8-2-2003

Modified Selection Mechanisms Designed to Help Evolution Strategies Cope with Noisy Response Surfaces

Sriphani Raju Gadiraju

Modified Selection Mechanisms Designed to Help Evolution Strategies Cope with Noisy Response Surfaces

By

Sriphani Raju Gadiraju

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Industrial Engineering
in the department of Industrial Engineering

Mississippi State, Mississippi

August 2003

Modified Selection Mechanisms Designed to Help Evolution

Strategies Cope with Noisy Response Surfaces

By

Sriphani Raju Gadiraju

Approved:

Royce O. Bowden, Jr.
Professor of
Industrial Engineering
(Director of Thesis)

Allen G. Greenwood
Associate Professor of
Industrial Engineering
(Committee Member)

Stanley F. Bullington
Professor of Industrial Engineering
(Graduate Coordinator of the
Department of Industrial Engineering
And Committee Member)

Mingzhou Jin
Assistant Professor of
Industrial Engineering
(Committee Member)

Wayne Bennett
Dean of the College of Engineering

Name: Sriphani Raju Gadiraju

Date of Degree: August 2, 2003

Institution: Mississippi State University

Major Field: Industrial Engineering

Major Professor: Dr. Royce O. Bowden, Jr.

Title of Study: Modified Selection Mechanisms Designed to Help Evolution Strategies Cope with Noisy Response Surfaces

Pages in Study: 91

Candidate for Degree of Master of Science in Industrial Engineering

With the rise in the application of evolution strategies for simulation optimization, a better understanding of how these algorithms are affected by the stochastic output produced by simulation models is needed. At very high levels of stochastic variance in the output, evolution strategies in their standard form experience difficulty locating the optimum. The degradation of the performance of evolution strategies in the presence of very high levels of variation can be attributed to the decrease in the proportion of correctly selected solutions as parents from which offspring solutions are generated. The proportion of solutions correctly selected as parents can be increased by conducting additional replications for each solution. However, experimental evaluation suggests that a very high proportion of correctly selected solutions as parents is not required. A proportion of correctly selected solutions of around 0.75 seems sufficient for evolution strategies to perform adequately.

Integrating statistical techniques into the algorithm's selection process does help evolution strategies cope with high levels of noise. There are four categories of techniques: statistical ranking and selection techniques, multiple comparison procedures, clustering techniques, and other techniques. Experimental comparison of indifference zone selection procedure by Dudewicz and Dalal (1975), sequential procedure by Kim and Nelson (2001), Tukey's Procedure, clustering procedure by Calsinki and Corsten (1985), and Scheffe's procedure (1985) under similar conditions suggests that the sequential ranking and selection procedure by Kim and Nelson (2001) helps evolution strategies cope with noise using the smallest number of replications. However, all of the techniques required a rather large number of replications, which suggests that better methods are needed. Experimental results also indicate that a statistical procedure is especially required during the later generations when solutions are spaced closely together in the search space (response surface).

DEDICATION

I would like to dedicate this research to my family and friends.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major professor and director of my thesis, Dr. Royce O. Bowden, Jr., for his guidance, expertise, support and time. I also would like to thank Dr. Bowden for sharing his insights and interesting discussions throughout the course of my studies. Also, I would like to express my sincere appreciation to Dr. Allen G. Greenwood, Dr. Stanley F. Bullington and Dr. Jin for their valuable suggestions as members of my thesis committee. I would also like to thank the Department of Industrial Engineering at Mississippi State University for accommodating me with computer resources to conduct my research.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

This thesis studies the optimization of simulated systems using heuristic evolutionary search algorithms. Simulation optimization is the process of linking simulation with an optimization method to determine the appropriate settings for user-controlled inputs that maximize or minimize the output responses of interest from a simulation model. Optimization algorithms have been developed that are capable of finding optimal or near optimal solutions by evaluating only a fraction of the possible solutions. These techniques may be broadly classified into direct search techniques, gradient based techniques and statistical techniques. Optimization algorithms called evolutionary algorithms (EA) are direct search techniques and have been successfully applied to a variety of optimization problems characterized by high dimensions and complex search spaces. Evolutionary Algorithms are heuristic search and optimization techniques based on the theory of evolution. The major classes of evolutionary search algorithms are genetic algorithms, evolution strategies, and evolution programming. Evolutionary algorithms have been successful in solving difficult optimization problems where other traditional techniques fail. For this reason, some commercial simulation optimization software packages are based on evolutionary algorithms.

Evolutionary algorithms were originally designed for optimization of deterministic problems. Many real world optimization problems contain stochastic variation in their response function, which poses further difficulty in optimization. This is the case of simulation optimization problem. Evolutionary algorithms (EA) are found to be relatively robust in handling variation in the response surface being searched. However, EA's become less effective in locating the optimal solution as the level of variation in the response surface increases.

Evolutionary algorithms are direct search techniques. As such, EA’s need good estimates of the expected value of the simulation model's response surface that composes the objective function used by the EA. The objective function evaluation is used only in the selection mechanism (the identification of better solutions) for most implementations of evolutionary algorithm. Stochastic variation causes the observed objective function value to be distorted and hence affects the selection mechanism’s accuracy. With this in mind, there are three objectives for this research. The first objective is to gain a better understanding of the level of variation in the response surface that an EA called evolution strategies (ES) can tolerate before its performance deteriorates. The evolution strategies algorithm is used in two commercial simulation optimization packages, one by PROMODEL and another by AutoSimulations. The second objective is to identify potential statistical techniques that could be integrated into an ES's selection process that may improve the algorithm's performance on response surfaces characterized by a high level of variation. Performance is measured by the number of the times the simulation model is called by the algorithm and the closeness of the average fitness of the parent solutions to the optimal solution in the final generation. And the third objective is to evaluate the effectiveness of the different statistically based selection techniques within the context of the performance of the ES. Though, the primary focus of this thesis is on problems involving simulation optimization, the results are applicable to any stochastic optimization problem using evolution strategies.

This thesis is organized as follows. A review of the literature on optimization using evolutionary algorithms in the presence of noise is presented in Chapter 2. In Chapter 3, optimization using evolution strategies is discussed and the potential statistical techniques that easily mesh with evolution strategies are presented. The experiments conducted to evaluate the effectiveness of these statistical techniques identified are described in Chapter 4. The results of the experiments are presented in Chapter 5. The final conclusions and recommendations for future research are presented in Chapter 6.

CHAPTER II

EVOLUTIONARY ALGORITHMS FOR OPTIMIZATION IN THE PRESENCE OF NOISE

2.1 Introduction

In Evolutionary Algorithms, the selection mechanism serves a critical role in evolving solutions towards more favorable search spaces on a response surface. The evaluation of the objective function, or the output from a simulation model in this research is used as a fitness measure for EA's. In a deterministic setting, the fitness of a solution can be obtained accurately, whereas in a stochastic setting, we obtain only estimates of the fitness of a solution, which could be very inaccurate. Hence, the selection mechanism is affected by the presence of stochastic variance (noise) (Boesel, 1999).

There is evidence in literature suggesting that evolutionary algorithms are robust in the presence of noise, especially low levels of noise [(Grant, 1998), (Hammel and Back, 1994), (Boesel, 1999), (Hall, 1997)). Biethahn and Nissen (1994) opine that an evolutionary algorithm's use of a population of solutions rather than a single solution to conduct a search makes them robust for optimization in the presence of noise. Since EA's use a population of solutions to conduct their search, it is likely to visit the same solution, or nearby solutions, multiple times. The authors explain this internal ability to resample as the justification for EA's robustness in the presence of noise. Furthermore, it is observed that noise can be helpful in optimization for some fitness functions (Rana et. al., 1996).

Boesel (1999) explains this robustness of EA's for optimization in the presence of noise in a different perspective. Generally, the solutions are assigned selection probabilities based on the fitness evaluation or rank of the solution. The better solutions are assigned higher selection probabilities of being selected. A slight change in the assignment of selection probabilities does not affect the overall performance of the evolutionary algorithm (Boesel, 1999). This small variability in the selection mechanism is a desirable characteristic of the evolutionary algorithms (Boesel, 1999). An erroneous classification of a poor solution as good or a good solution as a poor solution in stochastic environments does not necessarily lead to absolutely wrong search directions (Stagge, 1998). Hence, a close enough assignment of selection probabilities to the true selection probability (noise less case) would suffice in a stochastic environment. This robustness of EA's to slight variation in selection probabilities explains the good performance of an EA in the presence of low levels of noise (Boesel, 1999). Marrison and Stengel (1997) rightly observe that if the variation due to stochastic effects is smaller than the differences between the true fitness of solutions, then the selection mechanism is almost unaffected.

Although no one to our knowledge has quantified the level of variation that can adversely affect the performance of an EA, there surely exists a level of variation that will render these algorithms ineffective. For such cases, the selection mechanism needs to be adjusted for noisy conditions. In the following sections, a review of the literature that discusses the effect of noise on an EA search is presented followed by a review of various methodologies and techniques proposed in the literature that help EA's cope with noisy environments.

2.2 Effects of Noise on the Performance of EA

In deterministic environments, where there is no stochastic variance, we can conclusively rank all competing solutions. However, in stochastic environments, it becomes increasingly difficult to determine the actual ranking of the solutions, based on a single evaluation, with increasing levels of noise (Boesel, 1999). At very high levels of noise, the measured fitness of a solution based on a single evaluation of the objective function may be very inaccurate, and thus can cause the selection mechanism to pick inferior solutions as the parent solutions. Incorrect ranking of the solutions can cause wrong directions of search and thus ultimately render the algorithms to be ineffective (Boesel, 1999).

Hammel and Back (1994) conducted experiments to gain insights about convergence velocity and convergence reliability of ES in the presence of noise. Presence of noise reduces the convergence velocity and deteriorates the quality of the final solution found by the search (Hammel and Back, 1994; Beyer, 2000). Beyer (2000) provides theoretical results, which suggests that increasing noise deteriorates the performance of ES based on the N-dimensional sphere model. Boesel and Nelson (2000) opine that evolutionary algorithms, in their original form designed for deterministic environments, may "deteriorate into an aimless random search in the presence of very high levels of noise".

2.3 Methodologies for Optimization with EA's in the Presence of Noise

Beyer (2000) broadly classified the techniques that help EA’s, especially ES, cope with noise into three categories.

1. Resampling or Replications.

2. Increasing the population size.

3. Novel self-adaptive mutation operation.

Most of the methodologies found in the literature for optimization with EA's in the presence of noise employ one of the first two approaches or a combination of both.

In a noisy environment, multiple replications are required in order to obtain a more accurate estimate of the objective function (or simulation model output). The fitness of a solution is estimated by averaging the fitness values of a solution across different replications. Resampling improves the search procedure by obtaining more accurate estimates of the fitness of a solution. The second technique is implemented by simply increasing the population size, allowing more exploration of the search space and; therefore, evaluating more solutions. The third technique is to design novel self-adaptation schemes that direct evolutionary algorithms towards more favorable search spaces (Beyer, 2000). Studies of the third technique are mostly focused on self-adaptation of mutation step sizes in ES. The idea is to adapt these step sizes in such a manner that they are not fooled by noise and to utilize information from both the superior and inferior solutions to guide the EA. Beyer (2000) opines that more techniques that unify all the above three approaches will be developed in the future.

Sano and Kita (2000) classify different techniques for optimization with genetic algorithms in the presence of noise into two categories. One approach uses resampling, while the other approach uses the history of the search. Approaches that use history of the search are proposed by Tamaki and Arai (1997) and Tanooka *et al.* (1999). In these approaches, fitness of a solution is estimated as a weighted average of the sampled fitness estimate of parent solutions and the sampled fitness of the evaluated solution. Sano and

Kita (2000) propose a genetic algorithm where the fitness of an individual is estimated as a weighted average of the sampled fitness of the solution, and the sampled fitness of all the solutions previously visited by the search algorithm.

2.4 Resampling vs. Increasing Population Size in the Presence of Noise

There is a trade-off between the number of replications at each solution, which translates into the accuracy of evaluation of a solution, and the number of solutions evaluated, which translates into exploration of solution space (Fitzpatrick and Grefenstette, 1988). Different attempts are made to identify the best approach: increase the population size or increase the number of replications per solution. This gives rise to the question "Is it best to increase the population size or increase the number of replications, given a limited number of fitness evaluations (or simulation calls)" (Beyer, 2000). Different researchers attempted to answer the above question on a variety of problems using different methodologies and obtained conflicting answers. Fitzpatrick and Grefenstette (1988) conducted experiments on noisy fitness functions with genetic algorithms. Their results suggest that increasing the population size rather than increasing the number of replications per solution improves the performance of the search algorithm. In experiments conducted by Hammel and Back (1994) with $(1, \lambda)$-ES in the presence of noise, increasing the number of replications per solution resulted in better performance than increasing the population size. The above observation disputes the observation made by Fitzpatrick and Grefenstette (1988).

R.C. Grant (1998) performed some empirical research regarding allocation of available simulation calls. He conducted experiments at four levels of noise on various test functions ranging from single modal to multi modal and low dimensional to high dimensional functions. He considered three different population sizes (7, 28, and 49), seven levels of replications (1,4,7,10,13,16,19) and four levels of available simulation calls (100, 500, 1000, and 2000). Three optimization techniques, genetic algorithms, evolution strategies and scatter search are examined in his research. His results suggest that allowing the algorithm to search longer is favorable against increasing the number of replications when a limited number of simulation calls are available. He recommends increasing the population size rather than increasing the number of replications per solution, when the number of available simulation calls is very limited. Given a fixed number of available simulation calls, how to best allocate the available simulation calls to resampling and population size for optimum performance of the algorithm remains an open question (Beyer, 2000).

2.5 Self-Adaptive Mutation Operators in the Presence of Noise

Arnold and Beyer (2000) opine that increasing the population size would be favorable to increasing the sample size under the precondition that self-adaptive mutative scheme and the μ/λ ratio are suitably modified for (μ,λ)-ES. They attribute the inferior performance of the search algorithm with increased population size compared to increased sample size, as observed by Hammel and Back (1994), to discarding of information from inferior solutions in ES.

Kumar and Fogel (1999) focus on the mutation operator rather than the cross over operator for optimization with EP, both in the presence and absence of noise. They emphasize on fitness distribution analysis, where expected improvement and probability of improvement statistics are estimated for specific mutation operators in a few trials, both in the presence and absence of noise.

Matsumura, Ohkura and Ueda (2001) propose an extended evolutionary programming procedure for optimization in the presence of noise. They call their algorithm as Robust-EP (REP) and compare it with two other standard evolutionary programming algorithms, Fogel's EP and Yao and Liu's EP. The authors propose using Cauchy mutation instead of the traditional Gaussian mutation and new mutation mechanisms for changing strategy parameters. Their experimental results indicate that their proposed algorithm is favorable and robust in the presence of noise comparatively.

2.6 Statistical Procedures for Optimization with EA's

As the output from a simulation model has stochastic variance, it is prudent to employ some statistical technique to differentiate the outputs of different solutions before selecting parents. One way would be to perform as many replications as required by using some statistical technique to conclusively rank all the solutions and assign selection probabilities based on these means. However, such a methodology would require a fairly high number of replications. It is required to find an optimal sample size that expends only enough replications at each solution without sacrificing the objective of the selection mechanism.

Aizawa and Wah (1994) address two objectives for optimization with GA in the presence of noise: duration sizing and sample-allocation problem. In duration sizing, the termination of a generation is determined under the conditions of constant population size and equal assignment of replications to each solution with the assumption of infinite available replications. Sample allocation addresses the issue of allocating replications to solutions when the number of total available replications is constant per generation, with the goal of maximizing the probability of identifying good solutions, where the number of replications allocated to different solutions in a population may vary. Different solutions in a population may be assigned different sample sizes. Assuming that the fitness evaluations are normally distributed, they derived equations for the two objectives. This adaptive procedure of allocating replications performs better than the static procedures in which each solution is assigned a predetermined number of replications. The allocation of replications is based on the idea of assigning more replications to superior and high variance solutions.

Marrison and Stengel (1997) combine genetic algorithms with a statistical procedure for optimization in the presence of noise. They employ tournament selection as the selection mechanism within their algorithm and use within solution fitness variance to determine the number of replications required. Their methodology is based on the idea that if the error due to noise is smaller than the actual differences between fitness of solutions, then the selection method is unaffected. In order to make this error small enough, replications are allocated based on the ratio of the observed fitness variance between the top 25% of the solutions and the average within-solution fitness variance of these solutions.

Stagge (1998) proposed combining a statistical procedure and (μ, λ) selection mechanism with GA for optimization in the presence of noise. The author rightly observes that the number of evaluations per solution need not be equal for all the solutions in the population. Some solutions may be easy to detect as clearly inferior. In such a case, clearly inferior solutions may be eliminated for further consideration as potential parent solutions and thus reduce the total number of evaluations significantly (Stagge, 1998). A one-sided t-test was used to decide the number of evaluations per solution. In this test, it is hypothesized that one solution is superior to another solution. Replications are added to either one of the solutions or both until the hypothesis is rejected. In this manner, the order of the solutions is deduced. Allocating replications to the best solutions and eliminating the clearly inferior solutions from competition significantly reduced the total number of replications.

Tomick, Arnold and Barton (1995) combined single factor one-way analysis of variance (ANOVA) with Nelder-Mead simplex algorithm for simulation optimization. They used ANOVA to assist in determining the number of replications required per solution. In each iteration, the population of solutions is tested for the hypothesis of equality of solutions using single factor ANOVA. If the hypothesis is accepted then the number of replications to be performed in the next iteration is increased by some factor chosen by the user, else the number of replications is decreased by the same factor.

Olafsson(1999) developed algorithms that combine statistical ranking and selection techniques with a optimization method designed for deterministic objective functions. The optimization method, nested partitions (NP), was combined with Rinnott’s two-stage ranking and selection procedure. Rinott's indifference zone procedure is used to determine the number of replications per solution. Rinott’s two-stage ranking and selection procedure is applied in each iteration of the optimization procedure. The author presented theoretical evidence of convergence of the algorithm to the optimum under some assumptions.

An efficient GA in the presence of noise should have a population size that takes into account both the selection pressure of the selection mechanism employed and the amount of noise (Miller, 1997). Miller (1997) derived selection intensity models that would predict the impact of noise on the convergence velocity of GA's for various popular selection mechanisms such as tournament selection, linear ranking selection, (μ, λ) selection and stochastic universal selection. Miller (1997) derived models for GA that determines the optimal sample size and developed techniques to determine the lower bound and upper bound. Additionally Miller (1997) derived population-sizing models and extended these models to quantify the population-sizing requirement at various noise levels under different selection pressures for a given domain.

Boesel (1999) proposed grouping the competing solutions into a small number of groups for optimization with GA's in the presence of noise. The groups are arranged in the order of superiority using a statistical technique. Each member of the group is assigned the group's average selection probability. By assigning the groups average selection probability to each member in the group, the error in the probability of selection due to misranking of solutions is reduced. The grouping of the solutions is obtained by using a clustering procedure given by Calsinki and Corsten (1985).

Baesler and Sepulveda (2000) used Tukey's multiple comparison procedure with a GA for stochastic optimization. Using Tukey's procedure, groups of solutions are formed where solutions within a group are considered to be statistically indifferent. All the solutions within a group are assigned the same selection probability, which is equal to the group's average selection probability.

Hughes (2001) present an algorithm in which the solutions are assigned a probability of selection based on the probability that a solution dominates other solutions, for optimization with GA, in the presence of noise. The probability of a solution being superior to another solution is calculated based on the difference between the two fitness means, assuming that the means are normally distributed, with variance equivalent to the sum of the two variances. To make the calculation of this probability easier he formulated an alternate equation, which closely approximates the normal standard probability equation. Thus, if there is no noise, then it is possible to conclude the rank of each solution explicitly. As the noise increases, the assigned probabilities to each solution get closer to each other, since there is less evidence of the dominance of one solution over another. Increasing the sample size reduces this effect of noise, as we gain more evidence about the superiority of a solution. That is, if noise approaches infinity, all the solutions are assigned the same probability of selection, which is equivalent to 1/k (k is the number of solutions in competition), which is equivalent to random search. They do not provide any method to guide on the allocation of replications to solutions.

Pitchitlamken and Nelson (2001) combined a statistical ranking and selection method such as the sequential selection with memory (SSM) with an optimization algorithm such as hill climbing algorithm for optimization in the presence of noise. They experimentally compared the performance of SSM with three other approaches for optimization with hill climbing at different levels of noise in terms of the number of convergent paths and the average number of evaluations. Their empirical investigation suggests that SSM is superior to other approaches considered.

2.7 Other Procedures for Optimization in the Presence of Noise

Markov, Arnold, Back, Beielstein and Beyer (2001) propose a (1+1)-ES with thresholding operation for optimization in the presence of noise. A parent solution is replaced only if the fitness of the child solution exceeds the parent fitness by a certain amount τ, known as the threshold. The parent solution is reevaluated in every generation. The authors compare experimentally the progress of the proposed algorithm for a non-zero threshold and zero-threshold at various noise levels. The results favor considerably to a non-zero threshold, however the correct choice of this parameter remains an open question. Choosing a very high value for τ could stagnate the search algorithm, hence it is important to make a good choice of τ to obtain positive progress.

Stroud (2001) presented an optimization algorithm based on genetic algorithms in non-stationary and noisy environments. They call their algorithm the Kalman-extended genetic algorithm in which the solutions are resampled based on uncertainty. A population of solutions is generated, which contains a specified proportion of new solutions and the remaining population is filled with reevaluation of solutions from the previous generations. They propose that the solution having the highest uncertainty among the competing solutions, whose estimated means are greater than the population mean minus the population standard deviation, be selected for resampling. The proportion of solutions to be reevaluated and the proportion of solutions to be newly generated are usually pre-specified by the user.

