Use of the WFG Toolkit and PISA for Comparison of MOEAs

Lucas Bradstreet, Luigi Barone, Lyndon While, Simon Huband
School of Computer Science & Software Engineering
The University of Western Australia
35 Stirling Highway
Crawley, Australia, 6009
Email: {lucas, luigi, lyndon, huey}@csse.uwa.edu.au

Philip Hingston
School of Computer & Information Science
Edith Cowan University
2 Bradford Street
Mount Lawley, Australia, 6050
Email: p.hingston@ecu.edu.au

Abstract— **Understanding the behaviour of different optimisation algorithms is important in order to apply the best algorithm to a particular problem. The WFG toolkit was designed to aid this task for multi-objective evolutionary algorithms (MOEAs), offering an easily modifiable framework that allows practitioners the ability to test different features by "plugging" in different forms of transformations. In doing so, the WFG toolkit provides a set of problems that exhibit a variety of different characteristics.**

This paper presents a comparison between two state of the art MOEAs (NSGA-II and SPEA2) that exemplifies the unique capabilities of the WFG toolkit. By altering the control parameters or even the transformations that compose the WFG problems, we are able to explore the different types of problems where SPEA2 and NSGA-II each excel. Our results show that the performance of the two algorithms differ not only on the dimensionality of the problem, but also by properties such as the shape and size of the underlying Pareto surface. As such, the tunability of the WFG toolkit is key in allowing the easy exploration of these different features.

I. INTRODUCTION

There have been several attempts to define test suites and toolkits for testing multi-objective evolutionary algorithms (MOEAs) [1], [2], [3], [4], [5]. However, existing multi-objective test problems do not test a wide range of characteristics and problem features, and are often designed in a hard-wired manner. As such, they do not easily allow an MOEA researcher to easily manipulate problem features, such as shape, dimensionality and complexity. Creation of new test problems to suit the desires of researchers is thus fraught with difficulty as it is not simple to alter existing problems or build new problems within these test suites.

Addressing this problem is the Walking Fish Group (WFG) Toolkit [6], [7] which places an emphasis on allowing test problem designers to construct scalable test problems with any number of objectives, where features such as modality and separability can be customised as required. Uniquely, test problems in the WFG Toolkit are defined in terms of a simple underlying problem that defines the fitness space and a series of composable, configurable transformations that allow the test problem designer to add arbitrarily levels of complexity to the test problem. Problems created by the WFG Toolkit are well defined, are scalable with respect to both the number of objectives and the number of parameters, and have known Pareto optimal sets.

The ultimate goal of this paper is not to provide a comparison between MOEAs that informs researchers of which MOEA to use. Rather, we intend to show how a researcher can use the unique capabilities of the WFG toolkit to test the interaction between specific problem features and MOEAs. The WFG toolkit is suited for this goal as it allows problem designers to easily modify problem features and observe their effect on the MOEA performance. Towards this goal, we have integrated the WFG toolkit into the PISA [8] framework to ease the task of problem generation and analysis.

The next section of the paper introduces the multi-objective terminology used throughout. Section III briefly examines multi-objective test suites and specifies the WFG Toolkit, a configurable toolkit that allows for the construction of scalable, well-behaved test problems. Section IV then describes how the WFG Toolkit can be used to test MOEAs and analyse their results using the PISA framework. We demonstrate how to explore the properties of a test problem through the use of the toolkit.

II. TERMINOLOGY

Consider a multi-objective optimisation problem given in terms of a search space of allowed values of n parameters $x_1, \ldots, x_n$, and a vector of M objective functions $\{f_1, \ldots, f_M\}$ mapping parameter vectors into fitness space. The mapping from the search space to fitness space defines the *fitness landscape*.

In multi-objective optimisation, we aim to find the set of optimal trade-off solutions known as the *Pareto optimal set*. The Pareto optimal set is the set of all Pareto optimal parameter vectors, and the corresponding set of objective vectors is the *Pareto optimal front*. The Pareto optimal set is a subset of the search space, whereas the Pareto optimal front is a subset of the fitness space.

The following types of relationships are useful because they allow us to separate the convergence and spread aspects of sets of solutions for a problem. A *distance parameter* is one that when modified only ever results in a dominated, dominating, or equivalent parameter vector. A *position parameter* is one that

when modified only ever results in an incomparable parameter vector. All other parameters are *mixed parameters*.

When the projection of the Pareto optimal set onto the domain of a single parameter, the parameter optima, is a single value at the edge of the domain, then we call the parameter an *extremal parameter*. If instead the parameter optima clusters around the middle of the domain, then it is a *medial parameter*. Extremal parameters can be unduly favoured by truncation based mutation correction strategies, whereas medial parameters can be favoured by EAs that employ intermediate recombination [9].

III. WFG Toolkit Background

The WFG Toolkit [6], [7] defines a problem in terms of an underlying vector of parameters **x**. The vector **x** is always associated with a simple underlying problem that defines the fitness space. The vector **x** is derived, via a series of transition vectors, from a vector of working parameters **z**. Each transition vector adds complexity to the underlying problem, such as multi-modality and non-separability. The EA directly manipulates **z**, through which **x** is indirectly manipulated.

