Monograph:
Pursehouse, R.C. and Fleming, P.J. (2002) Elitism, Sharing and Ranking Choices in Evolutionary Multi-Criterion Optimisation. Research Report. ACSE Research Report 815 . Department of Automatic Control and Systems Engineering

ELITISM, SHARING, AND RANKING CHOICES IN EVOLUTIONARY MULTI-CRITERION OPTIMISATION

R. C. PURSHOUSE

P. J. FLEMING

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 815
January 2002

Elitism, Sharing, and Ranking Choices in Evolutionary Multi-Criterion Optimisation

R. C. Purshouse and **P. J. Fleming**
Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.
E-mail: {r.purshouse, p.fleming}@sheffield.ac.uk

Abstract

Elitism and sharing are two mechanisms that are believed to improve the performance of an evolutionary multi-criterion optimiser. The relative performance of the two most popular ranking strategies is largely unknown. Using a new empirical inquiry framework, this report studies the effect of elitism, sharing, and ranking design choices using a benchmark suite of two-criterion problems. Performance is assessed, via known metrics, in terms of both closeness to the true Pareto-optimal front and diversity across the front. Randomisation methods are employed to determine significant differences in performance. Informative visualisation of results is achieved using the attainment surface concept. Elitism is found to offer a consistent improvement in terms of both closeness and diversity, thus confirming results from other studies. Sharing can be beneficial, but can also prove surprisingly ineffective. Evidence presented herein suggests that parameter-less schemes are more robust than their parameter-based equivalents (including those with automatic tuning). Very little performance difference is evident between the two ranking strategies. A multi-objective genetic algorithm combining both elitism and parameter-less sharing is shown to offer very good performance across the test suite.

1 Introduction

The number of multi-objective evolutionary algorithm (MOEA) schemes proposed in the literature has accelerated dramatically over the last few years. As a result, evolutionary multi-criterion optimisation (EMO) practitioners are faced with a number of design choices beyond those encountered in a standard evolutionary algorithm (EA). In order to exploit the true potential of the evolutionary meta-heuristic, the optimiser should be tailored to the application rather than simply used as a black-box. The implementer should perhaps be encouraged to develop a bespoke MOEA rather than resort to a complete algorithm such as MOGA, NSGA-II, or SPEA-2. Thus, the nature of a design choice would be, for example, *'What mechanisms should be used to promote diversity in this application?'* rather than *'Should NSGA-II be used instead of SPEA-2?'*.

Existing performance comparisons available in the literature tend to compare different *brands* of algorithm, thus obscuring the underlying components and interactions that are attributable to performance (a notable exception to this is the paper by Laumanns *et al* [2001b]). It would be helpful to understand how the performance of a component changes in particular circumstances, such as the nature of the underlying problem and remaining optimiser configuration. It is also useful to know when performance is largely insensitive to a particular design choice within a certain set of bounds. The aim of this report is to expose the effect of the following EMO strategies using a rigorous and tractable experimental procedure:

- *Elitism* – the preservation and exploitation of known good solutions.
- *Sharing* – the modification of selection probabilities to account for distribution.
- *Ranking* – methods for scalarising performance using relative Pareto dominance.

The experimental framework is introduced in Section 2. The benchmark suite of test problems used in the study is described (equations are provided in the Appendix), together with suitable performance metrics. A method for statistical significance testing is introduced, as is an appropriate visualisation technique. A baseline MOEA is developed in Section 3, and its performance is established. The effects of elitism, sharing, and ranking strategies are then considered with reference to this baseline. An elitist strategy is developed and tested in Section 4. Sharing methodologies for the promotion of diversity are discussed in Section 5, in which both parameter-based and parameter-less techniques are investigated. The performances of the established Epanechnikov method and a new ranking-based method are compared. The two most popular multi-criterion ranking strategies in the literature are contrasted in Section 6. In Section 7, a new MOGA incorporating both elitism and parameter-less sharing is developed. Conclusions are offered in Section 8, together with recommendations for future work.

2 Experimental framework

2.1 Overall methodology

Evolutionary algorithms are complicated non-linear systems that have proved difficult to analyse. The large number of variables, and the interactions between them, can make a confident interpretation of results difficult. The *EMO empirical inquiry framework* presented in this section seeks to increase the benefit of empirical testing of algorithms. The following attributes of the methodology are emphasised:

- Modular, traceable, configuration changes.
- Transparent, understandable, test problems with realistic properties.
- Appropriate, accurate, performance measures.
- Rigorous, informative, analysis, including tests for statistical significance and visualisation.

The test suite, performance metrics, and statistical and visual analysis techniques utilised in this study are discussed in detail in the following sub-sections.

2.2 Test suite

The established set of test problems developed by Zitzler *et al* [2000] is used in this study. The suite consists of six, tractable, two-criterion functions, with varying characteristics as summarised in Table 1. The corresponding mathematical definitions are provided in the Appendix.

NAME	ATTRIBUTES
ZDT-1	Convex front
ZDT-2	Non-convex front
ZDT-3	Piece-wise continuous convex front
ZDT-4	Many local fronts, single global front
ZDT-5	Deceptive problem, discrete front
ZDT-6	Non-uniform distribution across a non-convex front

Table 1: Test function characteristics

These functions cover many of the features that may be found in real-world problems and are comparatively easy to analyse. The main concern is that the first criterion is a function of only a single decision variable (mapped without modification in the first four test problems). In particular, this may cloud the issues surrounding diversity preservation. It should also be borne in mind that these test functions consist of two criteria only. Much care should be taken before transferring conclusions drawn from these functions to problems with a higher, and more realistic, number of criteria.

2.3 Measuring performance

The performance of an MOEA can be decomposed into two, interacting, criteria:

- **Closeness** – the nearness of the identified non-dominated solutions to the true Pareto-optimal front, and
- **Diversity** – the distribution of the identified solutions across the trade-off surface. This distribution is commonly expressed in criterion-space.

The ideal outcome, at least for these comparatively low-dimensional test cases, is a final population with a uniform distribution of globally non-dominated solutions spread across the entire trade-off surface. This aim is largely feasible for two-criterion problems, but may prove problematic as the number of criteria is increased (where the trade-off surface becomes larger with respect to the size of the total search space).

Various performance metrics have been proposed to measure accuracy, diversity, and in some cases both simultaneously. Some of these metrics involve measurements made with respect to the true trade-off surface, whilst others involve a purely relative comparison of two sets of results. The former approach requires that the true surface be known and can be sampled but is advantageous in that conventional statistical tests can be straightforwardly applied. A review of performance metrics is provided by Deb [2001, pp306-324].

This study utilises three known performance metrics: *generational distance* to measure accuracy, *spread* to measure diversity, and *attainment surfaces* to provide visualisations of the results. These metrics are described in further detail below.

2.3.1 Generational distance

The accuracy of each non-dominated point produced by an MOEA can be measured in terms of its distance to the closest part of the global trade-off surface. These distances can be averaged to provide a measure of accuracy of a single MOEA run. The definition of distance is dependent on the problem domain. Euclidean distance is an obvious choice for the ZDT test problems, and has been adopted in this work. Note that the objective values must be normalised if they are not of the same scale. The generational distance metric is formalised in Equation 1 [Veldhuizen, 1999].

$$GD = \frac{1}{|Q|}\sum_{i=1}^{|Q|} d_i \qquad (1)$$

where GD is the generational distance,

Q is the obtained set of non-dominated criterion vectors,

$|\ |$ is the number of vectors in the set and,

d is the closest distance between vector q in Q and any vector in $P*$, where

$P*$ is the set of globally non-dominated criterion vectors.

The main advantages of this metric are its simplicity and its amenability to statistical analysis. The disadvantage is that the set $P*$ must be obtained. This set should be sufficiently numerous and should be uniformly distributed across the trade-off surface in order to avoid bias. Fortunately, for the ZDT problems, the global Pareto-optimal front is explicitly defined in each case. For the continuous and piece-wise continuous trade-off curves, uniform parametric sampling is quite straightforward (achieved, for example, using equations for curvature). The discrete trade-off surface of ZDT-5 can easily be enumerated.

2.3.2 Spread

Consider the distribution of distances between nearest-neighbour criterion vectors. In the case of a uniform distribution, all such distances will be identical and will equal the mean of the distribution. In the general case, uniformity can thus be measured by considering the difference between a nearest-neighbour distance and the mean of all such distances. Schott [1995] originally formulated the sum of all these differences as an indication of the uniformity of the identified trade-off surface. This was extended by Deb *et al* [2000] to include a measure of the extent of the obtained distribution. The resulting metric is shown in Equation 2.

$$\Delta = \left[\sum_{m=1}^{M} d_m^e + \sum_{i=1}^{|Q-1|} \left| d_i - \bar{d} \right| \right] \Bigg/ \left[\sum_{m=1}^{M} d_m^e + |Q-1| \bar{d} \right] \tag{2}$$

where Δ is the spread,

d_m^e is the Euclidean distance from the extreme point in Q to the extreme point in P^* for the mth criterion.

M is the number of criteria,

d_i is the Euclidean distance between consecutive criterion vectors in Q and,

$\bar{d}$ is the mean of all d_i.

The first term in the numerator of Equation 2 describes the extent of trade-off surface not included in Q. The second term describes the non-uniformity of the Q distribution. The denominator seeks to normalise these measures with respect to the total magnitude of the trade-off surface. Clearly, smaller values of Δ indicate superior diversity to larger values.

The spread metric is a suitable diversity measuring metric for two-criterion problems. In the formulation given above, only the extreme members of P^* must be known: therefore complete knowledge of P^* is not required. However, the denominator in Equation 2 uses information in Q in order to normalise the metric. It is possible that this may overestimate the length of the trade-off surface and thus falsely reduce Δ. Hence, it may be more appropriate to calculate the normalisation factor using P^* (in which case much more of P^* would have to be sampled). Also, in order to reduce the influence of the size of set Q, perhaps the second term in the numerator should be normalised with respect to this factor.

In the form presented above, the spread metric cannot be used in problems with more than two criteria. This is because the concept of consecutive criterion vectors does not apply in higher dimensions. However, since an ordering *can* be performed on a criterion-wise basis, a neighbouring vector can be defined as one that is immediately adjacent in any criterion. A neighbour connection can then be established between the two vectors (in any suitable space, such as Euclidean). Repetition of neighbour connections is avoided. Thus, the distances used in the modified spread metric are the length of these neighbour connections. The measure of uniformity then proceeds as before. The extreme points on an n-dimensional trade-off surface can be obtained by performing all lexicographic optimisation permutations on the P^* data. Again, repetition of extreme points should be avoided. The error in the extent of the front can then be computed as before. The suggested form of the metric is shown in Equation 3. Criterion vectors should be normalised prior to application of the metric.

$$\Delta = \sum_{ex=1}^{EX} d_{ex}^{e} + \frac{1}{|Q|} \sum_{ne=1}^{NE} \left| d_{ne} - \overline{d} \right| \tag{3}$$

where d_{ex}^{e} is the closest Euclidean distance between Q and extremum *ex*,

EX is the total number of unique extrema in P^*,

d_{ne} is the Euclidean distance between two neighbours and,

NE is the total number of unique neighbour connections.

If this new spread metric were to be normalised with respect to the magnitude of the trade-off surface, Equation 3 would be very similar to the original Equation 2 for two-criterion problems.

2.3.3 Attainment surfaces

Fonseca and Fleming [1996] introduced the concept of an *attainment surface*. Given a set of non-dominated vectors produced by a single run of an algorithm, the attainment surface is the boundary in criterion-space that separates the region that is dominated by or equal to the set from the region that is non-dominated. Note that this is fundamentally different to interpolating between the vectors. This latter approach is not, in general, correct because there is no guarantee that any intermediate vectors actually exist and, even if this were the case, the corresponding solutions are unknown. The concept of the attainment surface is illustrated in Figure 1.

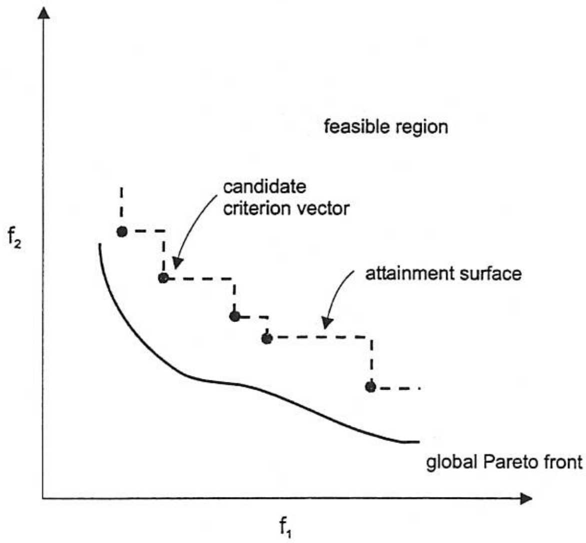

Figure 1: Example attainment surface

Attainment surfaces serve two very useful purposes. One the one hand, they provide a convenient means of visualising the results from multiple runs of an optimiser. On the other, through the use of *auxiliary lines*, they allow for algorithm comparisons using well-known univariate statistical tests. In this study, the attainment surfaces are used purely for visualisation. Refer to Fonseca and Fleming [1996] and Knowles and Corne [2000] for examples of the comparative statistics work.