2.8 Summary of Literature Review

The presence of noise has deteriorating effects on the performance of an evolutionary algorithm. Noise affects the selection mechanism in an EA; hence the selection mechanism has to be modified to take into account the noise (Boesel, 1999). Taking multiple observations (replications or samples) at each solution reduces the effect of noise and improves the selection process, however at the expense of increased computational cost. There is a trade off between the selection accuracy and number of replications to be performed. Evolutionary algorithms are robust to small changes in the assignment of selection probability (Boesel, 1999). Taking advantage of this fact, it is required to devise selection procedures that do not deviate much from their deterministic counter parts in the presence of noise. Hence, it is required to perform minimum number of replications that achieve the goal of "stochastic equivalence" to their deterministic counterparts in the presence of noise (Boesel, 1999). In this direction, we identify different statistical ranking and selection procedures and other statistical clustering procedures that could be used in place of the traditional selection technique used in evolution strategies. We confine our research to combining statistical techniques with ES for optimization of stochastic systems as most of the published research in this area is focused on combining statistical techniques with genetic algorithms. Furthermore, ES is used in at least two simulation optimization packages.

CHAPTER III

STATISTICAL TECHNIQUES FOR (μ, λ) SELECTION IN ES

3.1 Introduction

A (μ, λ) selection is traditionally employed as the guidance mechanism within an ES. In (μ, λ) selection, μ denotes the number of parents and λ denotes the number of offspring solutions. The μ parent solutions are selected by identifying the best solutions among the λ offspring solutions. A desirable characteristic of the selection mechanism in an ES is to drive the algorithm into favorable search spaces by exploiting good solutions while maintaining population diversity by exploring different regions of the search space. Diversity of the solutions is required in order to avoid convergence of the algorithm at a local optimum. Exploitation of solutions and exploration of solutions correspond to the convergence velocity and convergence reliability of the algorithm respectively. Another variant of the selection mechanism frequently employed with an ES is the ($\mu + \lambda$) selection, where the best μ solutions among $\mu + \lambda$ solutions are selected as parents. This technique has a higher selective pressure and there is a chance of premature convergence. Hence in order to avoid convergence at a local optimum, the (μ, λ) selection mechanism is recommended by Back, Hoffmeister and Schwefel (1991). Furthermore, the (μ, λ) selection mechanism with μ>1 is recommended for stochastic problems (Arnold and Beyer, 2001). The (μ, λ) selection mechanism is found to provide a good balance at both exploiting and exploring solutions, though exploitation and exploration of the solutions can be varied by changing the values assigned to μ and λ.

In the noise less case, the λ solutions can be ranked conclusively based on deterministic fitness evaluations and the top μ solutions are selected as parents for the next generation upon which recombination and mutation operations are performed to generate the offspring solutions for the next generation. Typically, each of these selected μ solutions is assigned an equal probability of participating in recombination and mutation. Therefore, it is not necessary to rank the solutions from best to worst, as in the case of GA, but only to identify the top μ solutions, irrespective of the within ranking of these μ solutions. Increasing the number of parents, μ, while keeping the number of offspring constant, allows the algorithm to derive information from a large number of solutions, thus, increasing the convergence reliability at the expense of decreasing convergence velocity. On the other hand, retaining only the single best individual as a parent speeds up the convergence at the cost of convergence reliability. In general, increasing the ratio of μ/λ increases the convergence reliability whereas decreasing the ratio of μ/λ increases the convergence velocity.

3.2 Selection and Noise

Typically, $\mu/\lambda \approx 1/7$ is used in experiments concerning optimization with ES in deterministic environments, which provides a balanced exploration and exploitation of the search space. However, it is not clear if this ratio is suitable in stochastic environments and remains an open question. Beyer (2000) has provided some theoretical evidence, which suggests that $\mu/\lambda \approx 1/2$ be used in stochastic environments, based on the sphere model. In noisy environments, increasing the ratio of μ/λ up to a value of 0.5 with increasing noise levels allows progress of the search algorithm towards favorable search spaces (Arnold and Beyer, 2000). This is based on the idea that, by incorporating information from more solutions (increased μ), we can compensate for the lack of accurate estimates of the objective function values (Arnold and Beyer, 2001).

The ratio of μ/λ may be increased by decreasing the offspring population size λ, or by increasing both the parent population size μ and the offspring population size λ, or by increasing μ keeping λ constant. Increasing the offspring population size, λ, excessively is very undesirable because mutation step sizes become very high and hinders the self-adaptation of mutation step sizes (Arnold and Beyer, 2000). Decreasing λ is not recommended as the exploration of the solution space decreases with decreasing population size. Since increasing λ very high or decreasing λ too low is not desirable, the ratio of μ/λ should be increased by increasing μ keeping λ constant. Thus, Arnold and Beyer believe the hindering effects of noise can be partially overcome by using a larger parent population size than would normally be used in the noise less case.

3.3 Statistical techniques and (μ, λ)-ES in the Presence of Noise.

A statistical methodology can be incorporated within the selection mechanism of the ES to identify the μ solutions, which are to serve as parents for the next generation. Since, a μ/λ ratio of 1/7 is found to be robust in deterministic settings, a statistical technique that significantly guarantees the top μ solutions being selected with a pre-specified probability is one possible methodology. There are a variety of statistical techniques, which perform the above stated goal of selecting the top μ solutions approximately with some specified probability. The following sections describe the different potential statistical techniques and the respective goals achieved.

Different ranking and selection procedures exist in statistical literature that achieves the goal of selecting a subset of size μ, which contains the μ best solutions with a pre-specified probability. One example is the technique proposed by Dudewicz and Dalal (1975) for selecting the best set of solutions of fixed size pre-specified by the user, where the within ordering of the subset of solutions is immaterial. A potential issue is that the selection procedure will require a large number of observations or simulation calls.

A less ambitious goal is to select a subset of size μ, which contains the best solution, popularly known as subset selection, with a pre-specified probability. Examples of statistical techniques that achieve the goal of subset selection are procedure by Kim and Nelson (2001), procedure by Gupta (1965) and procedure by Sullivan and Wilson (1989). There are also ranking and selection techniques available that return a subset of solutions of random size where this subset of solutions includes the 'μ' best solutions with a pre-specified probability. Caroll, Gupta and Huang (1975) propose one such technique. The subset, which is of varying size, is dependent on the number of observations obtained per solution.

The goal of (μ, λ) selection mechanism can also be stated as simply dividing the solutions into two groups of solutions where solutions in one group are superior to solutions in the other group. Statistical multiple-comparisons procedures may be used to group the solutions where one group of solutions is statistically different from the other group of solutions. Examples of statistical multiple comparison procedures are Fisher's Least Significant Difference (Fisher, 1935), Duncan's multiple range test (Duncan, 1955), Student-Newman Keuls test (Keuls, 1952), Scheffe's Procedure (Scheffe, 1959), Welsch' s procedure (Welsch, 1977), and Tukey's procedure (Tukey, 1949). Tukey’s multiple-comparison procedure was used by Baesler and Sepulveda (2000) with genetic algorithms for optimization in the presence of noise.

Another approach for grouping solutions is to use statistically based clustering techniques to select the μ best solutions. Many statistically based clustering techniques may not guarantee the superiority of solutions but split the solutions into groups such that solutions in a group may be considered internally homogenous. From this information, one may be willing to infer that the group of solutions with the highest group mean, in the case of a maximization problem, contains the best solutions. Cluster analysis techniques for means separation where solutions are grouped into non-overlapping sets of solutions are given by Bautista, Smith and Steiner (1997), Calsinki and Corsten (1985), Scott and Knott (1974). Clustering technique given by Calsinki and Corsten (1985) was used by Boesel (1999) in the selection mechanism for stochastic optimization problems with GA. Although such an approach does not guarantee that the top μ solutions are identified, it may be accurate enough for the ES to effectively conduct its search.

A very less ambitious goal would be to compare the average of the estimated means of the solutions corresponding to the top μ means to the average of the estimated means of the remaining solutions. If the top μ means is hypothesized as significantly different than the remaining λ-μ means, then the top μ means are selected as parents. Owing to the robustness of the selection mechanism of an ES, such a statistical procedure might be sufficient to correctly direct the search algorithm. Scheffe (1959) has proposed a method for comparing any set of contrasts among means. Though, Scheffe procedure is a multiple comparison procedure, it has been categorized under other technique since it is not used to conduct a multiple comparison procedure.

In summary, the techniques that could be used within the selection mechanism for an ES may be broadly classified into four categories. They are:

1. Ranking and Selection Procedures
2. Multiple Comparison Procedures
3. Cluster Analysis Procedures
4. Other Procedures.

3.4 Techniques for selection of parent solutions

In light of the above discussion, it can be seen that the number of possible statistical techniques that can be applied within the selection mechanism of an ES are very large. Hence, a few techniques that cover the various categories of procedures are selected for further experimental analysis. The procedure given by Dudewicz and Dalal (1975) is explained in a popular simulation textbook by Law and Kelton (1998). The procedure given by Kim and Nelson (2001) is sequential in nature and has very few assumptions compared to other procedures. Tukey's multiple comparison procedure is widely popular among all the multiple comparison procedures and is covered in most of the statistical textbooks. Clustering procedure given by Calsinki and Corsten (1985) was implemented within a genetic algorithm by Boesel (1999). Scheffe procedure to compare contrasts represents a new approach. The following five techniques are chosen for experimental analysis in such a way that there is at least one procedure from each category.

1. Procedure by Dudewicz and Dalal (1975)
2. Procedure by Kim and Nelson (2001)
3. Tukey's Multiple Comparison Procedure (1949).
4. Clustering with studentized range test by Calsinki and Corsten (1985).
5. Scheffe's Procedure (1959).

The above techniques are introduced in the next section and the details of their implementation are given in Chapter 4.

3.4.1. Procedure by Dudewicz and Dalal (1975)

This is a very straightforward procedure that selects the μ best of λ competing solutions with specified confidence level 1-α and indifference zone δ. Indifference zone is the minimum difference worth detecting. The observations from each solution are assumed to be independent and normally distributed. The variance is assumed unknown and the variance of the observations across solutions is allowed to be unequal. This is a two-stage procedure, which uses the first stage sample variance information of each of the solutions to determine the number of additional replications required at each of the solutions. The solutions corresponding to the top μ means are selected as parents. This procedure guarantees that the selected solutions are the top μ solutions with specified confidence level.

3.4.2. Procedure by Kim and Nelson (2001)

Kim and Nelson (2001) have proposed a sequential ranking and selection procedure for selection of the best and subset selection, where observations are obtained incrementally. The procedure assumes that the observations are normally distributed. The procedure allows unequal variances across solutions. In this procedure, the clearly inferior solutions are screened from competition at early stages. Additional observations are obtained for the solutions that remain in competition and are further screened until a subset of the desired parent population size μ is obtained.

3.4.3. Tukey's procedure (1949)

Baesler and Sepulveda (2000) used this procedure with genetic algorithms for proportionate selection. In Tukey's multiple comparison procedure, solutions are grouped based on the range. The observations from each solution are assumed to be independent and normally distributed. The variance is assumed to be unknown and equal across the solutions. Groups of solutions are obtained, which are significantly indifferent within a group. Tukey's procedure could produce overlapping groups of solutions, where a solution may be contained in more than one group. The critical distance measure, which determines the grouping for this procedure is a function of the number of replications. As the number of replications is increased, more groups are obtained. We consider the solutions in the top group to be superior to the solutions in the remaining groups.

3.4.4. Clustering with Studentized Range test by Calsinki and Corsten (1985)

This is a grouping procedure, where solutions are grouped into non-overlapping sets of solutions. The solutions in a group are considered to be homogenous or significantly indifferent. This procedure is used by Boesel et al. (1999) for simulation optimization with genetic algorithms. The observations are assumed to be independent and normally distributed. The variance across solutions is assumed constant. In this procedure, the solutions are ranked based on the sample fitness means and groups are formed based on the smallest mean fitness difference between solutions. As the number of replications is increased, more groups are obtained. The solutions in the top group may be considered to be superior to the solutions in the remaining groups.

3.4.5. Scheffe' Procedure (1959)

Scheffe's procedure is designed to compare any set of contrasts. A contrast is constructed as a linear function of the fitness means. The observations are assumed to be independent and normally distributed. The variance across solutions is assumed to be equal. The average of the mean fitness of the first μ solutions in rank order is compared against the average of the mean fitness of the remaining λ - μ solutions. A one-tail test is conducted on the hypothesis to determine if the average of the mean fitness of the μ solutions is significantly different than the average of the mean fitness of the λ - μ solutions. This procedure does not group solutions or identify the top μ solutions but merely provides a statistical test to evaluate the stated hypothesis.

CHAPTER IV

METHODOLOGY

4.1 Introduction

In this chapter, the experiments conducted are described. Also, the performance measures for these experiments are defined. All the experiments are conducted within the framework of an (μ, λ)-ES. The offspring population size remains constant throughout the experiments and is chosen to be 28. The parent population size used is 4, whereas for some implementations of the selection mechanism of ES, the parent population size varies in the range of λ/7 and λ/2, which is in the range of 4 to 14. The offspring solutions are generated by discrete recombination of the parent solution's decision variables and intermediate recombination of their strategy parameters. The offspring solutions are then subject to mutation using Schwefel's mutation method (Back, 1996).

Experiments are performed to gain a better understanding of the level of variation in the response surface that an evolution strategy can tolerate before its performance deteriorates. Further experiments are conducted to investigate the effectiveness of modified selection mechanisms identified in Chapter 3 in the presence of high levels of noise. The results of these experiments are presented in Chapter 5.

4.2 Test Functions

Conducting experiments using actual simulation models for testing would require an enormous amount of computational effort. For simplicity, two test-functions are chosen to represent the output from a simulation model. The test-functions are two-dimensional meaning that the test functions have two decision variables. The decision variables are continuous and range from 0 to 10 for both test functions. The algorithm is required to identify a solution that minimizes the objective functions. The optimum fitness value for both test function-1 and test function-2 is 1.

Test function-1 is a unimodal function and is defined as below.

$$f(x,y) = (x-5.0)^2 + (y-5.0)^2 + 1.0$$

Test function-1 has the optimal solution located at $x = 5.0$ and $y = 5.0$ and has no other local optimum. Figure 4.1 shows the response surface plot of test function-1 and Figure 4.2. shows the contour plot of test function-1.

Test function-2 is a tetra modal test function and is defined as below.

$$f(x,y) = cos(\pi.x / 2.5) + cos(\pi.y / 2.5) - 0.964001 * EXP(-(x - 2.5)^2 - (y - 2.5)^2) + 3.964$$

Test function-2 has the global optimal solution located at x=2.5 and y=2.5. Figure 4.3 shows the response surface plot of test function-2 and Figure 4.4 shows the contour plot of test function-2. It has three other attractive local optimal solutions, which correspond to the valleys in Figure 4.3. The global optimal fitness value of the objective function is 1.0 and the local optimal fitness values are 1.96.

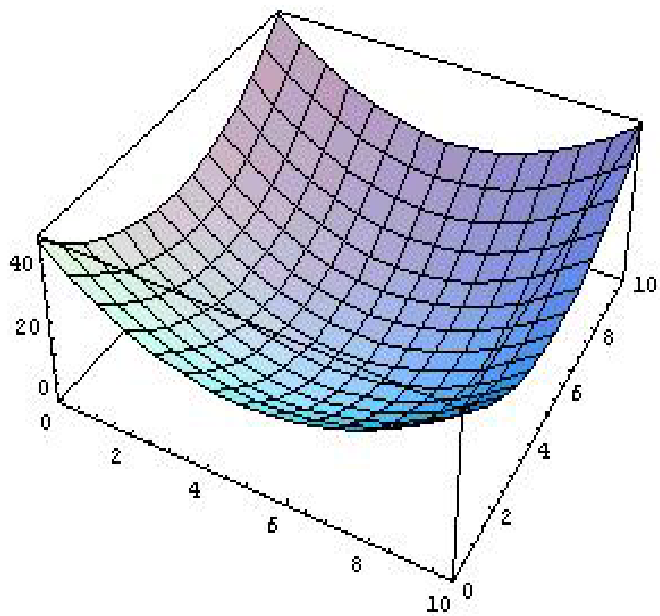

Figure 4.1 Response Surface for Test function-1

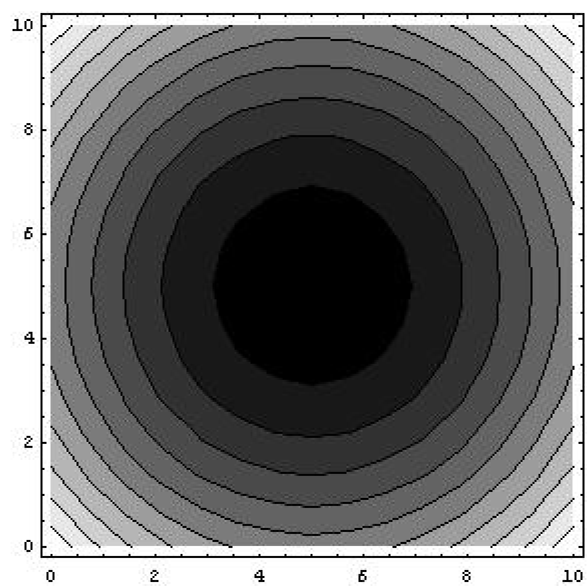

Figure 4.2 Contour Plot for Test function-1

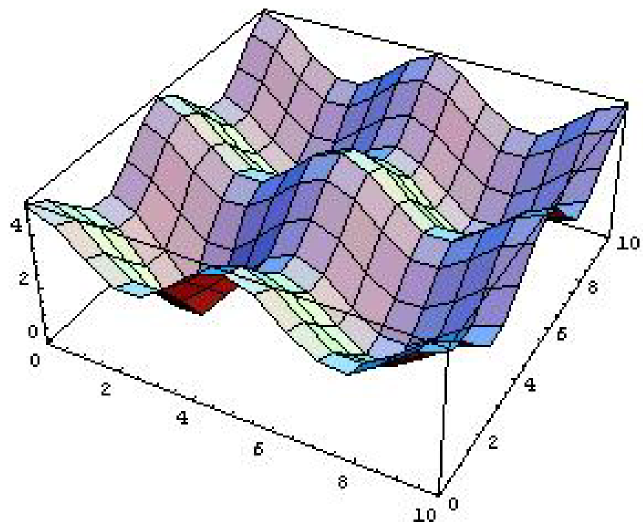

Figure 4.3 Response Surface for Test function-2

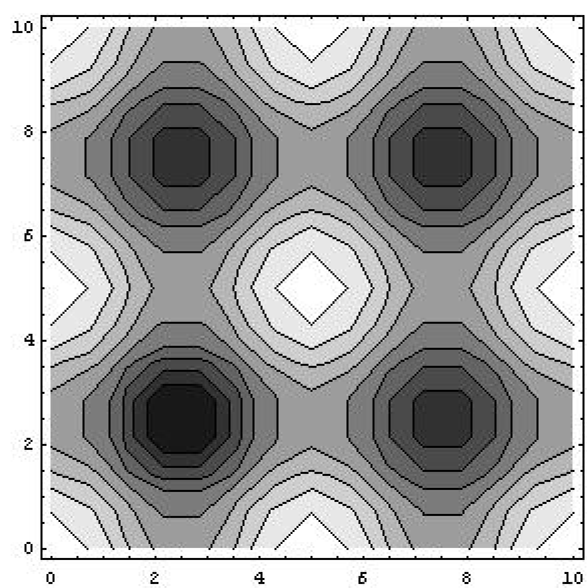

Figure 4.4 Contour Plot for Test function-2

4.3. Initial Number of Replications

The initial number of replications needs to be chosen in such a way that it is neither too small nor too large. If the initial number of replications is chosen to be very large, there exists the danger of spending excessive amount of time evaluating solutions wastefully. Choosing a very small initial number of replications may result in obtaining very unrealistic estimates of the variability. Law and Kelton (1998) recommend conducting at least three to five replications per solution and moreover most of the statistical techniques recommend conducting at least 5 replications per solution. In our experiments the initial number of replications is chosen to be 5 irrespective of the noise level.

4.4. Noise Levels

The noise levels added to the test functions to simulate a stochastic response surface are chosen in such a way that they range from low to very high. Test function-2 is used to derive the noise levels. Test function-2 is multi-modal, which implies that it contains local optimum in addition to the global optimum. Usually, multi-modal functions pose further difficulty for the search algorithm because of the presence of attractive local optimum. The local optimum can appear better than the global optimum at increasingly higher levels of noise. The amount of noise that can make a local optimum appear equivalent or better than the global optimum is identified by using a simple t-test. Let D denote the difference between the fitness of the global optimum solution and the fitness of the local optimum solution as shown in Figure 4.5, which is equal to 0.96 for test function-2. Let *f(x)* and *f(x₁)* denote the global optimum fitness and local optimum fitness respectively. Let *n* denote the initial number of replications, which is equal to 5.

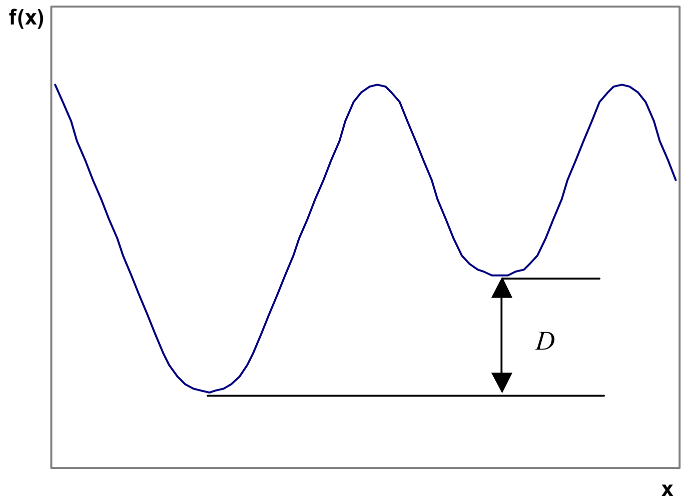

Figure 4.5 Calculation of Noise

A two-sample pooled t-test is conducted to calculate the variance, which makes the local optimum and the global optimum appear not significantly different. Let *s* denote variance. Let '*t*' denote the calculated test statistic, which is compared to $t_{\alpha/2,2n-2}$, where $t_{\alpha/2,2n-2}$ is the upper critical value of the studentized t-distribution with *2n-2* degrees of freedom at a significance level of α. We fail to reject the hypothesis that *f(x)* and $f(x_1)$ are equal if the calculated t-statistic is less than $t_{\alpha/2,2n-2}$.