Unlike previous test suites, such as DTLZ [10] in which complexity is "hard-wired", the WFG Toolkit allows a test problem designer to control, via a series of composable transformations, which features will be present in the test problem. To create a problem, the test problem designer selects several shape functions to determine the geometry of the fitness space, and employs a number of transformation functions that facilitate the creation of transition vectors. Transformation functions must be designed carefully such that the underlying fitness space (and Pareto optimal front) remains intact with a relatively easy to determine Pareto optimal set. The WFG Toolkit provides a variety of predefined shape and transformation functions to help ensure this is the case.

For convenience, working parameters are labelled as either distance or position-related parameters (even if they are actually mixed parameters), depending on the type of the underlying parameter being mapped to.

A. Shape Functions

Shape functions determine the nature of the Pareto optimal front, and map parameters with domain $[0, 1]$ onto the range $[0, 1]$. Each of $h_{1:M}$ must be associated with a shape function. For example, letting $h_1 = \mathrm{linear}_1$, $h_{m=2:M-1} = \mathrm{convex}_m$, and $h_M = \mathrm{mixed}_M$ indicates that h_1 uses the linear shape function, h_M uses the mixed shape function, and all of $h_{2:M-1}$ use convex shape functions.

Table I presents five different types of shape functions.

B. Transformation Functions

Transformation functions map input parameters with domain $[0, 1]$ onto the range $[0, 1]$. All transformation functions take a vector of parameters (called the primary parameters) and map them to a single value. Transformation functions may also employ constants and secondary parameters that further influence the mapping. Primary parameters allow us to qualify working parameters as being position- and distance-related.

There are three types of transformation functions: bias, shift, and reduction functions. Bias and shift functions only ever employ one primary parameter, whereas reduction functions can employ many.

Bias transformations have a natural impact on the search process by biasing the fitness landscape. Shift transformations move the location of optima. In the absence of any shift, all distance-related parameters would be extremal parameters, with optimal value at zero. Shift transformations can be used to set the location of parameter optima (subject to skewing by bias transformations), which is useful if medial and extremal parameters are to be avoided. We recommend that all distance-related parameters be subjected to at least one shift transformation.

Some transformation functions are specified in Table II.

C. The WFG Test Problems

Huband et al. [6], [7] propose the WFG multi-objective test problems (WFG1–WFG9) that focuses on some of the more pertinent problem characteristics.

Table III specifies the properties of WFG1–WFG9. The full construction of these problems can be found in Huband et al. [6], [7]. We will be using these test problems to get performance baselines for MOEAs included in PISA. After doing so, we will derive new test problems from the results of these experiments in order to test hypotheses about MOEA performance on these problems.

IV. Experiments

The aim of a test problem is to determine how different algorithms perform on problems with a variety of characteristics. In particular, it is useful to examine the classes of problems where optimisation algorithms perform badly. These problems can then be used to improve these algorithms or avoid their use when dealing with similar problems.

Fortunately, the WFG toolkit makes this task easy. As the WFG toolkit is scalable in the number of parameters and objectives, and provides a separation between distance and position parameters, we can tune the problem in order to make it more difficult in a variety of ways. Some examples include increasing the number of distance parameters (making the problem more difficult), increasing the number of position parameters (making it more difficult to find a good variety of solutions along the Pareto front) and increasing the number of objectives. Additionally, after determining how different MOEAs perform on these problems, the problems can altered to attempt to determine which properties cause the observed behaviour.

We took the implementation of the WFG toolkit used in Huband et al. [6], [7] and extended it for use as a variator within the PISA framework. The PISA framework allows for easy performance assessment of experimental results and use of a variety of included MOEAs (known in PISA as selectors). Using this setup, we will experiment on the WFG1–WFG9 test

TABLE I

SHAPE FUNCTIONS. IN ALL CASES, $x_1, \ldots, x_{M-1} \in [0, 1]$. A, α, AND β ARE CONSTANTS.

Linear
$$\text{linear}_1(x_1, \ldots, x_{M-1}) = \prod_{i=1}^{M-1} x_i$$
$$\text{linear}_{m=2:M-1}(x_1, \ldots, x_{M-1}) = \left(\prod_{i=1}^{M-m} x_i\right)(1 - x_{M-m+1})$$
$$\text{linear}_M(x_1, \ldots, x_{M-1}) = 1 - x_1$$
When $h_{m=1:M} = \text{linear}_m$, the Pareto optimal front is a linear hyperplane, where $\sum_{m=1}^{M} h_m = 1$.

Convex
$$\text{convex}_1(x_1, \ldots, x_{M-1}) = \prod_{i=1}^{M-1}(1 - \cos(x_i \pi/2))$$
$$\text{convex}_{m=2:M-1}(x_1, \ldots, x_{M-1}) = \left(\prod_{i=1}^{M-m}(1 - \cos(x_i \pi/2))\right)(1 - \sin(x_{M-m+1}\pi/2))$$
$$\text{convex}_M(x_1, \ldots, x_{M-1}) = 1 - \sin(x_1 \pi/2)$$
When $h_{m=1:M} = \text{convex}_m$, the Pareto optimal front is purely convex.

