UNCERTAINTY BASED MULTI-OBJECTIVE OPTIMIZATION OF GROUNDWATER REMEDIATION DESIGN

BY

ABHISHEK SINGH

B. E., Birla Institute of Technology and Science, 2001

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Civil and Environmental Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2003

Urbana, Illinois

ABSTRACT

Management of groundwater contamination is a cost-intensive undertaking filled with conflicting objectives and substantial uncertainty. A critical source of this uncertainty in groundwater remediation design problems comes from the hydraulic conductivity values for the aquifer, upon which the prediction of flow and transport of contaminants are dependent. For a remediation solution to be reliable in practice it is important that it is robust over the potential error in the model predictions. This work focuses on incorporating uncertainty within a state-of-the-art multi-objective optimization approach. Previous research has shown that small amounts of sampling within a single-objective genetic algorithm can produce highly reliable solutions. However with multiple objectives the noise can interfere with the basic operations of a multi-objective solver, such as determining non-domination of individuals, diversity preservation, and elitism. This work proposes several approaches to improve the performance of noisy multi-objective solvers for groundwater remediation problems. These include a simple averaging approach, taking samples across the population (which we call *extended averaging*), and a probabilistic selection approach. All of the approaches are tested on standard multi-objective benchmark problems and a hypothetical groundwater remediation case study. These approaches strike a balance between finding the most optimal and the most reliable solution to the problem, thus giving decision makers and designers a practical and robust optimization tool.

To Nani, Papa, Mummy, and Sunayana

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Dr. Barbara Minsker for giving me this exciting and challenging project. Throughout the last two years, she has been a constant source of guidance, support, insight and inspiration, and has often had more faith in me than I myself! Words are not enough to express the deep sense of gratitude and respect I feel for my mentor.

I take this opportunity to thank Dr. David Goldberg, from the Department of General Engineering, for opening out the world of evolutionary computation to me. His insights and the many discussions I had with him on this topic, have contributed very significantly to this work.

It would have been impossible for me to finish this project without the enthusiastic support of my research group. I want to thank all the members of the Minsker Research Group, specially Dr. Felipe Espinoza, Shengquan Yan, Meghna Babbar, Eva Sinha, and Xiaolin Ren for their help and support throughout this project. I have indeed been very lucky to have met and interacted with such wonderful people.

I would also like to thank the faculty members of the Civil and Environmental Engineering Department at the University of Illinois, Urbana-Champaign, especially Prof. Valocchi for his advise on the groundwater issues related to this project; faculty members of the Civil Engineering Department at the Birla Institute of Technology & Science, Pilani (India); and all those teachers who through inspiration and guidance led me to this point in my career.

I would also like to acknowledge my funding source, the U. S. Army Research Office under Grant No. DAAD19-001-1-0025. The computational resources of NCSA were used in the later stages of this project.

Finally I would like to take this opportunity to thank my parents for their care and affection, my wife for her unflinching love and faith, my friends for always being there when I needed them most, and my Grandma, who if she were here today would have been the happiest of all, for whatever I have achieved through my work and research.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

The pure and simple truth is rarely pure and never simple.

~Oscar Wilde

Most real life problems are complex and inherently uncertain. Finding the optimal solutions to such problems can thus be a rather challenging task. The field of water resources management is one in which such complex and uncertain problems are often encountered. One approach that has been increasingly used to solve such problems in the past few years is genetic algorithms. Genetic algorithms have been applied to many water resources applications, including groundwater remediation design, optimal reservoir system operation, calibrating rainfall-runoff models, remediation policy selection, and solving multiple objective groundwater remediation and monitoring problems (e.g., Wang 1991; Ritzel et al, 1994; Wang and Zheng, 1997; Wardlaw and Sharif, 1999; Smalley et al, 2000; Reed, 2002). One of the reasons that genetic algorithms (GAs) have been chosen is that they have been shown to handle non-convex, discrete, discontinuous, noisy, and multi-objective problems that frequently arise in the field of water resources management.

In solving these problems typically a simulation model is used to represent the problem domain and test possible solutions and an optimization algorithm is used to find feasible and acceptable design solutions using the simulation model. In the field of groundwater remediation design, *Aguado et al* (1974), *Ritzel et al* (1994), *Karatzas and Pinder* (1996), *McKinney and Lin* (1996), *Rizzo and Dougherty* (1996), *Culver and Shoemaker* (1997), , *Minsker and Shoemaker* (1998), *Yu et al* (1998),*Zheng and Wang* (1999), *Chan Hilton and Culver* (2000), *Erikson, et al (2002), Maskey et al* (2002), are among the many who have used optimization methods such as nonlinear and linear programming, optimal control, genetic algorithms, simulated annealing and cutting plane methods.

Recently the Department of Defense Environmental Security Technology Certification Program (ESTCP) undertook a project investigating the use of such state-of-the-art optimization methods to field-scale groundwater remediation design at three field sites. Three teams were formed, one to investigate a trial-and-error based solution to the problem (in which the simulation model is run by hand to experiment with alternative solutions) and the others to investigate the use of automated global optimization methods like genetic algorithms, simulated annealing, and tabu search. Three single-objective optimization formulations were created for each site, which were then solved using these approaches. Results showed that the use of such optimization algorithms resulted in significant improvements over the trial-and-error results. See ESTCP (2003) for more details on the formulation and results of this project.

While both the fields of multiobjective and “noisy” (uncertain) optimization have recently been the focus of intensive research in the evolutionary computation field, few researchers have considered combining these two approaches. Some initial work in this area was undertaken by *Hughes* (2001) and *Teich* (2001) who used estimates of the uncertainty distribution to deal with noise (uncertainty) within a multiobjective genetic algorithm. Another study was undertaken by *Chan Hilton and Culver* (2000) on groundwater remediation design in problems with uncertain aquifer properties. In this study we develop several new noisy multiobjective optimization approaches and then use them to investigate the effects of uncertainty and multiple objectives on a hypothetical groundwater remediation problem.

The paper is organized as follows. First a brief background is given on relevant research for the noisy GA and the multi-objective GA. Next two previously developed methodologies for combining uncertainty with multiple objectives are discussed in detail - a simple averaging-based approach and a more sophisticated probabilistic dominance approach. Several enhancements to these approaches are then proposed and tested on a multi-objective benchmark problem and a hypothetical groundwater remediation test case. Finally we present our conclusions with a brief overview of the salient features and findings of this research.

2. BACKGROUND

The two basic approaches that need to be understood before we delve deeper into the subject of uncertainty-based multi-objective evolutionary optimization are the basics of simple genetic algorithms, noisy evolutionary optimization (in particular noisy GAs) and multi-objective evolutionary algorithms. The motivation of this work is to combine and test various elements from these two approaches.

2.1 Simple Genetic Algorithms

GAs are search algorithms based on the Darwinian theory of ‘survival of the fittest’. The search is usually done with a group or population of chromosomes or strings (candidate solutions) that are (usually binary) encoded forms of the decision variables. After initializing the population with a random set of such strings three basic operators are used in the simple GA to evolve these strings to better and better performance, where performance is measured in terms of the objective function and constraints of the problem (the “fitness function”). The evolutionary process is modeled after the process of natural selection using three basic operations: selection, crossover, and mutation. First, the fitness function is evaluated for each string. Selection occurs when the individuals with higher fitness values are assigned higher probabilities of producing offspring for the next generation. Crossover occurs when selected individuals are mated by randomly dividing their two chromosomes in one or more locations and then exchanging the pieces.

After crossover, mutation randomly switches a bit or bits in the chromosome with a specified probability.

The way the GA assembles optimal solutions can be described by building block theory. Building blocks are short, low order, and highly fit groups of binary digits in the chromosome (Goldberg, 1989). More intuitively building blocks can be thought of as the implicit sub-solutions to a given problem. GAs are more efficient than random searches due to the effects of the operations of selection, crossover and mutation, which search the decision space by decomposing the problem and then finding the set of optimum building blocks, in parallel across the population. These building blocks are subsequently combined (mixed) through crossover to give the final optimal solution. In most real cases there is no way of explicitly identifying such building blocks, but their search and evolution is intrinsically noisy and is handled through sampling of sub-solutions by adequate population sizing (Harik, et al 1997).