The superposition of multiple attainment surfaces, as shown in Figure 2, provides a qualitative indication of the performance of a particular MOEA configuration. The regions of criterion-space created by the surfaces can be interpreted probabilistically. Given that both criteria are to be minimised, the region below all the attainment surfaces contains performance vectors that were not matched by the MOEA in any run. The region above all the surfaces contains vectors that were exceeded by all runs. In the intermediate regions, the performance vectors were exceeded on a number of occasions. Thus, it is possible to obtain a family of vectors that, individually, would be obtained in a given percentage of runs. The heavy line in Figure 2 shows the 50%-attainment surface (akin to the median statistic). Similarly, the grey lines indicate the 25% and 75% surfaces (quartiles). The 0% and 100% surfaces are shown as the dotted lines. This means of visualisation is employed throughout this study. The attainment surfaces provide information on location, dispersion, and skewness, in a similar manner to the box plot [Cleveland, 1993]. This methodology provides more reliable information than the unification-of-runs approach adopted by Zitzler *et al* [2000] and Purshouse and Fleming [2001].

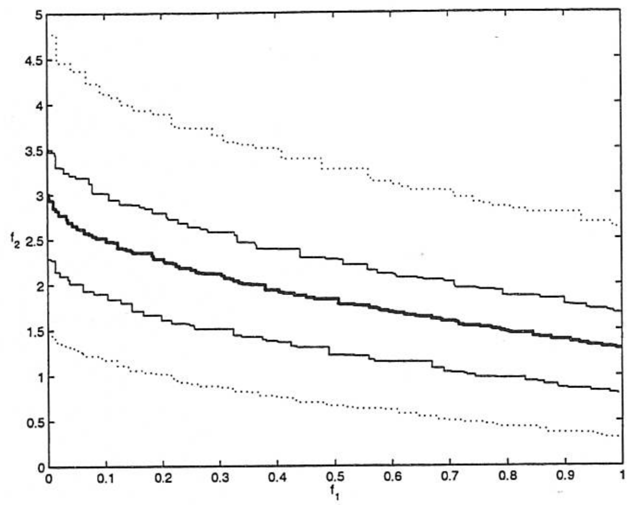

Figure 2: The superposition of multiple attainment surfaces

The attainment surface concept is extendable to any number of objectives, although visualisation becomes problematic at any dimension higher than three. Computational complexity also increases significantly.

2.4 Analysing performance

Upon completion of a single run of a specific MOEA configuration on a particular problem, three sets of non-dominated criterion vectors (and associated solutions) are obtained, namely:

- **final population** – the non-dominated vectors in the final population of the algorithm,
- **on-line archive** – the final elite set of vectors, and
- **off-line archive** – the complete set of non-dominated vectors identified by the algorithm.

The first of these sets is used for analysis and comparison purposes in this study since it provides the most appropriate measure of the on-line trade-off surface *maintenance* capabilities of an algorithm.

An evolutionary algorithm is a stochastic process and, thus, multiple runs (samples) are required in order to infer reliable conclusions as to its performance. Hence, 35 runs have been conducted for each MOEA configuration when applied to a particular test problem. The performance of the algorithm is expressed in the resulting distributions of generational distance and spread. A statistical comparison of two configurations is then possible through use of a test statistic.

In this study, the mean difference between two generational distance (or, alternatively, spread) distributions is taken as the test statistic. The significance of this observed result is then assessed using *randomisation testing*. This is a simple, yet effective, technique that does not rely on any assumptions concerning the attributes of the underlying processes, unlike conventional statistical methods [Manly, 1991]. The central premise of the method is that, if the observed result has arisen by chance, then this value will not appear unusual in a distribution of results obtained through many random relabellings of the samples. The randomisation method proceeds as follows:

1. Compute the difference between the means of the samples for each algorithm: this is the observed difference.

2. Randomly reallocate half of all samples to one algorithm and half to the other. Compute the difference between the means as before.

3. Repeat Step 2 until 5000 randomised differences have been generated, and construct a distribution of these values.

4. If the observed value is within the central 99% of the distribution, then accept the null hypothesis. Otherwise consider the alternative hypotheses. This is a two-tailed test at the 1%-level.

The null hypothesis is that the observed value has arisen through chance and so there is no performance difference between the two configurations. The alternative hypotheses are that the difference is unlikely to have arisen through chance and that one configuration has outperformed the other (depending on which side of the distribution the observed difference falls, and the direction in which the difference has been calculated). By demanding a 1%-level of significance, the probability of making a Type II error (accepting the null hypothesis when it is false) is increased. However, given the rising popularity of EMO research, this increased danger of rejecting a truly significant improvement is unlikely to be damaging.

Note that the observed value is included as one of the random relabellings since, if the null hypothesis is true, then this value is one of the possible randomisation results. 5000 randomisations is regarded as an acceptable quantity for a test at the 1%-level [Manly, 1991].

The results of randomisation testing are simple to visualise, as shown by the example in Figure 3. The randomised results are described by the grey histogram, whilst the observed result is depicted as a filled black circle. Each row shows the performance on a particular test function (from ZDT-1 at the top, to ZDT-6 at the bottom). The left-hand column indicates the relative performance regarding closeness, and the right-hand column shows the corresponding difference in diversity. It is usually clear from the figure whether or not the observed result is statistically significant, although it may occasionally prove necessary to resort to a closer analysis of the underlying randomisation data.

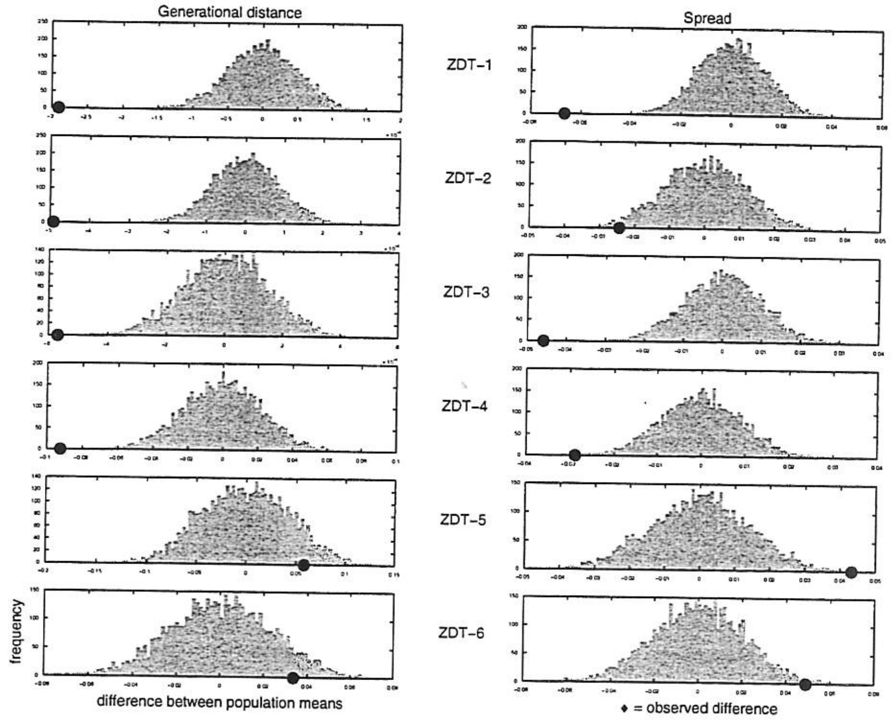

Figure 3: Randomisation testing – example results

In the example in Figure 3, given data for two algorithms *A* and *B* together with a test statistic of *mean(B) – mean(A)*, then the following results are observed:

- *B* obtains fronts closer to the true front than *A* for ZDT-1, 2, 3, and 4.
- *B* produces a superior distribution of criterion vectors to *A* for ZDT-1, 3, and 4.
- *A* offers a superior distribution to *B* on ZDT-5.
- No other results are significant at the 1%-level, although clearly some of these cases offer more evidence against the null hypothesis than others.

3 Baseline MOGA

The baseline optimiser used in this study has been developed according to the holistic design principles championed by Michalewicz and Fogel [2000] and has previously been shown to be effective at solving the ZDT test problems [Purshouse and Fleming, 2001]. A summary of the algorithm is provided in Table 2.

The multi-criterion performance of a solution is scalarised using Fonseca and Fleming's [1998] Pareto-based ranking procedure. A solution is ranked according to the number of solutions in the population that are *preferred* to it. If the entire Pareto-optimal front is to be identified, the preference relation collapses to a test for pure Pareto dominance.

EMO COMPONENT	STRATEGY
General Population size Total generations	 100 per generation 250
Elitism	None (zero generational gap).
Evaluation	[1] Fonseca and Fleming [1998] Pareto-based ranking. [2] Linear fitness assignment with rank-wise averaging. [3] No modification of fitness to account for population density.
Selection	Stochastic universal sampling [Baker, 1987]
Representation Real parameter Binary function	 Concatenation of real number decision variables. Accuracy bounded by machine precision. Binary string, 80 bits in length. Defined by the problem.
Search operators For real representations For binary representations	 [1] Naïve crossover. Probability = 0.8. [2] Gaussian mutation (initial search power of 40% of variable range; sigmoidal scaling set to 15; feasibility requirement of one standard deviation). Probability = Expected value of 1 phenotype per chromosome. [1] Single-point two-parent crossover. Probability = 0.8. [2] Simple bit-flipping mutation. Probability = 1/80.

Table 2: Baseline configuration

When ranking is complete, initial fitness values can be prescribed. The population is sorted according to rank and fitnesses are assigned by interpolating between the highest fitness value for the best rank and the lowest fitness value for the worst rank. In the baseline algorithm, linear interpolation is used and fitness is varied between the population size (highest) and unity (lowest). The ratio of these two fitnesses is a definition of the *selective pressure* of the assignment mechanism. Solutions of the same rank then have their fitnesses amended to the average of the original assignments for that rank. Since part of this study is concerned with the effect of diversity-preserving mechanisms, no manipulation of the above fitnesses through sharing is undertaken.

Stochastic universal sampling has been chosen as the selection mechanism [Baker, 1987]. This method achieves maximum spread with minimal bias, but is non-parallelisable. As part of this procedure, the above fitness values are normalised to provide an expected number of selections for each solution. In total, 100 selections are required since the chosen reinsertion strategy is that all offspring replace all parents (no generational gap) and since for the chosen recombination operators two parents are required to produce two offspring.

Since five of the test problems feature real number decision variables, it is logical to use a real number representation for these problems. Hence, a candidate solution is described by a concatenation of phenotypic decision variables. This representation offers a number of advantages over a binary encoded approach [Michalewicz, 1992]. The other test problem explicitly uses binary variables, thus a binary representation is natural for this problem.

Different representations require different search operators. For the binary chromosome case, the familiar single-point two-parent crossover and bit-flipping mutation operators are employed. Good results are known to be achievable using this simple approach [Zitzler *et al*, 2000]. Various operators for real representations have been suggested [Herrera *et al*, 1998]. This study uses the so-called *naïve crossover* in conjunction with a *Gaussian mutation* operator. The former of these search tools is a very simple two-parent single-point crossover operator, where the crossover sites are limited to points between decision variables. This offers quite a low-power search, since it cannot generate any values for decision variables that were not present in the original population. However, when coupled with a complementary high-power search tool, the resulting search capabilities are considerable[1]. Gaussian mutation is one such operator. Its main benefit is that it provides tuneable search power in the form of the standard deviation. This can be exploited to provide on-line adaptation that avoids the generation of infeasible solutions and controls convergence speed by varying the search from near global early on to very local towards the end. *Sigmoidal* scaling, as a function of the percentage of generations completed, of the standard deviation is useful because it allows concentrated periods of high- and low-power search [Purshouse and Fleming, 2001].

Attainment surfaces illustrating the performance of the baseline algorithm are shown in Figures 4 through 9.

---

[1] Coincidentally, the incorporation of naïve crossover largely prevents the convergence failures encountered by Ikeda *et al* [2001], thus showing that MOEA failure cannot be solely blamed on the use of Pareto ranking in these cases.