Concave
$$\text{concave}_1(x_1, \ldots, x_{M-1}) = \prod_{i=1}^{M-1} \sin(x_i \pi/2)$$
$$\text{concave}_{m=2:M-1}(x_1, \ldots, x_{M-1}) = \left(\prod_{i=1}^{M-m} \sin(x_i \pi/2)\right)\cos(x_{M-m+1}\pi/2)$$
$$\text{concave}_M(x_1, \ldots, x_{M-1}) = \cos(x_1 \pi/2)$$
When $h_{m=1:M} = \text{concave}_m$, the Pareto optimal front is purely concave, and a region of the hyper-sphere of radius one centred at the origin, where $\sum_{m=1}^{M} h_m^2 = 1$.

Mixed convex/concave ($\alpha > 0$, $A \in \{1, 2, \ldots\}$)
$$\text{mixed}_M(x_1, \ldots, x_{M-1}) = \left(1 - x_1 - \frac{\cos(2A\pi x_1 + \pi/2)}{2A\pi}\right)^\alpha$$
Causes the Pareto optimal front to contain both convex and concave segments, the number of which is controlled by A. The overall shape is controlled by α: when $\alpha > 1$ or when $\alpha < 1$, the overall shape is convex or concave respectively. When $\alpha = 1$, the overall shape is linear.

Disconnected ($\alpha, \beta > 0$, $A \in \{1, 2, \ldots\}$)
$$\text{disc}_M(x_1, \ldots, x_{M-1}) = 1 - (x_1)^\alpha \cos^2(A(x_1)^\beta \pi)$$
Causes the Pareto optimal front to have disconnected regions, the number of which is controlled by A. The overall shape is controlled by α (when $\alpha > 1$ or when $\alpha < 1$, the overall shape is concave or convex respectively, and when $\alpha = 1$, the overall shape is linear); and β influences the location of the disconnected regions (larger values push the location of disconnected regions towards larger values of x_1, and vice versa).

TABLE II

TRANSFORMATION FUNCTIONS. THE PRIMARY PARAMETERS y AND $y_1, \ldots, y_{|\mathbf{y}|}$ ALWAYS HAVE DOMAIN $[0, 1]$. A, B, C, α, AND β ARE CONSTANTS. FOR b_param, $\mathbf{y}'$ IS A VECTOR OF SECONDARY PARAMETERS (OF DOMAIN $[0, 1]$), AND u IS A REDUCTION FUNCTION.

Bias: Polynomial ($\alpha > 0$, $\alpha \neq 1$)
$$\text{b\_poly}(y, \alpha) = y^\alpha$$
When $\alpha > 1$ or when $\alpha < 1$, y is biased towards zero or towards one respectively.

Bias: Flat Region ($A, B, C \in [0, 1]$, $B < C$, $B = 0 \Rightarrow A = 0 \wedge C \neq 1$, $C = 1 \Rightarrow A = 1 \wedge B \neq 0$)
$$\text{b\_flat}(y, A, B, C) = A + \min(0, \lfloor y - B \rfloor)\frac{A(B-y)}{B} - \min(0, \lfloor C - y \rfloor)\frac{(1-A)(y-C)}{1-C}$$
Values of y between B and C (the area of the flat region) are all mapped to the value A.

Bias: Parameter Dependent ($A \in (0, 1)$, $0 < B < C$)
$$\text{b\_param}(y, \mathbf{y}', A, B, C) = y^{B + (C-B)v(u(\mathbf{y}'))}$$
$$v(u(\mathbf{y}')) = A - (1 - 2u(\mathbf{y}'))|\lfloor 0.5 - u(\mathbf{y}') \rfloor + A|$$
A, B, C, and the secondary parameter vector $\mathbf{y}'$ together determine the degree to which y is biased by being raised to an associated power: values of $u(\mathbf{y}') \in [0, 0.5]$ are mapped linearly onto $[B, B + (C-B)A]$, and values of $u(\mathbf{y}') \in [0.5, 1]$ are mapped linearly onto $[B + (C-B)A, C]$.

Shift: Linear ($A \in (0, 1)$)
$$\text{s\_linear}(y, A) = \frac{|y - A|}{|\lfloor A - y \rfloor + A|}$$
A is the value for which y is mapped to zero.

Shift: Deceptive ($A \in (0, 1)$, $0 < B \ll 1$, $0 < C \ll 1$, $A - B > 0$, $A + B < 1$)
$$\text{s\_decept}(y, A, B, C) = 1 + (|y - A| - B) \times$$
$$\left(\frac{\lfloor y - A + B \rfloor(1 - C + \frac{A-B}{B})}{A - B} + \frac{\lfloor A + B - y \rfloor(1 - C + \frac{1-A-B}{B})}{1 - A - B} + \frac{1}{B}\right)$$
A is the value at which y is mapped to zero, and the global minimum of the transformation. B is the "aperture" size of the well/basin leading to the global minimum at A, and C is the value of the deceptive minima (there are always two deceptive minima).

Shift: Multi-modal ($A \in \{1, 2, \ldots\}$, $B \geq 0$, $(4A + 2)\pi \geq 4B$, $C \in (0, 1)$)
$$\text{s\_multi}(y, A, B, C) = \frac{1 + \cos\left[(4A+2)\pi\left(0.5 - \frac{|y-C|}{2(\lfloor C - y \rfloor + C)}\right)\right] + 4B\left(\frac{|y-C|}{2(\lfloor C - y \rfloor + C)}\right)^2}{B + 2}$$
A controls the number of minima, B controls the magnitude of the "hill sizes" of the multi-modality, and C is the value for which y is mapped to zero. When $B = 0$, $2A + 1$ values of y (one at C) are mapped to zero, and when $B \neq 0$, there are $2A$ local minima, and one global minimum at C. Larger values of A and smaller values of B create more difficult problems.

