Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

# Population Distributions in Biogeography-Based Optimization Algorithms with Elitism

Dan Simon, Mehmet Ergezer, and Dawei Du
Cleveland State University
Department of Electrical and Computer Engineering
Cleveland, Ohio, USA
{d.j.simon, m.ergezer, d.du}@csuohio.edu

***Abstract*—Biogeography-based optimization (BBO) is an evolutionary algorithm that is based on the science of biogeography. Biogeography is the study of the geographical distribution of organisms. In BBO, problem solutions are represented as islands, and the sharing of features between solutions is represented as migration between islands. This paper develops a Markov analysis of BBO, including the option of elitism. Our analysis gives the probability of BBO convergence to each possible population distribution for a given problem. We compare our BBO Markov analysis with a similar genetic algorithm (GA) Markov analysis. Analytical comparisons on three simple problems show that with high mutation rates the performance of GAs and BBO is similar, but with low mutation rates BBO outperforms GAs. Our analysis also shows that elitism is not necessary for all problems, but for some problems it can significantly improve performance.**

***Index Terms*—biogeography-based optimization, evolutionary algorithms, probability, combinatorics, Markov analysis**

## I. INTRODUCTION

Mathematical models of biogeography describe the migration, speciation, and extinction of species. The science of biogeography began with empirical descriptions by 19th century naturalists such as Alfred Wallace [1] and Charles Darwin [2]. Eugene Munroe was the first to introduce mathematical models of biogeography in 1948 [3], [4], and Robert MacArthur and Edward Wilson were the first to extensively develop and publicize them in the 1960s [5], [6].

Islands that are well suited as habitats for biological species are said to have a high island suitability index (ISI). Features that correlate with ISI include rainfall, topographic diversity, area, temperature, etc. The variables that characterize these features are called suitability index variables (SIVs). SIVs are the independent variables of the island, and ISI is the dependent variable.

Islands with a high ISI tend to have a large number of species, and those with a low ISI have a small number of species. Islands with a high ISI have many species that emigrate to nearby islands because of the accumulation of random effects on its large populations. Emigration occurs as animals ride flotsam, fly, or swim to neighboring islands.

Biogeography-based optimization (BBO) was first presented in [7] and is an example of how a natural process can be modeled to solve optimization problems. This is similar to what has occurred in the past few decades with genetic algorithms, artificial immune systems, simulated annealing, particle swarm optimization, and other areas of computer intelligence. Suppose that we have some problem, and that we also have several candidate solutions. A good solution is analogous to an island with a high ISI, and a poor solution is like an island with a low ISI. High ISI solutions are more likely to share their features with other solutions, and low ISI solutions are more likely to accept shared features from other solutions. This approach to problem solving is called biogeography-based optimization. As with every other evolutionary algorithm (EA), each solution also typically has some probability of mutation, although mutation is not an essential feature of BBO.

The goals of this paper are three-fold. Our first goal is to present an overview of BBO, which we do in Section II. Our second goal is to use Markov analysis to obtain the limiting distribution of BBO populations, which we do in Section III. Our third goal is to confirm the theory with simulation, compare BBO Markov theory with GA Markov theory, and analyze the effect of elitism on BBO performance, which we do in Section IV. We provide some concluding remarks and directions for future work in Section V.

This work was supported by Grant 0826124 in the CMMI Division of the Engineering Directorate of the National Science Foundation.

## II. BIOGEOGRAPHY-BASED OPTIMIZATION

### A. Biogeography

In this subsection we first give a brief overview of biogeography. Figure 1 illustrates a model of species abundance in a single island. The emigration rate $\mu$ out of the island, and the immigration rate $\lambda$ into the island, are functions of the number of species $S$ on the island. The maximum possible immigration rate $I$ occurs when there are zero species on the island. As the number of species increases, the island becomes more crowded, fewer species are able to successfully survive immigration, and the immigration rate decreases. The largest possible number of species that the island can support is $S_{max}$, at which point the immigration rate is zero.

If there are no species on the island then the emigration rate is zero. As the number of species increases, the island becomes more crowded, representative individuals of species are more likely to leave the island, and the emigration rate increases. The maximum emigration rate $E$ occurs when the island contains the largest number of species that it can support.