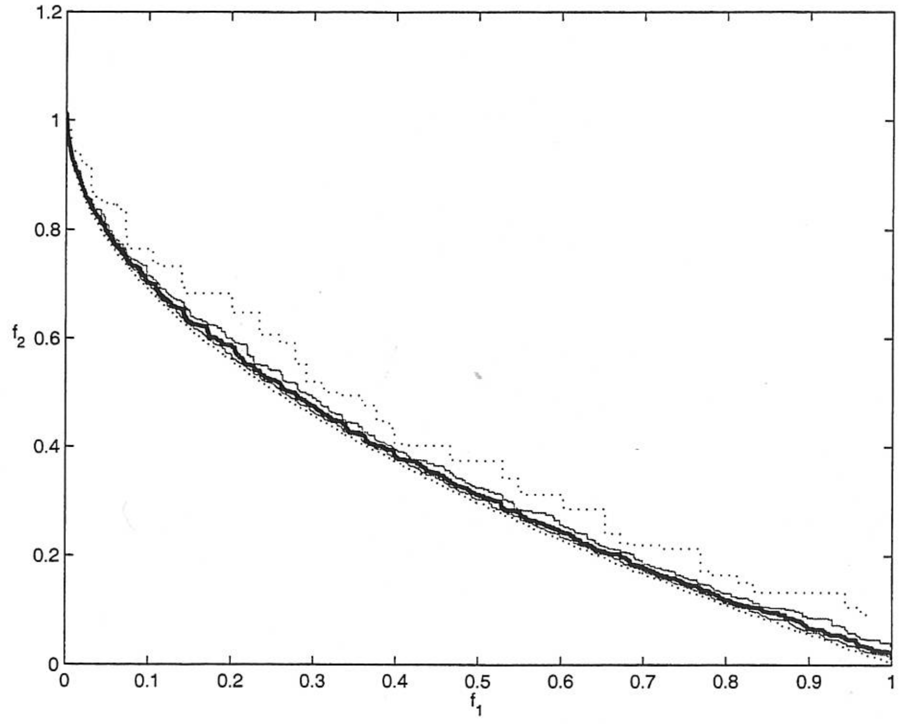

Figure 4: Attainment surface – baseline MOGA solving ZDT-1

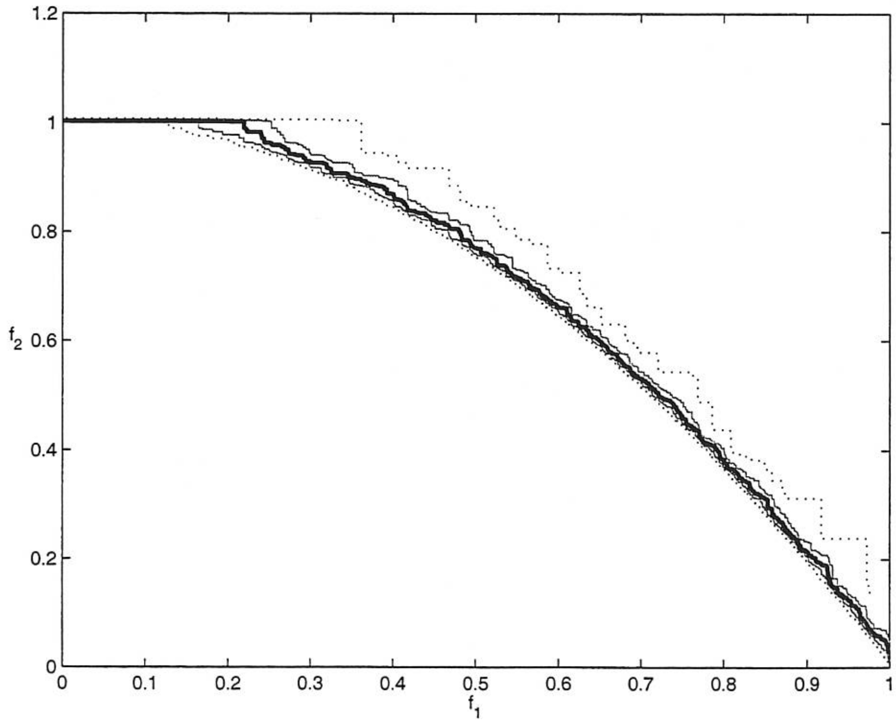

Figure 5: Attainment surface – baseline MOGA solving ZDT-2

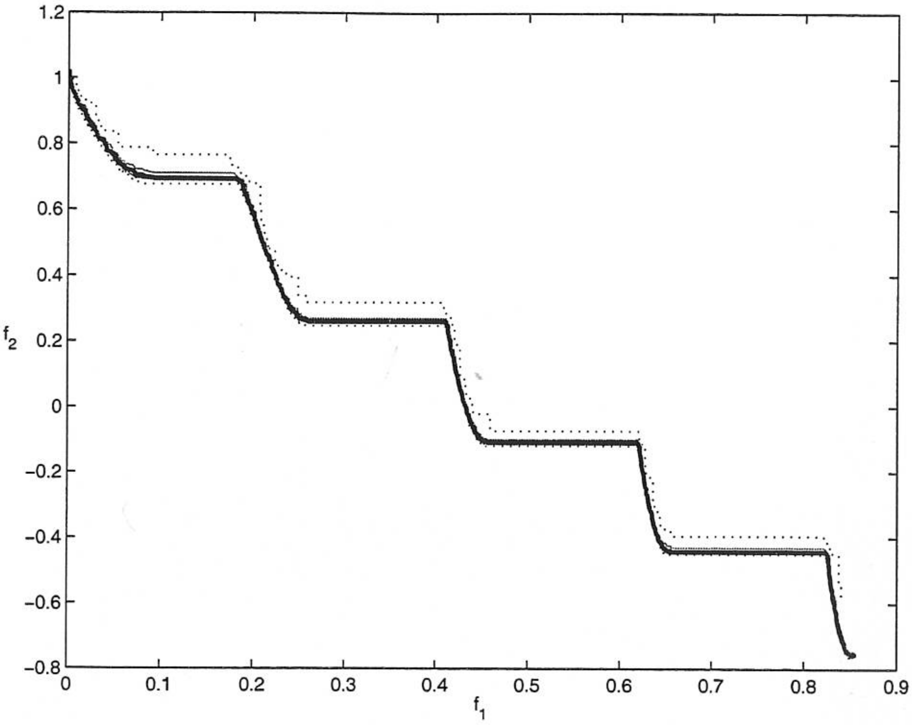

Figure 6: Attainment surface – baseline MOGA solving ZDT-3

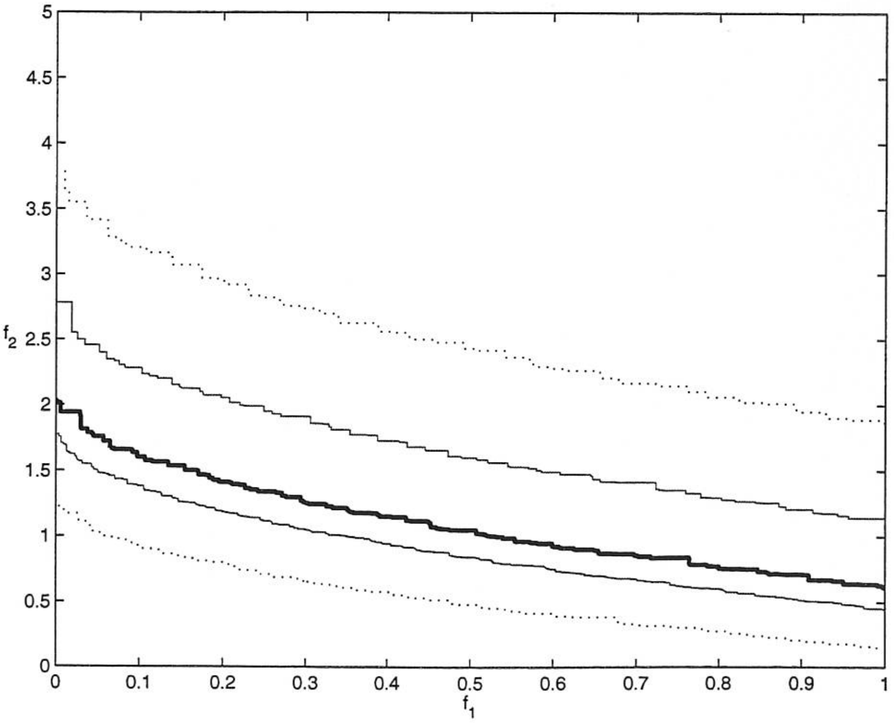

Figure 7: Attainment surface – baseline MOGA solving ZDT-4

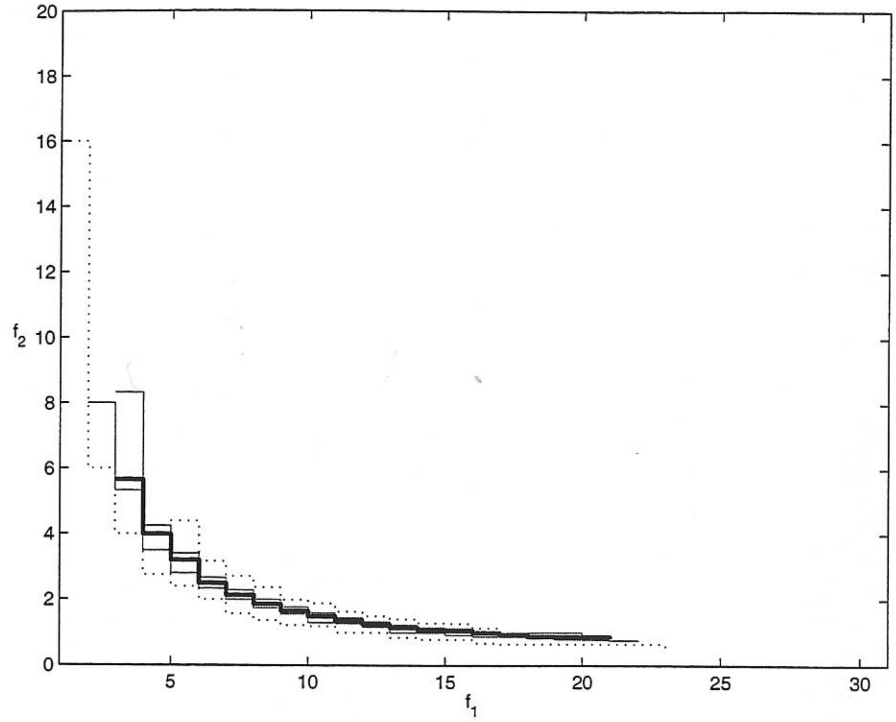

Figure 8: Attainment surface – baseline MOGA solving ZDT-5

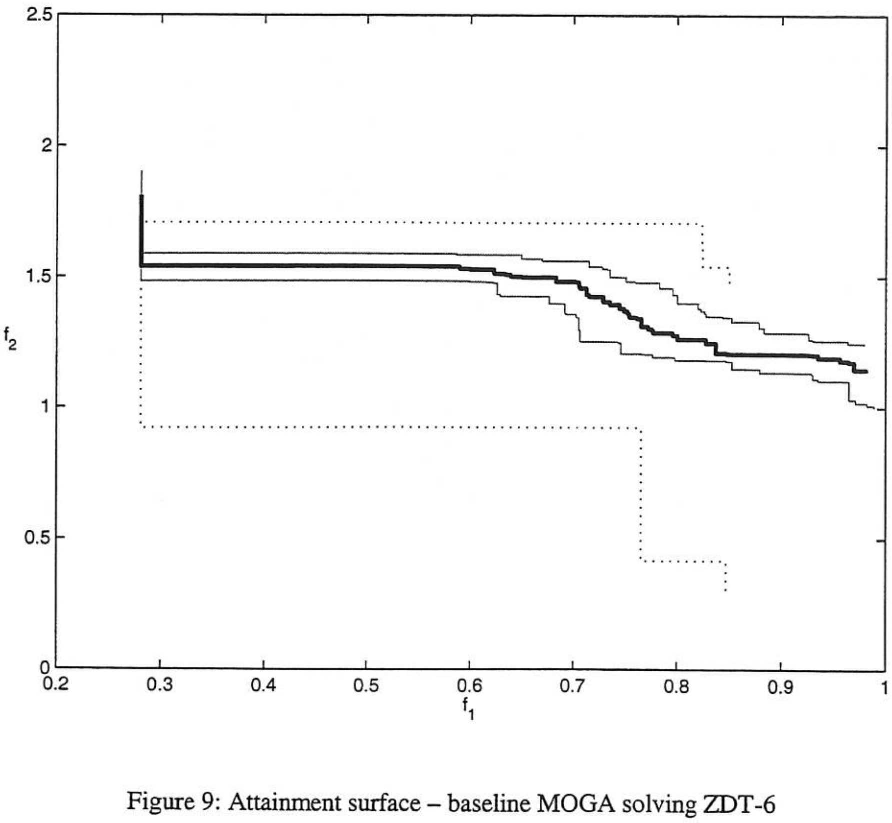

Figure 9: Attainment surface – baseline MOGA solving ZDT-6

Particularly good results were achieved for ZDT-1, ZDT-2, and ZDT-3 (Figures 4, 5, and 6 respectively) in terms of both closeness to the global Pareto front and diversity across the front. The tight envelopes of attainment indicate the high level of consistency achieved in these cases. Closeness was especially good for ZDT-3. As has been previously observed by Purshouse and Fleming [2001], the MOGA struggled to achieve good coverage of the surface as f_1 approaches zero on ZDT-2. Note that this is a region where there is little trade-off between the objectives.

As shown in Figure 7, the wider envelopes of attainment produced for the multi-fronted ZDT-4 signify entrapment in a locally non-dominated front. On no occasions did the MOGA converge to the global trade-off surface although coverage across the identified fronts was good.

The baseline MOGA achieved reasonable closeness to the global front on ZDT-5. Performance on this deceptive test function is depicted in Figure 8. Note that on no occasions was the algorithm able to identify the extreme right-hand section of the discrete trade-off surface.

Rather poor performance was observed on the non-uniform ZDT-6, as shown in Figure 9. Coverage is especially poor on the less dense area of the front. This, together with the missing part of the ZDT-5 front, is the strongest indication that density-based sharing would be beneficial. Closeness to the true Pareto front is also not good. Only the 0%-attainment surface lies on the global front, where coverage is particularly poor. Furthermore, the position of this front with respect to the median and quartiles suggests that this result is something of an outlier.

The study now progresses to consider the effects of elitism, sharing, and ranking choices, using this baseline algorithm.

4 Elitist strategy

Elitism is the process of preserving previous high-performance solutions from one generation to the next. This is conventionally achieved by simply copying the solutions directly into the new generation. Elitism has long been considered an effective method for improving the efficiency of an EA [De Jong, 1975]. Various recent studies in the EMO community have indicated that the inclusion of an elitist element can considerably improve the performance of an MOEA [Zitzler *et al*, 2000; Deb *et al*, 2000].

The implementation of elitism is straightforward for single criterion problems, since there is typically only a single best individual in a population. For multi-criterion problems, the elitist strategy requires greater intricacy. An elite sub-population of currently non-dominated solutions can be defined, but the magnitude of this set can potentially reach the size of the population itself. This becomes a particular problem during the later stages of the search. Copying large numbers of solutions, without modification, into the new population will hamper the search process. Thus, the two main issues are (1) how to manage the size of the elite sub-population, and (2) how to use elitism to drive the search effectively.

Zitzler and Thiele [1999] proposed a simple and effective elitist strategy for MOEAs, and implemented it in their Strength Pareto Evolutionary Algorithm (SPEA). The success of this algorithm across a diverse set of two-criterion problems has led to the widespread adoption of elitist schemes in the EMO community.