2.2 Noisy Genetic Algorithms

A GA that operates in a noisy or uncertain environment is referred to as a “Noisy GA”. The “noise” which exists in certain environments is defined as any factor that hinders the accurate evaluation and assessment of a given trial design. These factors can include the use of approximate fitness functions, the use of noisy data, knowledge uncertainty, modeling and numerical approximations, sampling and human error. Much of the theory and practice in this area can be understood if we remember that even in non-noisy domains genetic algorithms are stochastic algorithms, which operate under a certain amount of inherent ‘noise’. In the case of noisy fitness functions, the additional noise adds to the intrinsic (building block) noise described in the previous section and can be dealt with by using larger population sizes or adequate fitness sampling. Fitness sampling is done by drawing samples of each solution’s fitness from an appropriate probability distribution of the noise. In most cases the sampling is used to find the average values for the solutions performance on a given objective, which through the central limit theorem would increasingly tend to the actual mean of the noise distribution with increased sampling (the standard deviation of the mean is reduced by a factor of $\sqrt{n}$ for a sample size of n, Miller and Goldberg, 1996).

Miller and Goldberg (1996) were among the first to analyze and suggest design methodologies for single-objective noisy genetic algorithms. *Miller* (1997) suggested an adequate population size and optimal sampling size for single-objective noisy problems. His methodology was tested on a groundwater remediation design problem by *Gopalakrishnan, et al* (2003). They found that highly reliable solutions could be obtained for the problem considered with low amounts of sampling (as few as 5 samples per solution). This capability is critical in cases such as ours where it is impossible to afford large sample sizes given the expense of fitness evaluations. More recently, *Peralta and Kalwij* (2003) applied a similar approach for the Umatilla Chemical Depot site, varying the hydraulic conductivity values by a fixed percentage.

Another relevant study was undertaken by *Chan Hilton and Culver* (2000) to find robust solutions to groundwater remediation problems with uncertain aquifer properties using the past performance of a particular solution to assess its robustness over uncertainty. In this case, one of the objectives – reliability – was defined in terms of violation of an environmental standard. In our case this constraint is not predetermined allowing decision makers to decide the optimum environmental quality standard (in terms of concentrations or risks) from the final set of optimal solutions (Pareto front) by choosing a *reliable* level with good cost and cleanup tradeoffs.

Another important approach used in this study is the explicitly stochastic scheme suggested by *Hughes* (2001), called the probabilistic selection algorithm. This method works by estimating the noise distribution parameters of each individual and then using this information for the selection of individuals. The main advantage of this approach is the reduction in the noise during selection, which leads to better convergence. This technique, which is used specifically in the context of a multi-objective problem, is discussed in more detail in subsequent sections.

2.3 Multi-Objective Genetic Algorithms

Most problems in the real world have several (possibly conflicting) objectives to be satisfied. Traditionally such problems have been handled by converting the multiple objectives to one using weighting functions or using one objective to optimize and the others as constraints. Such an approach has many problems, including the loss of significant tradeoff information and the inability to search the true objective space for a global optimum. Changing the weights or the constraints leads to convergence to a different optimum point on the Pareto front and thus the algorithm must be rerun many times whenever trade-off based decision making needs to be done.

An alternative population-based approach is to simultaneously find the set of solutions that represent the optimal tradeoffs for the solution (globally non-dominated solutions). These comprise the non-dominated frontier (or Pareto front) for the problem and the objective now becomes to find solutions to the problem such that there are no feasible solutions that would be better on one criterion without simultaneously worsening at least one other criterion. This set of non-dominated solutions is referred to as the Pareto optimal set, after the famous mathematician Vilfredo Pareto, who generalized the concept of this kind of optimality. The Pareto front thus generated gives us valuable information about the various tradeoffs among the objectives that can and should be taken into consideration when making multi-criterion policy decisions.

Thus, unlike single objective optimization, multi-objective optimization (MO) has to work to find an entire set of good solutions. Since genetic algorithms manipulate populations of candidate solutions, they are a natural approach for this kind of optimization. Such evolutionary multi-objective (EMO) methods were introduced through the seminal work by *Schaffer* (1984). More pioneering work in this area was undertaken by *Fonseca and Fleming* (1993), *Srinivas and Deb* (1994), *Coello* (1999), and *Van Veldhuizen, et al* (1999) among others. *Cieniawski* (1993), *Ritzel et al* (1994) were among the earliest applications of EMO methods in water resource management. More recently *Reed et al* (2001, 2003, in press) developed guidelines for competent EMO and applied them to a groundwater-monitoring problem to find the optimal tradeoffs among the cost and errors in a monitoring system.

Multi-objective optimizers typically use different schemes to give a measure of local (within the existing set of solutions in the generation) 'non-dominance' to the solutions. Some methods, like those proposed by *Fonesca and Fleming* (1993), rank each individual as one more than the total number of individuals it dominates. Others, like the non-dominated sorted genetic algorithm (NSGA) proposed by *Srinivas and Deb* (1994), do layer-wise Pareto ranking. Here each individual is compared to all others in the population and those that are strictly non-dominated are marked as the locally Pareto (or rank 1) solutions. These are then removed from the list of considered solutions; rank 2 solutions are defined as those that are strictly non-dominant with respect to the remaining solutions and so on until all solutions in the population are ranked.

This layered ranking approach is shown in Figure 1 for the minimization of two conflicting objectives: cost and human health risk for a groundwater remediation design problem. In general, the objective of the EMO now becomes to minimize the rank. As a starting point for our analysis and implementation we use the elitist non-dominant sorting algorithms based on crowding distance (NSGA-II) proposed by *Deb, et al* (2000), which is an efficient algorithm that gives good results over a large range of problem types.

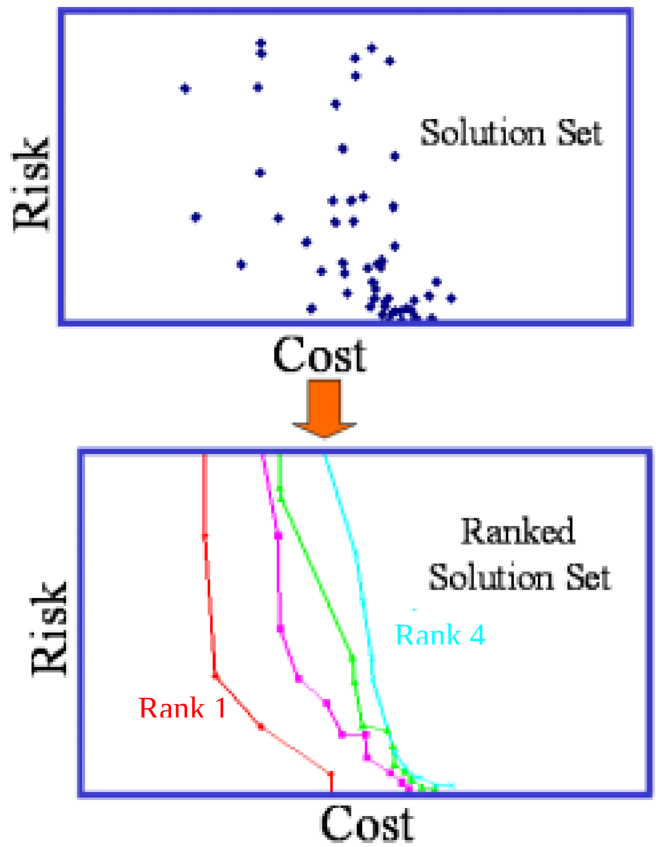

Figure 1. Layered ranking approach used in the non-dominated sorting algorithm.

Another issue that must be considered in multiobjective optimization with genetic algorithms is loss of diversity (convergence to one solution on the Pareto front instead of the entire set). To maintain good spread over the Pareto front and prevent loss of diversity, various methods can be used like fitness sharing, mating restrictions, and crowding distance among others. The basic motivation for all of these methods is to give preference to individuals in less densely represented areas of the Pareto front, and less preference to those that are in crowded areas. A good description of these approaches can be found in *Laumanns*, *et al* (2001). In the NSGA-II scheme, the crowding distance is used as a measure of a solution's uniqueness. The crowding distance for a particular individual is defined as the cumulative distance between its neighbors summed over all dimensions.

