An Alternative Archiving Technique for Evolutionary Polygonal Approximation

José Luis Guerrero, Antonio Berlanga and José Manuel Molina
Computer Science Department, Group of Applied Artificial Intelligence
University Carlos III of Madrid
Colmenarejo. Spain
jguerrer@inf.uc3m.es, aberlan@ia.uc3m.es, molina@ia.uc3m.es

Abstract—**Archiving procedures are a key parameter for Multi-objective evolutionary algorithms, since they guarantee the algorithm convergence and the good spread of the obtained solutions in the final Pareto front. For many practical applications, the cost of the algorithm is clearly dominated by the computational cost of the underlying fitness functions, allowing complex processes to be incorporated into the archiving procedure. This work presents a study of the archiving technique for evolutionary polygonal approximation (the division of a given curve into a set of n segments represented by a linear model) based on the epsilon-glitch concept, highlighting the cost of the technique compared to the fitness computation, and proposing a novel alternative archiving procedure, which yields statistically significant better results compared to available approaches.**

Keywords-Archiving; Polygonal approximation; Segmentation; Multi-objective evolutionary algorithms.

I. INTRODUCTION

Multi-objective evolutionary algorithms (MOEA) have become a proficient tool dealing with multi-objective problems [1]. This can be shown by the growing number of real world applications related to their use in the last years [2], with examples such as marine vehicle design [3], groundwater monitoring applications [4], or architectural design [5].

The phases of a MOEA can be stated as follows [1]: initialization, dominated individuals removal, use of density estimators, evolutionary operations, next generation selection, check termination criterion, and archive updating. Archive handling is a phase which is also very common for a wide range of metaheuristics [6]. Archiving can be defined as the process of storing individuals from a population to preserve them from the usual cycles of the evolutionary process applied to the consecutive generations. Some of the initial MOEAs including explicit archiving techniques were SPEA [7], PAES [8] and MOMGA [9].

The initial concern regarding archiving procedures was to guarantee a good distribution of the Pareto optimal solutions, since convergence was considered guaranteed [10]. However, as established in [11], under standard Pareto-based selection schemes, deterioration may occur, and, thus, convergence is no longer guaranteed. Examples of algorithms suffering this effect are PAES [8] or SPEA [7].

One of the most successful MOEAs introduced up to date, regarding its impact in the research community and

extended use, is probably SPEA2 [12]. This algorithm introduced a novel archiving technique based on an environmental selection process, according to the concept of strength: for every individual, how many individuals dominate and are dominated by it. This process provided the algorithm with statistically significant improvements versus relevant coetaneous algorithms, such as NSGA-II [13]. It must be noted, though, that these comparisons measured the computational cost for the stopping criterion in terms of function evaluations, disregarding the cost of the algorithms procedures (including the archiving technique).

Polygonal approximation techniques [14] are segmentation processes (the division of a given series of data into *n* segments, usually represented by a linear model) applied to closed curves as an offline process. The multi-objective nature of these processes has been explicitly faced in recent works [15], [16], as opposed to traditional heuristic, single objective approaches [17].

In [16], a MOEA for polygonal approximation is presented based on the SPEA2 algorithm. Its results proved to be statistically significant compared to a set of traditional heuristic techniques in terms of quality. This work will present an alternative archiving strategy comparing the cost of the environmental selection process versus a specific archiving technique based on the discrete nature of one of the objectives, inspired by the ε-dominance technique introduced in [11]. ε-dominance describes the ε value required for solution to dominate another one (and can be presented in additive or multiplicative terms). The presented proposal can be considered on terms of a general novel archiving procedure and also as a specific proposal for the multi-objective polygonal segmentation domain, which benefits from its improvements.

The remaining sections are organized according to the following structure: Section II focuses on the general presentation of the novel archiving procedure, analyzing some of the literature regarding archiving techniques and highlighting the basis for the presented proposal. Section III will present the MOEA approach for polygonal approximation, which is the specific problem than inspires the presented technique and also the one it is applied to in order to verify its results, followed by Section IV, where the proposed archiving technique is presented and detailed. Finally, Section V presents the experimental results obtained, and Section VI the obtained conclusions from these results and possible future lines for the developed research.

II. ARCHIVING TECHNIQUES AND THE ENVIRONMENTAL SELECTION PROCESS

In [11], Laumanns et al. proved that many MOEAs based on standard Pareto-based selection schemes could suffer deterioration, not guaranteeing convergence. Deterioration occurred when elements of a solution set at a given time were dominated by a solution set which the algorithm maintained some time before. Based on these observations, they presented new archiving strategies based on the ε-dominance concept, attempting to provide both convergence and good distribution properties.

However, in [18], some of the issues with this approach were highlighted, mainly the choice of the ε initial parameter. This parameter could be chosen by either a preset value or by an adaptative procedure. In the former case, the number of points in the archive is bounded by a function of the (possibly unknown) objective space ranges. In the latter case, ε may become arbitrarily large, providing a poor final archived set compared to the sequence of points presented to the archiving algorithm.