Reduction: Weighted Sum ($|\mathbf{w}| = |\mathbf{y}|$, $w_1, \ldots, w_{|\mathbf{y}|} > 0$)
$$\text{r\_sum}(\mathbf{y}, \mathbf{w}) = \left(\sum_{i=1}^{|\mathbf{y}|} w_i y_i\right) / \sum_{i=1}^{|\mathbf{y}|} w_i$$
By varying the constants of the weight vector $\mathbf{w}$, EAs can be forced to treat parameters differently.

Reduction: Non-separable ($A \in \{1, \ldots, |\mathbf{y}|\}$, $|\mathbf{y}| \bmod A = 0$)
$$\text{r\_nonsep}(\mathbf{y}, A) = \frac{\sum_{j=1}^{|\mathbf{y}|}\left(y_j + \sum_{k=0}^{A-2}|y_j - y_{1+(j+k) \bmod |\mathbf{y}|}|\right)}{\frac{|\mathbf{y}|}{A}\lceil A/2 \rceil(1 + 2A - 2\lceil A/2 \rceil)}$$
A controls the degree of non-separability (noting that $\text{r\_nonsep}(\mathbf{y}, 1) = \text{r\_sum}(\mathbf{y}, \{1, \ldots, 1\})$).

problems. These problems exhibit a variety of characteristics that should stress optimisation algorithms. NSGA-II [11] and SPEA2 [12], two standard MOEAs implemented in PISA, were run on the WFG test problems.

PISA was used for the performance assessment of selectors on all problems. We used the epsilon and hypervolume metrics to evaluate the final Pareto fronts generated by the MOEA selectors. These metrics use different properties of the final Pareto front found by a selector and provides a single value measure for the set. The hypervolume measure

TABLE III

PROPERTIES OF THE WFG PROBLEMS. ALL WFG PROBLEMS ARE SCALABLE, HAVE NO EXTREMAL NOR MEDIAL PARAMETERS, HAVE DISSIMILAR PARAMETER DOMAINS AND PARETO OPTIMAL TRADEOFF MAGNITUDES, HAVE KNOWN PARETO OPTIMAL SETS, AND CAN BE MADE TO HAVE A DISTINCT MANY-TO-ONE MAPPING FROM THE PARETO OPTIMAL SET TO THE PARETO OPTIMAL FRONT BY SCALING THE NUMBER OF POSITION PARAMETERS.

Problem	Obj.	Separability	Modality	Bias	Geometry
WFG1	$f_{1:M}$	separable	uni	polynomial,flat	convex, mixed
WFG2	$f_{1:M-1}$	non-separable	uni	—	convex, disconnected
	f_M	non-separable	multi		
WFG3	$f_{1:M}$	non-separable	uni	—	linear, degenerate
WFG4	$f_{1:M}$	separable	multi	—	concave
WFG5	$f_{1:M}$	separable	deceptive	—	concave
WFG6	$f_{1:M}$	non-separable	uni	—	concave
WFG7	$f_{1:M}$	separable	uni	parameter dependent	concave
WFG8	$f_{1:M}$	non-separable	uni	parameter dependent	concave
WFG9	$f_{1:M}$	non-separable	multi,deceptive	parameter dependent	concave

(or S-metric), due to Zitzler and Thiele [13], calculates the volume 'contained' by the solutions in n-dimensional space. The epsilon metric, due to Zitzler et al. [14], calculates the shift necessary for one solution set to be worse in all objectives than another solution set (i.e. dominated). Mann-Whitney U statistical significance tests were used where necessary. As we do not have separate samples for each indicator, the resulting p-values do not reflect true probabilities. See [15] for details.

A. Experimental Setup

All experiments were run using the PISA framework. The WFG toolkit was implemented as a variator within PISA.

The WFG PISA variator used real-parameter SBX [16] crossover with probability 1.0 and $\eta_c = 10$ and variable-wise polynomial mutation probability of $\frac{1}{n}$ where n is the sum of the number of position and distance parameters. MOEAs (referred to as selectors in the PISA framework) were run using a (100+100) population scheme for 750 generations. All experiments were repeated 40 times using the NSGA-II and SPEA2 selectors.

B. Experiment 1

1) Aim and Experimental Setup: The initial experiment was run using the WFG toolkit with 20 distance parameters and 4 position parameters in 2 objectives under the standard WFG1–WFG9 test problems. This experiment is intended to determine the baseline performance for the NSGA-II and SPEA2 selectors to provide a comparison for later experiments.

2) Results: Figures 1(a)-1(r) show box-and-whisker plots that demonstrate that in most cases NSGA-II bests SPEA2. These plots show the hypervolume and epsilon values for the different selectors relative to the hypervolume and epsilon values of the entire set of solutions achieved by all of the selectors. Therefore, smaller values are better.