Then, if $\frac{D}{s\sqrt{\frac{2}{n}}} \prec t_{\alpha/2,2n-2}$, then *f(x)* and $f(x_1)$ are not significantly different.

$$\Rightarrow \quad s > \frac{D}{t_{\alpha/2,2n-2}\sqrt{2/n}}$$

$$\Rightarrow \quad s > \frac{0.96}{1.860\sqrt{2/5}}$$

$$\Rightarrow \quad s > 0.81807$$

In other words, the local optimum $f(x_1)$ and the global optimum $f(x)$ appear to be not significantly different when the standard deviation is greater than 0.81807 at a significance level of 0.10 and 8 degrees of freedom. Hence, if the amount of noise added is greater than 0.81807, there is a significant chance that the local optimum can appear superior to the global optimum. Let the noise be represented by σ_{noise}.

Hence a noise of $\sigma_{noise}/2$ = 0.409 may be considered low and similarly noise levels of σ_{noise} = 0.818, σ_{noise}*1.5 = 1.227 and σ_{noise}*2= 1.636, may be considered to be moderate, high and very high respectively for test function-2 with 5 initial number of replications. Noisy fitness function values are obtained by adding normally distributed random variate with a mean of zero and standard deviations equal to 0.409, 0.818, 1.227 and 1.636 to the objective function values. The objective function can be defined as $O(x,y) = f(x,y) + N(0, \sigma_{noise})$, where $N(0, \sigma_{noise})$ is a normally distributed random variate with a mean of zero and standard deviation equal to σ_{noise} and $f(x,y)$ is the objective function of the test function. Experiments are conducted at these four levels of noise. The noise corresponding to 0.409, 0.818, 1.227 and 1.636 are represented as $0.5\sigma_{noise}$, $1\sigma_{noise}$, $1.5\sigma_{noise}$ and $2\sigma_{noise}$ respectively through out the remaining part of this thesis.

4.5 Common Experimental Conditions

In order to conduct a fair and effective comparison, the ES is run under similar conditions. ES starts with the same initial population in all the experiments. Similarly, recombination and mutation functions remain the same in every experiment. This is achieved by assigning separate random number seeds to each specific purpose of the ES, which include initial population, recombination, and mutation. Synchronization is achieved by reinitializing the random number seed values for each experiment. This would ensure that the ES starts with the same initial population in every experiment and identical random numbers are used for each specific purpose of the algorithm across replications. Each experimental condition is repeated 25 times and the results are obtained by averaging over 25 replications. Since, the purpose of this thesis is to study the effect of noise on evolution strategies, noise is generated by different streams of random numbers for each replication, while keeping the remaining elements of the ES the same. In other words, common random numbers are used for the ES but not for the noise. Independent observations are obtained by allocating different streams of random numbers for each replication. This allows a fairer comparison since any differences in the performance measures are only due to the various selection mechanisms employed and not due to changes in the experimental conditions.

4.6. Number of generations

The number of generations is constant and is chosen to be equal to 10. It is observed that the standard ES converges completely in 10 generations for the two test functions in the absence of noise. Hence, the number of generations is limited to 10 for all the experiments to see how it is affected as the noise is increased.

4.7 Effect of Noise

The presence of variation in the output of a simulation model affects the selection mechanism of an evolutionary algorithm by potentially causing the solutions to be incorrectly ranked. In other words, the proportion of correctly selected solutions as parents for the next generation decreases with increasing levels of noise. We define proportion of correct selection as the ratio of the number of solutions correctly classified as parents to the number of parent solutions. The standard ES with 5 initial number of replications per solution is evaluated under the four levels of noise $0.5\sigma_{noise}$, $1\sigma_{noise}$, $1.5\sigma_{noise}$ and $2\sigma_{noise}$. The proportion of correct selection for each noise level is captured to gain more insights on the effect of noise on the algorithm. Based on these experiments the level of noise that deteriorates the performance of the standard evolution strategies algorithm significantly is identified and is used as the noise level for remaining experiments. For sake of discussion, let us denote the noise level that deteriorates the performance of the algorithm to be σ_{high}.

4.8 Controlled Proportion of correct selection

An open research question is what proportion of correct selection is required by the ES to effectively conduct a search for the optimum. Therefore, we attempt to quantify the proportion of correct selection desired and gain some insights on the proportion of correct selection and its effect on the performance of evolution strategies. The performance of evolution strategies is studied experimentally at various levels of proportion of correct selection in the presence of noise equivalent to σ_{high}. Note that σ_{high} is based on previous experiments, where σ_{high} is the amount of noise that deteriorates the performance of the standard ES significantly. The proportion of correct selection is controlled by estimating the fitness of the solutions based on 2 initial replications, selecting μ parents and computing the proportion of correct selection, and then adding replications until the desired proportion of correct selection is achieved in each generation. The proportions of correct selection experimented with are 0.25,0.5,0.75 and 1.0. A proportion of correct selection of 0.25 corresponds to a correct selection of one parent among the four solutions selected as parents. Similarly, a proportion of correct selection equivalent to 0.5 corresponds to a correct selection of two parents among the four parent solutions. Correct selection of all the parents would be equivalent to a proportion of correct selection of 1. The minimum desired proportion of correct selection is estimated, where the performance of evolution strategies is not affected.

4.9 Modified Selection Mechanisms

The statistical techniques identified in Chapter 3 are incorporated within the standard (μ,λ) selection methodology to derive the modified selection methodologies. The original techniques are modified to fit within the evolution strategies selection mechanism. Experiments are conducted on five techniques, which are described in detail below.

4.9.1 Indifference Zone procedure by Dudewicz and Dalal (1975)

Dudewicz and Dalal (1975) have designed a two-stage indifference zone procedure for the selection problem. Indifference zone may be defined as the minimum fitness difference worth detecting. The procedure is later modified and extended by Koenig and Law (1985). Let m = number of systems, s = subset size, and p= number of best systems to be selected or identified, then, they give a generalized procedure that would allow selection of a subset of size s, which contains the p best systems at the specified confidence level. Specifically, the procedure addresses three goals, which are selection of best system, selection of a subset that contains the best system and the selection of the best subset of systems. Note that $p \leq s \leq m$. In this procedure, the user specifies the initial number of replications to be performed on each system. Using the information about the competing systems from the initial set of observations, the number of additional replications required to attain the stated goal is determined. Thus, if $s=p=1$, then the goal is selection of the best solution. If $p=1$ and $s>1$, then the goal is selection of a subset of size s that contains the best system. This corresponds to the goal of subset selection. If $s=p$, then the goal is to select a subset that contains the superior p solutions.

The procedure assumes that observations are independent, identical and normally distributed. One advantage of the procedure is that the procedure does not assume equality of variance across systems. This is a two-stage procedure. The subset size to be selected is pre-specified. The procedure for the three goals is similar except that the statistical constant 'h' (critical value) changes as applicable. So, h values are lower, which requires fewer replications, for the goal of selecting a subset that contains the best than for the goal of selecting the best subset, which requires more replications, as the latter goal is superior to the former goal. The procedure is outlined below.

1. Specify indifference zone δ, number of initial replications n, confidence level 1- α. Let the number of systems in competition be m. Then, let $X_{i1}, X_{i2}, \ldots\ldots X_{i,n}$ denote independent and identically distributed random output from system i.
2. Calculate the first stage m sample means, $\overline{X}_i^{(1)} = \sum_{j=1}^{n} \frac{X_{ij}}{n}$, for i =1,2… m.

Now, calculate the sample variance $S_i = \frac{\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2}{n-1}$. For i =1,2…m.

3. Compute the total sample size required for each system i as

$$N_i = \max\left\{n+1, \left\lceil \frac{h^2 S_i^{\,2}}{\delta^2} \right\rceil\right\}$$, where h value is obtained from tables.

4. Conduct N_i-n additional replications of each system and obtain the second stage sample means.

$$\overline{X}_i^{\,(2)} = \frac{\sum_{j=n+1}^{N_i} X_{ij}}{N_i - n}$$

5. Define the weights for i=1 to m.

$$W_{i1} = \frac{n}{N_i}\left[1+\sqrt{1-\frac{N_i}{n}\left(1-\frac{(N_i-n)\delta^2}{h_2^{\,2} S_i^{\,2}}\right)}\right]$$

$$W_{i2} = 1\text{-}W_{i1}$$

6. Compute the weighted sample means *for i =1…m*

$X_i = W_{i1}\overline{X}_i^{\,(1)} + W_{i2}\overline{X}_i^{\,(2)}$ and arrange the weighted means in ascending order.

The above selection procedure is accommodated within the selection mechanism of ES by choosing $m = \lambda$, $s = p = \mu$ and n = 5, where μ= 4 and λ=28. The solutions that correspond to the top μ means are selected as parents. The indifference zone, δ, is chosen to be a constant, which was initially, and somewhat arbitrarily, set equal to 0.1.

4.9.2 Sequential Procedure by Kim and Nelson (2001)

Kim and Nelson (2001) have proposed a sequential indifference zone selection method for selecting the best or a subset containing the best. Sequential sampling methods eliminate the clearly inferior systems at an early stage and henceforth reduce the number of observations (replications) required (Kim and Nelson, 2001). The observations are assumed to be independent, identical and normally distributed with unknown variance. The procedure is described below.

1. Specify indifference zone δ, initial number of replication n, confidence level 1- α. Let m represent the number of systems in competition. Let s represent the required subset size. Then, let $X_{i1}, X_{i2} \ldots X_{i,n}$ denote independent and identically distributed output from system i. Let I denote the systems still in competition, so $I=\{1,2,3,\ldots m\}$ initially.

2. Calculate the m sample means, $\overline{X}_i = \sum_{j=1}^{n} \frac{X_{ij}}{n}$, for i =1,2…m and

   For all $i \neq j$, calculate the sample variance of the difference between systems i and j.

   So, $S_{ij}^2 = \frac{1}{n-1} \sum_{l=1}^{n} (X_{il} - X_{jl} - [\overline{X}_i - \overline{X}_j])^2$

3. Calculate $N_{ij} = \left\lfloor \frac{h^2 S_{ij}^2}{\delta^2} \right\rfloor$, where $h^2 = 2c\eta \times (n-1)$.

   Kim and Nelson recommend c = 1, where $\eta = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n-1)} - 1\right]$

   Let $N_i = \max(N_{ij})$ for $i \neq j$. Then N_i+1 is the maximum number of observations from system i. If n>max N_i for $i=1$ *to* m, then stop the procedure and select the system with the largest $\overline{X}_i$ as the best, else go to step 4.

4. Set I^{old} =I, then

   $I = \{i : i \in I^{old}$ and $\overline{X}_i \geq \overline{X}_j - W_{ij}, \forall j \in I^{old}, j \neq i\}$ where

$$W_{ij} = \max\left\{0, \frac{\delta}{2cn}\left(\frac{h^2 S_{ij}^2}{\delta^2} - n\right)\right\}$$

5. If I <=s, then stop the procedure and select the systems whose index are in I as the subset of systems containing the best, else take one additional observation $X_{i,n+1}$ from each system $i \in I$, set n=n+1. If n = $\max_i N_i + 1$, then stop the procedure and select the system which has the largest mean as the best else go to step 4.

The above selection procedure is accommodated within the selection mechanism of ES by choosing $m = \lambda$, $s = \mu$ and n = 5. The solutions returned by this procedure are selected as parents for the next generation. If the subset of solutions returned by the selection procedure is less than the desired parent subset size, then the remaining parent solutions are selected by picking the offspring solutions with the lowest sample mean fitness.

4.9.3 Tukeys procedure (1949)

Tukey's multiple comparison procedure is very widely popular and is covered in most of the basic statistics textbooks. Tukey's procedure separates the solutions into groups of solutions, where solutions in a group are internally homogenous. The observations are assumed to be normally distributed. The observations are also assumed to be independent within and across systems. A common variance is assumed for all the systems in competition. Tukey's multiple comparison procedure involves the following steps with a specified confidence level 1-α.

1. Calculate the m sample fitness means, $\overline{X}_i = \sum_{j=1}^{n} \frac{X_{ij}}{n}$, for i =1,2…m.

Compute the common pooled sample variance estimate $S^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2}{m(n-1)}$.

2. Arrange the means in ascending order and let $\bar{X}_{[1]}, \bar{X}_{[2]}, \bar{X}_{[3]}......, \bar{X}_{[m]}$ denote the ordered sample means.
3. Obtain the upper critical value of the studentized range *q* (α, m, v), where *v* is the number of degrees of freedom for the sample variance and is equal to *m(n-1)*. *m* is the number of competing solutions. Calculate '*w*', the critical distance measure for the Tukey's procedure, where

   $$w = q\ (\alpha,\ m,\ v)\sqrt{S^2/n}\,.$$

4. If the means of the solution output differ by less than '*w* ', they are grouped together. That is if $\bar{X}_{[2]} - \bar{X}_{[1]} < w$ then solutions corresponding to $\bar{X}_{[2]}, \bar{X}_{[1]}$ are grouped together. Conduct all pair wise comparisons as above and declare the solutions as significantly different where the hypothesis of equality of the means is rejected.
5. After all pair wise comparisons are conducted, we obtain groups of solutions where, solutions in a group may be considered significantly indifferent.

Tukey's procedure could produce overlapping groups of solutions, where a solution might be included in more than one group. Tukey’s procedure is incorporated within the selection mechanism of evolution strategies by modifying the above procedure, where $m = \lambda$, and $n = 5$. Overlapping groups of solutions are combined together for adapting the procedure into the selection mechanism of ES. Moreover, the parent population size μ is not constant and is allowed to vary in the range of $\lambda/7$ to $\lambda/2$. Hence, if the number of solutions in the combined top groups fall in the range of

$\lambda/7$ to $\lambda/2$, the solutions are selected as parents. If the number of solutions in the combined groups of solutions obtained do not fall in the range of $\lambda/7$ to $\lambda/2$, then additional observations are obtained and the above procedure is repeated until the criteria is satisfied.

4.9.4 Cluster Analysis Procedure by Calsinki and Corsten (1985)

Cluster Analysis procedures separate the solutions into distinct groups of solutions. Calsinki and Corsten (1985) propose two clustering methods, where one is based on the studentized range and the other is based on the *F* test. Both these methods separate the solutions into non-overlapping sets of solutions. The observations are assumed to be independent and normally distributed. The variance is assumed to be equal across competing alternative solutions. The procedure based on the studentized range is described below.

1. Let *m* denote the number of systems in contention and let *n* denote the number of replications performed. Calculate the *m* sample fitness means $\overline{X}_i = \sum_{j=1}^{n} \frac{X_{ij}}{n}$, where $i=1$ to m.

   Compute the common pooled sample variance, $S^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2 / m(n-1)$.

2. Arrange the sample fitness means in ascending order and let $\overline{X}_{[1]}, \overline{X}_{[2]}, \overline{X}_{[3]}......, \overline{X}_{[m]}$ denote the ordered sample fitness means.
3. Obtain the upper critical value of the studentized range $q\ (\alpha, m, v)$ from tables, where v is the number of degrees of freedom for the sample variance and is equal to $m(n-1)$. *m*

is the number of competing solutions. Calculate R_α, the critical distance measure for this procedure, where

$$R_\alpha = q\ (\alpha,\ m,\ v)\sqrt{S^2/n}\ .$$

4. The two fitness means that result in the smallest ranges are combined together as a group or cluster and compared with R_α. If the smallest range is less than R_α, then the procedure is continued and the two means are grouped. The number of solutions to be grouped is reduced by 1 and the average of the means clustered together represents the output of this group. If the smallest range exceeds R_α, the procedure is stopped.

5. In each next step, the smallest range is compared with R_α and the means are combined if the range is smaller than R_α. If the smallest range exceeds R_α, then the procedure is stopped and the groups obtained in the previous step would be the final grouping.

The clustering procedure is incorporated within the selection mechanism of evolution strategies by modifying the above procedure, where $m = \lambda$, and n = 5. Groups of solutions are obtained, where there is no overlap of solutions between groups. If the number of solutions in the combined top groups fall in the range of $\lambda/7$ to $\lambda/2$, the solutions are selected as parents. If the groups of solutions obtained do not meet this criterion for grouping, then additional observations (replications) are obtained and the above procedure is repeated until the termination criteria is satisfied.

4.9.5 Scheffe Procedure (1959)

Scheffe' Procedure (1959) allows the analysis of all possible comparisons of competing solutions. The method is designed to compare any set of contrasts. The observations are assumed to be independent and normally distributed. The variance is assumed to be equal across competing alternative solutions. A contrast is constructed as a linear function of all or any of the fitness values of solutions. The steps involved in this procedure are described below.

1. Define the contrast as a linear function of the sample fitness. Calculate the m sample fitness means, $\overline{X}_i = \sum_{j=1}^{n} \frac{X_{ij}}{n}$, for i =1,2…m. Compute the common pooled sample variance estimate $S^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2}{m(n-1)}$.

2. Arrange the means in ascending order and let $\overline{X}_{[1]}, \overline{X}_{[2]}, \overline{X}_{[3]}......, \overline{X}_{[m]}$ denote the ordered sample means.

3. Compute the estimated value of the contrast, $\hat{L} = \sum_{i=1}^{m}\left(a_i \overline{X}_i\right)$, where a_i is the coefficient for fitness mean *i*. For linear contrasts, $\sum_{i=1}^{m} a_i = 0$. The a_i values are chosen to be equal to $1/\mu$, for i =1 to μ top ranked solutions, and a_i = -(1/($\lambda - \mu$)), for $i = \mu + 1$ to λ remaining solutions.

   Null hypothesis may be stated as $H_0 : \hat{L} = 0$.

   Alternate hypothesis $H_a : \hat{L} \neq 0$.

4. Compute the critical value $C = \sqrt{(m-1)F_\alpha \sum_{i=1}^{m} a_i^2 \left(\frac{S^2}{n}\right)}$, where F_α is the α level critical value of the F distribution with (m-1) and $m(n\text{-}1)$ degrees of freedom.

5. Compare the value of $\hat{L}$ with critical value C. If $\hat{L}$>C, then we reject the null hypothesis at the stated significance level.

6. If $\hat{L} < S$, then we fail to reject the null hypothesis.

Scheffe's procedure is implemented within the ES selection method by placing the top μ solutions in the hypothesized "best group" and the remaining $\lambda - \mu$ solutions in the hypothesized "second best group". The procedure is repeated with an additional replication for each solution until the mean fitness of the two groups is determined to be significantly different.

Table 4.1 shows the summary of the assumptions for each of the above techniques. An "YES" indicates that the procedure assumes the respective condition.

Table 4.1. Summary of assumptions for statistical techniques

Statistical Selection Techniques	**Independent**	**Normally Distributed**	**Equal variance across solutions**
Dudewicz and Dalal Selection Procedure	YES	YES	NO
Kim and Nelson's Sequential Procedure	YES	YES	NO
Tukey's Multiple Comparison Procedure	YES	YES	YES
Calsinki and Corsten's Clustering Procedure	YES	YES	YES
Scheffe's Procedure	YES	YES	YES

4.10 Criteria for Evaluation of Performance

Two performance measures are used to evaluate the effectiveness of these modified selection mechanisms in conjunction with evolution strategies under various experimental conditions. The two performance measures are the average fitness of the parent solutions and the total number of simulation calls. The average fitness of the parent solutions is computed by averaging the actual fitness values (objective function values without noise) of the solutions that are identified as parents. In order to make a fair comparison, the actual values are used rather than the estimated fitness values. The average fitness of the parent solutions is a measure of the quality of the solutions identified (smaller is better) and the total number of simulation calls is a measure of the computational effort required for the algorithm (smaller is better). It is required to identify good quality solutions with minimum computational effort in the presence of noise. In addition to average fitness of the parents and the total number of simulation calls, the proportion of correct selection is reported.

4.11 Probability of correct selection (1-α)

Three α significance levels will be examined for the modified selection methodologies. The three α levels are 0.1,0.2, and 0.4. It is hypothesized that small α values are not necessary owing to the fact that an ES can tolerate some imperfect selection of the best solutions as parents.

4.12 Indifference Zone

As the search algorithm progresses towards the optimum region, the solutions in the population get closer and closer together. In other words, the solutions are spaced farther apart in the earlier generations and are closely spaced in the latter generations. In order to exploit this property to decrease the number of simulation calls, the indifference zone could be set high in the earlier generations and then decreased in the latter generations, as solutions get closer together.

One measure of the distance between solutions is the average fitness distance between all the solutions in the population. In other words, the average fitness distance is computed as the average of the all pair wise fitness distance. Since, the actual fitness is not known, the estimated fitness based on the initial replications is used to compute the average distance. Half the average distance is used as the indifference zone in each generation. Indifference zone selection methodologies, DD and KN are modified to incorporate this methodology, where indifference is dynamic and is equivalent to half the average fitness distance.

Tukey's procedure, Clustering procedure and Scheffe procedure are not based on an indifference zone methodology. Hence, as solutions get closer to the optimum, the techniques might prescribe excessively high number of simulation calls wastefully. In order to avoid an excessive number of simulation calls, the number of simulation calls per solutions in each generation is restricted to twice the number of simulation calls in the previous generation. These methodologies are tested with the hope that the search algorithm finds the optimum solution with a much smaller number of simulation calls.