The SPEA maintains an on-line archive of currently non-dominated solutions and uses this in the processes that generate new candidate solutions. The archive should be a representative subset of all non-dominated solutions found thus far. Note that MOEAs prior to SPEA (including MOGA) generally maintain an up-to-date off-line archive of *all* non-dominated solutions found, but these results are not explicitly used in the generation of new candidate solutions.

The on-line archive requires a clustering mechanism in order to control the number of elite solutions. This set of solutions should represent the characteristics of the underlying off-line archive. *Characteristics* generally refer to the criterion vectors, although decision-space discrimination is also possible. The *truncation procedure* described in Zitzler *et al* [2001] is an effective means of elitism control for two-criterion problems. In these cases, this method can reduce an oversized archive without losing boundary solutions. This attribute is desirable in the search for diverse trade-off solutions. However, for problems with more than two criteria, extreme trade-off solutions can be lost by this method. This may be a particular problem if the focus is on a preferred region of the trade-off surface. It is also possible for the procedure to remove globally non-dominated solutions whilst retaining currently non-dominated, yet sub-optimal, solutions [Laumanns *et al*, 2001a]. Both solutions are non-dominated from the perspective of the truncation process, but the sub-optimal solution may be in a less-dense area of criterion-space. The truncation procedure is outlined below:

To remove one member of the over-sized archive follow the subsequent procedure with *k* initialised to 1:

1. Find the set of solutions, *S*, with the shortest distance between themselves and their *k*th nearest neighbours.

2. If the size of this set is greater than one, increment *k* and repeat Step 1 using only solutions in *S*, otherwise select the individual in *S* for removal.

If nearest neighbour information is exhausted, select randomly from the current set *S*. The above procedure should be repeated until the archive has been reduced to an acceptable size.

The elitist strategy adopted in this study is a variant on the *universal elitism* approach developed by Zitzler [1999] and is illustrated by the schematic in Figure 10. The key difference is that the archive size is allowed to vary within pre-defined limits, whilst the number of newly generated candidate solutions is varied such that the total population size (elites plus new solutions) is held constant.

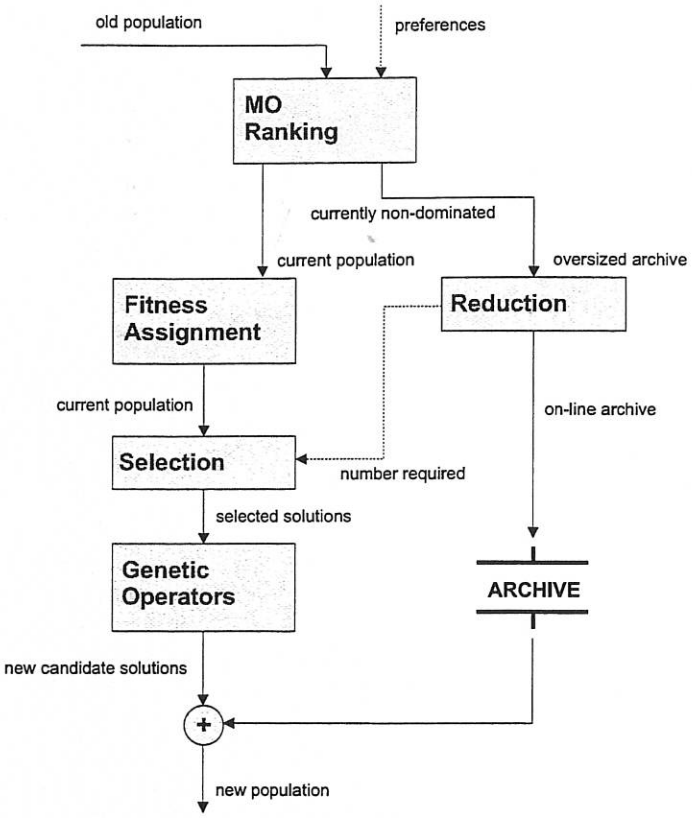

Figure 10: Elitist strategy employed in this study

The on-line archive is initialised to the empty set, whilst the initial population is initialised to a random set of candidate solutions (possibly seeded with information provided by the decision-maker). The populations at subsequent iterations of the algorithm are the combination of new solutions and current elite solutions. The currently non-dominated solutions in the population are identified and are stored as the new, potentially over-sized, archive. Over-represented solutions are then eliminated from the archive, if necessary, using the truncation procedure defined above. For the test problems used in this study, the neighbourhood distance measure is defined as the Euclidean distance between two criterion vectors.

When the new elite set has been finalised, the size of this set is known, and thus the number of new candidate solutions required to fill the population can be calculated. These solutions are created through the selection and genetic manipulation of members of the current population. The new solutions are then combined with the elite set to form the subsequent total population, which completely replaces the old population.

This elitist strategy has been integrated within the baseline MOEA described in the previous section and has been applied to the six benchmark problems. The results of the randomisation testing between the elitist algorithm and the baseline algorithm are shown in Figure 11. Observed differences to the left of the randomisation distribution offer evidence in favour of the elitist version outperforming the baseline case.

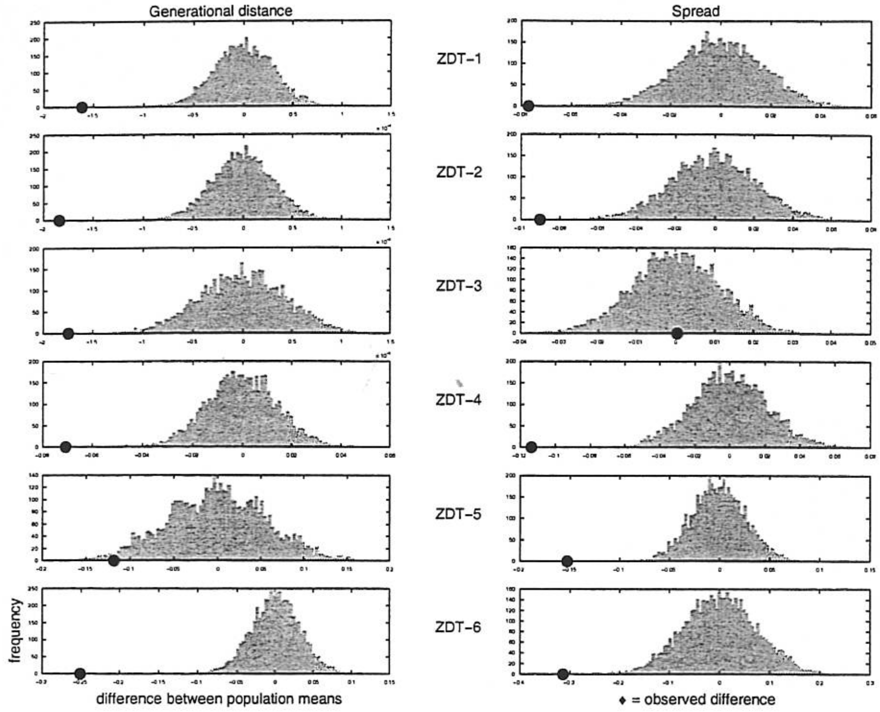

Figure 11: Randomisation test: elitist versus baseline

There is considerable evidence, clearly shown by the results in Figure 11, that the elitist algorithm produces results closer to the true front than the baseline for ZDT-1, 2, 3, 4, and 6. The observed result for ZDT-5 is not significant at the 1%-level, although it would have been significant at the 5%-level. Superior performance in terms of diversity is strongly suggested for ZDT-1, 2, 4, 5, and 6.

The inclusion of elitism increases the convergence speed of the algorithm. The danger of sub-optimal convergence is somewhat reconciled by the distributed nature of the elite set. High-power search operators, such as the Gaussian mutation operator used in this work, can also reduce the risk of premature convergence. Hence, the increased successful convergence exhibited in this study is expected.

The elitism scheme also maintains the characteristics of the currently identified trade-off surface within the on-line population. Thus, diversity of non-dominated solutions in the population is maintained and encouraged (through the thinning of similar criterion vectors) by the truncation mechanism. This helps to explain the improvement in diversity seen in the results. However, the truncation process only represents the current distribution: it does not, directly through fitness, drive the search towards a superior distribution. Despite this, the inclusion of elitism did lead to improved diversity on the non-uniformly distributed ZDT-6. Modifications to the fitness, such as those arising through sharing, may assist further in improving diversity across the trade-off surface. These issues receive further consideration in the next section.

5 Sharing strategy

5.1 Introduction

One of the aims of a multi-objective evolutionary algorithm is to obtain a suitable *distribution* of candidate solutions in regions of interest to the decision-maker. In an evolutionary algorithm, this can be achieved through the formation of sub-population clusters – known as *niches* – within the global population. *Fitness sharing* is the most popular method for fostering this niching process [Goldberg and Richardson, 1987]. In this approach, the raw fitness value of a candidate solution is reduced by a factor dependent on the local population density. This measure should be made in the domain over which a good distribution is of interest: usually criterion-space.

5.2 Parameter-based methods

Fitness sharing has been shown to combat the problem of *genetic drift* (population convergence to a single point due to stochastic selection errors), thus helping to attenuate the possibility of sub-optimal convergence and to enhance coverage of trade-off surfaces. However, the power law equation on which the technique is based requires a definition of *closeness* in order to calculate the population densities. This can be difficult to estimate in practice. Furthermore, the method is sensitive to choice of this *niche size* parameter. Several methods have been proposed in order to estimate the niche size, such as those of Deb and Goldberg [1989] and Fonseca and Fleming [1993], of which the dynamic approach presented by Fonseca and Fleming [1995] is particularly interesting.

Fonseca and Fleming [1995] noted the similarity between the power law sharing function and the *Epanechnikov* kernel density estimator used by statisticians. The kernel smoothing parameter used in the estimator was found to be directly analogous to the fitness sharing niche size parameter. The key benefit of this is that statisticians have developed successful techniques for estimating the value of this parameter [Silverman, 1986]. Furthermore, the approach is amenable to update at each generation of the EA population. This approach can be regarded as parameter-based sharing with automatic tuning. Epanechnikov sharing has been used in several contemporary MOGA applications, although sharing has generally been performed in the decision-space in these instances [Griffin *et al*, 2000; Schroder *et al*, 2001].

Epanechnikov sharing has been added to the baseline MOEA used in this study and has been applied to the benchmark problems. Sharing is performed using the Euclidean distance metric in the criterion domain. Results of a randomisation comparison with the baseline algorithm are shown in Figure 12. Observed values that favour the sharing scheme will lie to the left of the randomisation distribution.

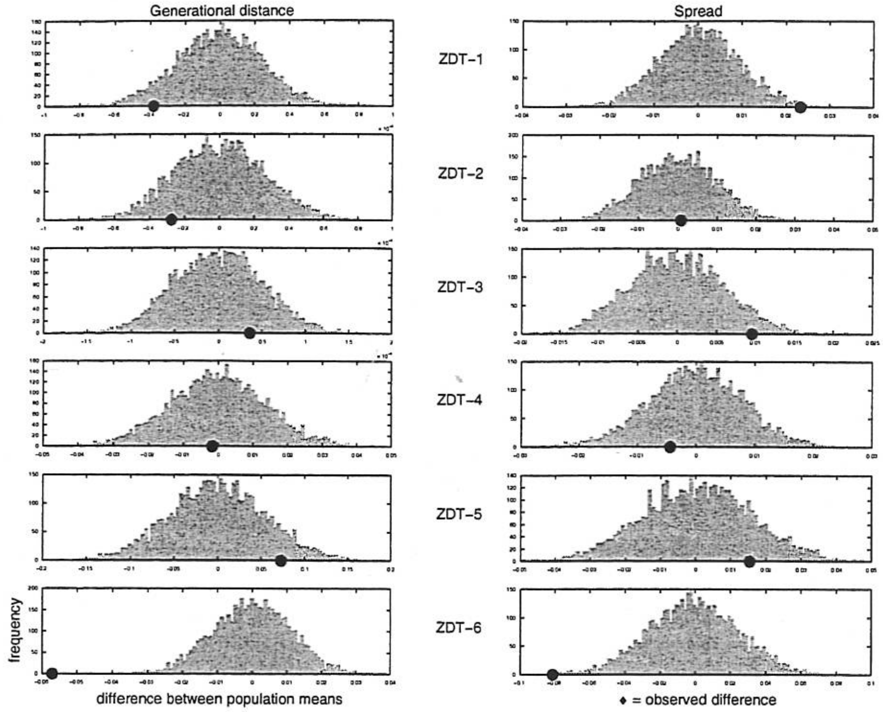

Figure 12: Epanechnikov versus baseline

The inclusion of Epanechnikov sharing has improved both aspects of performance on ZDT-6. The non-uniform nature of this problem should particularly highlight the benefits of a sharing scheme. Note in particular that a method designed to improve diversity has also helped to improve convergence, thus suggesting the strong interaction between the two performance criteria. However, no improvements in either diversity or closeness have been achieved for any other test function. Indeed there is some evidence to suggest a deterioration in diversity on ZDT-1, although this is not significant at the 1%-level. The lack of improvement to diversity is of particular concern, since the elitist results in Section 4 have indicated that diversity *can* be greatly improved on these problems.

5.3 Parameter-less methods

5.3.1 Discussion

The difficulty and inconvenience involved in determining the niche size value has led many researchers to investigate parameter-less methods for achieving niching [Deb *et al*, 2000; Zitzler and Thiele, 1999].

A rank-based niching technique is described herein. No definition of closeness is required. The new ranking increases the resolution of an existing multi-objective ranking. This latter ranking can be obtained using any method, including the popular Fonseca and Fleming [1993, 1998] and Goldberg [1989] methods.