We can see that in almost all cases NSGA-II is as good or better than SPEA2. For WFG4, under the epsilon metric, SPEA2 has a smaller median value and it is difficult to visually determine which MOEA is the best. In this case, we use the Mann-Whitney statistical test which concludes that NSGA-II is better than SPEA2 for the epsilon metric with a 0.235 confidence. Even though several of the comparisons do not have high statistical confidence, it does not seem a stretch to conclude that NSGA-II is able to provide least as good results

for the WFG problems with these parameters and in many cases is able to provide better results.

We can see that the results for the hypervolume metric on the WFG9 problem are over a large range. This could reflect that WFG9 is difficult problem, given its high degree of non-separability and parameter dependence.

C. Experiment 2

1) Aim and Experimental Setup: As Section IV-B provides a baseline for the performance of SPEA2 and NSGA-II, we would now like to experiment with some of the parameter options provided by the WFG toolkit. In this case we will change the number of position parameters which relate to a solutions position on the Pareto front. Unlike the DTLZ [4] test problems, WFG has additional parameters that allow us to experiment with how different selectors perform and stresses a selectors ability to find a good spread of solutions along the Pareto front. These experiments were run with 20 distance parameters and 12 position parameters in 2 objectives.

2) Results: An increase in the number of position parameters does not lead to a dramatic improvement for SPEA2 when compared to NSGA-II. The results for experiment 2 largely mimic those of experiment 1 and therefore we omit most plots from this paper. However, one exception is found in WFG6. Box plots shown in Figures 2(a) and 2(b) demonstrate that as a result of additional position parameters SPEA2 is able to beat NSGA-II. However, the Mann-Whitney U statistical test shows that SPEA2 only performs better than NSGA-II on WFG6 with 0.182 confidence when compared using the hypervolume metric. For the remaining test problems NSGA-II is able to thoroughly beat SPEA2, although in the case of WFG1 and WFG4 it does not do so with an acceptable confidence. Despite this, it seems reasonable to conclude that the number of position parameters does not change the conclusion that NSGA-II is the algorithm of choice for the WFG problems for 20 distance parameters in 2D.

Though these results are similar to the results in experiment 1, this is unsurprising as NSGA-II and SPEA2 selectors operate in the objective space. These findings suggest that despite an increase in the difficulty in achieving good coverage of the Pareto front, NSGA-II's selection mechanism still works well in comparison to SPEA2.
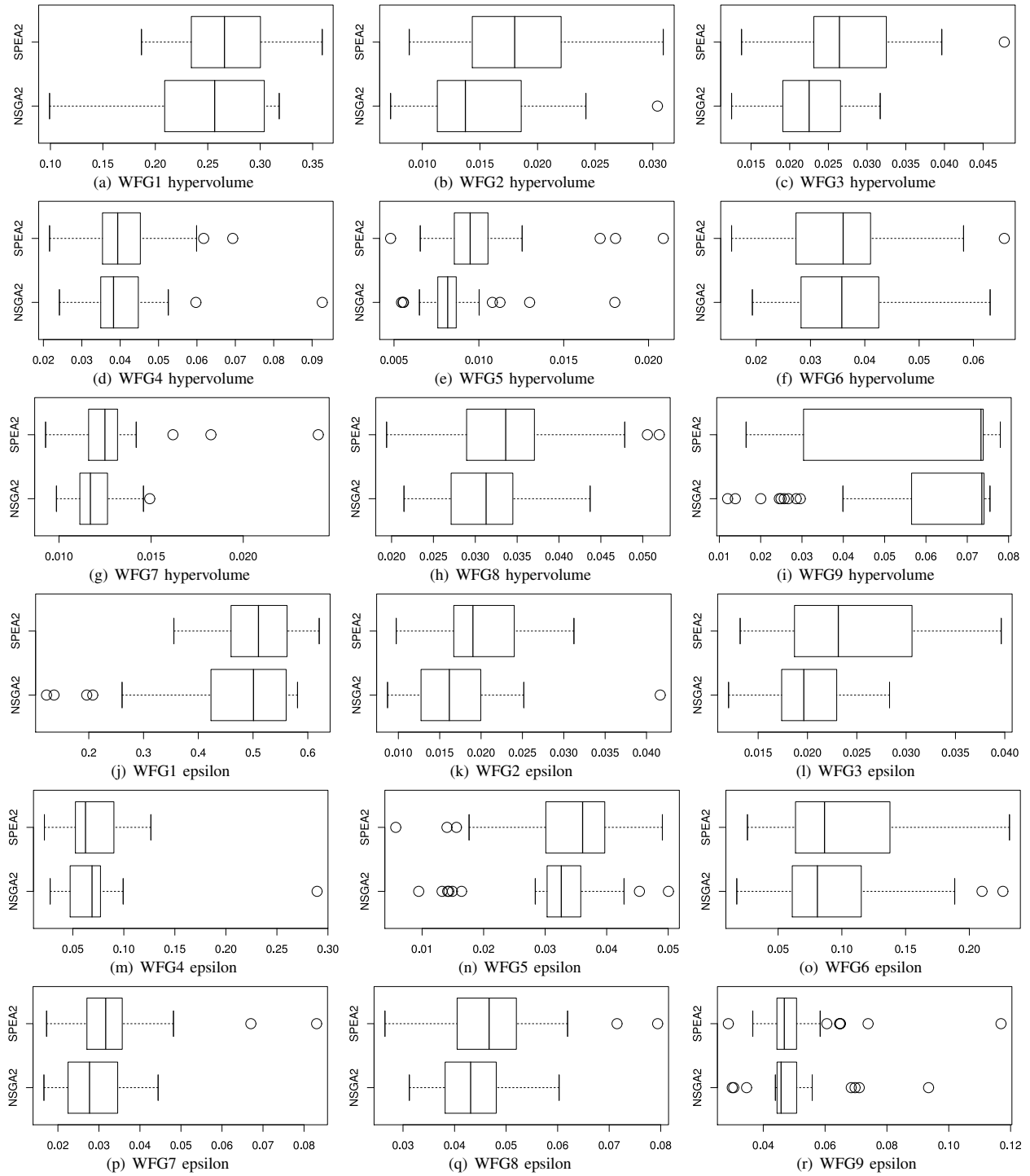
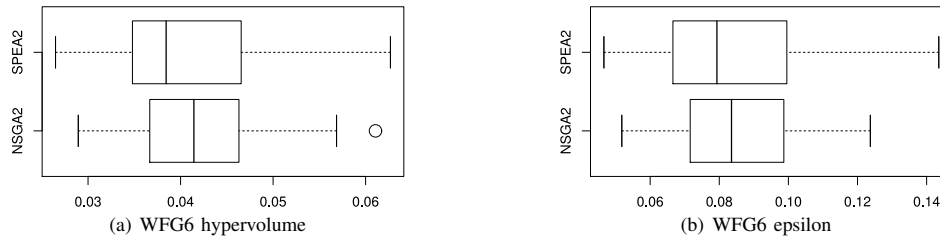Fig. 1. WFG: 4 position parameters 2D, hypervolume and epsilon plots