4.13 Comparison of modified selection mechanisms with the standard.

The best performing modified selection methodologies experimented are compared to the standard selection methodology used in ES. The standard ES is compared with these modified selection mechanisms by allocating an equivalent number of simulation calls expended by the modified selection methodology and comparing the average fitness of the parent solutions. Let, T denote the number of simulation calls expended by the algorithm using modified selection mechanism in 10 generations. The population size of the standard ES is increased by a factor in such a way that the number of simulation calls available is equivalent to T. The population size is computed by dividing T simulation calls by the number of generations times the initial number of replications. The modified standard ES used for comparison is denoted as *SD-C*.

CHAPTER V

RESULTS

5.1. Overview

In this chapter, the results of the experiments described in Chapter 4 are presented. Plots of the results are included where appropriate. The primary performance measures of interest reported are the average fitness (actual fitness without noise) of the parent solutions and the total number of simulation calls (or function calls). In addition the proportion of correct selection is reported, where appropriate. The averaged results of 25 macro replications for each experimental configuration are presented. For convenience, the techniques, selection of the *s* best by Dudewicz and Dalal, subset selection by Kim and Nelson, Tukey's multiple comparison procedure, clustering procedure by Calsinki and Corsten, and Scheffe's procedure are denoted as DD, KN, TP, CC and SP, respectively. For comparison, the performance of standard ES with constant number of replications equal to 5 in the absence of noise is denoted as OCP, which stands for optimum convergence path. The OCP is included in plots where appropriate.

5.2 Standard ES in the presence of Noise

In this section, the results of the experiments conducted to gain a better understanding of the effect of noise are presented. The amount of noise that an evolution strategy can tolerate before its performance deteriorates is identified.

5.2.1 Effect of noise

Table 5.1 and Table 5.2 show the results of the average actual fitness of the parent solutions in each generation for test function-1 and test function-2, respectively. A plot of the average fitness of the parent solutions in each generation at various levels of noise is shown in Figure 5.1 and Figure 5.2 for test function-1 and tests function-2, respectively. The performance of the algorithm deteriorates at increasingly higher levels of noise for both test function-1 and test function-2. Comparison of Figure 5.1 and Figure 5.2 show that noise has a more deteriorating effect with test function-2 than test function-1, which may be explained owing to the fact that test function-2 is more complex than test function-1. The influence of $2\sigma_{noise}$ on the algorithm is rather significant for test function-2. Hence, a noise level of $2\sigma_{noise}$ is set as the noise level for the remaining experiments.

Table 5.1 Average Fitness of Parents at various noise levels for test function-1

	Noise				
Gen	0	0.5S	1S	1.5S	2S
1	5.196	5.221	5.230	5.258	5.258
2	3.240	2.881	2.456	2.599	2.533
3	1.135	1.683	1.639	1.665	1.876
4	1.027	1.292	1.324	1.447	1.516
5	1.037	1.227	1.278	1.296	1.410
6	1.027	1.149	1.168	1.249	1.342
7	1.006	1.099	1.164	1.190	1.226
8	1.002	1.066	1.115	1.172	1.198
9	1.003	1.078	1.131	1.200	1.232
10	1.003	1.063	1.127	1.163	1.245

Table 5.2 Average Fitness of Parents at various Noise Levels for test function-2

	Noise				
Gen	0	0.5S	1S	1.5S	2S
1	2.260	2.267	2.326	2.488	2.634
2	1.585	1.617	1.793	1.888	2.048
3	1.480	1.530	1.654	1.830	2.014
4	1.124	1.260	1.355	1.590	1.742
5	1.090	1.172	1.281	1.439	1.615
6	1.063	1.132	1.230	1.317	1.532
7	1.042	1.099	1.199	1.266	1.417
8	1.006	1.082	1.162	1.227	1.406
9	1.015	1.071	1.201	1.229	1.417
10	1.006	1.073	1.158	1.203	1.367

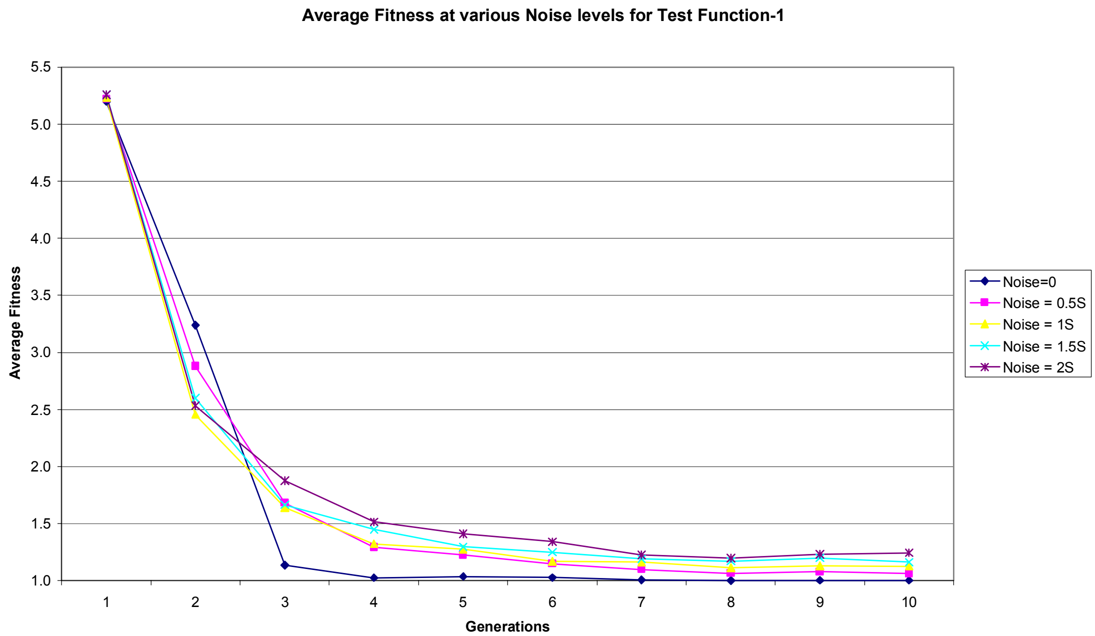

Figure 5.1 Average Fitness of Parents at various noise levels for test function-1

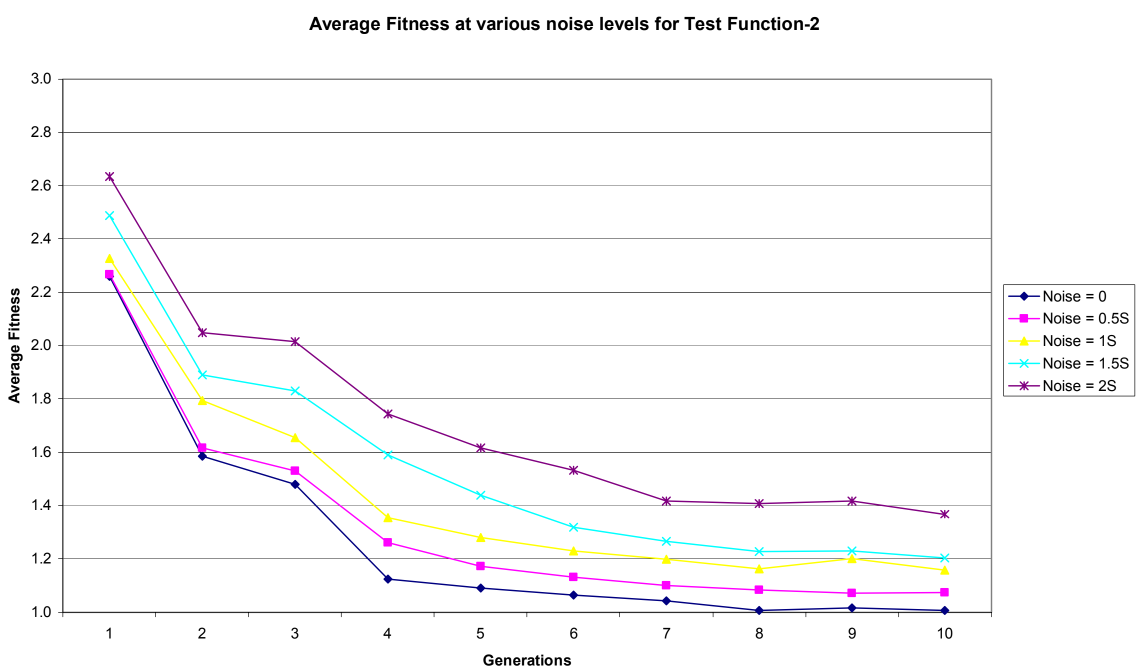

Figure 5.2 Average Fitness of Parents at various Noise Levels for test function-2

5.2.2 Proportion of correct selection

The decrease in the performance of evolution strategies at increasingly higher levels of noise can be attributed to the decrease in the proportion of correct selection. Figures 5.3 and 5.4 show the results of the proportion of correct selection in each generation at various levels of noise for test function-1 and test function-2, respectively. It can be clearly seen that the proportion of correct selection decreases with increasing levels of noise for both test function-1 and test function-2. The proportion of correction selection decreases with the number of generations at various levels of noise. The high proportion of correct selection during the early generations is due to the fact that solutions are widely spaced and the algorithm can easily detect superior solutions in spite of the presence of noise. The decrease in the proportion of correct selection as the search algorithm progresses may be explained owing to solutions being closely spaced in later generations and henceforth making the selection mechanism hard to detect the superior solutions in the presence of noise. Because of the decrease in the proportion of correct selection, the algorithm fails to converge to the optimum within 10 generations at increasingly higher levels of noise.

Knowledge about the required proportion of correct selection for the algorithm to do well is gained by evaluating the performance of ES at various levels of proportion of correct selection under very highly noisy conditions. The noise level for the remaining of the experiments is chosen at $2\sigma_{noise}$. Plots of the average fitness of the parents at various proportions of correct selections for test function-1 and test function-2 are shown in Figure 5.5 and Figure 5.6, respectively. The number of simulation calls expended to obtain the desired proportion of correct selection for test function-1 and test function-2 is tabulated in Table 5.3 and 5.4, respectively. A target proportion of correct selection was obtained by adding replications (simulation calls) to estimate the fitness of solutions until the estimates yielded the desired number of correctly selected parents. Plots 5.5 and 5.6 indicate that a very high proportion of correct selection is not required by the algorithm to perform adequately. For test function-1, a proportion of correct selection greater than 0.5 seems necessary for the algorithm to perform adequately. However for test function-2, a proportion of correct selection greater than 0.75 seems necessary. There is significant impact on test function-2 compared to test function-1.

Table 5.3 and 5.4 show the number of simulation calls required to attain the desired proportion of correct selection. The number of simulation calls required increases with increase in the desired proportion of correct selection. Maintaining a high proportion of correct selection during the later generations, where solutions are closer together, requires a very large number of simulation calls. It is desired to find or develop methodologies that allow the solutions to follow the optimum convergence path with a minimum number of simulation calls.

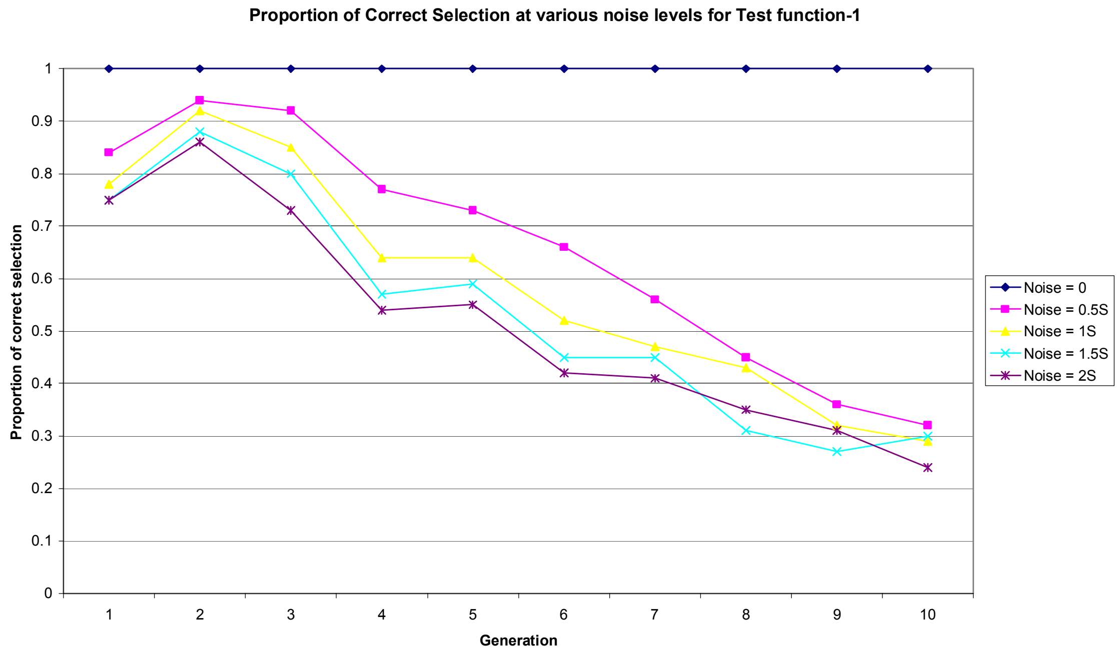

Figure 5.3 Proportion of correct selection at various Noise Levels for test function-1

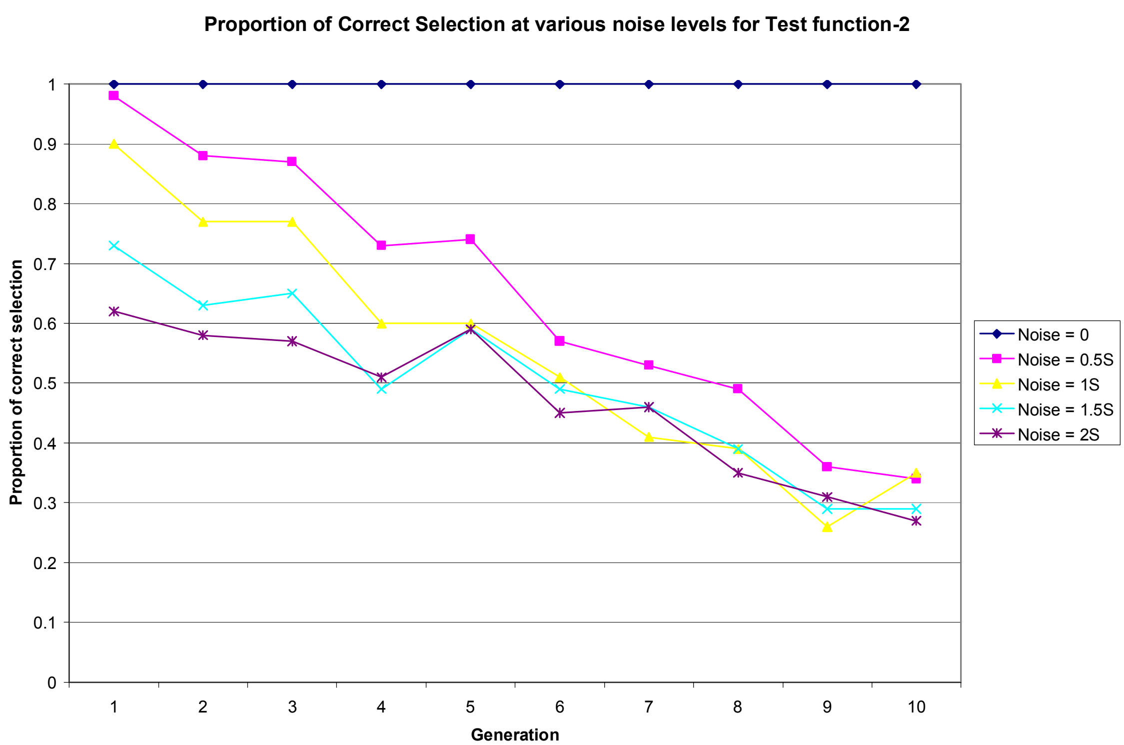

Figure 5.4 Proportion of correct selection at various Noise Levels for test function-2

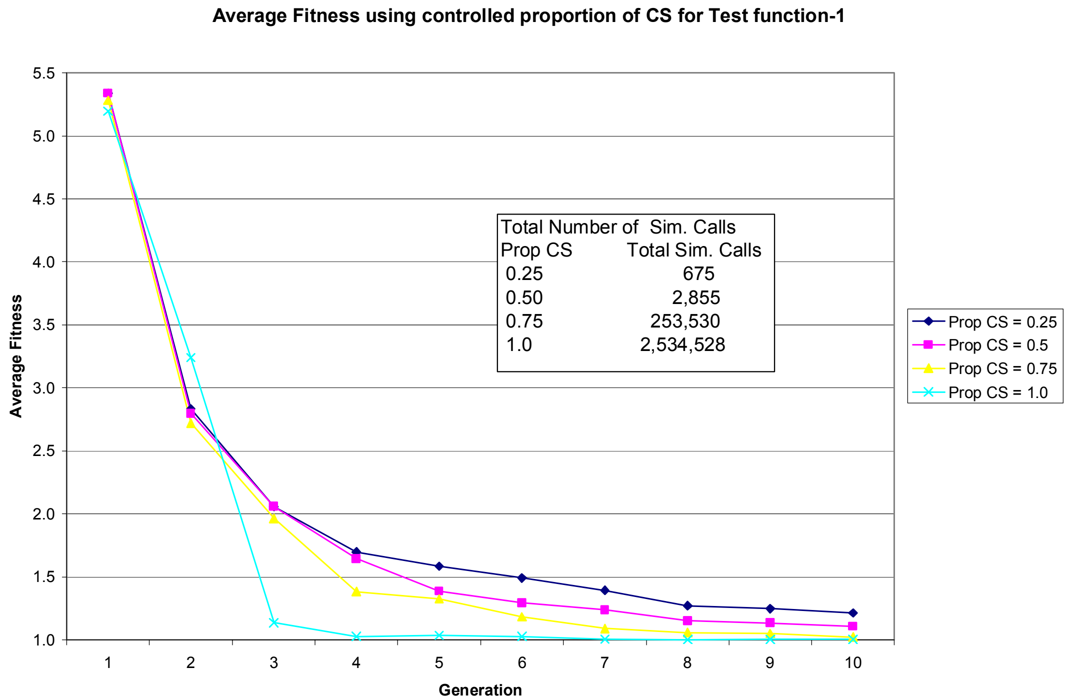

Figure 5.5. Average fitness of parents using controlled proportion of correct selection for test function-1

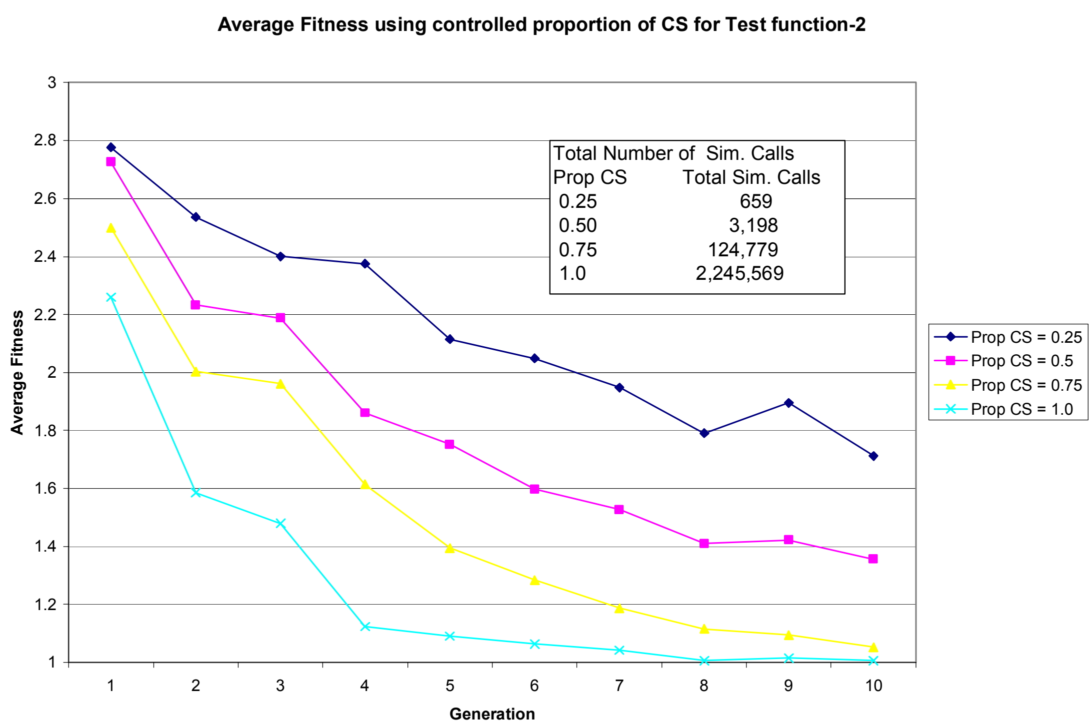

Figure 5.6. Average fitness of parents using controlled proportion of correct selection for test function-2

Table 5.3 Number of simulation calls in each generation for various levels of proportion of correct selection for test function-1

	Proportion of Correct Selection			
Gen	0.25	0.5	0.75	1.0
1	56	56	73	1,800
2	56	57	81	768
3	56	57	81	547
4	58	86	358	10,236
5	59	138	3,624	69,162
6	63	517	2,361	233,512
7	93	167	4,948	289,824
8	72	597	33,999	492,399
9	64	775	74,346	672,373
10	99	404	133,660	763,907

Table 5.4 Number of simulation calls in each generation for various levels of proportion of correct selection for test function-2

	Proportion of Correct Selection			
Gen	0.25	0.5	0.75	1.0
1	56	64	141	572
2	60	102	207	3,714
3	58	69	211	1,751
4	62	120	422	10,702
5	76	137	697	28,245
6	63	190	1,172	145,650
7	58	240	10,020	230,016
8	69	304	12,443	521,852
9	75	1,575	13,755	601,511
10	81	398	85,712	701,557

5.3 Evaluation of modified selection mechanisms

The performance of modified selection mechanisms is evaluated in the presence of very high levels of noise equivalent to $2\sigma_{noise}$ (normally distributed), where the performance of the algorithm has deteriorated significantly. The experiments are conducted at three different levels of probability of correct selection equivalent to 0.9, 0.8 and 0.6, respectively. The results of these modified selection mechanisms at significance levels of 0.1, 0.2 and 0.4 are presented in sections 5.3.1, 5.3.2 and 5.3.3, respectively. For convenience, the techniques, standard ES with constant number of replications equal to 5, selection of the s best by Dudewicz and Dalal, subset selection by Kim and Nelson, Tukey's multiple comparison procedure, clustering procedure by Calsinki and Corsten, and Scheffe's procedure are shortly denoted as SD, DD, KN, TP, CC and SP, respectively. A fixed indifference zone of 0.1 is used for DD and KN.

