing winds and currents, and those winds and currents have a nonzero time correlation. This is quantified by biodiffusion theory, the telegraph equation, and the equation of diffusion [Oku01]. If a species migrates from island A to island B, it is likely to continue in the same direction to the next island in the chain at the next time step. In BBO this means that if a solution feature migrates from one solution to the next, it is likely to continue migrating in that direction (with respect to the solution space) at the next evolutionary generation. Note that this is not the same as changing the migration rate for an entire island (solution), but biases the migration direction of migrating solution features.

5.9 Software Development

The result of this work will be made available to the engineering community as Internet-based software. CSU offers a Master of Science in Software Engineering (MSSE). CSU's MSSE program is the first in the state of Ohio, and the PI of this proposal is actively involved in teaching in this program. This gives the PI access to students and resources to accomplish this task in a systematic way.

General-purpose BBO software will be written and placed on a dedicated web site at CSU. A graphical user interface (GUI) will allow the user to input relevant information, and the software will output the results of the optimization progress. This will allow users to explore the effect on BBO of the various algorithmic modifications discussed above.

The software will be written in the quasi-open source language Flex [Fle08]. Flex is a web-based GUI that includes many of the features of the Matlab GUI. Flex will be accessible from any web browser with Flash Player installed. A Flex-based BBO GUI will allow users to experiment with different models and see how those models affect BBO performance and results. The GUI could also be used by classroom instructors with an Internet connection. Flex is very similar to HTML and JavaScript, but is much easier to work with. HTML and JavaScript are also limited to the low-bandwidth communication method Ajax [Hin06], while Flex communication can be implemented through .NET remoting [McL02], which is very simple to use and can interface with any web-based communication protocol.

Source code will be made available on the BBO web site for researchers who want to extend this work. The source code will be written in a high-level, popular language such as Matlab, C, or Java.

6. Project Timeline

Progress will be monitored with weekly meetings between the PIs and the graduate students. The PIs will supervise one engineering graduate student at CSU and one biology graduate student at UA. The CSU and UA teams will meet every other week, either in person or via telecon. More frequent meetings may be scheduled as circumstances dictate. (CSU and UA are only about 30 miles apart.) The time schedule for this research is outlined below. The task numbers specified in the timeline correspond to subsections in Section 5 of this proposal.

- (1) **Fall 2011** – The PIs will supervise graduate students on the fundamentals of biogeography and BBO, and will begin **Tasks 1–3**. This will give the students grounding in biogeography and BBO and the analytical tools that we propose to develop for BBO.
- (2) **Spring 2012** – The PIs will continue supervising the graduate students as in fall 2011. This will conclude the students' education in the fundamentals of this research, and will position them to make adjustments to biogeography models, and to apply BBO to benchmark problems.

- (3) **Summer 2012** – The graduate students will focus on **Tasks 4–5**. This will allow them to begin extending BBO to include more features from biogeography theory.
- (4) **Fall 2012** – The PIs will continue supervising the graduate student tasks as in summer 2012.
- (5) **Spring 2013** – The PIs will supervise the graduate students on **Task 6**. Note that this task has many permutations and is therefore allotted more time than the other tasks.
- (6) **Summer 2013** – The PIs will supervise the graduate students on **Tasks 6–8**.
- (7) **Fall 2013** – The PIs will supervise the graduate students on **Task 6** and **Task 9**, which is the software/web page implementation of BBO research results, and which comprises a major part of the dissemination portion of this research.
- (8) **Spring 2014** – The PIs will supervise the graduate students on **Task 9** and promising directions from **Tasks 4–8**, depending on the research results of the previous two years.
- (9) **Summer 2014** – The PI will supervise the graduate students on **Task 9** and one or more of **Tasks 4–8**, as in spring 2014.

7. Dissemination

The results of this work will be disseminated in several ways.

- (1) **Software** – The algorithms and software developed under this project will be packaged as a software tool with a user-friendly interface. This software will be available on a dedicated web site at the PI's university. Source code for various BBO algorithms will be available on the web site.
- (2) **Publications** – This work will be submitted to journals and conferences as results become available. Some typical conferences include the *IEEE Congress on Evolutionary Computation*, the *American Control Conference*, and the *Genetic and Evolutionary Computation Conference*. Typical journals include *Evolutionary Computation* and the *IEEE Transactions on Evolutionary Computation*. Results will also be of interest to the biogeography community and will be submitted to professional biogeography journals and at meetings for groups such as the *International Biogeography Society*.
- (3) **Workshops** – As BBO conference and journal publications continue to proliferate, BBO will become popular enough so that special sessions, tutorials, and workshops at conferences such as those listed above, will become attractive options to conference organizers. This will help publicize BBO to students and researchers. We plan to propose a special session at the 2012 *American Control Conference*, which will reach a broad audience of academic and industrial researchers.
- (4) **Integration into Course Work** – The PI of this proposal teaches a graduate course called Population Based Optimization [Sim08b] which covers evolutionary algorithms. The results of this research will be integrated into the lecture material and assignments for the course. This will help disseminate the research results, and will also encourage graduate students to pursue BBO for their master's and doctoral research. BBO will also be incorporated into a required methods course as part of UA's Integrated Bioscience program. Recognizing that future major advances in understanding complex systems will be made across levels of biological organization, this new program is designed to train students in ways that cross the boundaries of disciplines.