978-1-4244-2794-9/09/$25.00 

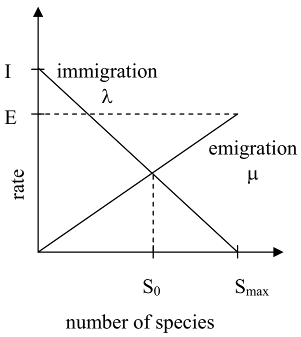



Fig. 1. Linear migration relationships for an island, adapted from [6]. The equilibrium number of species is $S_0$, at which point the immigration and emigration rates are equal.

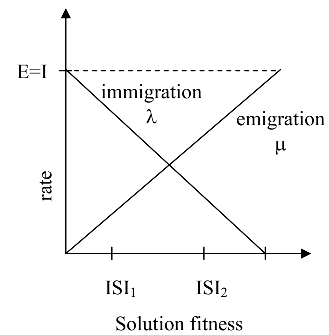



Fig. 2. Illustration of two candidate solutions to some problem using symmetric immigration and emigration curves. $\mathrm{ISI}_1$ is a relatively poor solution and $\mathrm{ISI}_2$ is a relatively good solution. $\mathrm{ISI}_1$ has a high immigration and a low emigration rate, and $\mathrm{ISI}_2$ has a low immigration and a high emigration rate.

We have shown the immigration and emigration curves in Figure 1 as straight lines, but in general they might be nonlinear. Nevertheless this simple model gives a general description of the process of immigration and emigration.

### *B. Biogeography-Based Optimization Algorithms*

Suppose that we have a problem and a population of candidate solutions that are represented as vectors. Further suppose that we have some way of assessing the goodness of the solutions. Good solutions are analogous to islands with a high ISI, and poor solutions are analogous to islands with a low ISI. Note that ISI is the same as "fitness" in other population-based optimization algorithms.

In biogeography, species migrate between islands. However, in BBO we instead migrate solution features (SIVs) between islands. We base the migration probabilities on a curve similar to that shown in Figure 1, but for the sake of simplicity we assume that all solutions (islands) have identical migration curves with $E = I$. Figure 2 illustrates the migration curves along with two solutions. $\mathrm{ISI}_1$ represents a poor solution and $\mathrm{ISI}_2$ represents a good solution. The immigration rate for $\mathrm{ISI}_1$ will therefore be higher than the immigration rate for $\mathrm{ISI}_2$, and the emigration rate for $\mathrm{ISI}_1$ will be lower than the emigration rate for $\mathrm{ISI}_2$.

We use the migration rates of each solution to probabilistically share features between solutions. This can be implemented in several different ways, but in this paper we use the original BBO formulation [7], which is called partial immigration-based BBO in [8]. In this approach, for each feature in each solution, we use the immigration curve to probabilistically decide whether or not to immigrate. If immigration is selected for a given solution feature, then the emigrating island is selected probabilistically (e.g., using $\mu$-based roulette wheel selection). Figure 3 is a conceptual description of one generation of this approach, where we use the notation $y_k(s)$ to denote the $s$th feature of the $k$th population member. Migration and mutation of the entire population take place before any of the solutions are replaced in the population, which requires the use of the temporary population vector $w$ in Figure 3. A qualitative comparison between BBO and other EAs for some common benchmark functions is given in [7].

---

Let $y$ denote the population of solutions
$w \leftarrow y$ (note that $w$ is a temporary population vector)
For each island $w_k$
  For each SIV $s$
    Use $\lambda_k$ to probabilistically decide whether to immigrate to $w_k(s)$.
    If immigrating then
      Use $\mu$ to probabilistically select the emigrating island $y_j$.
      $w_k(s) \leftarrow y_j(s)$
    end if
    Probabilistically decide whether to mutate $w_k(s)$
  next SIV
next island
$y \leftarrow w$

---

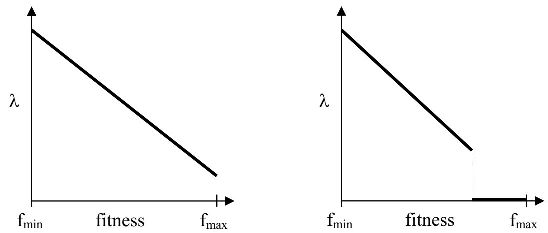
Fig. 3. One generation of the partial immigration-based BBO algorithm. $y$ is the entire population of solutions, $y_k$ is the $k$th solution, and $y_k(s)$ is the $s$th feature of $y_k$.