5.3.1 Performance of Modified selection mechanisms at significance level = 0.1

The average fitness of the parent solutions in each generation for each of the modified selection mechanisms under very highly noisy conditions equivalent to $2\sigma_{noise}$ at a significance level of 0.1 for test function-1 and test function-2 are plotted in Figure 5.7 and Figure 5.8, respectively. The optimum convergence path in the absence of noise is also included in the plot and is denoted as SD. Similarly, Table 5.5 and Table 5.6 show the corresponding number of simulation calls expended for test function-1 and test function-2, respectively, for each of the modified selection mechanisms at a significance level of 0.1. In addition, corresponding plots of the proportion of correct selection in each generation for test function-1 and test function-2 are shown in Figure 5.9 and Figure 5.10, respectively.

It is observed in Figure 5.7 and Figure 5.8 that DD and KN follow closely along the convergence path towards the optimum, whereas TP, CC and SP do not perform as well under the given conditions. Note that all the five techniques perform well on test function-1, whereas the techniques TP, CC and SP fail to converge to the optimum in 10 generations for test function-2. This is due to the fact that test function-2 is much more complex than test function-1. SP performs comparatively better than TP and CC in 10 generations. Looking at Figures 5.7, 5.8, 5.9 and 5.10, TP and CC do not perform well inspite of the relatively high proportion of correct selection. The under performance of TP and CC may be explained owing to the higher parent population size, which would decrease the selective pressure and slow the algorithm's convergence speed. It is also observed that TP performs slightly better than CC. Moreover, the proportion of correct selections remains relatively high for TP and CC until the 10^{th} generation. This explains the deterioration of the selection mechanism TP and CC due to the higher parent population size. So, we would expect that TP and CC might converge to the optimum if the algorithm is allowed to run longer than 10 generations.

Scheffe's procedure may perform better if it is allowed to run for longer number of generations or at a higher probability of correct selection. Note that the proportion of correct selection was relatively low compared to the other techniques. Scheffe's procedure could be improved by increasing the probability of correct selection. Note that TP, CC and SP consumed a very small number of simulation calls in the early generations for test function-1 and test function-2.

Table 5.5 and 5.6 show that DD consumed an excessively high number of simulation calls compared to other techniques. Figure 5.3 and Figure 5.4 show that DD and KN follow very closely along the optimum convergence path and KN performs as good as DD with far fewer simulation calls than required for DD. Table 5.5 and 5.6 show a general trend of increase in the simulation calls for generation 1 through generation 10 for all the techniques.

We would expect that DD and KN to have a high proportion of correct selection since they follow closely along the optimum convergence path. An interesting result is that the proportion of correct selection was high in the early generations for DD and KN, but was not maintained high in the later generations. This is because the difference between the fitness of the solutions is less than the indifference zone. In other words, the solutions are so closely spaced in the later generations that the procedures consider them to be indifferent. This low proportion of correct selection did not affect the search procedure using DD and KN. The low proportion of correct selection in the later generations did not have any affect on the performance of DD and KN because the solutions quickly reached the optimum region and the solutions are very closely spaced together. Hence, an improper selection of a solution as parent after the solutions have converged did not have any affect on the convergence of the algorithm. TP and CC maintain a very high proportion of correct selection, but still do not converge since the parent population size was high and also the solutions are spaced farther apart even at the end of 10 generations, which can be observed based on the average fitness at the end of 10 generations.

Table 5.5 No of sim calls for modified selection techniques at $\alpha = 0.1$ for Test function-1

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	130,588	7,190	140	140	140
2	131,553	5,656	141	149	140
3	129,095	8,700	162	153	140
4	129,564	34,720	243	236	140
5	129,705	47,043	335	299	172
6	128,878	75,632	486	563	547
7	128,069	81,229	1,025	383	400
8	128,957	105,016	3,734	563	976
9	128,215	103,738	4,517	755	1,816
10	130,663	115,496	8,637	965	6,225

Table 5.6 No of sim calls for modified selection techniques at $\alpha = 0.1$ for Test function-2

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	130,588	17,298	570	306	217
2	131,553	16,748	320	271	189
3	129,095	21,160	379	306	178
4	129,564	34,321	525	336	200
5	129,705	51,077	738	328	278
6	128,878	67,004	519	402	564
7	128,069	84,197	972	355	1,586
8	128,957	109,000	1,648	1,269	3,287
9	128,215	112,914	2,474	876	3,634
10	130,663	123,486	3,368	1,935	5,124

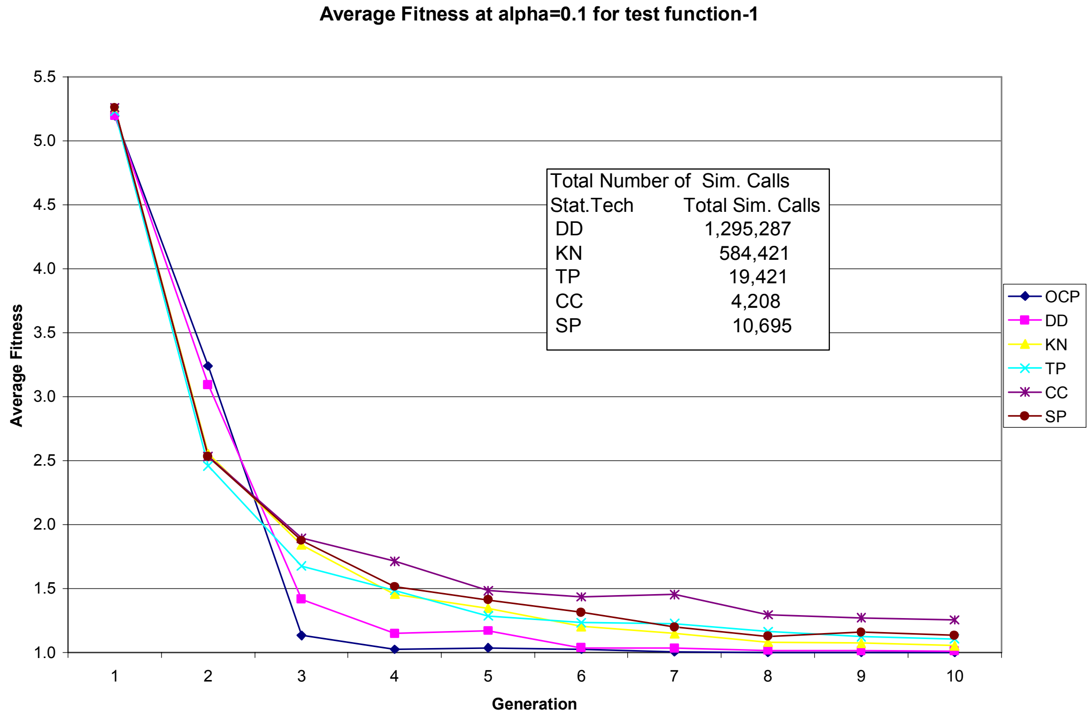

Figure 5.7 Average Fitness of parents for modified selection mechanisms at α=0.1 for Test function-1

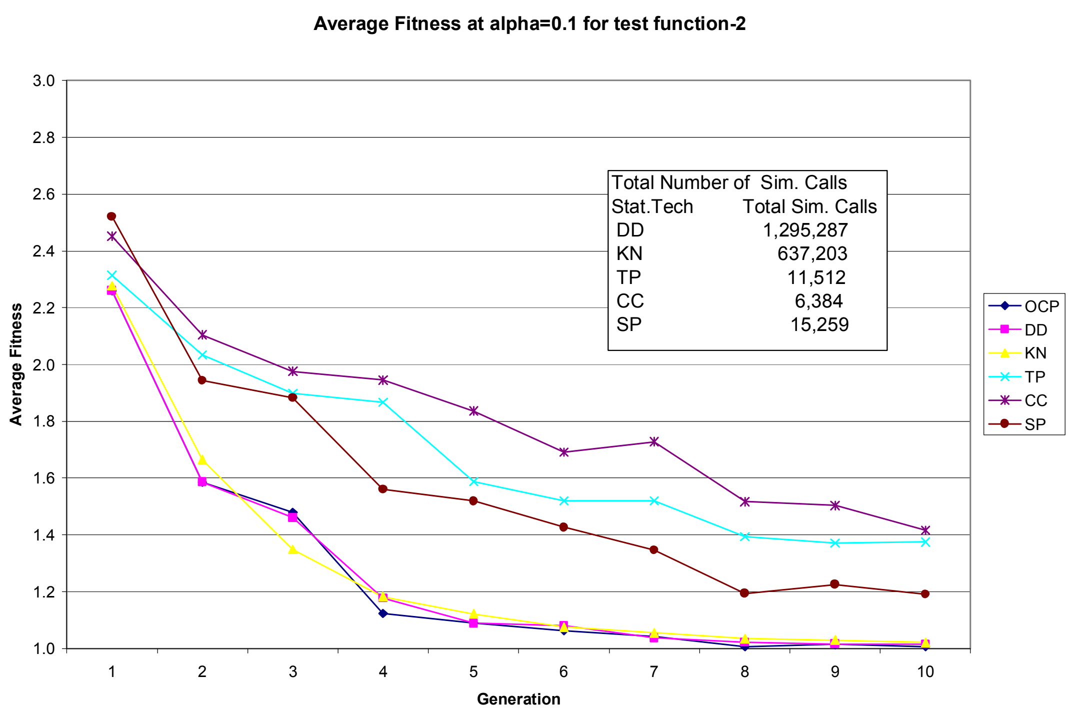

Figure 5.8. Average Fitness of parents for modified selection mechanism at α=0.1 for Test function-2

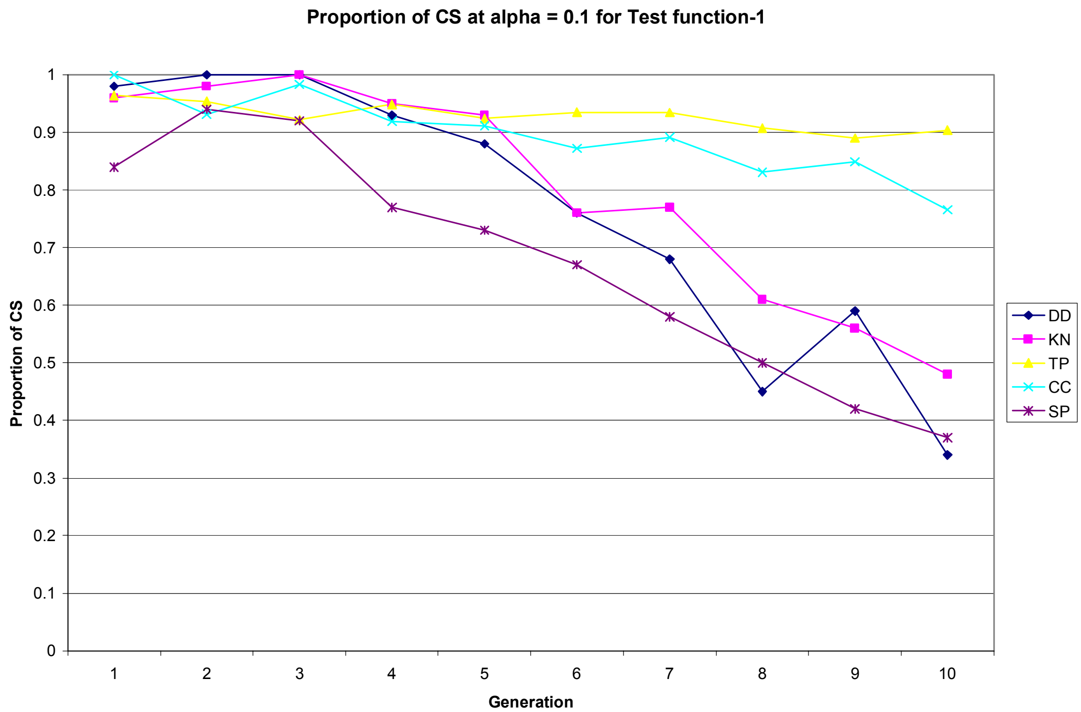

Figure 5.9. Proportion of CS for modified selection mechanisms at α=0.1 for Test function-1

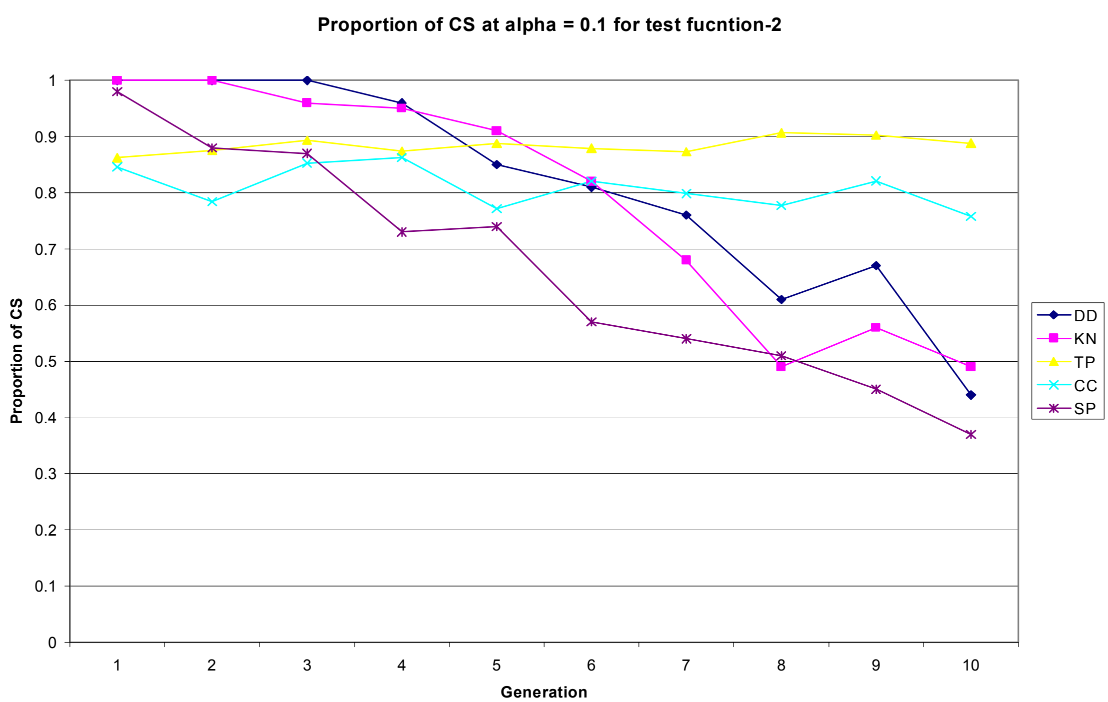

Figure 5.10 Proportion of CS for modified selection mechanisms at α=0.1 for Test function-2

5.3.2 Performance of Modified selection mechanisms at significance level = 0.2

The average fitness of the parent solutions in each generation for each of the modified selection mechanisms under very highly noisy conditions equivalent to $2\sigma_{noise}$ (normally distributed), at a significance level of 0.2 for test function-1 and test function-2 are plotted in Figure 5.11 and Figure 5.12, respectively. Similarly, Table 5.7 and Table 5.8 show the number of simulation calls expended for test function-1 and test function-2, respectively, for each of the modified selection mechanisms at a significance level of 0.2. In addition, corresponding plots of the proportion of correct selection in each generation for test function-1 and test function-2 is shown in Figure 5.13 and Figure 5.14, respectively. It can be seen that DD and KN follow closely along the convergence path towards the optimum, whereas TP, CC and SP do not perform as well under the given conditions.

The results of the average fitness of the parent solutions are similar to the case where the probability of correct selection is 0.9. Selection mechanism TP, CC and SP do not perform well under the given experimental conditions for test function-2. Note the quality of the solution is about the same at the end of search for both levels of α for test function-1. For test function-2, the average fitness of the parents is higher using TP, CC and SP at α = 0.2 compared to α=0.1 indicating the degradation in the quality of the solution with increase in the level of significance. The number of simulation calls expended decreased for all the modified selection mechanisms with increase in the level of significance from 0.1 to 0.2. The main difference between the results for the case of α=0.1 and α=0.2 is the decrease in the number of simulation calls, while the remaining output parameters of interest are fairly similar for DD and KN. Note that DD and KN still expended an excessively high number of simulation calls.

Table 5.7. No of sim calls for modified selection techniques at $\alpha = 0.2$ for Test function-1

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	104,075	4,688	140	140	140
2	102,124	3,394	140	149	140
3	105,458	6,312	156	144	140
4	101,170	25,142	197	216	140
5	101,395	34,579	242	199	151
6	101,640	55,652	551	580	496
7	103,112	62,056	878	386	314
8	101,652	86,142	2,675	687	1,992
9	101,124	87,801	4,604	825	2,415
10	102,074	98,148	7,615	1,023	2,558

Table 5.8. No of sim calls for modified selection techniques at $\alpha = 0.2$ for Test function-2

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	104,075	11,592	469	283	196
2	102,124	11,103	290	226	155
3	105,458	13,179	334	241	170
4	101,170	22,815	529	282	198
5	101,395	36,908	598	272	299
6	101,640	43,298	683	421	515
7	103,112	62,465	899	605	571
8	101,652	82,982	1,436	758	3,338
9	101,124	89,291	2,367	880	3,093
10	102,074	101,736	3,076	687	2,463

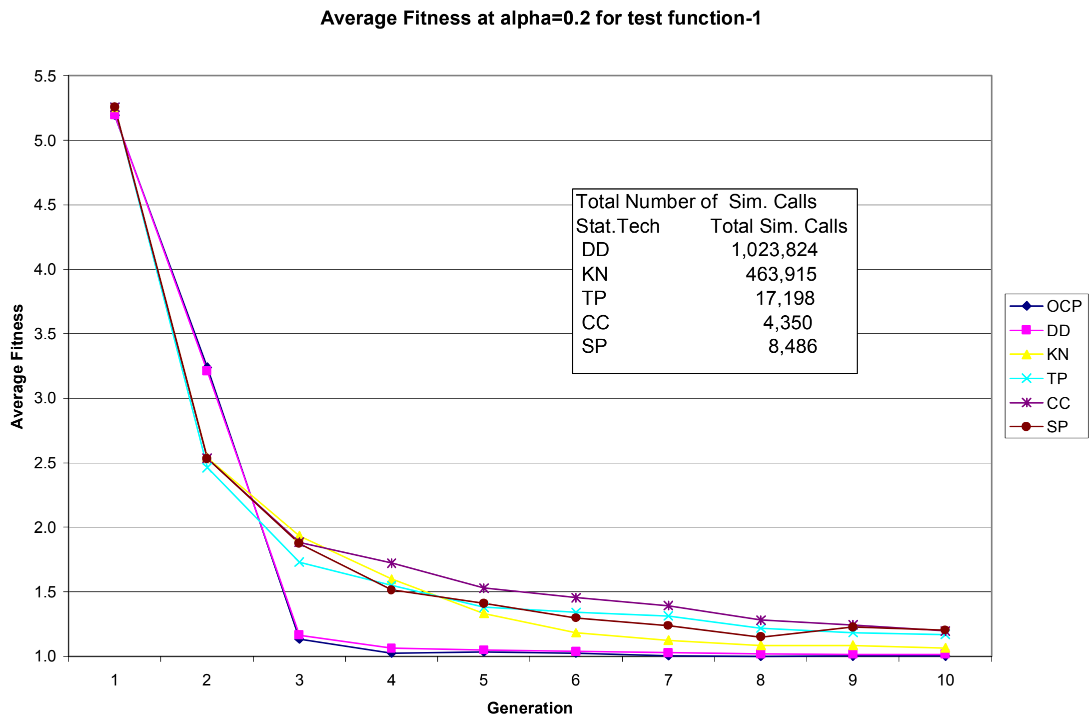

Figure 5.11 Average Fitness of the solutions for modified selection mechanism at α=0.2 for Test function-1