In addition to preservation of diversity, most EMO approaches have some type of elitist scheme so that the existing Pareto optimal solutions are not lost. In NSGA-II elitism is implemented by always using the $\mu + \lambda$ selection scheme. In this scheme the μ parents are combined with the λ children and the best (most non-dominant) individuals are put into the next generation. Thus typically all rank one solutions survive to the next generation; however if in the combined population the number of rank one solutions are more than μ then these individuals are ranked based on the crowding distance factor and the μ solutions with maximum distance (i.e. the μ most unique individuals) are selected.

3. TEST CASES

To test the various algorithms developed in this work a rigorous set of test problems was chosen. First standard test functions were taken from the multi-objective literature and noise was added to them to model noisy objective functions. These test cases helped in rapidly testing some of the proposed methods and quickly eliminating those that did not show much promise. To see how the remaining methods performed on more realistic problems these were then tested on a more intensive hypothetical groundwater remediation problem..

3.1 Multi-objective Standard Test Case

The algorithms were tested on a set of multi-objective problems taken from *Deb* (1998) to allow extensive and rapid testing on different types of problems before a water resources application was undertaken. Since groundwater remediation problems are typically multi-modal and discontinuous, only results on a surrogate mathematical objective function with similar properties are shown in subsequent sections. This function has the following two objectives, is multi-modal, and is discontinuous in objective space:

$$Min(g_1) = x \tag{1}$$

$$Min(g_2) = 1 - (\mathrm{g}_1/(1+10\mathrm{y}))^2 - (\mathrm{g}_1/(1+10\mathrm{y})) \times (\sin(2\Pi.4\mathrm{g}_1)) \tag{2}$$

where g_1 and g_2 are the two objectives of the standard test function and x and y are the two decision variables, which lie between 0 and 1.

Noise was considered in objective, decision, and parameter space. Previous work on population sizing (*Mahfoud*, 1995) was used to derive conservative population sizes (200) and the GA was run for 60 generations to ensure convergence. This is almost 4 times the string length, while normally convergence is expected in *2l* generations, where *l* is the string length. An artificial noise signal with a standard deviation of 10% of the range of objective values g_1 and g_2 was added to the each of these objectives. Sample sizes of 5, 10, and 20 were used for sampling fitness with respect to each of these objective functions.

3.2 Groundwater Remediation Test Case

This risk-based remediation design case study has been previously developed using data derived from the Borden site as detailed by *Smalley, et al* (2000). The aquifer was modeled using a coarse grid of 16 by 8 elements. This coarse grid was derived from a finer mesh that was used to generate conditional hydraulic conductivity realizations as described by *Feyen et al (2000)*. Benzene with an initial peak concentration of 133 mg/L is assumed to be present at the site. Multiple parameter sets are defined, with each set consisting of a single sample drawn randomly from the pool of generated hydraulic conductivity realizations and from each of nine variable human exposure parameter distributions. It is worth mentioning here that, unlike the benchmark multi-objective problem where all solutions have the same uncertainty, the uncertainty of a particular design solution can vary substantially from one design to another.

The remediation strategy is to use extraction wells from two possible sets of well locations to clean up the aquifer. A risk-based remediation plan was developed for the case study using a modified form of the risk management model developed by *Smalley et al (2000)*. The management model combines a genetic algorithm with a fate and transport simulation model and a risk assessment module to identify promising remediation designs. Pump and treat technology is used to treat the groundwater.

The objectives of the optimization are to minimize the total cost for the remediation system and minimize the risk from human exposure to the contaminants through household groundwater use. The cost consists of the capital and operating costs for the wells, the cost of on-site monitoring, and the additional capital and operating costs for the ex-situ treatment system. The risk is calculated from an assessment module, which uses an analytical model to predict contaminant concentrations at off-site exposure wells given concentrations in the source area, and estimates human health risk associated with the predicted concentrations. Existing simulation models MODFLOW (*McDonald and Harbaugh*, 1996) and RT3D (see *Clement et al, 1997* for details) are used to simulate contaminant concentration in the source area for each candidate solution. For further details, see *Smalley, et al* (2000) and *Gopalakrishnan, et al* (2003).

4. METHODOLOGY

In this section we present the various approaches we used to investigate the problem of noisy multi-objective domains. First, the simple averaging approach, which has already been successfully used for single objective problems (Gopalakrishnan, et al 2003), is presented. Within a multiobjective framework, this approach suffers from undesirable bias in the final population due to limited fitness sampling. An approach, called 'e*xtended averaging*', that tackles this problem without any additional sampling is then proposed. Finally an explicitly stochastic technique based on *Hughes* (2001) multi-objective probabilistic selection algorithm (MOPSEA) is discussed. A new enhancement to this approach is proposed to make it more robust to errors in estimation of noise characteristics. The results and improvements for each of these techniques are presented in section 4.

4.1 Simple Averaging-Based NSGA-II

The most straightforward approach for handling noise in multi-objective domains is to use the same approach as the single objective noisy genetic algorithm (Miller, 1997). In this scheme each individual is tested over *n* samples for each objective and the average value of each objective is used to optimize the problem using an efficient multi-objective genetic algorithm, like the NSGA-II.

There are a few issues that need to be addressed with this approach. First, the population sizing and optimal sample size are only available for single objective problems (see Miller, 1997 and Gopalakrishnan et al, 2003). At the writing of this thesis there were no significant studies published to obtain similar formulae for multiobjective problems. Secondly, in the case of reliably multi-objective optimization, since the solution space is often unconstrained with respect to the decision variables the concept of reliability cannot be extended to test individuals for their 'goodness'. For example, in the case of groundwater remediation design, if the first objective is cost and the second objective is some measure of the performance of the design (e.g. minimize risk), a range of performance levels are desired and there is no measure of 'reliability' in a design as there is no fixed performance standard. Instead it is desirable that solutions obtained have low uncertainty (if the noise is fitness dependent) and have estimated performance close to their true performance, so that they are reliable for decision making. Thus in addition to optimization based on the average fitness values, the standard deviation of each individual should also play a part in the optimization process. The following methods seek to address some of these issues.

4.2 Extended Averaging

To address bias problems seen the averaging based approach due to the small sample size, we utilize the population and convergence characteristics of the GA – to create a new approach called *extended averaging.* At any given time there can be many copies of the same individual within the population. If we extend our averaging to consider all identical solutions in the present and future generation, thus *averaging over all the samples of identical individuals* we get a much larger effective sample size for that individual. Moreover, as the GA converges the stable solutions in the Pareto front have more and more copies in future generation and are thus sampled more and more. By including all of these samples in the estimates of the objective value we get more and more reliability as the GA converges.

We implement this method by comparing each individual that is evaluated with all of the previous parent and children individuals. All identical solutions are marked. Then the average for *all of these identical* individuals are updated using the sampling from the new individual. This process is repeated over all generations. This approach can easily be extended to consider more previous generations, ostensibly giving more reliability in the estimates. However, this would require a storing a substantial database of solutions for problems with large populations, so in this work we examine the performance of this approach using only the parent and child population for the extended averaging, which requires no maintenance of the history of different solutions.

4.3 Probabilistic NSGA-II

The final approach we examine in this paper incorporates both the expected fitness of individuals and their standard deviation. This approach proposed by *Hughes* (2001), is called the multiobjective probabilistic algorithm. In this approach, instead of deciding which individual is better than the other based only on their average sampled value, a probabilistic model is used for selection. If we assume that each fitness value (in a multiobjective domain) belongs to a normal distribution, then the difference between these fitness values (convolution) is also a normally distributed random variable, such that

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n}+\frac{{}_2^2}{m}}} \qquad (5)$$

is a normally distributed random variable with mean 0, and standard deviation 1. Here, $\overline{X}$, $\overline{Y}$ and S_1, S_2 are the estimated means and standard deviations of the samples after averaging over *n* and *m* samples, respectively while μ_1 and μ_2 and are the true means.