(a) WFG6 hypervolume

(b) WFG6 epsilon

Fig. 2.   WFG: 12 position parameters 2D, hypervolume and epsilon plots

D. Experiment 3

1) Aim and Experimental Setup: As we have now tested SPEA2 and NSGA-II's performance on the WFG toolkit problems in two objectives, it is important to compare their performance in more than two objectives. Experiment 3 aims to compare the performance of NSGA-II and SPEA2 on WFG1–WFG9 in three objectives. 20 distance and 12 position parameters are used as in experiment 3.

2) Results: Figures 3(a)-3(j) show a dramatic change in the results compared to those achieved in two dimensions. Whereas NSGA-II was able to perform as well or better than SPEA2 for all problems in two objectives, in three objectives SPEA2 bests NSGA-II on almost all problems for both metrics. The exceptions are WFG3, where NSGA-II is able to outperform SPEA2 for both metrics, and WFG6 where NSGA-II wins for the epsilon metric. However, for WFG6 epsilon, the Mann-Whitney U test gives a p-value of 0.473 so any advantage that NSGA-II has for this problem is not significant.

Overall, these results suggest that NSGA-II's crowdedness comparison operator is not as effective on the WFG problems in 3 objectives as it is for 2 objectives. The results mimic those found by Zitzler et al. [12] which concluded that SPEA2 has advantages over NSGA-II in higher dimensional spaces when tested on the DTLZ problems. However, we will see in the next section that the WFG toolkit allows us to explore MOEA performance using techniques that are not available using the DTLZ test suite. Using the toolkit we can change the properties of the test problem to further explore observed behaviour.

E. Experiment 4

1) Aim and Experimental Setup: As seen in Section IV-D, the WFG3 problem was an exception to SPEA2's overall dominance on the WFG problems in three objectives. One theory that explains NSGA-II's ability to outperform SPEA2 on WFG3 with 3 objectives is that WFG3's linear shape favours the NSGA-II selection mechanism. In order to test this hypothesis, we would like to test a similar problem without a linear shape. By altering specific properties of a test problem we can determine which properties most influence an MOEA's performance for that problem. In this case, we believe that shape could be the primary factor causing NSGA-II to outperform SPEA2. Luckily the WFG toolkit makes this assertion easy to test.

As test problems in the WFG toolkit are composed of a series of transformations, including a module for the shape of the Pareto front, there are two simple ways to test the above hypothesis. The first way is by applying an alternative shape transformation to WFG3 to determine whether NSGA-II loses its advantage. The second is to use the linear shape from WFG3 on another problem with an alternative shape.

For these experiments, we have altered WFG3 to use the concave shape transformation as in WFG4, and have applied the linear shape transformation from WFG3 to WFG8. We use the same WFG parameters as in Section IV-D.

2) Results: Figures 4(a)-4(f) confirm our assertion that the linear shape favours NSGA-II. These plots show a total reversal in the results shown in figures 3(a)-3(j) for experiment 3, as a result of alteration of the shape transformation used.

The fact that we can easily perform such experiments is an excellent example of where the WFG toolkit excels compared to the DTLZ and ZDT test problems. The DTLZ and ZDT test problems are hard-wired and are not easily "tweaked" in this way and thus it is difficult to see how different problem properties affect MOEA performance.

V. CONCLUSION AND FUTURE WORK

The WFG toolkit is intended to make it relatively simple to test MOEAs on a wide range of problem types, characteristics, and problem parameters. We have tested NSGA-II and SPEA2 over a range of position parameters and objectives on the standard WFG1–WFG9 test problems.