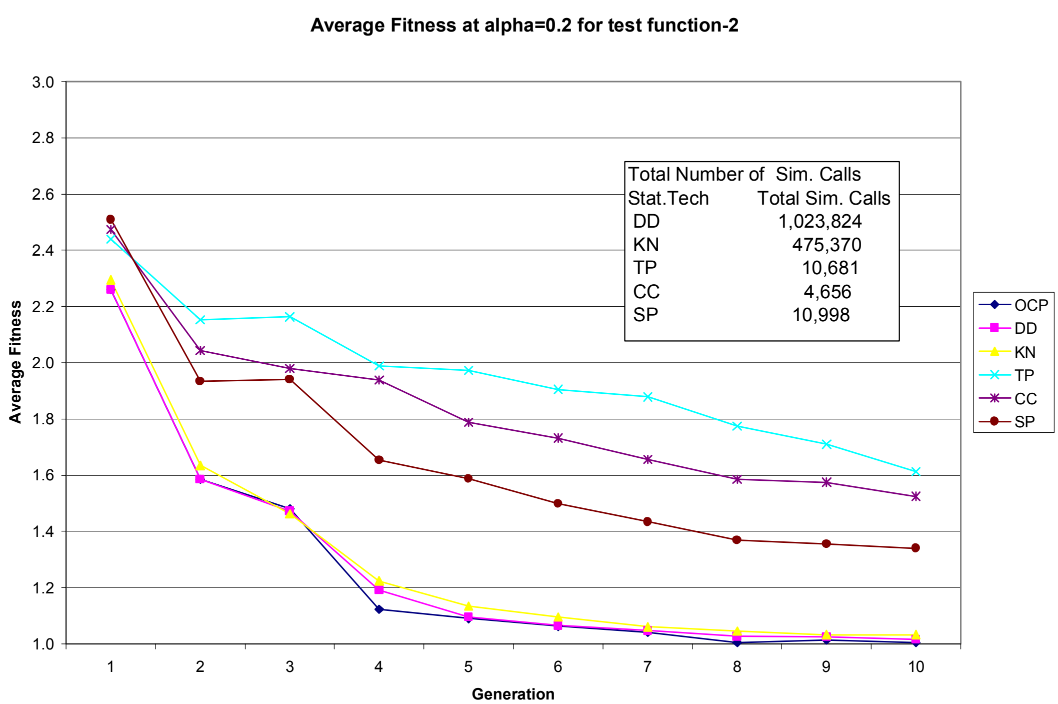

Figure 5.12 Average Fitness of the solutions for modified selection mechanism at α=0.2 for test function-2

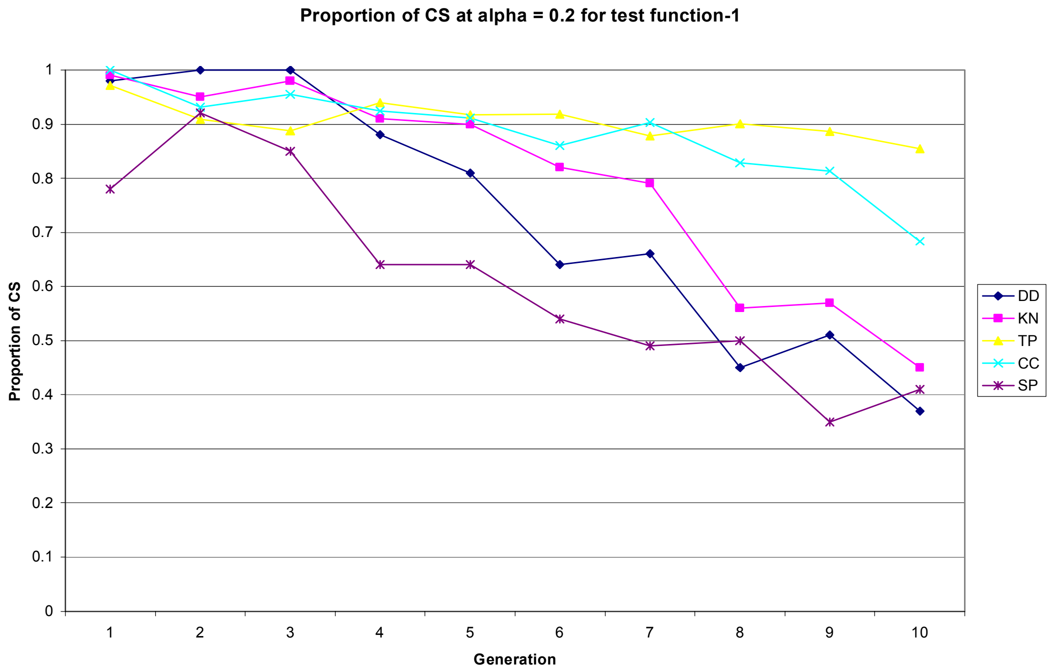

Figure 5.13 Proportion of CS for modified selection mechanisms at α=0.2 for Test function-1

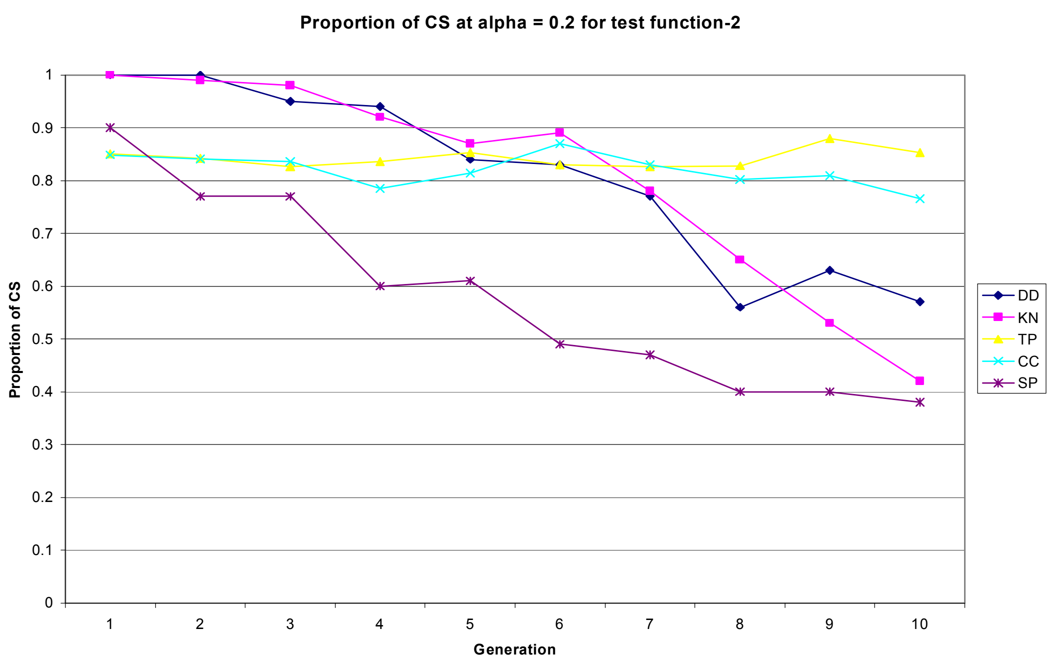

Figure 5.14 Proportion of CS for modified selection mechanisms at α=0.2 for Test function-2

5.3.3 Performance of Modified selection mechanisms at significance level = 0.4

The average fitness of the parent solutions in each generation for each of the modified selection mechanisms under very highly noisy conditions equivalent to $2\sigma_{noise}$ (normally distributed) at a significance level of 0.4 for test function-1 and test function-2 are plotted in Figure 5.15 and Figure 5.16, respectively. In addition, corresponding plots of the proportion of correct selection in each generation for test function-1 and test function-2 are shown in Figure 5.17 and Figure 5.18, respectively. Table 5.9 and Table 5.10 show the number of simulation calls expended for test function-1 and test function-2, respectively, for each of the modified selection mechanisms at a significance level of 0.4. It is observed that DD and KN follow closely along the convergence path towards the optimum, whereas TP, CC and SP do not perform as well under the given conditions. Similar results are obtained for significance level of 0.2 and 0.1

The results of the average fitness of the parent solutions are similar to the case where the probability of correct selection is 0.9 and 0.8. The performance of the search algorithm using SP is further degraded with the decrease in the probability of correct selection. However, the number of simulation calls expended is much lower than the number of simulation calls expended for the case where α is 0.1 or α is 0.2. The number of simulation calls expended decreased for all the modified selection mechanisms with increase in the level of significance from 0.2 to 0.4. The main difference between the results for the case of α=0.1, α=0.2 and α=0.4 is the decrease in the number of simulation calls, while the remaining parameters are fairly similar for DD and KN. The number of simulation calls is much lower compared to the case where the probability of correct selection is 0.8 without any degradation in the quality of the solutions for DD and

KN. Although, the number of simulation calls has decreased for DD and KN with decrease in the probability of correct selection, the total number of simulation calls expended is still very high. Especially, DD consumed a very large number of simulation calls compared to KN with relatively insignificant difference in the performance. Also, the number of simulation calls is very high even in the early generations for DD, which implies that a significant amount of simulation effort is utilized than necessary.

Table 5.9. No of simulation calls for modified selection techniques at $\alpha = 0.4$ for Test function-1

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	88,917	2,982	140	140	140
2	86,447	2,303	140	148	140
3	86,780	3,969	144	148	140
4	90,019	12,892	188	198	140
5	86,008	19,491	267	343	140
6	86,903	34,294	776	306	431
7	87,450	42,210	1,056	439	964
8	87,133	55,239	2,624	365	1,626
9	88,490	59,503	2,188	375	544
10	85,544	70,887	5,140	1,200	1,947

Table 5.10 No of simulation calls for modified selection techniques at $\alpha = 0.4$ for Test function2

	Modified Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	88,917	7,561	351	223	176
2	86,447	7,607	264	208	146
3	86,780	9,557	264	204	143
4	90,019	15,170	250	260	181
5	86,008	21,319	501	228	217
6	86,903	29,902	417	245	371
7	87,450	41,413	430	284	452
8	87,133	52,071	660	366	1,644
9	88,490	57,071	1,032	420	1,670
10	85,544	65,724	1,845	520	4,620

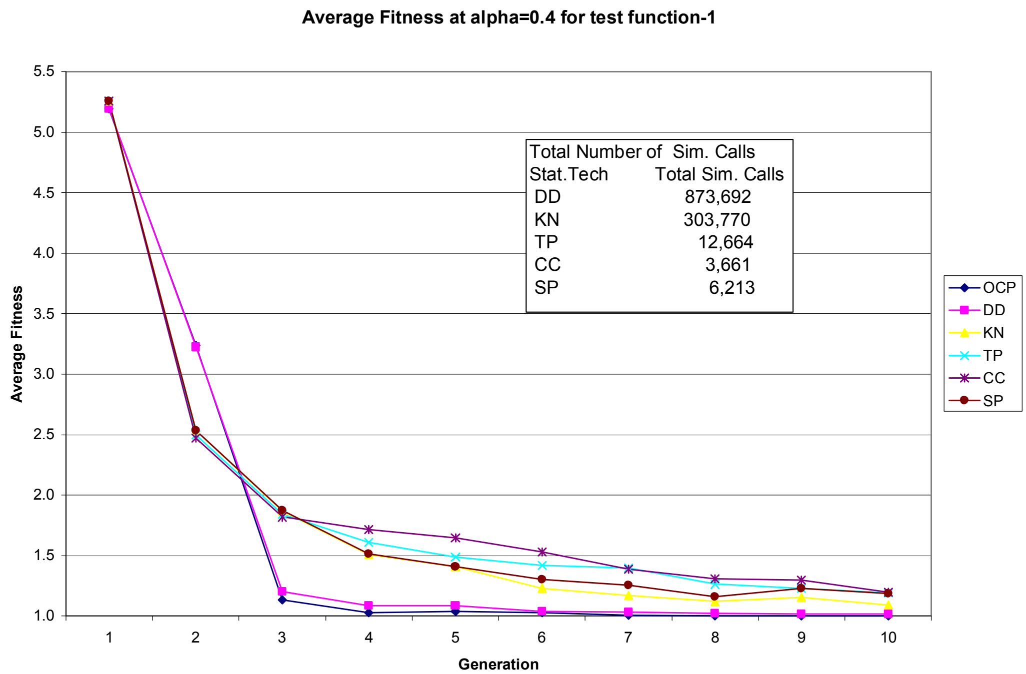

Figure 5.15 Average Fitness of the solutions for modified selection mechanism at α=0.4 for Test function-1

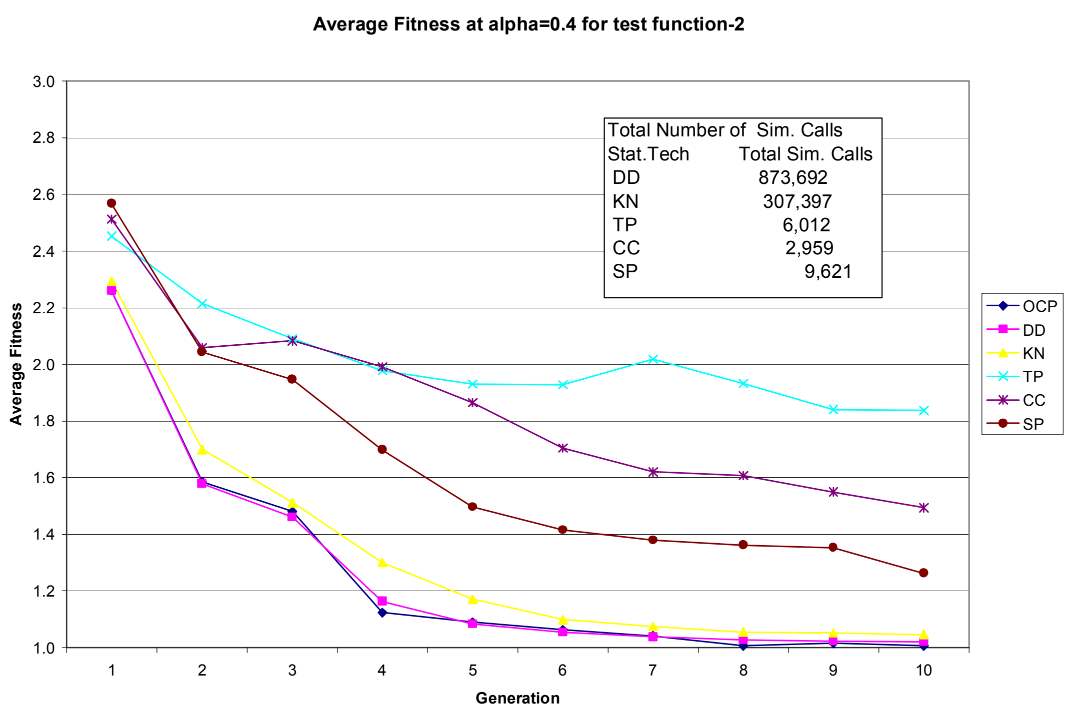

Figure 5.16 Average Fitness of the solutions for modified selection mechanism at α=0.4 for Test function-2

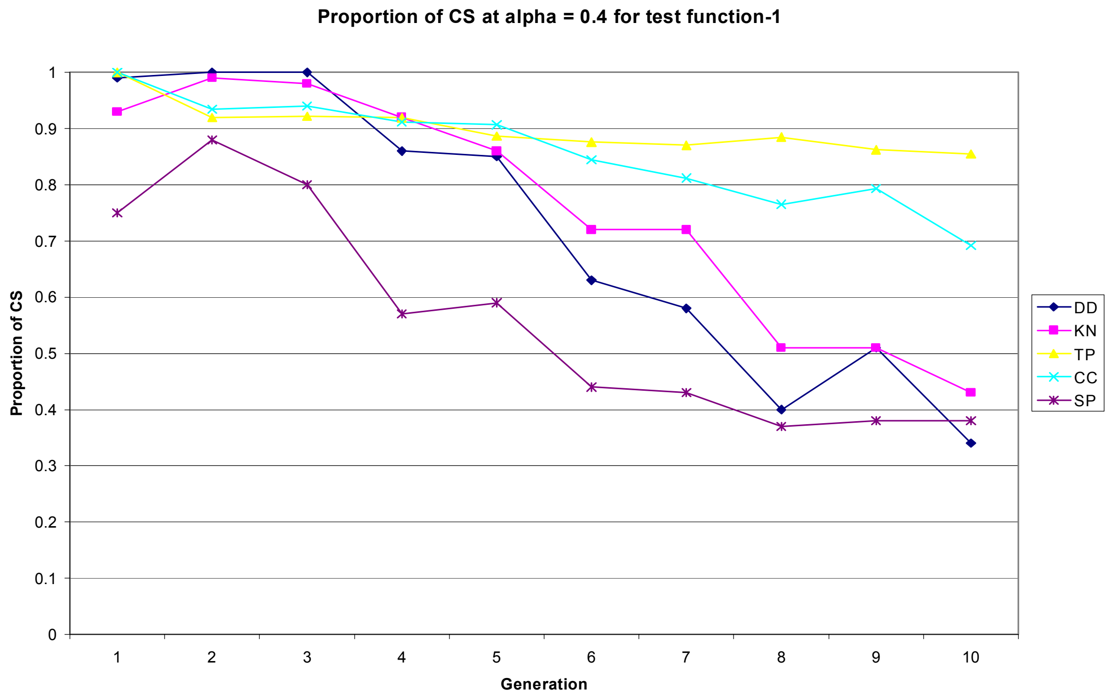

Figure 5.17 Proportion of CS for modified selection mechanisms at α=0.4 for Test function-1

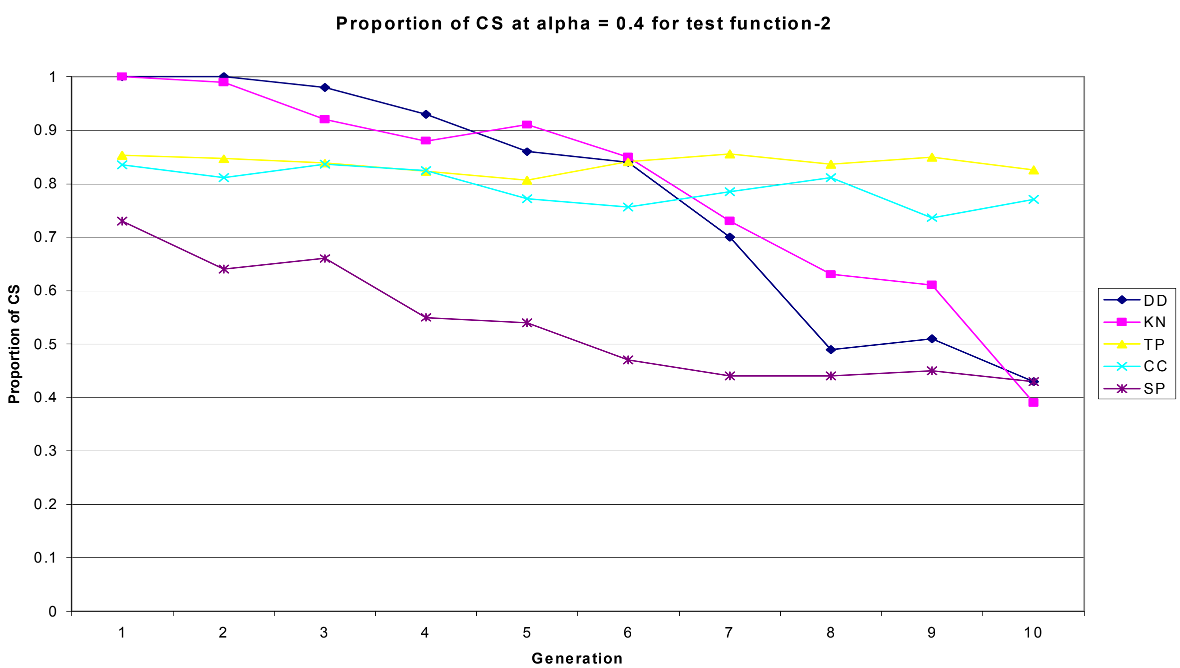

Figure 5.18 Proportion of CS for modified selection mechanisms at α=0.4 for Test function-2

5.4 Indifference zone techniques

A variable indifference zone methodology is employed within the two selection mechanisms, DD and KN in such a way that the indifference zone starts higher in the early generations and decreases as the search progresses towards to the optimum. This was tested to determine if it would allow the algorithm to perform adequately with far less number of simulation calls. Half the average fitness distance is used as the indifference zone, which is based on the initial estimates of the fitness from 5 replications. To further decrease the number of simulation calls required, DD and KN selection mechanisms are evaluated with the probability of correct selection equivalent to 0.6, since the quality of the search algorithm remains good even at a high significance level using DD and KN. Since, TP, CC and SP do not have an indifference zone kind of procedure inherently, the number of simulation calls in any generation is limited to twice the number of simulation calls expended in the previous generation. This would restrict the algorithm from spending excessive number of simulation calls especially when solutions are very closely spaced. TP and CC are evaluated at a probability of correct selection equivalent to 0.6. SP is evaluated at a probability of correct selection equal to 0.99 with the hope of improving the convergence of the algorithm. Note that the techniques are evaluated at a very high noise level equivalent to $2\sigma_{noise}$ (normally distributed).

Table 5.11 and Table 5.12 show the number of simulation calls expended in each generation for test function-1 and test function-2 respectively. The average actual fitness of the parent solutions in each generation for each of the selection mechanisms with modified indifference zone procedures under very highly noisy conditions at a significance level of 0.4 for test function-1 and test function-2 are plotted in Figure 5.19 and Figure 5.20, respectively. In addition, corresponding plots of the proportion of correct selection in each generation for test function-1 and test function-2 are shown in Figure 5.21 and Figure 5.22, respectively.

Note that from Table 5.11 and 5.12, the total number of simulation calls has decreased significantly by using the adaptive indifference zone method for DD and KN for both test function-1 and test function-2. Also note that the number of simulation calls has decreased significantly for the other techniques as well. (See Table 5.9 and Table 5.10). KN uses slightly lower number of simulation calls compared to the number of simulation calls expended by DD. The number of simulation calls increases as the search progresses towards the optimum region.

Comparison of Figure 5.19 and Figure 5.20 show that test function-2 proves more difficult. For test function-2, the algorithm does not converge using the selection techniques TP, CC under the given experimental conditions as seen in Figure 5.20. There is a significant improvement in the performance of SP (Compare Figure 5.16 with Figure 5.20). SP is approaching being competitive with DD and KN using significantly fewer simulation calls. The algorithm with selection techniques DD and KN converges towards the optimum utilizing much smaller number of total simulation calls than before. However, the solutions are not quite as good (Compare Figure 5.16 with Figure 5.20).