Using this basic formula, *Hughes* (2001) derived an equation to find the probability of the i^{th} objective value for one individual (A) being less than the i^{th} objective value for another individual (B). This is given by

$$\mathrm{P}(A_i < B_i) = 1 - \frac{1}{\sqrt{2\pi(r^2+1)}}\int_0^{\infty}\exp\left(\frac{-(x-m)^2}{2(r^2+1)}\right)dx \qquad (6)$$

where σ_a and σ_b are the true standard deviations, and μ_1 and μ_2 the true means of A and B, $m = (\mu_a-\mu_b)/\sigma_a$ and $r = \sigma_a/\sigma_{b.}$

The probability of one individual (A) dominating another (B) over all *n* objectives (if all objectives are posed as minimization problems) is given by

$$P(A\ Dom\ B) = \Pi_{i=1,n}\ P(A_i < B_i) \tag{7}$$

However these equations assume that the true mean and standard deviations are either known or can be estimated accurately and without bias. This is usually not the case with real world optimization applications, where the uncertainty is unknown and there are severe restrictions in terms of sample size leading to biased estimates of the mean and the standard deviation. We propose to handle this uncertainty in the estimates by performing a hypothesis test using the *T* distribution. For a given measurement of mean and standard deviation we know that

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{}{n+m-2}\left(\frac{}{n} + \frac{}{m}\right)}} \tag{8}$$

has a T distribution with n+m-2 degrees of freedom, where n and m are the number of samples considered for each individual to estimate the mean and standard deviations. Using this estimate, we can bound the difference between the measured mean values with a given level of confidence given by:

$$\left[\left(\bar{X} - \bar{Y}\right) - bR < \mu_1 - \mu_2 < \left(\bar{X} - \bar{Y}\right) + bR\right] = \alpha \tag{9}$$

where α is the percentage confidence level, b is found from *T* distribution tables such that $P(-b < T < b) = \alpha$ with degrees of freedom n+m-2, and R is the denominator of (8).

This then becomes a significance test for the null hypothesis H: $\mu_1 = \mu_2$. If the range of $\mu_1 - \mu_2$ given by (7) is less than the lower confidence bound (-*bR* from equation 7) then the fitness of individual 1 is less than 2 with confidence α; if this range is more than the upper confidence bound (+*bR* from equation 7) then the fitness of individual 1 is greater than 2 with confidence α. In both cases the null hypothesis is rejected and we can decide statistically which of the two values dominate the other. If the range is between the upper and lower confidence bounds then we cannot reject the null hypothesis and must conclude that the two samples are statistically similar. In this case *we prefer the individual with lower standard deviation*, thus ensuring that more reliable (less uncertain) solutions are preferred during selection.

Using MOPSEA as a basis, the above new method was implemented in NSGA-II to decide the dominance of one individual over the other. The solutions that were most likely to be non-dominated by all others go into rank one. Those that are most likely to be dominated by all except 'rank one' solutions go into rank two, and so on until the entire population is ranked probabilistically.

4.4 Performance Criteria

To compare performances of these approaches, performance criteria are needed. In noisy single objective problems, the quality of the solution has previously been measured by performing numerous Monte Carlo simulations on the final solution (Gopalakrishnan et al, 2003). In their constraint-based approach, all but one of the objectives are posed as constraints and the Monte Carlo samples are tested on this constraint to determine how many of the solutions satisfy the constraint. The percentage satisfying the constraint is the reliability of the solution. However, in multi-objective domains the problem is unconstrained in objective space, leading to an entire Pareto front of possible solutions. The question then becomes how to test the reliability of such solutions.

The approach that we follow in this work is to conduct Monte Carlo simulations using Latin Hypercube sampling of the performance of each solution on the final Pareto frontier to better estimate the true average as well as upper and lower confidence intervals. Upon establishing such bounds we can then compare the final Pareto frontier to see where the solutions obtained from the GA (with limited sampling) lie in relation to the upper and lower confidence regions. The closer the GA solution are to the true performance (in terms of the average obtained from full Monte Carlo analysis) the better. Solutions with this property can be thought of as solutions with *low bias* in the uncertain domain. Moreover, in problems such as ours where the uncertainty is dependent on the characteristics of the individual (unlike a constant uncertainty factor due to, say, measurement error) it is also desirable that the final solutions have as little uncertainty as possible, since less uncertainty corresponds to higher reliability of solutions. Solutions with less uncertainty would be those with *low standard deviations* in the final Monte Carlo simulation. Thus along with Pareto optimality we also desire solutions with low bias and low deviations in our final front.

5. RESULTS

This section presents the results of the various approaches discussed in the methodology section. For each of the approaches the results are shown for the benchmark multi-objective problem and the hypothetical groundwater remediation problem.

5.1 Results with the Simple Averaging Approach

The first approach examined is the simple averaging approach in which a candidate solution is sampled a number of times to get different estimates of the objective values under the effect of uncertainty. For the standard multi-objective problems this consisted of adding a scaled random Gaussian signal to the final fitness values. For the groundwater remediation test problems each candidate design is tested on various input parameter (e.g. hydraulic conductivity) realizations and the resulting fitness values for each realization are averaged over all of these. The average fitness values are then used within the multi-objective GA as discussed previously.

The results for the standard test case described in section 3.1 are shown below (Figure 2). The triangles show the final population from NSGA-II when a sampling size of 5 was used, while the crosses, blocks, and diamonds show the average, 5th, and 95th percentiles obtained from an offline Monte Carlo analysis on the final population with sample size 500.

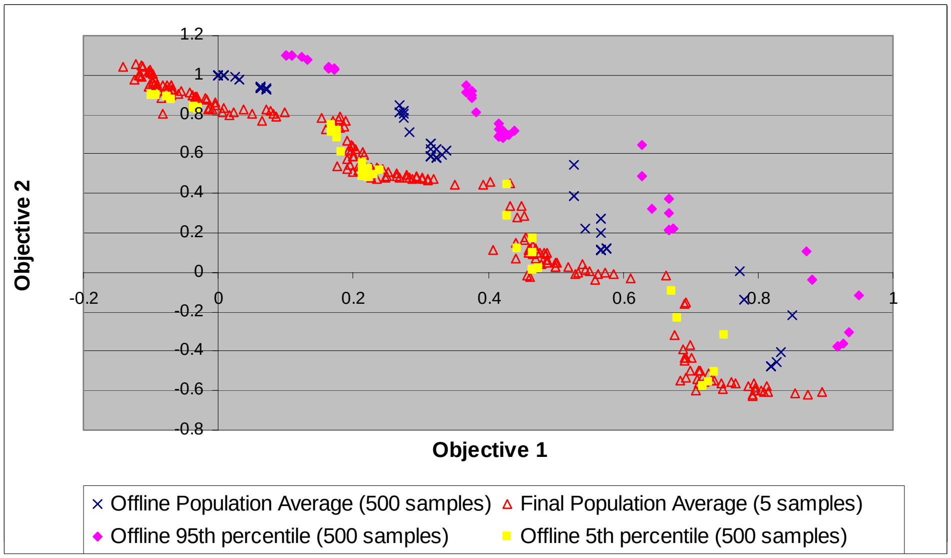

Figure 2. Results of simple averaging-based scheme on multi-modal, discontinuous multi-objective test function

Several observations can be made from the results shown in Figure 2. First, the Pareto front as generated from the final generation of NSGA-II averaged over 5 samples is very different from the actual Pareto front obtained from Monte Carlo analysis of the same solutions. Because of the small sample sizes used during the GA, solutions with biased sampling in the lower portions of their distributions tend to survive and take over the populations. This effect is especially enhanced due to the elitist $\mu + \lambda$ selection scheme being used by the NSGA-II, which causes solutions with low objective values due to biased sampling to always dominate other solutions and thus survive into future generations. One alternative that we could use to tackle this problem is not to use the $\mu + \lambda$ selection scheme, removing the source of elitism within the NSGA-II. However, this causes the Pareto frontier to become very unstable since every generation the good individuals in the parent population are replaced by their children, thus losing the existing Pareto front. Initial testing with this approach confirmed that it led to a very sparse final Pareto front as shown in Figure 3, thus this approach was abandoned for future testing. The other alternative is to increase the sample size so that the average values used in the GA are closer to the true average value. However increased sample sizes mean increased computation, which will be too costly for the much larger and time-consuming groundwater remediation problems.