### *C. Elitism*

As with other population-based algorithms, we often incorporate elitism in order to retain the best solutions in the population from one generation to the next. We use $z$ to denote the number of top individuals in the population that have a zero probability of immigration. If elitism is not used, then $z = 0$ and the immigration probability $\lambda$ is a linear function of fitness and is positive for all fitness values. This is shown in Figure 4(a). If elitism is used, then $z > 0$ and the immigration probability $\lambda$ is zero for the top $z$ individuals in the search space. This is shown in Figure 4(b).

## III. Markov Analysis

In this section we derive the probabilities for a BBO transition from some population distribution at one generation to another distribution at the next generation. First we set up the notation.




(a) Immigration probability with no elitism ($z = 0$). The immigration probability $\lambda$ is nonzero for all fitness values.

(b) Immigration probability with elitism ($z \neq 0$). The immigration probability $\lambda$ is zero for the $z$ most fit individuals in the search space.

Fig. 4. Immigration probability $\lambda$ as a function of fitness.

Suppose that we have a problem with a binary search space. The search space consists of $n$ possible bit strings $x_i$, each $x_i$ containing $q$ bits. The cardinality of the search space is $n = 2^q$. We use $N$ to denote the population size, and we use $v$ to denote the population vector. That is, $v_i$ is the number of $x_i$ individuals in the population. We see that

$$\sum_{i=1}^{n} v_i = N \tag{1}$$

The entire population consists of $N$ bit strings. We use $y_k$ to denote the $k$th individual in the population, and we order them in the same order as $x_i$. That is,

$$\begin{aligned} \text{Population} &= \{y_1, \cdots, y_N\} = \{\underbrace{x_1, x_1, \cdots, x_1}_{v_1 \text{ copies}}, \\ & \underbrace{x_2, x_2, \cdots, x_2}_{v_2 \text{ copies}}, \cdots \underbrace{x_n, x_n, \cdots, x_n}_{v_n \text{ copies}}\} \end{aligned} \tag{2}$$

This can be written more compactly as

$$\begin{aligned} y_k &= x_m \text{ for } k = 1, \cdots, N \\ m &= \min r \text{ such that } \sum_{i=1}^{r} v_i \geq k \end{aligned} \tag{3}$$

If we need to denote the generation number, we use an additional subscript. For example, $y_k(s)_t$ is the value of the $s$th bit of the $k$th individual at generation number $t$.

We use $\lambda_i$ to denote the immigration probability of $x_i$, and $\mu_i$ to denote the emigration probability of $x_i$. We use the notation $x_i(s)$ to denote the $s$th bit of solution $x_i$. We use the notation $\mathcal{J}_i(s)$ to denote the set of population indices $j$ such that the $s$th bit of $x_j$ is equal to the $s$th bit of $x_i$. That is,

$$\mathcal{J}_i(s) = \{j : x_j(s) = x_i(s)\} \tag{4}$$

*A. Migration*

For each SIV (bit), $y_k$ has $v_m$ chances of being selected for immigration, each chance with probability $\lambda_m$. If the $s$th SIV of $y_k$ is not selected for immigration during generation $t$, then

$$y_k(s)_{t+1} = x_m(s) \text{ (no immigration)} \tag{5}$$

That is, $y_k(s)$ does not change from generation $t$ to generation $t+1$. However, if the $s$th SIV of $y_k$ is selected for immigration during generation $t$, then the probability that $y_k(s)_{t+1}$ is equal to $x_i(s)$ is proportional to the combined emigration rates of all individuals whose $s$th feature is equal to $x_i(s)$. This probability can be written as

$$\Pr\left(y_k(s)_{t+1} = x_i(s)\right) = \frac{\sum_{j \in \mathcal{J}_i(s)} v_j \mu_j}{\sum_j v_j \mu_j} \text{ (immigration)} \tag{6}$$