These results indicate that a higher proportion of correct selection in the early generations helps the algorithm converge towards the optimum region quickly, and in the later stages, a moderate proportion of correct selection would be sufficient (Figure 5.21 and Figure 5.22). In other words, maintaining a reasonably good convergence velocity in the early generations helps the algorithm converge quickly towards the optimum. This would be advantageous since maintaining a high proportion of correct selection in the early generations requires far fewer simulation calls as compared to later generations.

Table 5.11 No of simulation calls using indifference zone technique for test function-1

	Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	140	140	140	140	140
2	142	140	140	148	140
3	144	140	142	148	140
4	173	144	166	184	140
5	443	254	194	175	148
6	1,106	923	289	245	205
7	1,096	985	391	234	284
8	2,768	2,317	616	353	464
9	2,044	1,508	811	396	608
10	3,658	2,608	1,397	445	1,146

Table 5.12 No of simulation calls using indifference zone technique for test function-2

	Selection Mechanisms				
GEN	DD	KN	TP	CC	SP
1	2,148	813	291	223	306
2	1,635	530	231	203	220
3	1,770	688	240	179	242
4	2,059	1,111	206	230	267
5	2,022	1,675	289	246	325
6	2,956	1,981	282	240	449
7	3,020	2,677	283	259	606
8	4,104	4,499	389	290	1,094
9	3,459	4,944	451	300	1,406
10	4,983	5,561	416	412	1,973

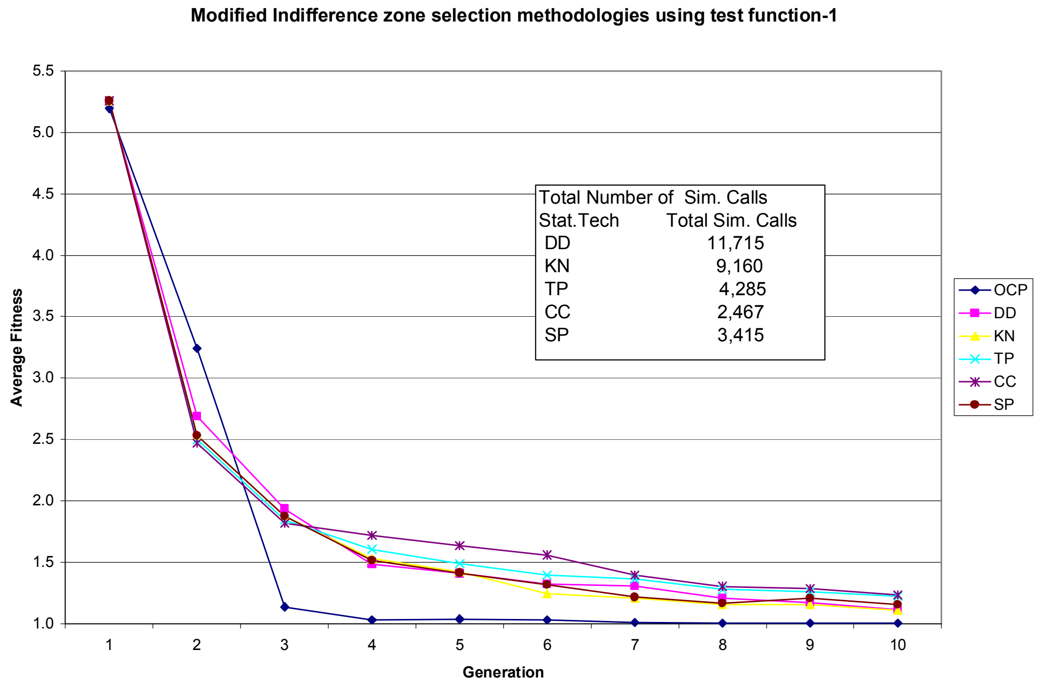

Figure 5.19. Average Fitness of the solutions for modified indifference zone selection mechanism for Test function-1

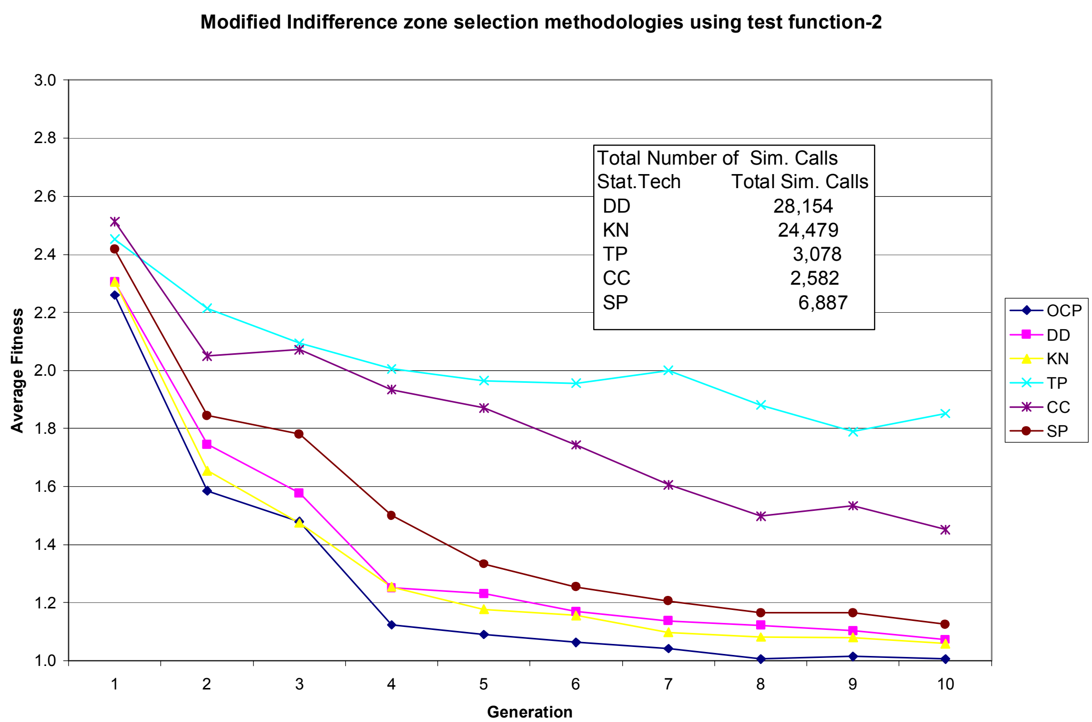

Figure 5.20. Average Fitness of the solutions for modified indifference zone selection mechanism for Test function-2

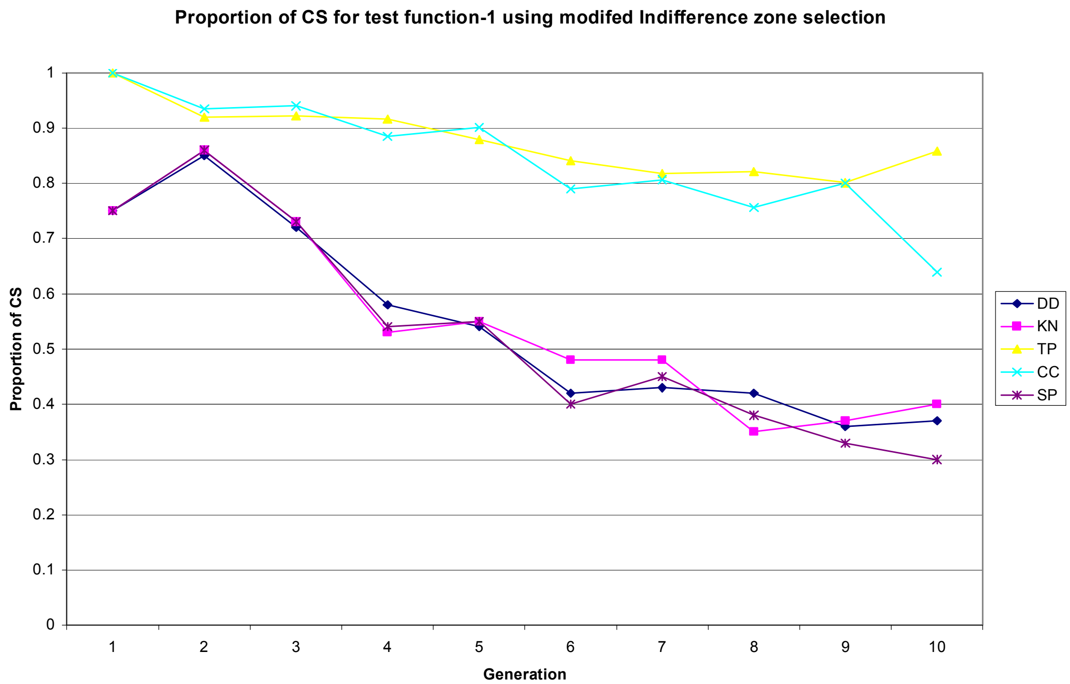

Figure 5.21 Proportion of CS for modified indifference zone selection mechanisms for Test function-1

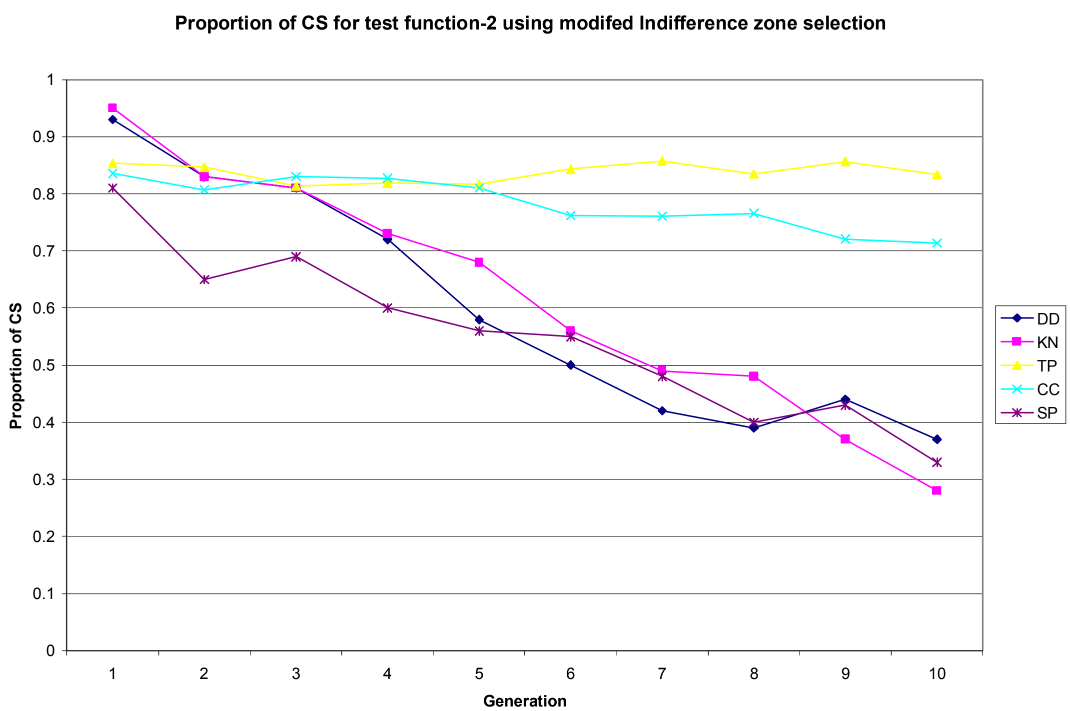

Figure 5.22 Proportion of CS for modified indifference zone selection mechanisms for Test function-2

A question that remains unanswered is if the deterioration in the quality of the solution in terms of the average fitness of the solutions is worth the savings in the number of simulation calls. Table 5.13 shows the comparison of the average fitness of the parents in generation 10 and the total number of simulation calls required by the search procedure on test function-1 for DD and KN using α=0.4 and dynamic indifference zone or the constant (static) indifference zone of 0.10. Similar results for test function-2 are presented in Table 5.14.

Table 5.13. Comparison of dynamic indifference zone and static indifference zone on test function-1

	Average Fitness		Simulation Calls	
Technique	DD	KN	DD	KN
Dynamic Indifference Zone	1.113	1.108	11,715	9,160
Static Indifference Zone	1.018	1.013	873,692	303,770

Table 5.14. Comparison of dynamic indifference zone and static indifference zone on test function-2

	Average Fitness		Simulation Calls	
Technique	DD	KN	DD	KN
Dynamic Indifference Zone	1.073	1.059	28,154	24,479
Static Indifference Zone	1.020	1.017	873,692	307,397

Note that an average fitness improvement of 0.095 for test function-1 using DD with static indifference zone of 0.10 required an additional 861,977 simulation calls (See Table 5.13). Similarly, an average fitness improvement of 0.053 for test-function-2 using DD with static indifference zone of 0.10 required an additional 845,538 simulation calls. Clearly, though a slight improvement can be seen, DD has used an excessively large number of simulation calls. An average fitness improvement of 0.095 for test function-1 using KN with static indifference zone of 0.10 required an additional 294,610 simulation calls. Similarly, an average fitness improvement of 0.0421 is obtained for test function-2 using an additional 282,918 simulation calls. It is concluded that the improvement in terms of the average fitness of the parent solutions is not worth the increase in the number of total simulation calls, when using a static indifference zone of 0.10. This analysis depends on the value chosen for the static indifference zone. Perhaps, a static indifference zone equal to 0.10 is too small.

Figures 5.23 and 5.24 show the dynamic indifference zone for DD for test function-1 and test function-2, respectively. The plots show that the indifference zone decreases as the search progresses for both test function-1 and test function-2. As expected the plots show that the indifference zone decreases as the search algorithm progresses towards the optimum region. Note that the indifference zone in the final generation is approximately 0.8 for test function-1, which is much higher than the static indifference zone of 0.1. Similarly, the indifference zone in the final generation for test function-2 is approximately 0.5. The chosen indifference zone of 0.1 is too small for this test function and that explains the excessive simulation effort spent for both test-function-1 and test function-2 using a static indifference zone. Given that a reasonable indifference zone value will be difficult to determine for most problems, the dynamic indifference zone procedure is appealing.

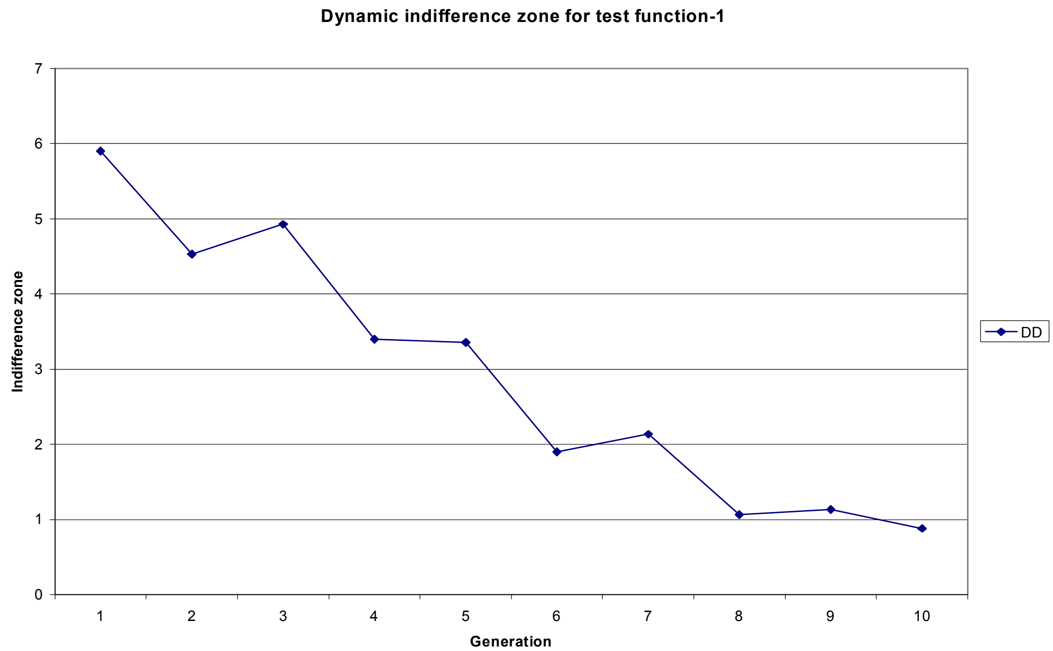

Figure 5.23. Dynamic Indifference zone for Test function-1 using DD at $\alpha = 0.4$.

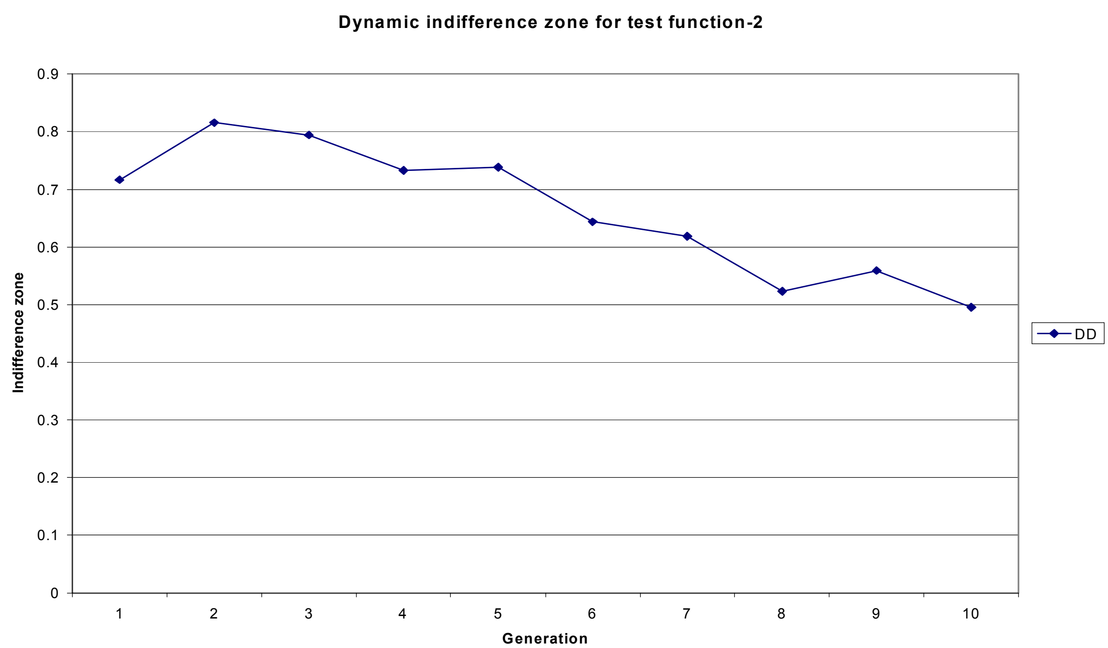

Figure 5.24. Dynamic Indifference zone for Test function-2 using DD at $\alpha = 0.4$.

5.5 Comparison of modified selection mechanisms with standard ES

Although, the number of simulation calls is very low for TP, CC and SP; the solutions fail to converge to the optimum for test function-2 under the given experimental conditions. Noting, however, that SP was much more competitive. Hence, DD and KN, which performed well, even at high levels of noise for the given experimental conditions are compared against the standard ES by allocating an equivalent number of simulation calls used by KN to the standard ES. The simulation calls are allocated to the standard ES by increasing the population size by a factor that makes use of allocated simulation calls when each solution's fitness is estimated from 5 replications. This modified standard ES designed for comparison is denoted as *SD-C*. Table 5.14 shows that 28,000+ simulation calls are expended by DD in 10 generations for test function-2. KN required a smaller number of simulation calls for both test function-1 and test function-2. Hence the population for SD-C is set equal to 28,000/(10*5) = 564 (approximately). Figure 5.25 shows a plot of the average actual fitness of the parent solutions for test function-1 at a noise level of $2\sigma_{noise}$ using the various selection mechanisms under comparison. A similar plot for test function-2 is shown in Figure 5.26.

The average fitness of the parent solutions using DD, KN, and SD-C procedures on test function-1 are not significantly different in the 10^{th} generation (See Figure 5.25). Figure 5.26 illustrates that KN reached a lower average fitness value than did SD-C using approximately the same number of simulation calls. An interesting observation is that the average fitness of the parents using SD-C is much lower than the other techniques in the beginning generations, which is primarily because of the very high population size.

However, a high population size did little to speed up SD-C's rate of convergence to the optimum as the search progresses in the presence of very high levels of noise. The KN with dynamic IZ more wisely distributed the simulation calls among the population of solutions and generations illustrating that the effects of high levels of noise cannot be overcome by simply increasing the size of the algorithm's population size.