8. Under-Represented Groups

Cleveland State University is a leader in enrolling and graduating minorities, and has been recognized nationally in the top 100 in *Diverse Issues in Higher Education*. In fall 2009, 30% of undergraduate students were under-represented minorities [CSU2, p. 60]. CSU is noted for its multicultural initiatives, including more than 100 courses with a cultural/ethnic focus. CSU ranks first in Ohio for graduating minority students with master's degrees [CSU1, p. 7]. CSU has two programs for increasing research opportunities for under-represented students: the state-funded STARS Program and the federally-funded McNair Program. This proposal thus has a good opportunity to attract minority students to research work. The PI of this proposal has a history of mentoring and publishing with minority undergraduates [Chu08,

Sch11], including an African-American doctoral graduate [Tum02]. Likewise, the University of Akron has a large minority population (16%) and similar initiatives in place to encourage minority students to pursue graduate studies. The PIs of this proposal are committed to increasing diversity, and consider diversity to be an essential component of their research. Several undergraduate students at CSU and UA who belong to under-represented groups (females and African-Americans) are qualified for graduate work, and one or more of them will be recruited to work on this research. This will enable these under-represented minorities to obtain advanced degrees, which will place them in a better position to take leadership roles during their engineering careers.

9. Results from Prior NSF Support

The PI of this proposal, Dan Simon, has received prior NSF awards. Following is a description of the award that is most closely related to the current proposal.

- a) The NSF award number is 0826124. The initial amount and period of support was \$295,879 from August 2008 to August 2011; including supplements, the award amount is currently \$505,206 from August 2008 to August 2013.
- b) Project title: “GOALI: Biogeography-Based Optimization of Multiple Related Complex Systems.”
- c) The work has involved the initial development of biogeography-based optimization (BBO), extensions to special types of optimization problems (e.g., constrained and multi-objective), and applications (e.g., power distribution, robot control, ECG analysis, and prosthetic leg control). The work has involved three doctoral students, three masters students, and three undergraduate students, including one under-represented minority. The work has resulted in several academic/industry collaborations, and two student internships.
- d) The work has thus far resulted in 13 peer-reviewed publications, including eight conference papers, four journal papers, and one book chapter.
- e) BBO source code and publications resulting from this research are available for download from the PI’s BBO web site.
- f) The current proposal is not for renewed support.

The co-PI of this proposal, Greg Smith, has received prior NSF awards. Following is a description of the award that is most closely related to the current proposal.

- a) The NSF award number is DBI-0851782. The funded amount is \$180,001 and the period of support is from June 1, 2009 to May 31, 2012.
- b) The project title is “REU Site: Ecology at the Urban-Rural Interface.”
- c) This program is designed to introduce undergraduate students to ecological research. To date, 13 students have participated in the program and conducted summer-long research projects ranging from pollination ecology to small mammal demography to lake food web structure. Students have also worked closely with Metro Parks Serving Summit County, Cleveland Metroparks, and the National Park Service, gaining valuable experience in the management of natural spaces in an urbanizing landscape. Of the students that have participated in the program, 60% have been women and 30% have been under-represented minorities. The program also has provided opportunities for two graduate students to serve as research mentors for the undergraduate participants.
- d) All students in the program presented their research at a local undergraduate research symposium and one student presented research at a national limnological conference.
- e) Survey data collected during summer work are housed at the University of Akron Field Station. These data include presence/absence and abundance data on the species that were the focus of student research (birds, mammals, insects, etc.).
- f) The current proposal is not for renewed work.

Budget Description

During the first year, the PIs will supervise two one engineering graduate student at CSU and one biology graduate student at UA. The CSU PI will be paid for 25% of his time during the summer, and the UA PI will be paid for 33% of his time during the summer. Travel costs include expenses for conferences. Fringe rates are 5.5% for CSU students, 19.7% for the CSU PI, 14.5% for UA students, and 29% for the UA PI. Indirect rates are 48.5% at UA and 42% at CSU. The budget request for the first year is summarized as follows.

\$ 34,510	Graduate student stipends and summer salary
\$ 8,447	Graduate student tuition
\$ 12,770	Faculty salary
\$ 2,000	Travel
\$ 6,419	Fringes
<u>\$ 35,719</u>	Indirects
\$ 99,865	First Year Total

The second year is similar to the first year, with allowances for 3% wage increases, 6% tuition increases, and page charges for publications. The budget request for the second year is summarized as follows.

\$ 35,713	Graduate student stipends
\$ 8,954	Graduate student tuition
\$ 13,207	Faculty salary
\$ 3,500	Travel
\$ 1,000	Page charges
\$ 6,652	Fringes
<u>\$ 27,188</u>	Indirects
\$ 96,214	Second Year Total

The third year budget is similar to the second year, with allowances for 3% wage increases and 6% tuition increases. The budget request for the third year is summarized as follows.

\$ 36,959	Graduate student stipends
\$ 9,491	Graduate student tuition
\$ 13,659	Faculty salary
\$ 3,500	Travel
\$ 1,000	Page charges
\$ 6,892	Fringes
<u>\$ 28,072</u>	Indirects
\$ 99,573	Third Year Total

This gives a total three-year project budget of **\$295,652**.