We can combine (5) and (6), along with the fact that the probability of immigration to $y_k(s)$ is equal to $\lambda_m$, to obtain

$$\begin{aligned} \Pr\left(y_k(s)_{t+1} = x_i(s)\right) &= (1 - \lambda_m)\mathbf{1}_0(x_m(s) - x_i(s)) + \\ & \lambda_m \frac{\sum_{j \in \mathcal{J}_i(s)} v_j \mu_j}{\sum_j v_j \mu_j} \end{aligned} \tag{7}$$

where $\mathbf{1}_0(\cdot)$ is the indicator function on the set $\{0\}$. Since there are $q$ SIVs in each island, the probability that immigration results in $y_{k,t+1}$ being equal to $x_i$, given that the population is described by the vector $v$, is denoted as $P_{ki}(v)$ and can be written as

$$\begin{aligned} P_{ki}(v) &= \Pr\left(y_{k,t+1} = x_i\right) \\ &= \prod_{s=1}^{q} \Bigg[(1 - \lambda_m)\mathbf{1}_0(x_m(s) - x_i(s)) + \\ & \lambda_m \frac{\sum_{j \in \mathcal{J}_i(s)} v_j \mu_j}{\sum_j v_j \mu_j}\Bigg] \end{aligned} \tag{8}$$

$P_{ki}(v)$ can be computed for each $k \in [1, N]$ (i.e., each solution in the population) and each $i \in [1, n]$ (i.e., each element of the search space) in order to form the $N \times n$ matrix $P(v)$. The $k$th row of $P(v)$ corresponds to the $k$th iteration of the outer loop in Figure 3. The $i$th column of $P(v)$ corresponds to the probability of obtaining island $x_i$ during an outer loop iteration.

The BBO algorithm entails $N$ trials (i.e., $N$ iterations of the outer loop in Figure 3), where the probability of the $i$th outcome on the $k$th trial is given as $P_{ki}(v)$. We use $u_i$ to denote the total number of times that outcome $i$ occurs after all $N$ trials have been completed, and $u = \left[\begin{array}{ccc} u_1 & \cdots & u_n \end{array}\right]^T$. Then the probability that we start with a population vector $v$ and obtain a population vector $u$ at the next generation is given by the generalized multinomial theorem [9] as follows.

$$\begin{aligned} \Pr(u|v) &= \sum_{Y} \prod_{k=1}^{N} \prod_{i=1}^{n} [P_{ki}(v)]^{J_{ki}} \\ Y &= \Bigg\{J \in \mathbf{R}^{N \times n} : J_{ki} \in \{0, 1\}, \\ & \sum_{i=1}^{n} J_{ki} = 1 \text{ for all } k, \sum_{k=1}^{N} J_{ki} = u_i \text{ for all i} \Bigg\} \end{aligned} \tag{9}$$

Some assumptions were made in the above analysis that we now explicitly note. First, all of the new islands are created before any islands are replaced in the population. This is clear from the use of the temporary population vector $w$ in

Figure 3. Second, an island can emigrate an SIV to itself. This is analogous to crossover in GAs when an individual is allowed to cross with itself. Third, the migration rates $\lambda$ and $\mu$ are independent of the population distribution. That is, absolute fitness values are used to obtain $\lambda$ and $\mu$, as opposed to rank-based fitness.

### B. Mutation

The previous section considered only migration. In this section we add the possibility of mutation. We use $U$ to denote the $n \times n$ mutation matrix, where $U_{ij}$ is the probability that $x_j$ mutates to $x_i$. The probability that the $k$th immigration trial followed by mutation results in $x_i$ is denoted as $P_{ki}^{(2)}(v)$. This can be written as

$$\begin{aligned} P_{ki}^{(2)}(v) &= \sum_{j=1}^{n} U_{ij} P_{kj}(v) \\ P^{(2)}(v) &= P(v) U^T \end{aligned} \tag{10}$$

where the elements of $P(v)$ are given in (8). We can write the probability of transitioning from population vector $v$ to population vector $u$ after one generation of both migration and mutation as

$$\Pr^{(2)}(u|v) = \sum_{Y} \prod_{k=1}^{N} \prod_{i=1}^{n} [P_{ki}^{(2)}(v)]^{J_{ki}} \tag{11}$$

where $Y$ is given in (9). We can use standard Markov tools [15] with this matrix to find the limiting distribution of the BBO population.