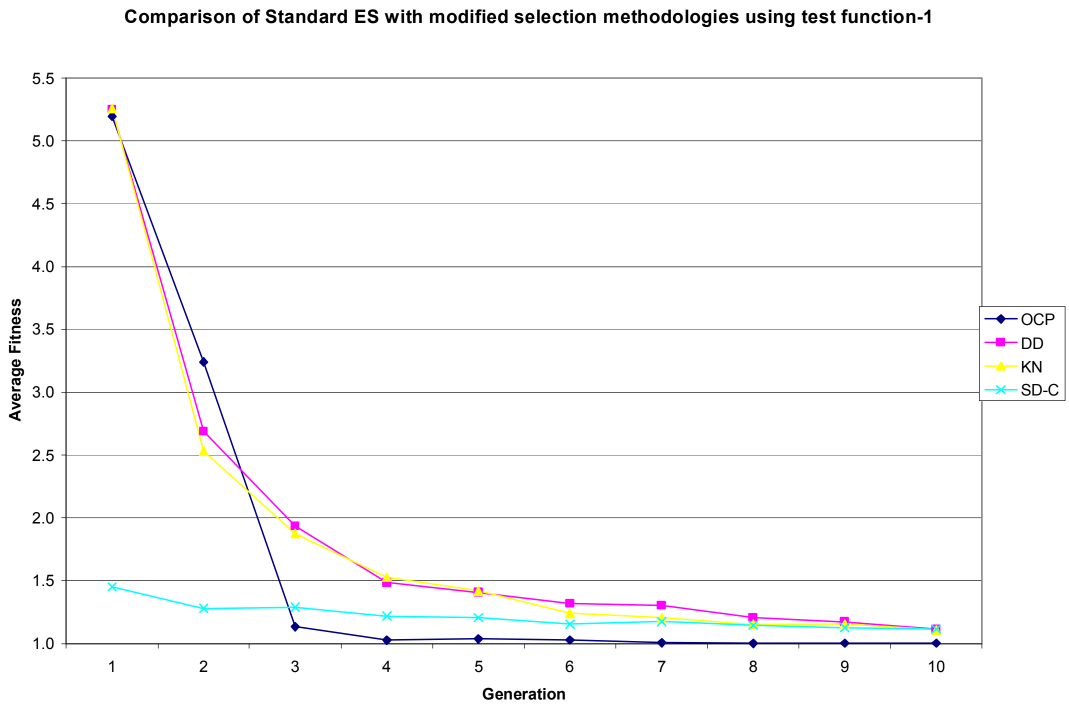

Figure 5.25 Comparison of modified selection mechanisms in terms of average fitness for test function-1

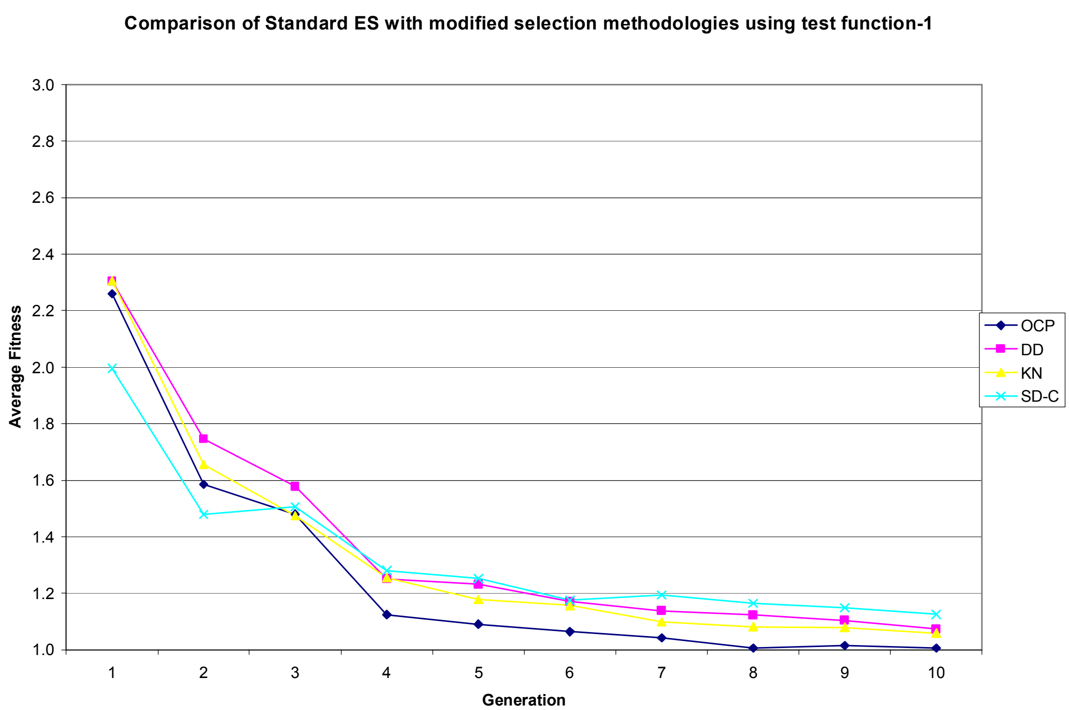

Figure 5.26 Comparison of modified selection mechanisms in terms of average fitness for test function-2

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The objectives of this thesis are to gain a better understanding of the effect of variation in a response surface on the performance of evolution strategies, identify potential statistical techniques that can be integrated into the ES to address the variation and to evaluate the effectiveness of these techniques within the context of evolution strategies. After conducting a series of experiments followed by analysis of results and review of the literature, the following conclusions have been reached. Also, recommendations for future research are presented.

Evolution strategies become less effective in locating the optimum solution at increasingly higher levels of noise. Noise affects the selection mechanism in an ES; hence the selection mechanism has to be modified to cope with high levels of noise. As the level of variation or noise increases, the proportion of solutions correctly identified as parents decreases. For any noise level, the proportion of correct selection is high early in the search phase and decreases as the search progresses towards the optimum. This is because the solutions are farther apart early in the search phase and noise has less affect on the selection mechanism. However, in the later stages of the search, solutions are more closely spaced and the affect of noise is felt more severely as evolution strategies converge towards the optimum. Experimental results suggest that a very high proportion of correct selection is not required for evolution strategies to cope with noise. A moderate proportion of correct selection approximately greater than 0.75 was sufficient to guide evolution strategies towards the optimum. A high proportion of correctly selected solutions can be achieved by obtaining large number of observations at each solution. Since simulation computation is very expensive there is a need to identify techniques that achieves the goal of guiding evolution strategies towards the optimum solution with minimum computational effort.

Three different methodologies are frequently employed to cope with noise, which include replications, increasing population size, and rescaled mutations (Beyer, 2000). This thesis focuses on the first methodology. It is predicted that techniques that unify some or all the above methodologies will be developed in the future. Statistical techniques are helpful in determining the appropriate amount of computational effort required to lessen the effect of noise on the search algorithm's performance.

There are several statistical techniques available that could be integrated into the selection mechanism of an ES. The statistical techniques studied in the research are broadly classified into four categories namely ranking and selection techniques, multiple comparison procedures, clustering procedures, and other statistical procedures. These statistical techniques vary with respect to their goals and assumptions. Specific techniques studied in this research are Dudewicz and Dalal's procedure that selects the 's' best among 'k' competing systems, Kim and Nelson's sequential procedure that selects a subset of size 's' that contains the best solution, Tukey's multiple comparison procedure, Calsinki and Corsten's Clustering procedure, and Scheffe's Procedure.

The scope of this thesis is limited to evaluating the effectiveness of the example techniques mentioned, as the selection mechanism within an ES. Experimental results suggest modified statistical selection procedures help to guide the search algorithm towards the optimum at high levels of noise. Experimental evaluations show that a statistical ranking and selection technique such as the sequential procedure by Kim and Nelson (2001) outperforms the other statistical techniques. The procedure given by Dudewicz and Dalal and the sequential procedure given by Kim and Nelson followed very closely along the optimum convergence path. However, the procedure given by Kim and Nelson achieved this close proximity to the convergence path utilizing a relatively smaller number of total simulation calls than did the procedure by Dudewicz and Dalal. Experimental results also indicated that the aid of a statistical technique is required during the later phase of the search. Sequential ranking and selection procedures based on indifference zone methodology such as the one given by Kim and Nelson (2001) are recommended (within the context of limited research conducted) since they eliminate the clearly inferior solutions and additional observations are obtained from only the competing solutions still in play. Additionally, it is recommended that the procedure be implemented at a low probability of correct selection in the range of 0.5 to 0.6, since the performance of the algorithm is not severely impacted while lowering the number of simulation calls required. Another important factor is to use an adaptive indifference zone, where the indifference zone is proportional to the distance between the solutions. In other words, a sequential statistical procedure, with low probability of correct selection, and an adaptable indifference zone is recommended. Moreover, sequential procedures are easily adaptable to simulation optimization since observations can be obtained sequentially. Incorporating the sequential statistical techniques with lower probability of correct selection and dynamically adjusting the indifference zone significantly decreased the simulation effort required without greatly compromising the quality of solutions found by the search algorithm.

Future efforts should be directed towards defining the optimal parameters, such as the initial number of replications per solution, the level of significance, and indifference zone for different statistical techniques for effective convergence with minimum computational effort. The selection methodologies proposed need to be evaluated for more test functions. Further research is required for the case of unequal variance across the population of solutions. All the statistical techniques examined in the research assume independent and normally distributed observations from each system. The sensitivity of these methods for deviations of the assumptions of independence and normality is to be examined. Further research can be conducted using other statistical techniques, which are not considered in this research. Experiments on dynamic parent population sizing methodologies combined with Tukey's procedure and Clustering procedures showed that the solutions were converging very slowly towards the converging path. Hence, a promising avenue for future research, which seems to have a great potential for improving the search algorithm, is to combine statistically based selection methodologies with dynamic parent population-sizing. Research on the relation between the proportion of correct selection and the probability of correct selection for the statistical techniques could lead to new insights that might be helpful is designing optimization methodologies to cope with noisy response surfaces. The interaction between the modified selection methodologies presented with other recombination and mutation mechanisms is another avenue for future research.

In conclusion, this research presents methodologies to help optimization algorithms such as evolution strategies to cope with problems characterized by highly noisy response surfaces. The research can be utilized to develop more efficient and effective simulation optimization methodologies, which can be incorporated into commercial simulation optimization software.

REFERENCES

Aizawa, A., and B. Wah (1993). Dynamic Control of Genetic Algorithms in a Noisy Environment. *Proceedings of the fifth International Conference on Genetic Algorithms*, 48-55.

Andradottir, S. (1998). Simulation Optimization. Chapter 9 in *Handbook of Simulation* (J. Banks, ed.). New York: John Wiley and Sons.

Arnold, D.V., and H-G. Beyer (2000). Efficiency and Mutation Strength Adaptation of the ($\mu / \mu_I , \lambda$)-ES in a Noisy Environment. *Parallel problem solving from nature-PPSN 6,* 39-48.

Arnold, D.V., and H-G. Beyer (2001). Investigation of the (μ, λ)-ES in the Presence of Noise. *Proceedings of the 2001 congress on evolutionary computation,* v1, 332-339.

Back, T. (1996). *Evolutionary Algorithms in Theory and Practice.* Oxford University press.

Back, T., and U. Hammel (1994). Evolution Strategies Applied to Perturbed Objective Functions. *Proceedings of the IEEE World Congress of Computational Intelligence*, Orlando, Florida, 40-45.

Back, T., F. Hoffmeister, and H.P. Schwefel (1991). A Survey of Evolution Strategies. *Proceedings of the 4th International Conference on Genetic Algorithms*, Ed R.K. Belew and L.B. Booker, Morgan Kaufmanns Publishers, San Mateo, CA, 2-9

Baesler, F., and J.A. Sepulveda (2000). Multi-response simulation optimization using stochastic genetic search within a goal programming framework *Proceedings of the 2000 Winter Simulation Conference*, 788-794.

Banks, J., J.S. Carson, and B.L. Nelson (1996). *Discrete-Event System Simulation.* 2nd ed., Prentice-Hall, Upper Saddle River, New Jersey.

Bautista, M.G., D.W. Smith, and R.L. Steiner (1997). A Cluster-Based Approach To Means Separation. *Journal of Agricultural, Biological, and Environmental Statistics*, Vol 2, Number 2, 179-197

Bechhofer, R., T.J. Santner, D.M. Goldsman (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. John Wiley & Sons, Inc, New York.

Beyer, H.G. (2000). Evolutionary algorithms in noisy environments: theoretical issues and guidelines for practice. *Computer methods in applied mechanics and engineering*, v 186, n2, 239-267.

Beyer, H.G., and D.V. Arnold (1999). Fitness Noise and Localization Errors of the Optimum in General Quadratic Fitness Models. *Proceedings of the Genetic and Evolutionary Computation Conference*, 817-824.

Biethahn, J., and V. Nissen (1994). Combinations of simulation and evolutionary algorithms in management science and economics. *Annals of Operations Research*, 52, 183-208.

Booker, L. (1987). Improving Search in Genetic Algorithms. Chapter 5 in *Genetic Algorithms and Simulated Annealing,* Davis, L., Ed. Pitman, London, and Morgan Kaufman: Los Altos.

Boesel, J. (1999). *Search and selection for large-scale stochastic optimization*. Ph.D. Dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Boesel, J., B.L. Nelson, and S. Kim (1999). Using Ranking and Selection to `Clean Up' After Simulation Optimization. *Technical Report, Dept. Of Industrial Engineering and Management Sciences*, Northwestern University.

Boesel, J., and B.L. Nelson (1998). Accounting for Randomness in Heuristic Simulation Optimization. *Proceedings of the 12th European Simulation Multiconference*, Society for Computer Simulation International, 634-638.

Bowden, R.O., and J.D. Hall (1998). Simulation Optimization Research and Development. *Proceedings of the 1998 Winter Simulation Conference*, Washington, D.C., 1693-1698.

Bramlette, M.F. (1991). Initialization, Mutation, and Selection Methods in Genetic Algorithms for Function Optimization. *Proceedings of the Fourth International Conference on Genetic Algorithms.* R. K. Belew and L. B. Booker, Eds. Morgan Kaufman: Los Altos, CA.

Calinski, T., and C.A. Corsten (1985). Clustering means in ANOVA by simultaneous testing. *Biometrics*. 41, 39-41.

Carroll, R.J., Gupta, S. S., and Huang, D. Y. (1975). On selection procedures for the *t* best populations and some related problems. *Commun. Stat.-Theory and Methods*, A4, 987-1008

Dudewicz, E.J., and S.R. Dalal (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya,* B 37, 28-78.

Duncan, D.B. (1955). Multiple Range and Multiple *F* Tests, *Biometrics*, 11, 1-42.

Fisher, R.A. (1951). The Design of Experiments (6th Ed.), Oliver and Boyd, London.

Fitzpatrick, J.M., and J.J. Grefenstette (1988). Genetic Algorithms in Noisy Environments. *Machine Learning*, 3, 101-120.

Gen, M., and R. Cheng (1997), *Genetic Algorithms And Engineering Design*. New York: John Wiley and Sons.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Inc. Reading, MA.

Goldsman, D., and W.S. Marshall (1999). Selection procedures with standardized time series variance estimators. *Proceedings of the 1999 Winter Simulation Conference*, 382-388.

Goldsman, D., and B.L. Nelson (1998). Statistical screening, selection, and multiple comparison procedures in computer simulation. *Proceedings of the 1998 Winter Simulation Conference*. 159-166.

Goldsman, D., and B.L. Nelson (1998). Comparing Systems via Simulation. Chapter 8 in *Handbook of Simulation* (J. Banks, ed.). New York: John Wiley and Sons.

Grant, R.S. (1998). *Evaluating the use of genetic algorithms, evolution strategies, and scatter search for performing simulation optimization*. M.S. Thesis, Department of Industrial Engineering, Mississippi State University, Mississippi.

Greffenstette, J.J. (1986). Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics SMC-16*, 1, 122-128.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, 7, 225-245.

Hall, J.D. (1997). *Investigation of a two-phased strategy for simulation optimization*. PhD Dissertation, Department of Industrial Engineering, Mississippi State University, Mississippi.

Hammel, U., and T. Back (1994). Evolution Strategies on Noisy Functions: How to Improve Convergence Properties. *Parallel problem solving from nature-PPSN, International Conference on Evolutionary Computation*, Berlin, Springer, 418-427.

Hsu, J.C., and B.L. Nelson (1998). Multiple Comparisons in the General Linear Model. *Journal of Computational and Graphical Statistics*, 7, 23-41.

Hughes, E.J. (2000). Evolutionary Algorithm with a Novel Insertion Operator For Optimizing Noisy Functions. *Proceedings of the 2000 congress on evolutionary computation,* v1, 790-797.

Hughes, E.J. (2001). Constraint Handling With Uncertain and Noisy Multi-Objective Evolution. *Proceedings of the 2001 congress on evolutionary computation,* v2, 963-970.

Kim, S., and B.L. Nelson (2001). A Fully Sequential Procedure for Indifference-Zone Selection in Simulation. *ACM Transaction On Modeling And Computer Simulation (TOMACS)*, v 11, 3, 251-273.

Keuls, M. (1952). The use of the 'Studentized Range' in Connection With an Analysis of Variance*, Euphytica*, 1, 112-122.

Koenig, L.W., and A.M. Law (1985). A Procedure for Selecting a Subset of Size *m* Containing the *l* best of *k* Independent Normal Populations. *Commun. Statist.-Simulation and Computation*, 14: 719-734 (1985).

Kumar, C., and D.B. Fogel (1999). Fitness distributions in evolutionary computation: Analysis of Noisy Functions. *Proceedings of the international society for optical engineering*, v 3722, 313-323.

Law, A.M., and W.D. Kelton (1998). *Simulation Modeling and Analysis*, 3rd ed., McGraw-Hill, New York.

Markon, S., D.V. Arnold, T. Back, T. Beielstein, and H-G, Beyer (2001). Thresholding-a Selection Operator for Noisy ES. *Proceedings of the 2001 congress on evolutionary computation,* v1, 465-472.

Marrison, C.I. and R.F. Stengel (1997). Robust Control System Design Using Random Search and Genetic Algorithms. *IEEE Transactions on Automatic Control*, Vol. 42, No.6, 835-839.

Matejcik, F.J., and B.L. Nelson (1993). Simultaneous ranking, selection and multiple comparisons for simulation. *Proceedings of the 1993 Winter Simulation Conference*, 386-392.

Michalewicz, Z. (1996). *Genetic Algorithm +Data Structures = Evolution Programs*. Third ed. Springer-Verlag: New York.

Miller, B.L. (1997). *Noise, Sampling and Genetic Algorithms*. Ph.D. Dissertation, Department of Computer Science, University of Illinois, Urbana Champaign, Illinois.

Miller, B.L., and D.E. Goldberg (1996). Genetic algorithms, selection schemes, and the varying effects of noise, *Evolutionary Computation* (2), 113-131

Miller, J.O., and B.L. Nelson, and C.H. Reilly (1998). Efficient Multinomial Selection in Simulation. *Naval Research Logistics*, 45, 459-482.

Nelson, B.L., and S. Banerjee (2001). Selecting a Good System: Procedures and Inference. *IIE Transactions*, vol.33, no.3, 149-166.

Nelson, B.L., J. Swann, D. Goldsman, and W. Song (2001). Simple Procedures for Selecting the Best-Simulated System when the Number of Alternatives is Large. *Operations Research,* vol.49, no.6, 95—963.

Nissen, V., and J. Propach (1998). Optimization with Noisy Function Evaluations. *Parallel problem solving from nature-PPSN,* 159-168.

Nissen, V., and J. Propach (1998). On the Robustness of Population-Based Versus Point-Based Optimization in the presence of Noise. *IEEE Transactions on evolutionary computation*, vol. 2,no.3, 107-119.

Olafsson, S. (1999). Iterative ranking-and-selection for large-scale optimization. *Proceedings of the 1999 Winter Simulation Conference*, 479-484.

Pitchitlamken, J., and B.L. Nelson (2001). Selection of the best procedures for optimization via simulation. Proceedings of the 2001 Winter Simulation Conference, 401-407.

Pierreval, H., and L. Tautou (1997). Using evolutionary algorithms and simulation for the optimization of manufacturing systems. *IIE Transactions*, 29, 181-189

Rana, S., L.D. Whitley, and R. Cogswell (1996). Searching in the presence of noise. *Parallel problem solving from nature-PPSN*, 198-207.

Sano, Y., and H. Kita (2000). Optimization of Noisy Fitness Functions by Means of Genetic Algorithms Using History of Search. *Parallel problem solving from nature-PPSN,* 571-580.

Scheffe, H. (1959). The Analysis of Variance (1st ed.), Wiley, New York.

Scott, A.J., and M. Knott (1974). A Cluster Analysis Method For Grouping Means in the Analysis of Variance. *Biometrics*, 30, 507-512.

Shannon, R.E. (1998). Introduction to the art and science of simulation. *Proceedings of the 1998 Winter Simulation Conference*, 7-14.

Shi, L. and S. Olafsson (1997). An integrated framework for deterministic and stochastic optimization. *Proceedings of the 1997 Winter Simulation Conference*, 358-485.

Stagge, P. (1998). Averaging Efficiently in the Presence of Noise. *Parallel problem solving from nature-PPSN,* v5, Berlin, Springer, 188-197.

Stroud, P.D. (2001). Kalman-extended genetic algorithm for search in nonstationary environments with noisy fitness evaluations. *IEEE Transactions on Evolutionary Computation*, v 5, n 1, 66-77.

Sullivan, D.W. and J.R. Wilson (1989). Restricted subset selection procedures for simulation. *Operations Research*, Vol.37, No.1, 52-70.

Swisher, J.R., and S.H. Jacobson (1999). A Survey of Ranking, Selection, and Multiple Comparison Procedures for Discrete-Event Simulation. *Proceedings of the 1999 Winter Simulation Conference*, 492-501.

Tamaki, H, and T. Arai (1997). A Genetic Algorithm Approach to Optimization Problems in an Uncertain Environment, *Proceedings of the 4 th International Conference on Neural Information Processing (ICONIP)*, 436-439.

Tanooka, K., H. Tamaki, S. Abe, and S. Kitamura (1999). A Continuous Age Model of Genetic Algorithms Applicable to Optimization Problems with Uncertainties. *Proceedings of the 1999 IEEE International Conference on Systems Man and Cybernetics*, Vol. 1, 637-642.

Tomick, J.J., S.F. Arnold, and R.R. Barton (1995). Sample size selection for improved Nelder-Mead performance. *Proceedings of the 1995 Winter Simulation Conference*, 341-345.

Tukey, J.W. (1949). Comparing Individual Means in the Analysis of Variance, *Biometrics*, 5,99-114.

Welsch, R.E. (1977), Stepwise Multiple Comparison Procedures, *Journal of the American Statistical Association*, 72, 56-575.