In Fonseca and Fleming's [1998] approach, a candidate solution is ranked according to how many other solutions in the current population are *preferred* to it. Given the minimum amount of preference information (a direction of monotonically increasing preference in each criterion), the comparison is made in terms of pure Pareto dominance. This concept is illustrated in Figure 13. In this simple example, both criteria are to be minimised. Criterion vectors for five candidate solutions {A, ..., E} are shown.

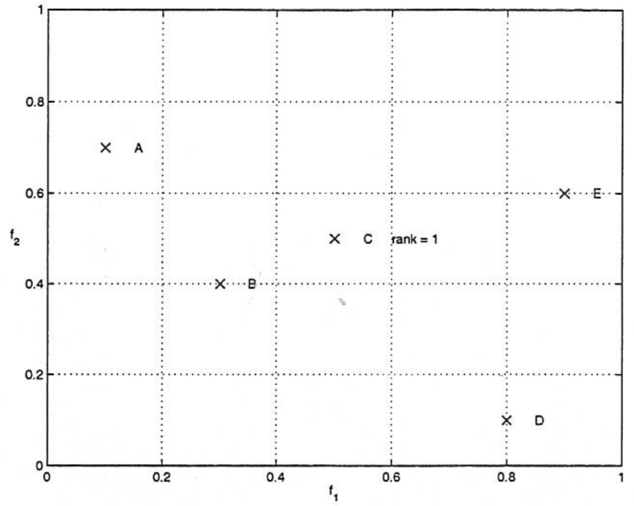

Figure 13: Example of Fonseca and Fleming [1998] multi-objective ranking

A solution is dominated by all other solutions within the hypercube defined by its own criterion vector and the utopian point (in this case $\{0,0\}$). In this example, the domination hypercube for solution C is indicated by the grey rectangle. Solution C is seen to be dominated by solution B alone, and thus receives a multi-objective ranking of 1. The rankings for all solutions are shown in Table 3.

CANDIDATE	CRITERIA $\{f_1, f_2\}$	RANK
A	{0.1 0.7}	0
B	{0.3 0.4}	0
C	{0.5 0.5}	1
D	{0.8 0.1}	0
E	{0.9 0.6}	3

Table 3: Multi-objective ranking for the solutions in Figure 13

The original multi-objective genetic algorithm [Fonseca and Fleming, 1993] uses the *stochastic universal sampling* selection mechanism because of its low stochastic error properties [Baker, 1987]. This technique requires a mapping between ranking and fitness value (in contrast to tournament selection). This is achieved by sorting the population according to rank, assigning fitness according to some function, and then averaging the fitnesses for solutions of the same rank. This process is illustrated in Figure 14. The narrower bars show the pre-averaged fitness values, whilst the wider bars indicate the post-averaged fitnesses.

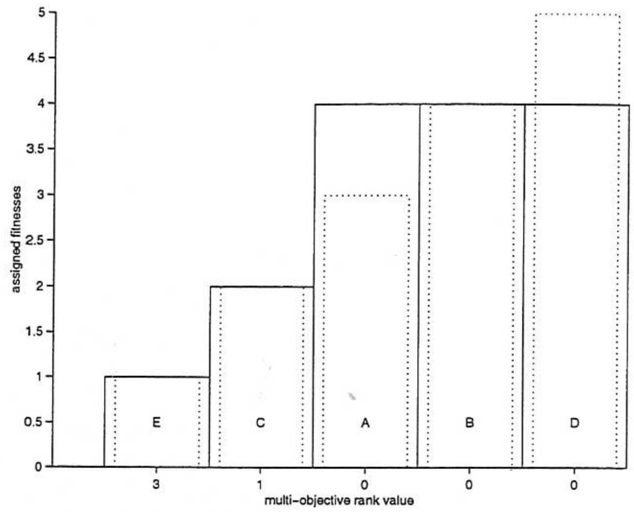

Figure 14: Rank to fitness assignment procedure

The functional mapping between the sorted list and fitness is often either linear or exponential, although other forms are possible. It generally includes a *selective pressure* term that can be used to vary the rate of convergence. Two linear mappings are shown in Equations 4 and 5 below.

$$f(r) = s_1 - (s_1 - 1)\frac{2r}{N-1} \tag{4}$$

$$f(r) = r\frac{s_2 - 1}{N-1} + 1 \tag{5}$$

where f is fitness,

r is the index in the sorted list (starting at zero),

N is the number of candidate solutions,

$s_{1,2}$ is the selective pressure: $1 \le s_1 \le 2$; $1 \le s_2 < \infty$.

The results in Figure 14 have been computed using Equation 5. The selective pressure has been set to the standard choice of $s_2 = N = 5$.

The niching approach presented here increases the resolution of the above ranking procedure through the inclusion of population density information. An *intra-ranking* is performed on candidate solutions of identical multi-objective rank, discriminating on the basis of population density. Solutions in less dense areas receive a superior intra-ranking to their counterparts in denser regions. This approach requires a definition of *distance* but does not require a definition of closeness. The distance metric is likely to be problem dependent and could conceivably contain decision-maker preference information. Following the new fine-grained ranking, the fitness assignment procedure remains unchanged. This new diversity preserving measure is illustrated in Figure 15.

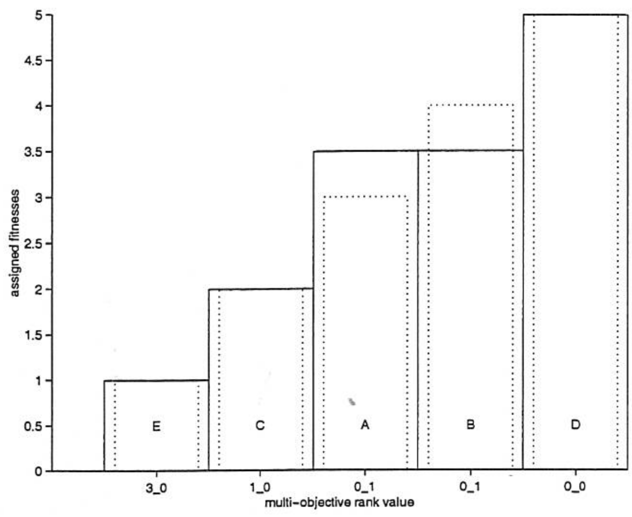

Figure 15: Fitness based on multi-obejctive ranking followed by diversity intra-ranking

The density measure selected is the Euclidean distance to the nearest neighbour in criterion-space (see Figure 13). Solutions A, B, and D all have the same Pareto rank, but solution D is the remotest and thus receives the highest fitness. Solutions A and B are an identical distance apart and thus share the next two available fitnesses equally. Note that A and B still receive a higher fitness than the solutions that they dominate (C and E). The associated fine-grained ranks are shown in Table 4, to the right of the original coarse-grained equivalents.

CANDIDATE	DISTANCE TO NEAREST NEIGHBOUR	COARSE RANK	FINE RANK
A	0.361	0	0_1
B	0.361	0	0_1
C	Infinite	1	1_0
D	0.583	0	0_0
E	Infinite	3	3_0

Table 4: Multi-objective and diversity-based ranking for Figure 15

The proposed new scheme has a number of important properties, namely:

- If one candidate solution is preferred over another, then the former is guaranteed to have a superior fitness value. This was not the case under the old fitness sharing scheme.
- The cumulative fitness assigned to each ranking group remains unchanged.
- When all solutions are currently non-dominated, discrimination is based purely on density.
- When all solutions are currently non-dominated and the population density measure is globally uniform, all fitness values are identical.

With any type of ranking scheme, information content is lost. The ranking indicates that one solution lies in a more densely packed region than another solution but the actual difference in density between the two is lost. This limits the amount of information available to the search procedure but protects against premature convergence to locally *superfit* solutions and removes the requirement for a niche size setting.

5.3.2 Results

The results of randomisation testing, when comparing the new sharing method to the baseline (non-sharing) system are shown in Figure 16. Observed differences between sample means to the left of the randomisation distribution provide evidence in favour of the new scheme.

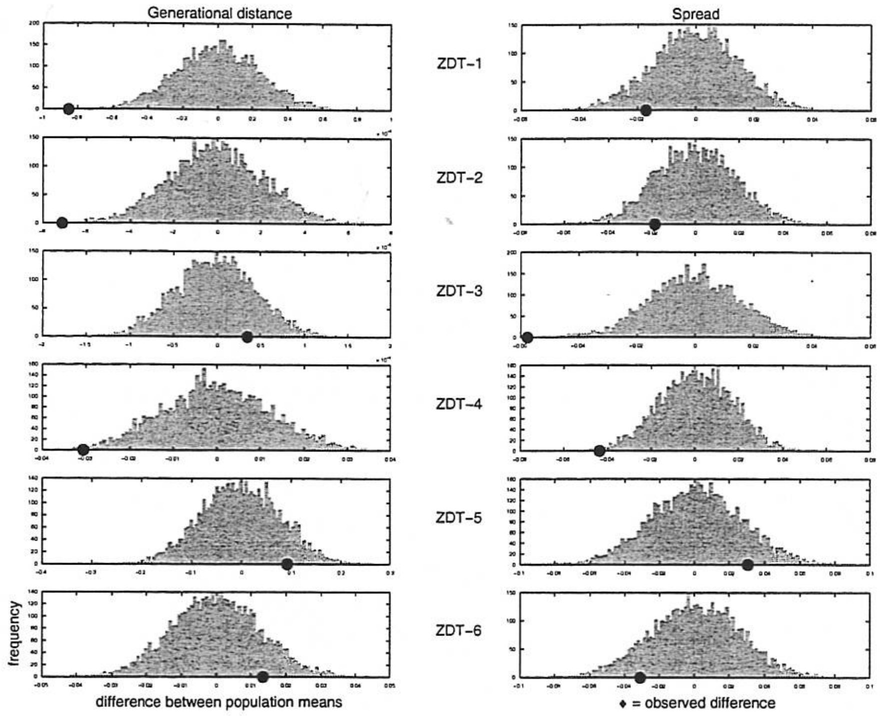

Figure 16: Randomisation testing – New sharing versus no sharing

The central aim of sharing is to improve the distribution of solutions in criterion-space and this should be primarily evident in the spread results. As shown in Figure 16, there is strong evidence to suggest that the new method improved spread on ZDT-3 and ZDT-4. The use of the Epanechnikov kernel, by contrast, did not improve results on these problems. In no cases, was the absence of a sharing mechanism shown to be preferable (whereas there was some evidence in Section 5.2 to suggest that the Epanechnikov kernel could cause a deterioration in diversity). However, there is no evidence to suggest that the use of sharing made any difference to the results for ZDT-6. This is particularly disappointing since this problem has a non-uniform distribution across its trade-off surface: a situation in which sharing is considered a highly appropriate strategy. On a more positive note, the new sharing scheme provided improved closeness on ZDT-1, 2, and 4.

6 Ranking strategy

6.1 Introduction

Many multi-objective evolutionary algorithms that are based on the concept of Pareto dominance use a derivative of one of the following multi-criterion ranking procedures:

- *Non-dominated sorting* (NDS) [Goldberg, 1989]
- *Multi-objective ranking* (MOR) [Fonseca and Fleming, 1993]

Despite some comparative analysis in the literature, there remains much uncertainty over the relative worth of the two methodologies. In this section the empirical performance of both procedures, embedded in a MOGA, is obtained for the Zitzler *et al* [2000] test suite. Previous published comparisons have been between algorithm *brands*, in which it is difficult to decide on exactly what is responsible for the observed discrepancies in performance. By focusing solely on the ranking method, it is hoped that clearer evidence will be produced. The discussion herein is based on pure Pareto dominance, but it should be noted that it is equally applicable to other dominance measures such as *preferability* [Fonseca and Fleming, 1998].

6.2 Non-dominated Sorting

Goldberg [1989, p201] proposed the first Pareto-based treatment of multi-criterion problems using an EA. Given a population of candidate solutions, each with its own criterion vector, Goldberg's method identifies non-dominated *waves* of solutions. Initially, the solutions corresponding to the global (in a population sense) non-dominated criterion vectors are assigned the best rank (denoted as rank zero for consistency herein, although the labelling began at one in the original text). This set of solutions is then temporarily removed from the population and a further check is made for non-dominated solutions. This next set of solutions is assigned the next best rank (rank one) and is again removed from the remaining population. This process is continued until all solutions have been ranked. This process is illustrated in Figure 17. NDS was used by Srinivas and Deb [1994] in the *NSGA*, and subsequently by Deb *et al* [2000] in the *NSGA-II*.

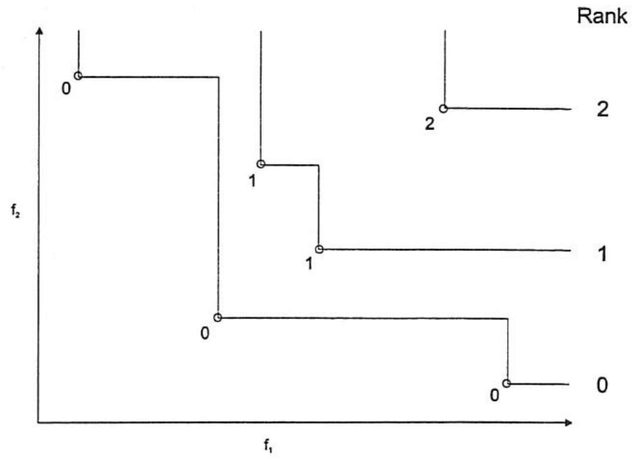

Figure 17: Non-dominated Sorting

6.3 Multi-objective Ranking

Rather than producing a series of non-dominated fronts, Fonseca and Fleming's [1993] method simply ranks a solution according to the number of solutions in the population by which it is dominated. Thus, non-dominated solutions are ranked as zero, whilst the worst possible ranking is the population size minus one. Note that not all possible ranks will necessarily be represented. The method is illustrated, using the same data as for the NDS example, in Figure 18. MOR was used in the original Pareto-based EA, Fonseca and Fleming's [1993] MOGA, many subsequent MOGA applications, and was also recently implemented in a multi-criterion *estimation of distribution* algorithm [Thierens and Bosman, 2001].