The Markov transition matrix $Q$ is obtained by computing (11) for each possible $v$ vector and each possible $u$ vector. $Q$ is therefore a $T \times T$ matrix, where $T$ is the total number of possible population distributions. That is, $T$ is the number of possible $n \times 1$ integer vectors $v$ whose elements sum to $N$ and each of whose elements $v_i \in [0, N]$. It is shown in [10] that

$$T = \binom{n+N-1}{N} = (n+N-1)\text{-choose-}N \tag{12}$$

Other mathematically equivalent expressions for $T$ are given in [11].

## IV. Results

The theory of the preceding section was confirmed with simulations in [12]. In this section we first compare the BBO population distribution with GAs, and then analyze the effect of elitism on BBO performance.

### A. Analytical Comparison with Genetic Algorithms

Consider a genetic algorithm with fitness-proportional (roulette wheel) selection, followed by mutation, followed by single point crossover. We use $v_i$ to represent the $i$th element of the population vector $v$, $f_i$ is the fitness of $x_i$, $G_j^s(v)$ is the probability of obtaining individual $x_j$ by selection alone, $U_{ij}$ is the probability of obtaining $x_i$ from $x_j$ by mutation, $G_i^{sm}(v)$ is the probability of obtaining individual $x_i$ by selection and mutation combined, $r(i,j,k)$ is the probability that $x_i$ and $x_j$ cross to form $x_k$, and $G_k^{smc}(v)$ is the probability of obtaining individual $x_k$ by selection, mutation, and crossover combined. These quantities are obtained in [10], [13], [14], [15] as

$$\begin{aligned} G_j^s(v) &= \frac{v_i f_i}{\sum_j v_j f_j} \\ G_i^{sm}(v) &= \sum_j U_{ij} G_j^s(v) \\ G_k^{smc}(v) &= \sum_j r(i,j,k) G_i^{sm}(v) G_j^{sm}(v) \end{aligned} \tag{13}$$

Equation (13) can be used with the multinomial theorem [16] to obtain the probability that population vector $v$ transitions to $u$ after one generation.

$$\Pr_G(u|v) = N! \prod_i \frac{[G_i^{smc}(v)]^{u_i}}{u_i!} \tag{14}$$

In this section we use Equation (11) to obtain the limiting population distributions of BBO, and (14) to obtain the limiting population distributions of a GA. Due to the exponential increase of matrix sizes with problem size, investigation was limited to three-bit problems ($n = 8$) with a population size of four ($N = 4$), results in 330 possible population vectors as shown by (12). The three fitness functions we investigated are

$$\begin{aligned} F_1 &= \begin{pmatrix} 2 & 4 & 6 & 8 & 6 & 4 & 2 & 1 \end{pmatrix} \\ F_2 &= \begin{pmatrix} 1 & 2 & 3 & 2 & 1 & 2 & 3 & 2 \end{pmatrix} \\ F_3 &= \begin{pmatrix} 3 & 2 & 1 & 4 & 1 & 2 & 3 & 3 \end{pmatrix} \end{aligned} \tag{15}$$

where the fitness values are in binary order of the population members. That is, for problem $F_1$, individual 000 has a fitness of 2, individual 001 has a fitness of 4, ..., and individual 111 has a fitness of 1. For the BBO, we use $\mu_i = F_{1i}/10$ for $F_1$, $\mu_i = 3F_{2i}/10$ for $F_2$, and $\mu_i = 2F_{3i}/10$ for $F_3$. We use $\lambda_i = 1 - \mu_i$ and no elitism for all of the problems.

Tables I-III show comparisons between theoretical GA and theoretical BBO results. The crossover probability used in the GA was 0.9 throughout in this paper. The tables show the probability of obtaining a uniform optimal population, and the probability of obtaining a population which does not have any optimal individuals.