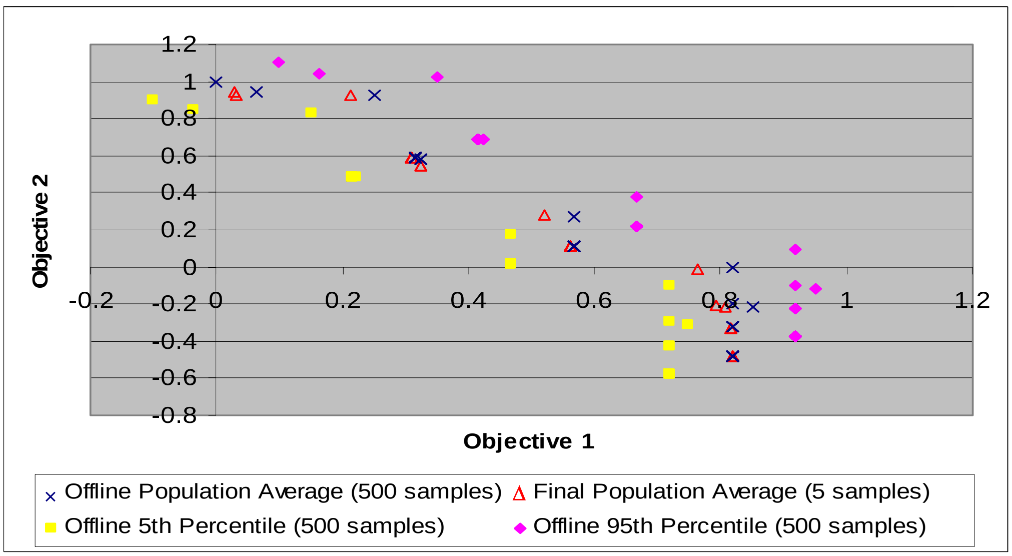

Figure 3. Results of simple averaging-based scheme on standard test function without the elitist $\mu+\lambda$ *selection scheme*

The experiments described above were repeated with the hypothetical groundwater remediation problem based on the Borden aquifer. The results for the averaging based NSGA-II with elitist $\mu+\lambda$ selection are shown below in Figures 4 and 5. Since the Borden aquifer is a sandy aquifer this test case was seen to be rather insensitive to the effects of uncertainty in hydraulic conductivity values. Hence, even with a small sample size of 5, the results from the NSGA-II were much closer to the true average when compared to the results on the standard multi-objective problems. However if one looks at the low risk solutions with higher uncertainty (shown in Figure 5), the same trend of bias and uncertainty observed with the standard test function is seen. The solutions are close to the 5th percentile and distant from the true average obtained from offline analysis.

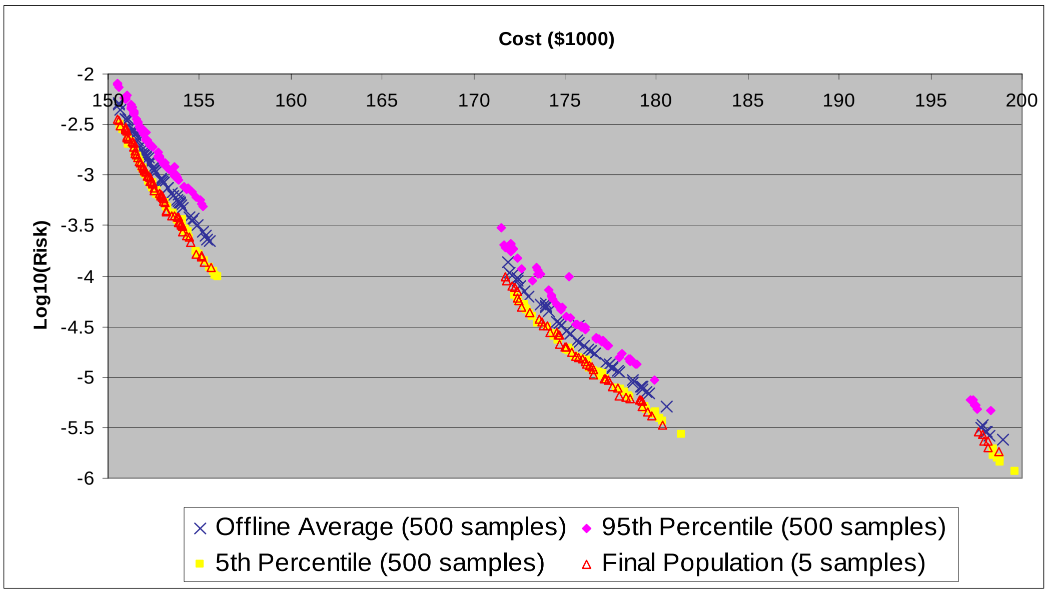

Figure 4. Results of simple averaging-based scheme on the hypothetical groundwater remediation problem

If we see these results from a groundwater remediation point of view we realize that the solutions obtained from the simple averaging based NSGA-II are non-conservative for decision-making. The actual performance of the solutions (as shown by the blocks) is *worse* than the predicted performance (shown by the crosses). Thus there is a need to improve the decision making process to improve the performance of the GA. Offline

Monte Carlo analysis as performed here can be conducted to find the true bounds on the solutions, however this can prove to be a very expensive affair for the entire Pareto front. More importantly, within the averaging-based GA framework solutions that would perform well on the offline Monte Carlo analysis are clearly not getting selected. It is desirable to obtain solutions with good reliability using only the limited sampling possible within the optimization so that the decision maker can choose just a few optimal solutions from these and conduct Monte Carlo analysis to confirm reliability. This is where the scheme of *Extended Averaging* discussed earlier can prove to be very useful. Next we test the other methods proposed on the same test cases. However to improve clarity of results for the hypothetical groundwater remediation test case, we will subsequently show only the results of the low risk and high uncertainty portion of the Pareto front as shown in Figure 5.

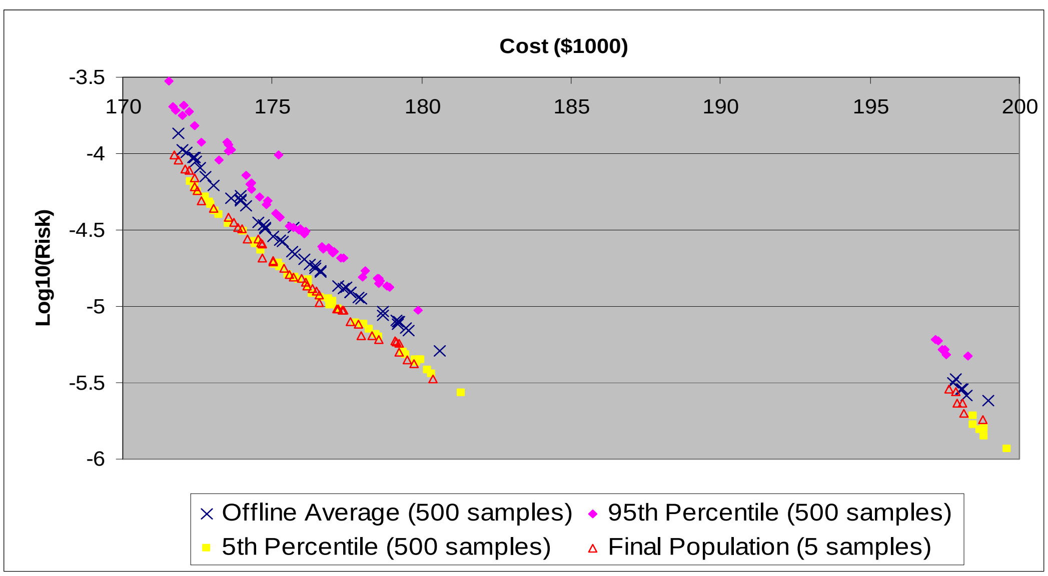

Figure 5. Solutions (from the previous graph) with low risk and high uncertainty

5.2 Results with the Extended Averaging Approach

Before presenting the results of the extended averaging approach on the test problems, we first present the results of a simple experiment on the standard multiobjective test problem to determine how much additional sampling we gain from this approach. For each individual the effective sample size was recorded and then an average taken for all individuals over the entire population for each generation, shown in Figure 6. As can be seen for the standard multiobjective problem, the effective sample size increases from around 5 in the first generation almost ten fold in the final generations. The actual increase in the effective sample size is dependent on the speed of convergence to the Pareto front and population size being used. Recall that the sample size (and thus the total fitness evaluations) for this approach is *the same* as that for the simple averaging approach.