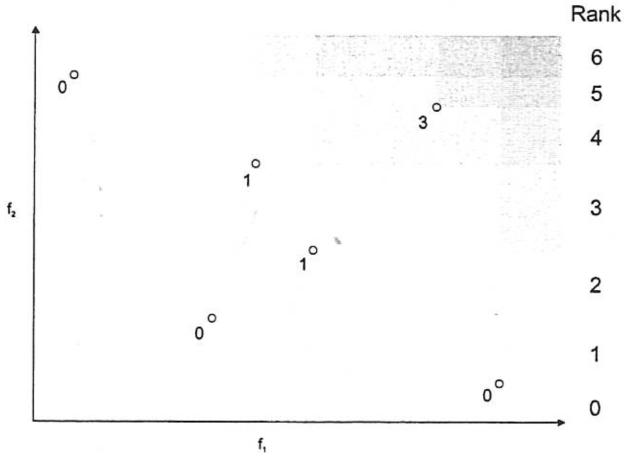

Figure 18: Multi-objective Ranking

6.4 Discussion

In their review of the EMO field, Veldhuizen and Lamont [2000] argue that there is no clear evidence to favour either ranking method overall. MOR is generally regarded as the more efficient method [Coello 1999; Veldhuizen and Lamont, 2000] and has been suggested to be easier to analyse [Fonseca and Fleming, 1997]. MOR has also been found to be the simpler method to extend [Hughes, 2001]. In the only direct empirical comparison of the two schemes, in the context of a single real-world problem, MOR was shown to provide a more accurate trade-off surface [Thomas, 1998].

In essence, MOR provides a more fine-grained ranking than NDS [Horn, 1997]. However, it is arguable whether or not this is a definite benefit. The MOR ranking of a solution describes how many other solutions in the population are preferable to itself, whereas NDS provides only a minimum number. Thus, current population density has more impact in the MOR scheme. This led Deb [2001] to suggest that MOR may be sensitive to the shape of the Pareto front and to the density of solutions in the search space.

Both methods meet the fundamental aims of a multi-criterion ranking strategy: (1) that all preferred individuals are assigned the same rank, and (2) that all individuals are ranked higher than those that they are preferable to. Note that both methods produce identical rankings for a single-criterion problem.

6.5 Evaluation

Randomisation testing results for the two ranking methodologies when integrated within the baseline MOGA are displayed in Figure 19. Similar results for the elitist, sharing MOGA described later in Section 7 are shown in Figure 20. Observed differences to the left of the randomisation distribution favour the NDS technique.

No significant evidence was found on any of the test problems for either performance metric to suggest that one of the ranking schemes was superior to the other. However, it should be noted that the remainder of the MOGA selection algorithm is that which was originally used with MOR, so it is possible that there may be implicit bias towards this ranking procedure.

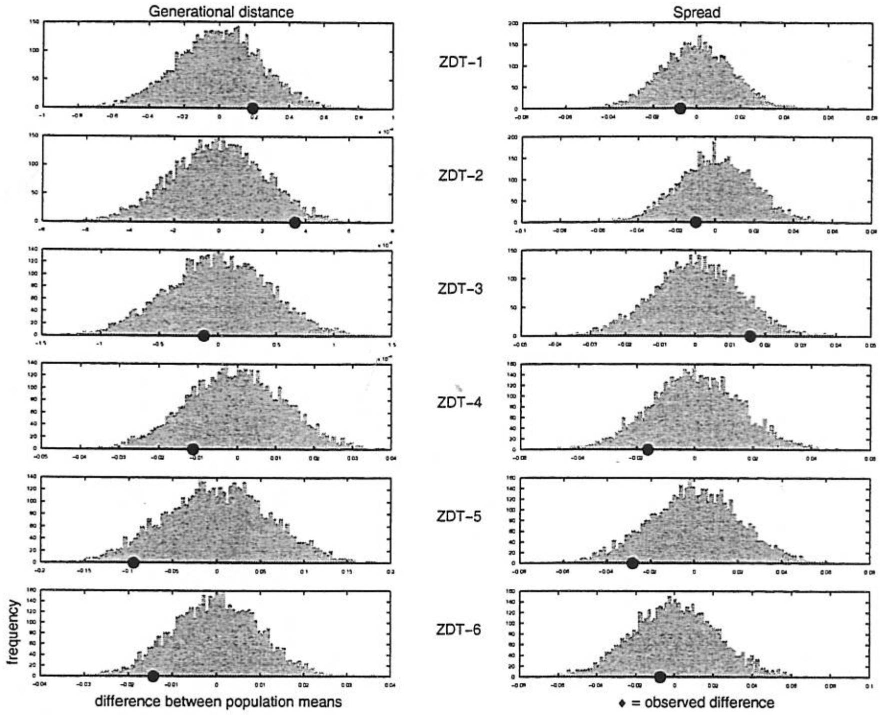

Figure 19: Randomisation testing: NDS MOGA versus MOR MOGA (baseline)

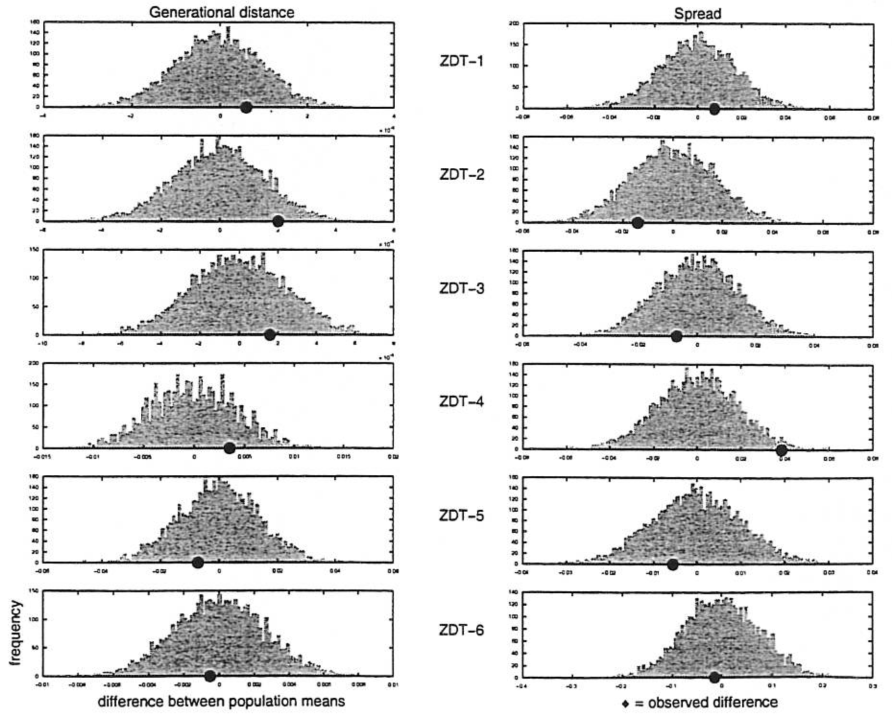

Figure 20: Randomisation testing: NDS MOGA versus MOR MOGA (elitist, sharing)

7 Elitist sharing MOGA

The use of an elitist strategy or a parameter-less sharing strategy in isolation has been shown to offer improved performance in terms of both closeness and diversity. It is instructive to now consider the effect of these schemes in combination. A schematic of the resulting algorithm is shown in Figure 21.

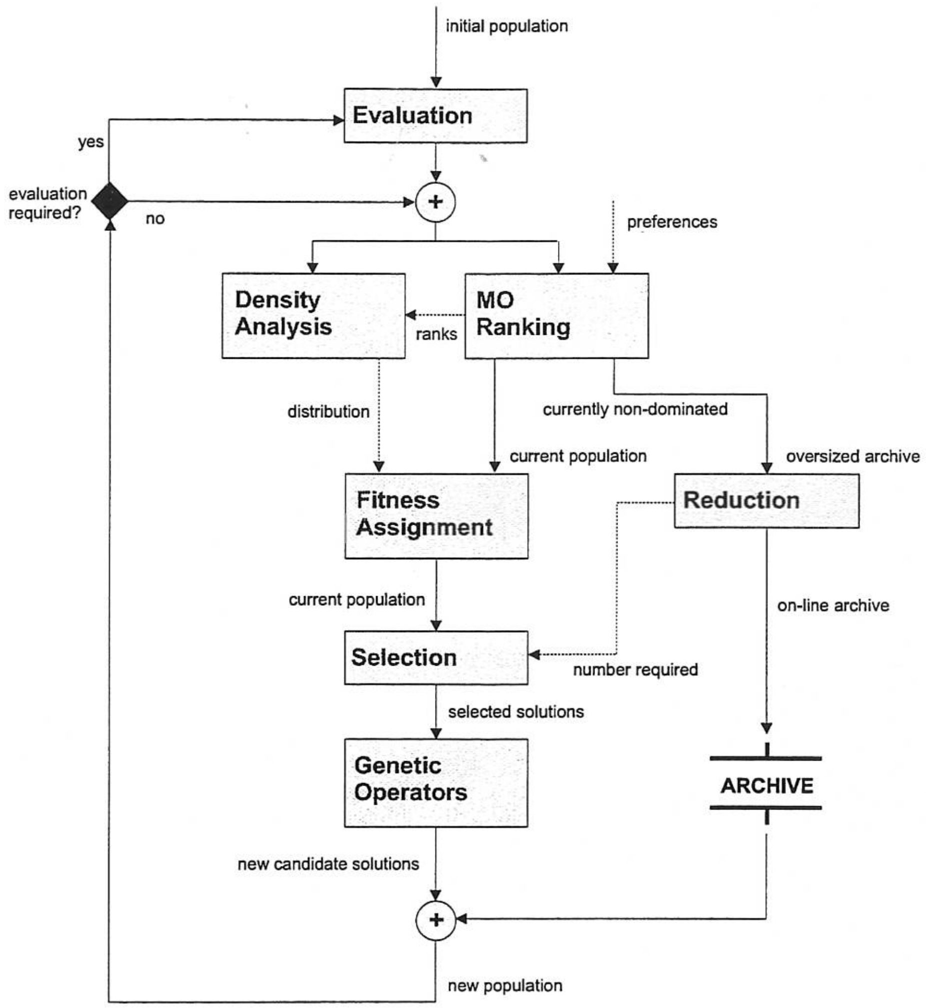

Figure 21: Elitist sharing MOGA schematic

This optimiser has been applied to the problems in the ZDT test suite. The resulting attainment surfaces are shown in Figures 22 through 27.