In doing so, we discovered in Section IV-B that NSGA-II performs better than SPEA2 on the WFG test problems in two objectives. Digging deeper, in Section IV-C we looked at how increasing the number of position parameters, an option provided by the WFG toolkit, affects an optimiser's ability to find and retain good coverage of the Pareto front. While this did not lead to a large change in the results from the first experiments, it did cause a reversal in results for the WFG6 test problem.

In Section IV-D we looked at how an increase in the number of objectives affects an MOEAs performance. In doing so, we saw a drastic change in SPEA2's performance in comparison to NSGA-II. Whereas SPEA2 previously lost to NSGA-II in almost all cases it now dominates on all problems except for WFG3.

SPEA2 and NSGA-II's performance on the WFG3 problem in Section IV-D gave starkly different results compared to
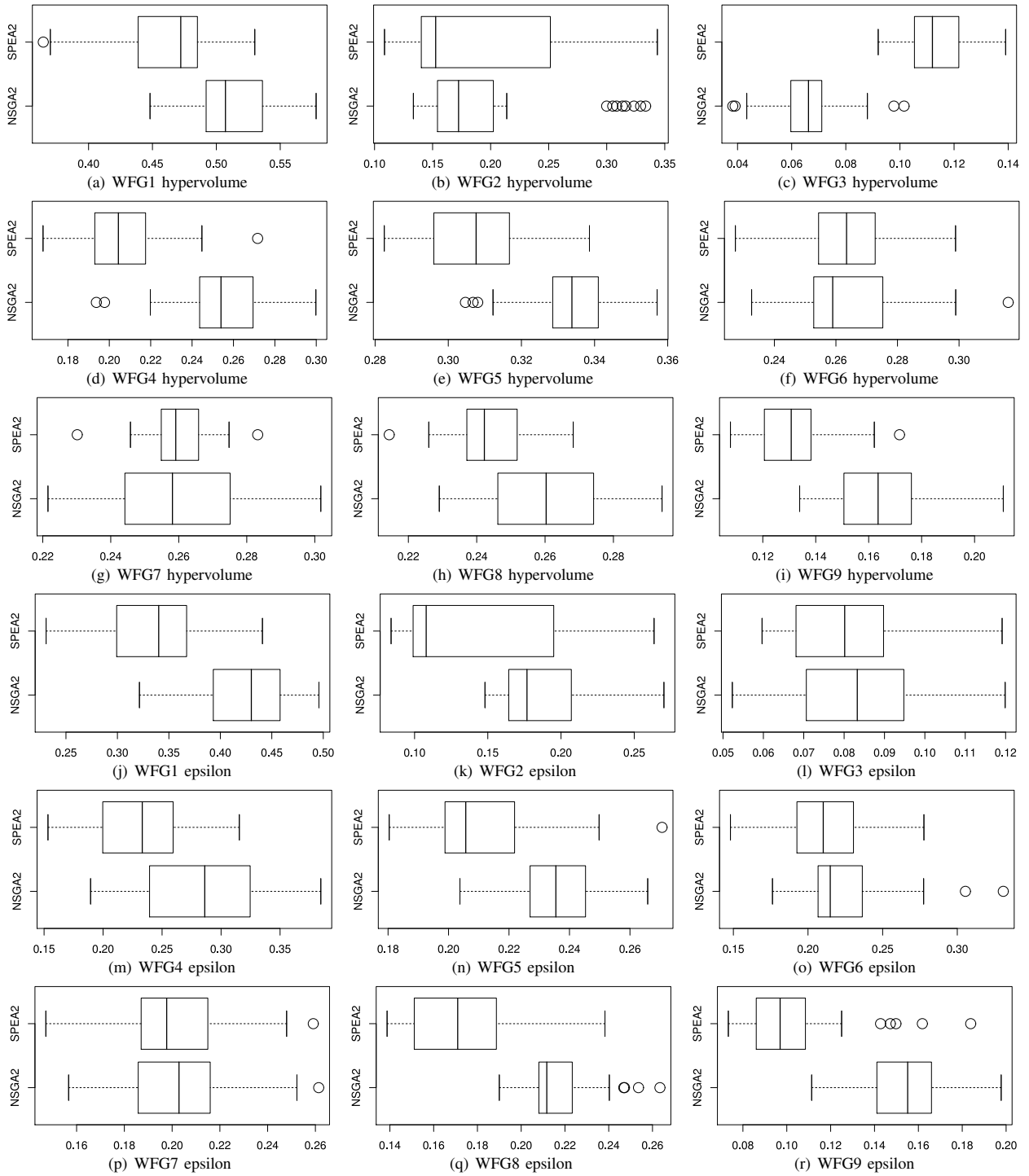
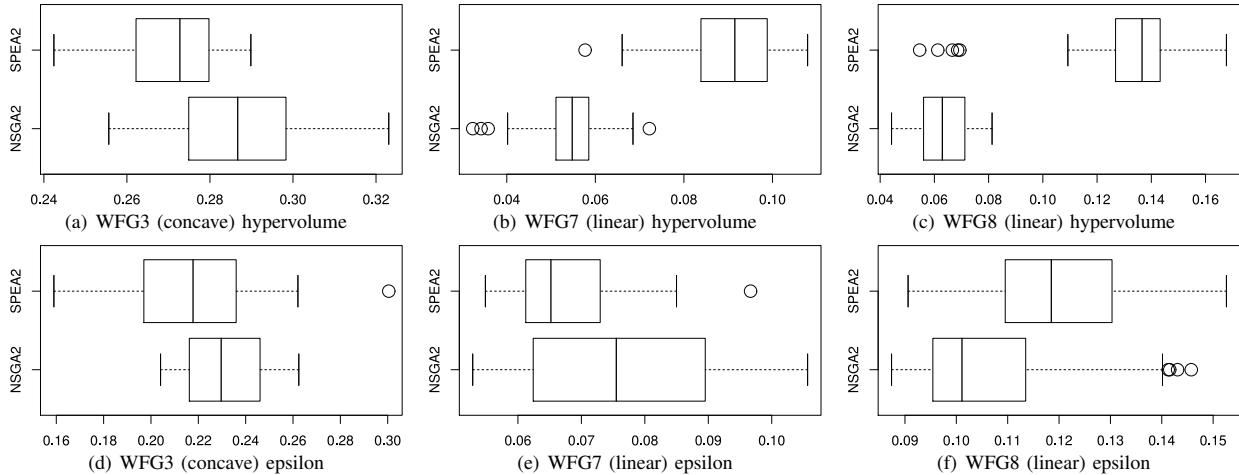Fig. 3.   WFG: 12 position parameters 3D, hypervolume and epsilon plots