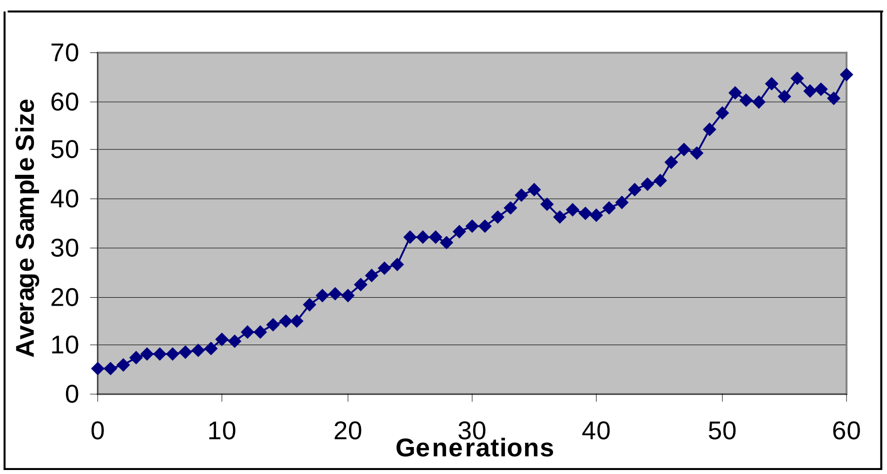

Figure 6. Growth of effective sample size with generations

The results for the extended averaging approach on the test problems are shown in Figures 7 and 8. As can be seen from the figures the performance of the final population shown by the triangles is closer to the actual offline average. Thus using extended averaging reduces the bias in our GA results in some cases with little additional computational effort.

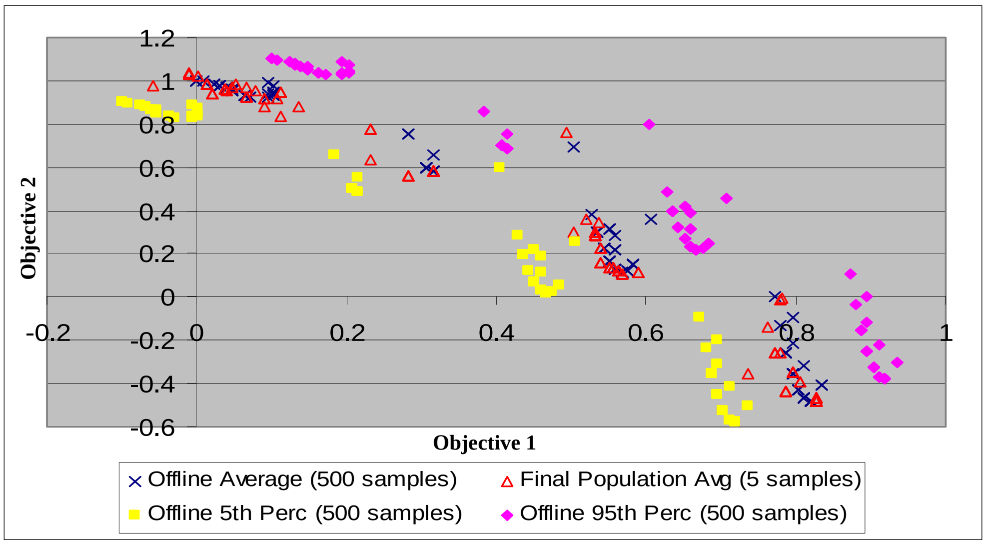

Figure 7. Results of extended averaging on multi-objective test function

However in this case we see that there are certain areas where biased solutions still exist. This is due to new individuals in the later generations that have not been sampled as much as the other reliable solutions. Rejecting these solutions would lead to loss of prospective good solutions; we thus need a way of comparing solutions with large effective sample sizes (and higher reliability) to those with smaller samples (low reliability). Moreover, this approach is driven by the way the GA converges. In the groundwater problem, since most of the low-risk high-uncertainty solutions are found later in the runs, these do not gain from the extended averaging approach and are still far from the expected performance as seen in Figure 8. To explicitly incorporate the effects of uncertainty and make the decision making process within the GA more considerate of the uncertainty in the system, we next examine the probabilistic dominance approach with extended averaging.

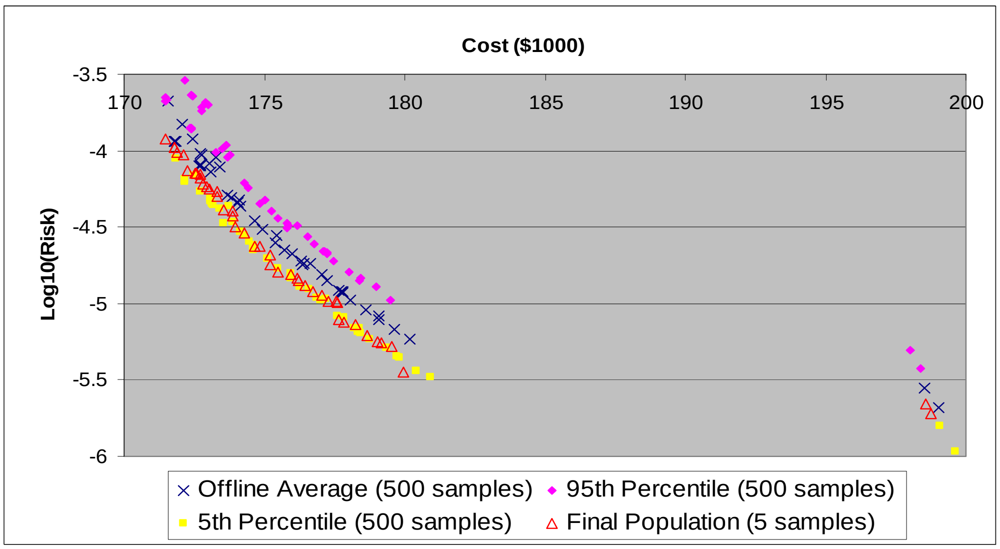

Figure 8. Results of extended averaging on the hypothetical groundwater remediation problem

5.3 Results with the Probabilistic NSGA-II and Extended Averaging

The results for the probabilistic NSGA-II with extended averaging on the test problems are shown in Figures 9 and 10. When compared to the previous best results for extended averaging (Figures 7 and 8) it can be seen that the final population has a predicted performance much closer to the actual average shown by the crosses. Moreover, there are almost no outliers and all parts of the final Pareto front are well represented. Thus using the probabilistic approach has made the decision-making within the GA more robust.