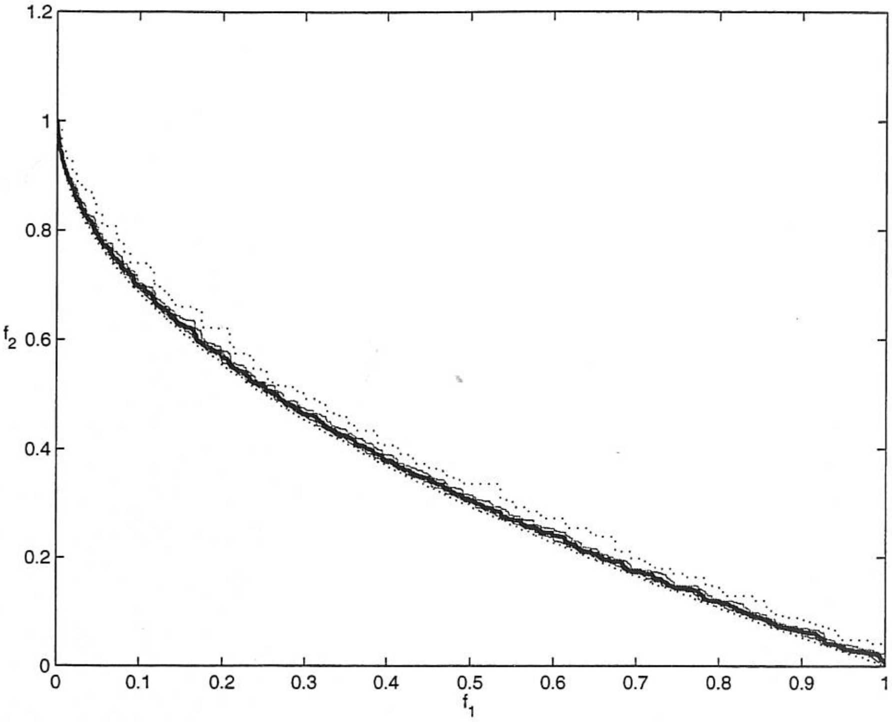

Figure 22: Attainment surfaces – MOGA solving ZDT-1

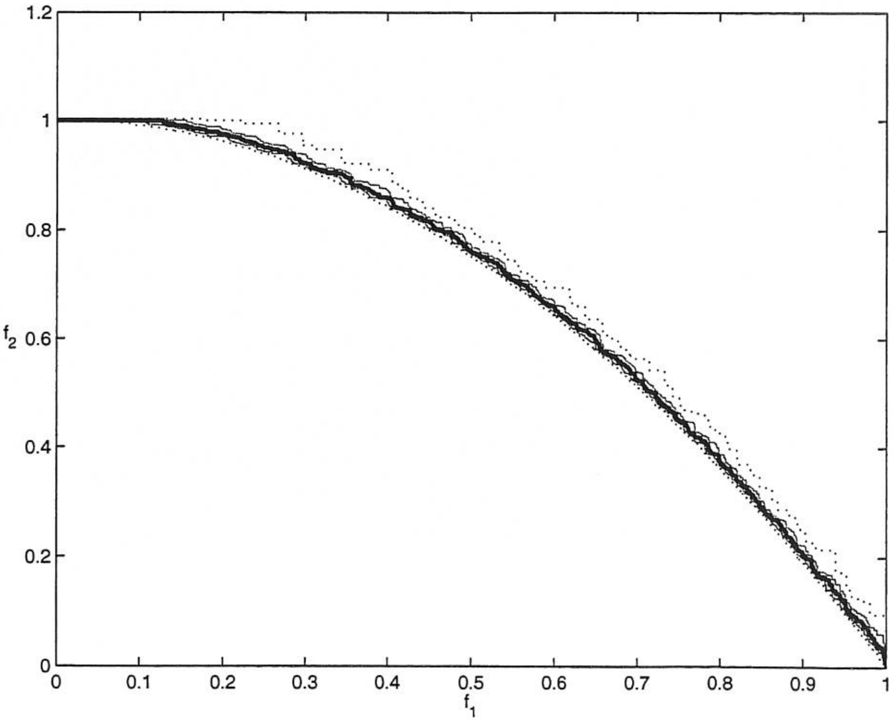

Figure 23: Attainment surfaces – MOGA solving ZDT-2

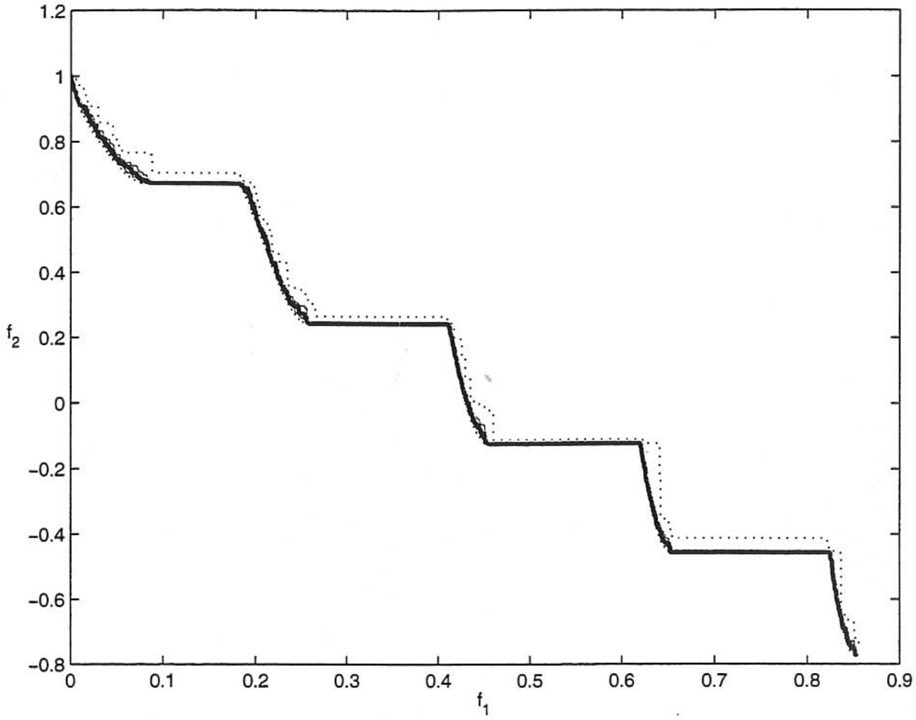

Figure 24: Attainment surfaces – MOGA solving ZDT-3

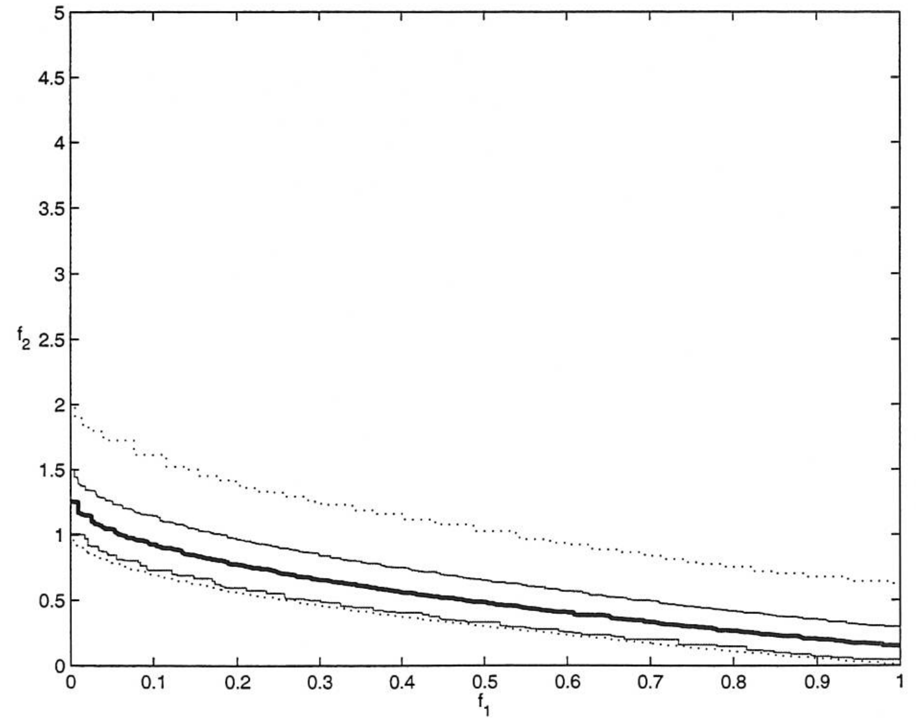

Figure 25: Attainment surfaces – MOGA solving ZDT-4

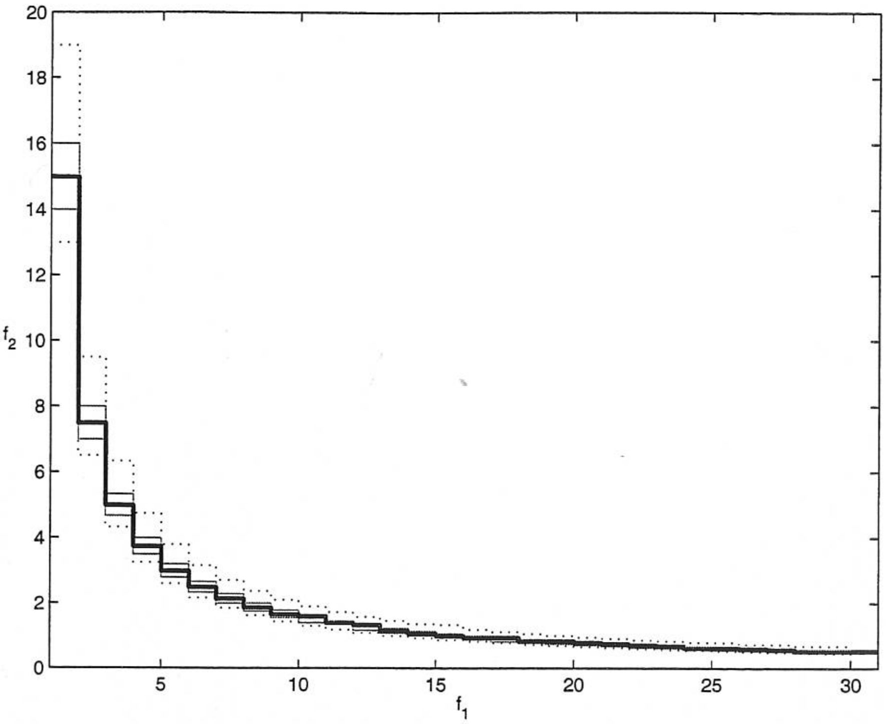

Figure 26: Attainment surfaces – MOGA solving ZDT-5

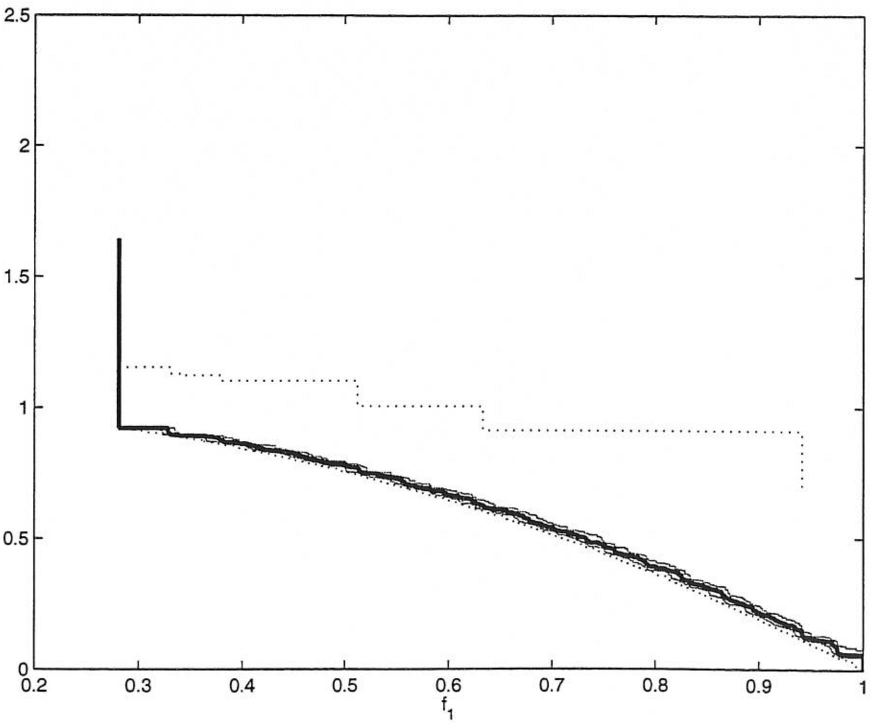

Figure 27: Attainment surfaces – MOGA solving ZDT-6

The envelopes of attainment are generally very tight, indicating good consistency in the results. As evident from Figure 25, closeness has been greatly improved on ZDT-4: indeed the 25%-attainment surface lies very close to the global front of this difficult test problem. Complete coverage of the right-hand section of the trade-off surface has been achieved for ZDT-5, as shown in Figure 26. Finally, closeness and diversity have been much improved on ZDT-6 (see Figure 27).

A comparison with the baseline algorithm is made, via randomisation testing, in Figure 28. Observed differences between the means of each metric that lie to the left of the randomisation distribution favour the elitist, parameter-less sharing algorithm.

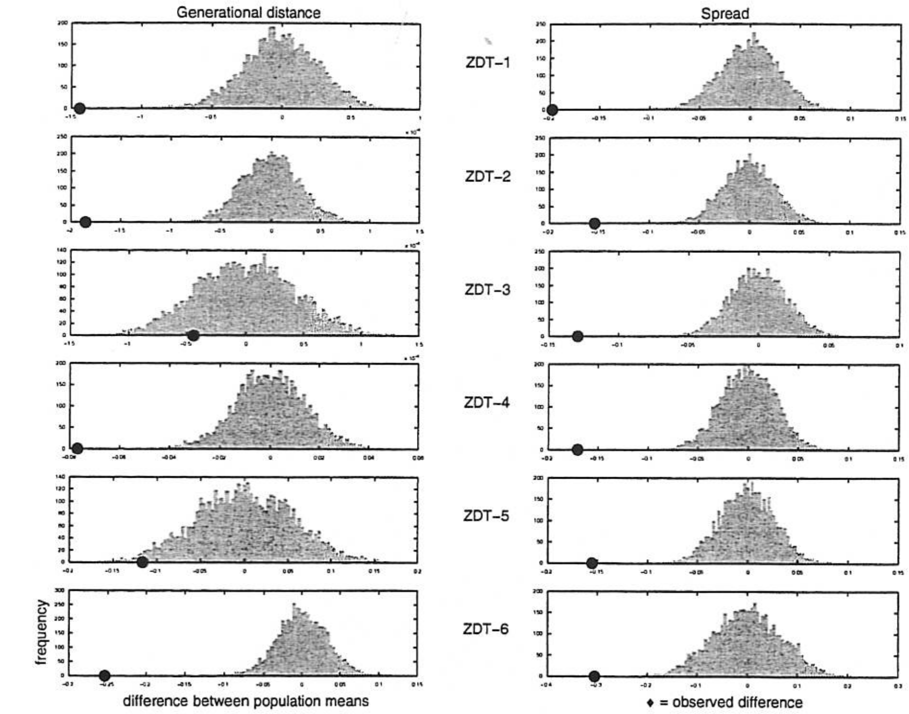

Figure 28: Randomisation testing: elitist sharing MOGA versus baseline MOGA

Compelling evidence points to the new algorithm outperforming the baseline in terms of diversity across all six benchmark problems. Improved closeness was observed for ZDT-1, 2, 4, and 6 (the result for ZDT-5 is not significant at the 1%-level). Of particular note is the improved diversity on ZDT-3. The *combination* of elitism and new sharing was required in order to achieve this significant improvement. Neither elitism nor sharing alone was capable of producing this result. Unfortunately, this benefit has been accompanied by degradation to the closeness results when compared the elitist-only algorithm described in Section 4.

A direct comparison of the combined scheme with the elitist-only MOGA is shown in Figure 29. Observed differences to the left of the randomisation distribution favour the combined scheme. There is substantial evidence that the incorporation of sharing has improved diversity still further on ZDT-1, 2, 3, and 4. As mentioned above, the improved diversity on ZDT-3 has been accompanied by attenuation of closeness but, as indicated in Section 3, the baseline results themselves were particularly good for this test problem.