Fig. 4.   WFG modified test problems: 12 position parameters 3D, hypervolume and epsilon plots

the other WFG problems. Such results warranted further investigation, and ultimately we would like to determine the properties of WFG3 that cause NSGA-II to outperform SPEA2 on this problem. We modified several of the test problems by using different shape transformations to demonstrate that it is WFG3's linear shape causing NSGA-II to dominate. This task was performed by swapping the shape modules used in WFG3 with those used in WFG7 and WFG8. This is a trivial example of how the WFG toolkit can be used to shed light on the causes of observed MOEA performance.

The integration of the WFG toolkit within the PISA framework makes it is easy to test new MOEAs on standard benchmark problems or to generate new test problems. In addition, one can refine problem characteristics that cause MOEAs difficulty. More substantially, the problems themselves can even be modified to obtain similar problems with alternative properties so we can further explore MOEA performance. These capabilities of the WFG toolkit allow us to easily determine where common MOEAs have problems and simplifies the creation of good benchmark test problems for the future.

REFERENCES

[1] K. Deb, "Multi-Objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems," *Evolutionary Computation*, vol. 7, no. 3, pp. 205–230, Fall 1999.

[2] D. A. V. Veldhuizen and G. B. Lamont, "Multiobjective Evolutionary Algorithm Test Suites," in *Proceedings of the 1999 ACM Symposium on Applied Computing*, J. Carroll, H. Haddad, D. Oppenheim, B. Bryant, and G. B. Lamont, Eds. San Antonio, Texas: ACM, 1999, pp. 351–357.

[3] E. Zitzler, K. Deb, and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, Summer 2000.

[4] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable Test Problems for Evolutionary Multi-Objective Optimization," Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, Tech. Rep. 112, 2001.

[5] T. Okabe, Y. Jin, M. Olhofer, and B. Sendhoff, "On Test Functions for Evolutionary Multi-objective Optimization," in *Parallel Problem Solving from Nature - PPSN VIII*, X. Y. et al., Ed. Birmingham, UK: Springer-Verlag. Lecture Notes in Computer Science Vol. 3242, September 2004, pp. 792–802.

[6] S. Huband, L. Barone, L. While, and P. Hingston, "A Scalable Multiobjective Test Problem Toolkit," in *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Guanajuato, México: Springer. Lecture Notes in Computer Science Vol. 3410, March 2005, pp. 280–295.

[7] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, October 2006.

[8] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, "PISA—A Platform and Programming Language Independent Interface for Search Algorithms," in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 494–508.

[9] D. B. Fogel, T. Bäck, and Z. Michalewicz, *Handbook of Evolutionary Computation*. Institute of Physics Publishing and Oxford University Press, 1997, pp. 491–495.

[10] K. Deb, "Nonlinear goal programming using multi-objective genetic algorithms," *Journal of the Operational Research Society*, vol. 52, no. 3, pp. 291–302, 2001.

[11] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II," Indian Institute of Technology, Kanpur, India, KanGAL report 200001, 2000.

[12] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm," in *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, K. Giannakoglou, D. Tsahalis, J. Periaux, P. Papailou, and T. Fogarty, Eds., Athens, Greece, 2002, pp. 95–100.

[13] E. Zitzler and L. Thiele, "Multiobjective Optimization Using Evolutionary Algorithms—A Comparative Study," in *Parallel Problem Solving from Nature V*, A. E. Eiben, Ed. Amsterdam: Springer-Verlag, September 1998, pp. 292–301.

[14] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance Assessment of Multiobjective Optimizers: An Analysis and Review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, April 2003.

[15] J. Knowles, L. Thiele, and E. Zitzler, "A tutorial on the performance assessment of stochastic multiobjective optimizers," Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich," 214, July 2005.

[16] K. Deb and A. Kumar, "Real-coded Genetic Algorithms with Simulated Binary Crossover: Studies on Multimodal and Multiobjective Problems," *Complex Systems*, vol. 9, pp. 431–454, 1995.