TABLE I
OPTIMIZATION RESULTS FOR $F_1$. THE BETTER-PERFORMING RESULT IS SHOWN IN **red bold font** IN EACH ROW.

| Mutation Rate | Population Vector | Probability GA | Probability BBO |
|---|---|---|---|
| 0.1 | Uniform Optimal | **0.0564** | 0.0509 |
| | No Optima | 0.3990 | **0.3285** |
| 0.01 | Uniform Optimal | 0.5429 | **0.6649** |
| | No Optima | 0.2873 | **0.0943** |
| 0.001 | Uniform Optimal | 0.6989 | **0.9057** |
| | No Optima | 0.2815 | **0.0652** |

Several things are notable about the results of Table I-III. First, as the mutation rate decreases, the probability of a uniform optimal population increases, and the probability of no optima decreases. This is true for both the GA and BBO.

TABLE II
OPTIMIZATION RESULTS FOR $F_2$. THE BETTER-PERFORMING RESULT IS SHOWN IN **red bold font** IN EACH ROW.

| Mutation Rate | Population Vector | Probability GA | Probability BBO |
|---|---|---|---|
| 0.1 | Uniform Optimal | **0.13274** | 0.1156 |
| | No Optima | 0.2427 | **0.1483** |
| 0.01 | Uniform Optimal | 0.69384 | **0.8004** |
| | No Optima | 0.1663 | **0.0137** |
| 0.001 | Uniform Optimal | 0.83198 | **0.9743** |
| | No Optima | 0.1526 | **0.0053** |

TABLE III
OPTIMIZATION RESULTS FOR $F_3$. THE BETTER-PERFORMING RESULT IS SHOWN IN **red bold font** IN EACH ROW.

| Mutation Rate | Population Vector | Probability GA | Probability BBO |
|---|---|---|---|
| 0.1 | Uniform Optimal | **0.0385** | 0.0377 |
| | No Optima | 0.4558 | **0.3752** |
| 0.01 | Uniform Optimal | 0.4667 | **0.6206** |
| | No Optima | 0.3772 | **0.1363** |
| 0.001 | Uniform Optimal | 0.6180 | **0.8771** |
| | No Optima | 0.3640 | **0.0938** |

Second, the GA outperforms BBO only when the mutation rate is high (10% per bit), and even then the probability of a uniform optimal population is only slightly higher in the GA than in BBO. Third, in every other performance comparison in the tables, BBO far outperforms the GA.

### B. Elitism Analysis

Next we explore the effect of elitism on BBO performance. As explained in Section II-C, elitism is implemented by setting the immigration probability of the $z$ most fit individuals in the search space to zero. Note that this differs from the implementation of elitism in practical problems. In practical problems, we do not know the most fit individuals in the search space, we only know the most fit individuals in the current population. However, the analysis in this section can be approximated in a practical problem by setting the immigration probability to zero if the fitness exceeds some problem-dependent threshold.

Another difference between our analysis and most practical implementations is that we use elitism only to prevent immigration, but not to prevent mutation. Our analysis could be extended with future work to prevent mutation, which would require a modification of the mutation matrix in Section III-B.

Figure 5 shows the theoretical probabilities that a BBO population converges to a uniform optimal population for problem $F_1$, for various mutation rates, and for various values of $z$ (number of elite individuals in the search space). Figures 6 and 7 show corresponding results for problem $F_2$ and problem $F_3$. We see that the addition of elitism significantly improves convergence results for $F_1$ and $F_3$, but only slightly for $F_2$. But if we use too many elites in the population, convergence becomes worse.

Figure 8 shows the theoretical probabilities that BBO converges to a population that does not have any optima for problem $F_1$, for various mutation rates, and for various values of the elitism parameter. Figures 9 and 10 show corresponding results for problem $F_2$ and problem $F_3$. We see that the addition of elitism significantly improves the probability that the population has at least one optima. In fact, with a low enough mutation rate, the use of elitism reduces the probability of no optima to essentially zero. But as noted above, if we use too many elites in the population, the performance worsens.

The MATLAB® code that was used to generate the results in this paper is available at http://academic.csuohio.edu/simond/bbo/markov.