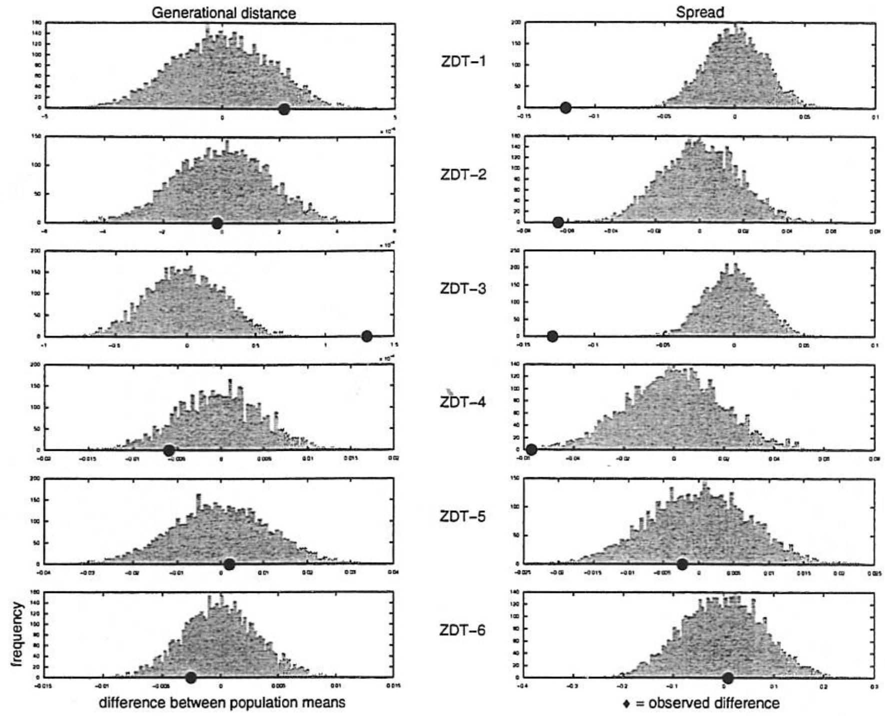

Figure 29: Randomisation testing: elitist sharing MOGA versus elitist MOGA

8 Conclusion

Using a progressive and tractable experimental approach, supported by appropriate statistical and visual analyses, this report has demonstrated that suitable elitist and sharing strategies can significantly improve the performance of an evolutionary multi-criterion optimiser.

An empirical inquiry framework has been introduced for EMO studies, with the aim of identifying and helping to explain the components and interactions that provide good performance. The cornerstone of the framework is that suggested MOEA innovations and modifications are applied in a structured and tractable manner. This will ensure that if, as is often the case, theoretical analysis is particularly difficult then some useful empirical evidence is obtained. Statistical analysis of suitable closeness and diversity performance metrics, obtained via the final populations of multiple optimiser runs, should then be performed. The use of attainment surfaces is suggested as a means for obtaining information-rich, reliable, visualisations.

A baseline MOGA has been introduced and its performance has been obtained. Of particular note is that the algorithm does not include an explicit diversity-promoting mechanism. Thus, a benchmark can be established from which such mechanisms can be assessed. Many existing studies simply compare the relative performance of sharing schemes, without an assessment of whether this represents any improvement over complete inaction. The baseline MOGA has been shown to struggle with diversity preservation on ZDT-5 and ZDT-6, and converged to local non-dominated fronts on every occasion when applied to ZDT-4.

The deployment of an elitist heuristic has again been shown to be highly beneficial, this time using a new experimental framework and in the context of MOGA. Zitzler's [1999] universal elitism scheme, and variants thereof, is both simple and effective. It is possibly the purest EMO elitism technique, with respect to single-criterion EA methodologies. Elitism in this form improves convergence and diversity-*preservation* capabilities. Regarding diversity-*enhancing* mechanisms, some possible shortcomings of the popular parameter-based sharing technique have been exposed, as have the dangers of relying too heavily on an automatic parameter-setting method. A new parameter-less method of sharing has been introduced and has been shown to be more reliable than the standard method. This method is designed to be used in conjunction with a multi-criterion ranking process. It would be interesting to study the similarity of this selection system with a two-step tournament selection mechanism (an initial dominance check, followed by a density comparison if required). Very good results were achieved when both elitism and parameter-less sharing were used together. These results generally improved still further on the elitist strategy in isolation.

As a final word of caution, these results have been obtained for two-criterion problems: further research is required to ascertain the effectiveness of these strategies as the dimension of the problem increases. Future investigations will use the scalable test suite proposed by Deb *et al* [2001] and will consider real-world applications.

The results detailed in this report are available for download from the following site:

http://www.shef.ac.uk/~acse/research/students/r.c.purshouse/

Acknowledgments

The first author gratefully acknowledges support from the UK Engineering and Physical Sciences Research Council. Both authors would like to thank Dr Nick Fieller from the Department of Probability and Statistics at the University of Sheffield for his advice.

References

Baker, J. E., 1987, *Reducing bias and inefficiency in the selection algorithm*, Proceedings of the 2nd International Conference on Genetic Algorithms, pp14-21.

Cleveland, W. S., 1993, *Visualizing Data*, Hobart Press, Summit, NJ.

Coello, C. A. C., 1999, *A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques*, Knowledge and Information Systems. An International Journal, Vol. 1, No. 3, pp269-308.

Deb, K., 2001, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons.

Deb, K., Agrawal, S., Pratab, A., and Meyarivan, T., 2000, *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II,* KanGAL report 200001.

Deb, K., Thiele, L., Laumanns, M., and Zitzler, E., 2001, *Scalable Test Problems for Evolutionary Multi-Objective Optimization*, TIK-Technical Report No. 112, ETH Zürich, Switzerland.

Deb, K. and Goldberg, D. E., 1989, *An Investigation of Niche and Species Formation in Genetic Function Optimization*, Proceedings of the 3rd International Conference on Genetic Algorithms, pp42-50.

De Jong, K. A., 1975, *An analysis of the behavior of a class of genetic adaptive systems*, PhD thesis.

Fonseca, C. M. and Fleming, P. J., 1993, *Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization*, Proceedings of the 5th International Conference on Genetic Algorithms, pp416-423.

Fonseca, C. M. and Fleming, P. J., 1995, *Multiobjective Genetic Algorithms Made Easy: Selection, Sharing and Mating Restriction*, Proceedings of GALESIA '95, pp45-52.

Fonseca, C. M. and Fleming, P. J., 1996, *On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers*, Parallel Problem Solving from Nature – PPSN IV, Lecture Notes in Computer Science, Vol 1141, pp584-593.

Fonseca, C. M. and Fleming, P. J., 1997, *Multiobjective Optimization*, in: Bäck, T., Fogel, D., and Michalewicz, Z. (Eds), Handbook of Evolutionary Computation, Vol. 1, ppC4.5:1-C4.5:9.

Fonseca, C. M. and Fleming, P. J., 1998, *Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms – Part I: A Unified Formulation* and *Part II: Application Example*, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 28, No. 1, pp26-37 and pp38-47.

Griffin, I. A., Schroder, P., Chipperfield, A. J., and Fleming, P. J., 2000, *Multi-objective optimization approach to the ALSTOM gasifier problem*, Proceedings of the Institution of Mechanical Engineers, Vol. 214, Part I, pp453-468.

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

Goldberg, D. E. and Richardson, J., 1987, *Genetic Algorithms with Sharing for Multimodal Function Optimization*, Proceedings of the 2nd International Conference on Genetic Algorithms, pp41-49.

Herrera, F., Lozano, M., and Verdegay, J. L., 1998, *Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioural Analysis*, Artificial Intelligence Review, Vol. 12, pp265-319.

Horn, J., 1997, *Multicriterion Decision Making*, in: Bäck, T., Fogel, D., and Michalewicz, Z. (Eds), Handbook of Evolutionary Computation, Vol. 1, ppF1.9:1-F1.9:15.

Hughes, E. J., 2001, *Evolutionary Multi-objective Ranking with Uncertainty and Noise*, Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001), Lecture Notes in Computer Science, Vol. 1993, pp329-343.

Ikeda, K., Kita, H., and Kobayashi, S., 2001, *Failure of Pareto-based MOEAs: Does Non-dominated Really Mean Near to Optimal?*, CEC 2001, Vol. 2, pp957-962.

Knowles, J. D. and Corne, D. W., 2000, *Approximating the non-dominated front using the Pareto archived evolution strategy*, Evolutionary Computation, Vol. 8, No. 2, pp149-172.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E., 2001a, *On the Convergence and Diversity-Preservation Properties of Multi-Objective Evolutionary Algorithms*, TIK-Report 108, ETH Zürich.

Laumanns, M., Zitzler, E., and Thiele, L., 2001b, *On the Effects of Archiving, Elitism, and Density Based Selection in Evolutionary Multi-objective Optimization*, Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001), Lecture Notes in Computer Science, Vol. 1993, pp181-196.

Manly, B. F. J., 1991, *Randomization and Monte Carlo Methods in Biology*, Chapman and Hall, London.

Michalewicz, Z., 1992, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag.

Michalewicz, Z. and Fogel, D. B., 2000, *How to Solve It: Modern Heuristics*, Springer-Verlag, Berlin.

Purshouse, R. C. and Fleming, P. J., 2001, *The Multi-Objective Genetic Algorithm Applied To Benchmark Problems – An Analysis*, ACSE Research Report 796, http://www.shef.ac.uk/~acse/research/students/r.c.purshouse/reports/report796.pdf.

Schott, J. R., 1995, *Fault Tolerant Design Using Single and Multi-Criteria Genetic Algorithms*, Master's Thesis.

Schroder, P., Green, B., Grum, N., and Fleming, P. J., 2001, *On-line evolution of robust control systems: an industrial active magnetic bearing application*, Control Engineering Practice, Vol. 9, No. 1, pp37-49.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Monographs on Applied Statistics and Probability, Vol. 26, Chapman and Hall.

Srinivas, N. and Deb, K., 1994, *Multi-objective function optimization using non-dominated sorting genetic algorithms*, Evolutionary Computation, Vol. 2, No. 3, pp221-248.

Thierens, D. and Bosman, P. A. N., 2001, *Multi-objective Mixture-based Iterated Density Estimation Evolutionary Algorithms*, Proceedings of GECCO 2001, pp663-670.

Thomas, M. W., 1998, *A Pareto Frontier for Full Stern Submarines via Genetic Algorithm*, PhD thesis.

Veldhuizen, D. A. V., 1999, *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*, PhD thesis.

Veldhuizen, D. A. V. and Lamont, G. B., 2000, *Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art*, Evolutionary Computation, Vol. 8, No. 2, pp125-147.

Zitzler, E., 1999, *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*, PhD thesis.

Zitzler, E., Deb, K., and Thiele, L., 2000, *Comparison of Multiobjective Evolutionary Algorithms: Empirical Results*, Evolutionary Computation, Vol. 8, No. 2, pp173-195.

Zitzler, E., Laumanns, M., and Thiele, L., 2001, *SPEA2: Improving the Strength Pareto Evolutionary Algorithm*, TIK-Report 103, ETH Zürich.

Zitzler, E. and Thiele, L., 1999, *Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach*, IEEE Transactions on Evolutionary Computation, Vol. 3, No. 4, pp257-271.

Appendix

Definitions of the Zitzler *et al* [2000] set of test problems are provided below:

Minimise $\mathrm{T}(\underline{x}) = \{f_1(x_1), f_2(\underline{x})\}$

Subject to $g(x_2,\ldots,x_m) \times h[f_1(x_1), g(x_2,\ldots,x_m)]$

Where $\underline{x} = (x_1,\ldots,x_m)$

ZDT1:

$$f_1(x_1) = x_1$$

$$g(x_2,\ldots,x_m) = 1 + \frac{9}{m-1}\sum_{i=2}^{m} x_i$$

$$h(f_1,g) = 1 - \sqrt{\frac{f_1}{g}}$$

where $m = 30$ and $x_i \in [0,1]$

ZDT2:

as ZDT1 except:

$$h(f_1,g) = 1 - \left(\frac{f_1}{g}\right)^2$$

ZDT3:

as ZDT1 except:

$$h(f_1,g) = 1 - \sqrt{\frac{f_1}{g}} - \left(\frac{f_1}{g}\right)\sin(10\pi\, f_1)$$

ZDT4:

as ZDT1 except:

$$g(x_2,\ldots,x_m) = 1 + 10(m-1) + \sum_{i=2}^{m}\left[x_i^2 - 10\cos(4\pi\, x_i)\right]$$

where $m = 10$, $x_1 \in [0,1]$, and $x_2,\ldots,x_m \in [-5,5]$

ZDT5:

$$f_1(x_1) = 1 + u(x_1)$$

$$g(x_2,\ldots,x_m) = \sum_{i=2}^{m} v[u(x_i)]$$

$$h(f_1,g) = \frac{1}{f_1}$$

where $u(x_i) = \sum_{k=1}^{5}(x_i[k] \cap 1)$,

$$v[u(x_i)] = \begin{cases} 2+u(x_i) & \text{if } u(x_i) < 5 \\ 1 & \text{if } u(x_i) = 5 \end{cases},$$

and $m = 11$, $x_1 \in \{0,1\}^{30}$, and $x_2,\ldots,x_m \in \{0,1\}^5$

ZDT6:

$$f_1(x_1) = 1 - e^{-4x_1}\sin^6(6\pi\, x_1)$$

$$g(x_2,\ldots,x_m) = 1 + \frac{9}{m-1}\left(\sum_{i=2}^{m} x_i\right)^{0.25}$$

$$h(f_1, g) = 1 - \left(\frac{f_1}{g}\right)^2$$

where $m = 10$ and $x_i \in [0,1]$