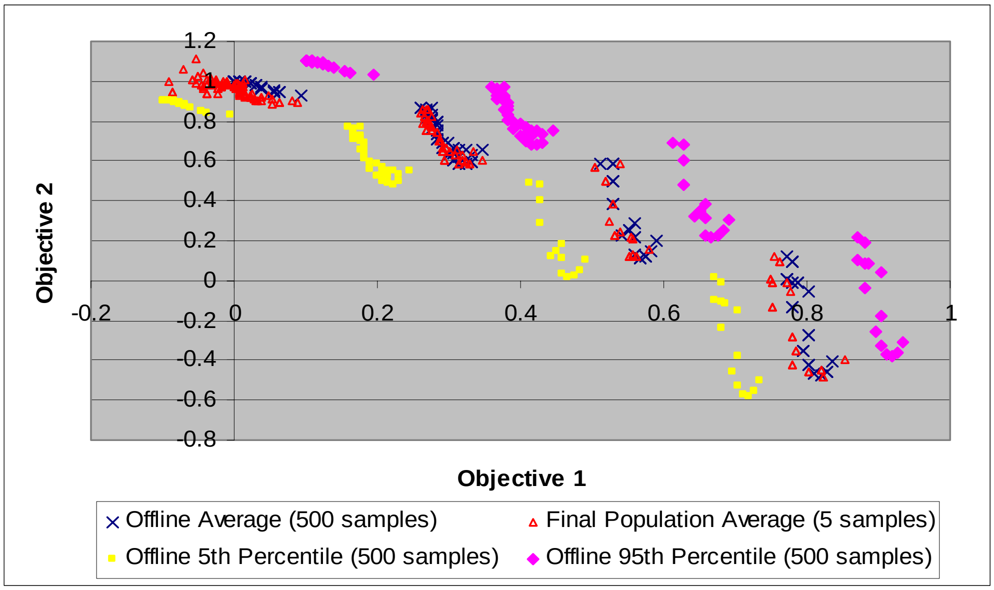

Figure 9. Results of the probabilistic approach on multi-objective test function

This approach also explicitly favors solutions with lower uncertainty during selection (see section 4.2 for details) leading to more reliable results. This effect cannot be seen in the standard test function results, since all solutions had the same artificial uncertainty, but can be observed in the hypothetical groundwater remediation test case. When the results with this approach (Figure 10) are compared with the previous results with simple averaging (Figure 5) there was a decrease of about 7% in the standard deviation averaged over all the solutions in the Pareto front, indicating that the probabilistic approach combined with extended averaging leads to more robust solutions.

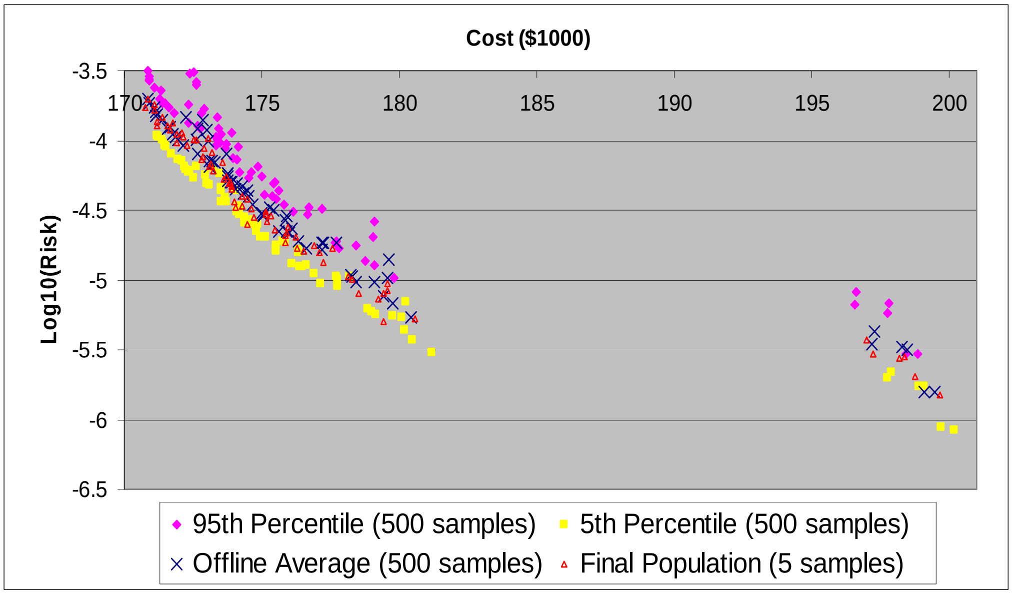

Figure 10. Results of the probabilistic approach on the groundwater problem

For further validation, the results from this multi-objective approach were compared with the results from an averaging-based noisy single-objective genetic algorithm for the same problem, which has been demonstrated to give reliable solutions with adequate sampling (*Gopalakrishnan, et al*, 2003). A comparison of the optimal multi-objective results (after line Monte-Carlo analysis) with the single objective case (also after offline Monte-Carlo analysis) is shown in Figure 11. As can be seen the optimal single-objective results are on the same Pareto-front as the multiobjective results. When the decision variables of the optimal solutions were compared it was found that the optimal well locations and pumping rates found by the single-objective approach were almost identical to that of the optimal well-locations and pumping rates found by the multi-objective approach. These findings indicate that the multi-objective approach has found the true Pareto optimal solutions.

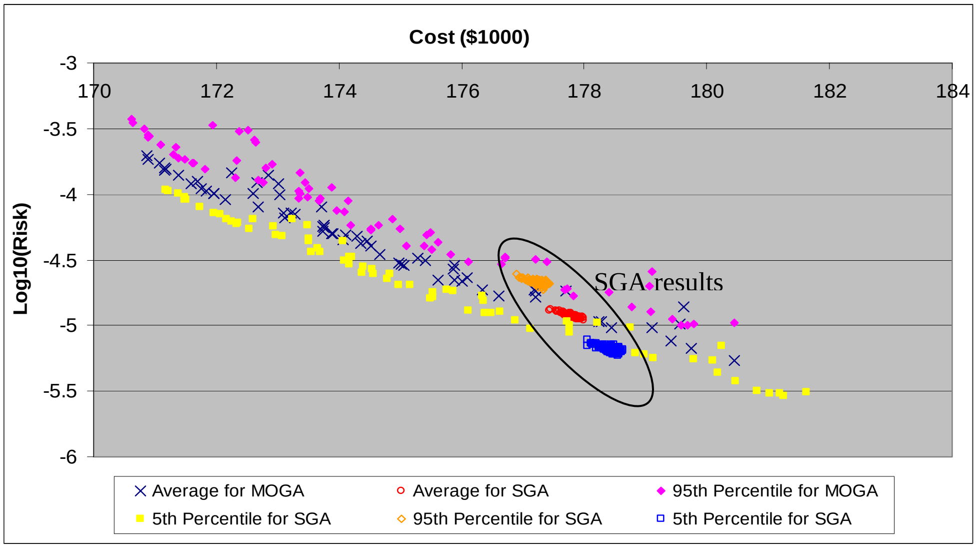

Figure 11. Comparison of Results of the Multi-objective probabilistic approach with the Noisy single-objective approach for the groundwater problem

6. CONCLUSIONS

This work investigates several methodologies for effectively solving multi-objective problems in uncertain domains that are commonly encountered in environmental and water resources management problems. Sources of such uncertainty can include the use of approximate fitness functions, the use of noisy data, knowledge uncertainty, numerical modeling approximations, and sampling or human error.

The first approach combines simple averaging scheme, which was effective for single objective problems, with a state of the art multi-objective genetic algorithm like the NSGA-II. This approach leads to bias in the final result, primarily due to the small sample size that most practitioners must use when evaluating uncertainty in real life computationally intensive problems. The second approach proposes extending the averaging across all identical members of the child and parent populations. This approach reduces the bias but some of the final designs still had poor performance in the offline analysis. The third approach combines extended averaging with probabilistic selection, to help make more robust decisions during the run of the GA so that potentially good solutions are not lost and the low uncertainty (high reliability) solutions persist. This approach gave the best results, producing reliable designs with offline performance as predicted within the GA. These results were verified on a standard multi-objective problem and the hypothetical groundwater remediation test case. It was seen that for the groundwater case where the noise is non-uniform throughout the Pareto front, using the probabilistic approach with extended averaging not only led to lower bias in the decision making but also a decrease in the average standard deviations of the final Pareto front.

The probabilistic NSGA-II with extended averaging has given very promising results on both the standard multi-objective problem and the hypothetical groundwater case. This technique is being applied to a field-scale case study at the Umatilla Chemical Depot, situated at Hermiston Oregon, which was the subject of a recent transport optimization demonstration (Minsker et al. 2003). It is believed that this approach can readily be applied to such real world problems, ensuring more reliable and robust Pareto optimal solutions for the decision maker.

REFERENCES

Aguado, E., Remson, I. Pikul, M. F., and Thomas, W. A. (1974). “Optimum pumping for aquifer dewatering.” *Journal Hydraulics Division*, ASCE, 100, 860-877.