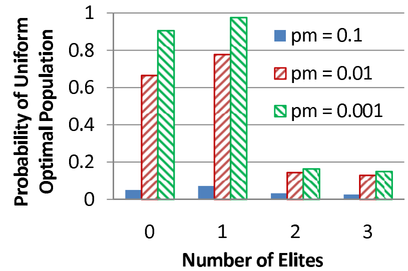



Fig. 5. Probability of BBO achieving a uniform optimal population for problem $F_1$, where $p_m$ is the mutation probability per bit. An elitism parameter of 1 significantly improves convergence to the optimum, but if the elitism parameter is too high then convergence becomes worse.

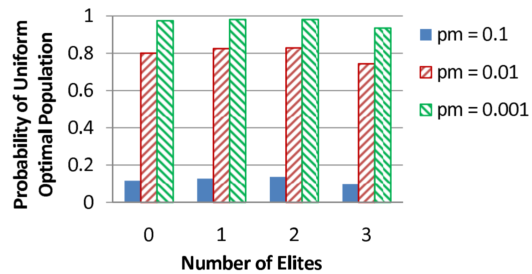



Fig. 6. Probability of BBO achieving a uniform optimal population for problem $F_2$, where $p_m$ is the mutation probability per bit. An elitism parameter of one slightly improves convergence to the optimum.

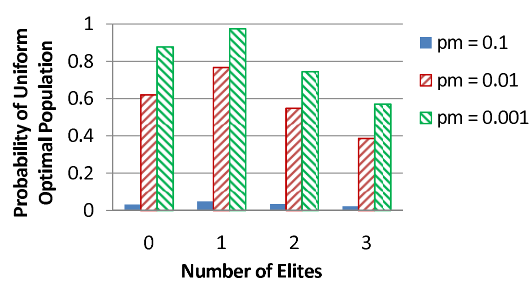



Fig. 7. Probability of BBO achieving a uniform optimal population for problem $F_3$, where $p_m$ is the mutation probability per bit. An elitism parameter of 1 significantly improves convergence to the optimum, but if the elitism parameter is too high then convergence becomes worse.

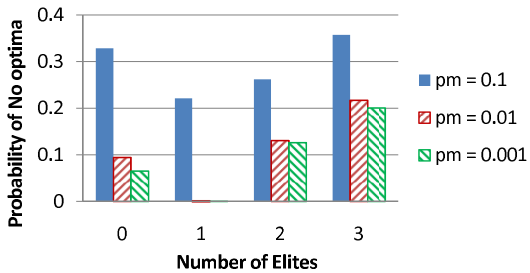



Fig. 8. Probability that BBO does not find any optima for problem $F_1$, where $p_m$ is the mutation probability per bit. An elitism parameter of 1 significantly improves the probability of finding an optimum, but if the elitism parameter is too high then the performance worsens.

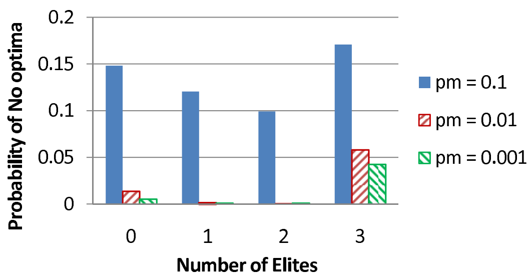



Fig. 9. Probability that BBO does not find any optima for problem $F_2$, where $p_m$ is the mutation probability per bit. An elitism parameter of one slightly improves the probability of finding an optimum.

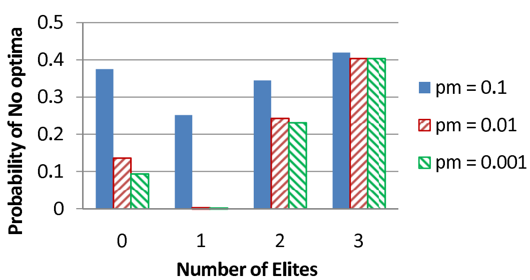



Fig. 10. Probability that BBO does not find any optima for problem $F_3$, where $p_m$ is the mutation probability per bit. An elitism parameter of 1 significantly improves the probability of finding an optimum, but if the elitism parameter is too high then the performance worsens.