Aly, A.H. and Peralta, R.C. (1999). Optimal Design of Aquifer Cleanup Systems under Uncertainty using a Neural Network and Genetic Algorithm, *Water Resource Research*, 35(8), 2523-2532.

Cieniawski, S. E. (1993). An Investigation of the Ability of Genetic Algorithms to Generate the Tradeoff Curve of a Multi-Objective Groundwater Monitoring Problem, Masters Thesis University of Illinois: Urbana, IL.

Chan Hilton, A. B. and Culver, T. B. (2000). Optimizing groundwater Remediation Design under Uncertainty, *Proceedings of the Joint Conference on Water Resource Engineering and Water Resources Planning and Management*, July 30-August 2, 2000, Minneapolis, MN.

Clement, T. P. (1997). RT3D - A modular computer code for simulating reactive multi-species transport in 3-Dimensional groundwater aquifers, *Battelle Pacific Northwest National Laboratory Research Report*, PNNL-SA-28967. (http://bioprocesses.pnl.gov/rt3d.htm).

Coello, C. A. C. (1999). An Empirical Study of Evolutionary Techniques for Multiobjective Optimization in Engineering Design. Doctoral dissertation Tulane University: New Orleans, L.A.

Culver, T. B., and Shoemaker, C. A. (1997). "Dynamic optimal groundwater reclamation with treatment capital costs." *Journal of Water Resources Planning and Management*, 123(1), 23-29.

Deb, K. (1998). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. Technical Report No: CI-49/98. Dortmund: Department of Computer Science/LS11, University of Dortmund, Germany.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan. T. (2000.) A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimisation: NSGA-II, *Proceedings of the Parallel Problem Solving from Nature VI Conference*, 849-858.

Erikson, M., Mayer, A., and Horn, J. (2002). "Multi-objective optimal design of groundwater remediation design systems: application of the niched Pareto genetic algorithm (NPGA)", *Advanced in Water Resources,* 25(1), 51-65.

Feyen, L., Bevan, K. J., De Smedt F., and Freer, J. (2001). Stochastic capture zone delineation within the generalized likelihood uncertainty estimation methodology: Conditioning on head observations, *Water Resources Research* 37(3), 625-638.

Fonseca, C. M. and Fleming, P. J. (1993). Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization, In Stephanie Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, 416-423.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, New York.

Gopalakrishnan, G., Minsker, B. S., and Goldberg, D. E. (2003). Optimal Sampling in a Noisy Genetic Algorithm for Risk-Based Remediation Design, *Journal of Hydroinformatics,* 05.1, IWA Publishing, 11-25.

Harik, G.R., Cantú-Paz, E., Goldberg, D. E., and Miller, B. L. (1997). "The gambler's ruin problem, genetic algorithms, and the sizing of populations." In *Proceedings of the 1997 IEEE Conference on Evolutionary Computation,* pp. 7-12, IEEE Press, New York, NY.

Hughes, E. J. (2001). Evolutionary Multi-Objective Ranking with Uncertainty and Noise, *Proceedings of the First International Conference Evolutionary Multi-Criterion Optimization*, Zurich, Switzerland, March 7-9, 2001, 329-343

Karatzas, G. P. and Pinder, G.F. (1993). Groundwater Management using Numerical Simulation and the Other Approximation Method for Global Optimization, *Water Resource Research*, 29(10), 3371-3378.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2001). On the Convergence and Diversity-Preservation Properties of Multi-Objective Evolutionary Algorithms, Technical Report 108, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland.

Mahfoud, S. (1995). Niching Methods for Genetic Algorithms, Doctoral Dissertation, University of Illinois: Urbana-Champaign, IL.

Maskey, S., Jonoski, A., and Solomatine, D. (2002). "Groundwater remediation strategy using global optimization algorithms." *Journal of Water Resources Planning and Management*, 128(6), 431-440.

McKinney, D.C. and Lin, M.D. (1994). Genetic Algorithms Solution of Groundwater Management Models, *Water Resource Research*, 31(3), 731-740.

Miller, B.L. (1997). Noise, Sampling and Efficient Genetic Algorithms, IlliGAL Rep. 97001.

Miller, B.L and Goldberg, D.E. (1996). Optimal Sampling for Genetic Algorithms, In Dagli, C.H, Akay, M., Chan, C.L.P Fernandez, B.R., and Ghosh J. (Eds.), *Intelligent Engineering systems through artificial neural networks* (ANNIE '96), Volume 6, pp 291-298, New York, ASME Press.

Minsker, B.S. and Shoemaker, C.A. (1998). Quantifying the Effects pf Uncertainty on Optimal Groundwater Bioremediation Policies, *Water Resource Research*, 34(12), 3615-3625.

Minsker, B., Y. Zhang, R. Greenwald, R. Peralta, C. Zheng, K. Harre, D. Becker, L. Yeh, and Yager, K., ESTCP (2003). Final Technical Report for Application of Flow and Transport Optimization Codes to Groundwater Pump and Treat Systems, Environmental Security Technology Certification Program. Available at http://www.frtr.gov/estcp/

Peralta, R. C. and Kalwij, Ineke (2003). Simulation/Optimization Tools for Robust Pumping Strategy Design, *Proceedings of 2003 AFCEE Technology Transfer Workshop,* Brooks City-Base, Texas.

Reed, P., Minsker, B. S., and Goldberg, D. E. (2001). A multiobjective approach to cost effective long-term groundwater monitoring using an Elitist Nondominated Sorted Genetic Algorithm with historical data, *Journal of Hydroinformatics* 3(2), 71-90.

Reed, P., Minsker, B., and Goldberg, D.E. (2003). Simplifying Multiobjective Optimization: An Automated Design Methodology for the Nondominated Sorted Genetic Algorithm-II, *Water Resources Research*, 39(7), July 2003, pp. TNN 2.1-2.5, 1196, doi: 10.1029/2002WR001483.

Reed, P., and Minsker, B. (in press). Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives, *ASCE Journal of Water Resources Planning and Management*.

Ritzel, B.J., Eheart, J.W. and Ranjithan, S. (1994). Using Genetic Algorithms to Solve a Multiple Objective Groundwater Remediation Problem, *Water Resources Research*, 30, (5), 1589-1603.

Rizzo, D.M. and Dougherty, D.E. (1996). Design Optimization for Multiple Management Period Groundwater Remediation, *Water Resource Research*, 32(8), 2549-2561.

Schaffer, J. D. (1984). *Some experiments in machine learning using vector evaluated genetic algorithms*, Doctoral dissertation, Vanderbilt University: Nashville, TN.

Smalley, J. B., Minsker, B. S., and Goldberg, D. E. (2000). Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resources Research*, 36(20), 3043-3052.

Srinivas, N. and Deb, K. (1994). Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3), 221-248.

Teich, Jurgen (2001). Pareto-Front Exploration with Uncertain Objectives, *Proceedings of the First International Conference Evolutionary Multi-Criterion Optimization*, Zurich, Switzerland, March 7-9, 2001, 314-328

Tompson, A.F.B., Ababou, R. and Gelhar, L.W. (1989). Implementation of the Three-Dimensional Turning Bands Random Field Generator, *Water Resource Research*, 25(10), 2227-2243.

Van Veldhuizen, D. A. and Lamont, G. B. (2000). Multiobjective evolutionary algorithms: analyzing the state-of-the-art. *Evolutionary Computation*, 8(2), 125-147.

Wang, Q.J. (1991). The genetic algorithm and its application to calibrating conceptual rainfall runoff models, *Water Resource Research*, 27(9), 2467-2471.

Wang, M. and Zheng, C. (1997). Optimal Remediation Quality Management under Parameter Uncertainty, *Water Resources Research*, 23(7), 1162-1174.

Wardlaw R. and Sharif M. (1999). Evaluation of genetic algorithms for optimal reservoir system operation", *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, 125(1), 25-33.

Yu, D. M., Rizzo, M., and Dougherty, D. E. (1998). “Multi-period objectives and groundwater remediation using samoa: tandem simulated annealing and external cutting plane method for containment with cleanup.” *XII International Conference on Computational Methods in Water Resources*, 332-340.