## V. Conclusion

We have presented a Markov analysis of biogeography-based optimization. This analysis shows the theoretical probability of each possible population for a BBO problem. Three small representative problems were investigated. The results show that GAs outperform BBO slightly if the mutation rate is large, but BBO far outperforms GAs if the mutation rate is small. Elitism improves performance, but if too many elites are used, then performance worsens.

The analysis in this paper is computationally expensive because the size of the Markov transition matrix increases combinatorially with the problem size. Computational savings can be obtained by grouping Markov states together and computing the probability that the population transitions from one group of populations to another [15], but this is left for further research. Computational savings could also be obtained by not allowing duplicate individuals in the population. This would change the Markov analysis and reduce the size of the transition matrix, but only by a small amount.

Other future work includes extending this analysis to other BBO variations. This paper investigated the original BBO algorithm with linear migration curves, which is called partial immigration-based BBO. An extension of our Markov analysis to other BBO variations would analytically show their advantages or disadvantages.

Finally, the Markov analysis developed here forms a foundation that can be used to develop a dynamic systems analysis of BBO. Dynamic systems analysis of EAs is used to find the proportion of each possible individual in a population as the population size tends to infinity. This is exemplified by the extension of Markov analysis for GAs to dynamic systems analysis [15].


## Acknowledgment

The authors are grateful to Rick Rarick for reviewing this paper and helping to clarify it.


## References


[1] A. Wallace, The Geographical Distribution of Animals, Boston, MA: Adamant Media Corporation, 2006. (*First published in 1876*)

[2] C. Darwin, The Origin of Species, New York, NY: Gramercy, 1995. (*First published in 1859*)

[3] E. Munroe, The geographical distribution of butterflies in the WestIndies, PhD Dissertation, Cornell University, Ithaca, New York, 1948.

[4] E. Munroe, "The size of island faunas," Seventh Pacific Science Congress of the Pacific Science Association, Volume IV: Zoology, Auckland, New Zealand, pp. 52-53, 1953.

[5] R. MacArthur and E. Wilson, "An equilibrium theory of insular zoogeography," Evolution, vol. 17, pp. 373-387, 1963.

[6] R. MacArthur and E. Wilson, The Theory of Biogeography, Princeton, NJ: Princeton University Press, 1967.

[7] D. Simon, "Biogeography-based optimization," IEEE Transactions on Evolutionary Computation, vol. 12, pp. 702-713, December 2008.

[8] D. Simon, "A probabilistic analysis of a simple biogeography-based optimization algorithm," unpublished. Available online at http://academic.csuohio.edu/simond/bbo/simplified.

[9] N. Beaulieu, "On the generalized multinomial distribution, optimal multinomial detectors, and generalized weighted partical decision detectors," IEEE Transactions on Communications, vol. 39, pp. 193-194, February 1991.

[10] A. Nix and M. Vose, "Modeling genetic algorithms with Markov chains," Annals of Mathematics and Artificial Intelligence, vol. 5, pp. 79-88, 1992.

[11] D. Simon, The dimension of an evolutionary algorithm transition matrix, http://academic.csuohio.edu/simond/bbo/markov, May 2009.

[12] D. Simon, M. Ergezer, and D. Du, "Markov models for biogeography-based optimization and genetic algorithms with global uniform recombination," unpublished. Available online at http://academic.csuohio.edu/simond/bbo/markov.

[13] T. Davis and J. Principe, "A simulated annealing like convergence theory for the simple genetic algorithms," in Proceedings of the Fourth International Conference on Genetic Algorithms, R. Belew and L. Booker, Eds. San Francisco, CA: Morgan Kauffman, 1991, pp. 174-181.

[14] T. Davis and J. Principe, "A Markov chain framework for the simple genetic algorithm," Evolutionary Computation, vol. 1, pp. 269-288, 1993.

[15] C. Reeves and J. Rowe, Genetic Algorithms: Principles and Perspectives, Norwell, MA: Kluwer, 2003.

[16] C. Chuan-Chong and K. Khee-Meng, Principles and Techniques in Combinatorics, Hackensack, NJ: World Scientific, 1992.