SPEA algorithm family, both SPEA [7] and SPEA2 [12], relies on the concept of strength for their archiving strategy: originally proportional to the number of solutions which an individual dominated (in the SPEA algorithm), it was improved in the SPEA2 algorithm by also including the number of solutions which dominate it. This led to the environmental selection update mechanism for the archive.

The original archive update mechanism was based on a clustering technique. This mechanism tended to lose boundary solutions when the archive size was too small for the required number of non-dominated solutions. The truncation technique presented in SPEA2 is an iterative process that eliminates at each stage the individual with the minimum distance to another individual (considering the following distances to the second, third... closest individuals in case of ties). This process continues until the maximum number of individuals, according to the archive size parameter, has been introduced.

The archive size in SPEA2 is fixed. If the number of non-dominated individuals is not sufficient to fill it, dominated ones are inserted. Also, the environmental selection mechanism dominates the complexity of the whole algorithm, with a worst case complexity of $O(M^3)$, where M is the population size plus the archive size. On average, that complexity is reduced to $O(M^2 log M)$.

III. MULTI-OBJECTIVE EVOLUTIONARY POLYGONAL APPROXIMATION

Polygonal approximation is the process of dividing a given closed curve into a set of n segments, represented each of them by a linear model. If we formalize a general closed curve according to (1), the segmentation process can be formalized with (2).

$$t = \{\vec{p}_i\}, \vec{p}_i = (x_i, y_i, i), i = 1,...,n \qquad (1)$$

$$S(t) = \{B_m\}, B_m = \{\vec{p}_i\}, i = k_{min},...,k_{max}$$
$$m \in [1,...,q], q < n \qquad (2)$$

In (2), t represents the closed curve, $\vec{p}_i$ each of its points, x_i and y_i its two dimensional components, i an ordering value, $S(t)$ the result of the segmentation process, B_m a segment resultant of that process and, for each of these resultant segments, k_{min} and k_{max} represent the points at its edges.

Two different problems have been traditionally associated with this process: Min-# and Min-ε problems. Min-# is based on the optimization of the representation error for a previously set value of q. Min-ε, on the other hand, finds the minimum number of segments such that the final representation error does not exceed a previously established error ε. Both problems rely on the two objectives of a possible multi-objective formalization for the problem, presented in (3).

$$\min\{E(S(t),t),q\}$$
$$s.that \begin{cases} E(S(t),t \leq total\_error \\ \forall m, E(S(B_m),B_m) \leq max\_seg\_error \end{cases} \qquad (3)$$

This formalization presents the two functions to be minimized jointly: the representation error $E(S(t), t)$ and the cost of that representation, its number of segments q, which must be lower than its original number of points n, as established in (2). Traditionally, restrictions have been considered for this issue [17], being these restrictions represented by the overall representation error (*total_error*) and the maximum representation error for each individual segment (*max_seg_error*)

In [16], this problem is faced using a SPEA2 MOEA algorithm. This proposal allows dealing with both minimization problems jointly. Also, as explained in [16], there is a great degree of possible information sharing between different solutions with different numbers of segments, which allows obtaining a Pareto front of possible solutions in an efficient way. Finally, Min-ε and Min-# problems require user configuration parameters, which may be hard to establish a-priori.

The archiving procedure of this algorithm has been thoroughly described in Section II. Also, as commented in the introduction, the performance comparison between different MOEAs does usually only take into account the number of function evaluations (or, similarly, the number of generations), under the consideration that the computational complexity of these function evaluations exceeds that of the algorithm itself. Recent procedures [19] have considerably reduced the computational complexity of fitness evaluations in evolutionary polygonal approximation.

Figure 1 shows a comparison of the running time of the different procedures of the algorithm using this enhanced fitness evaluation. These results have been obtained using the JMetal [20] environment with the general configuration established in [16] using a population size of 200 and the
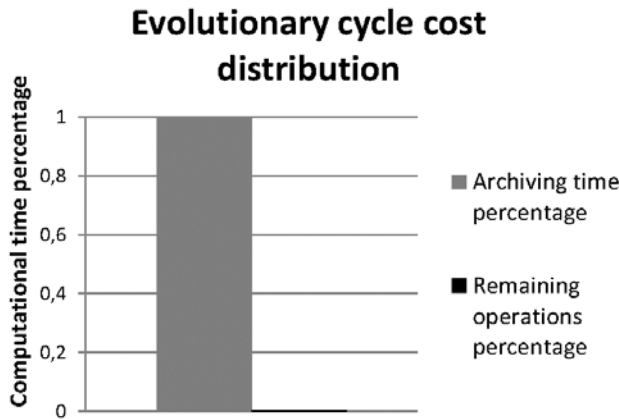
Figure 1.   Computational cost distribution in the SPEA2 algorithm for a polygonal approximation problem instance
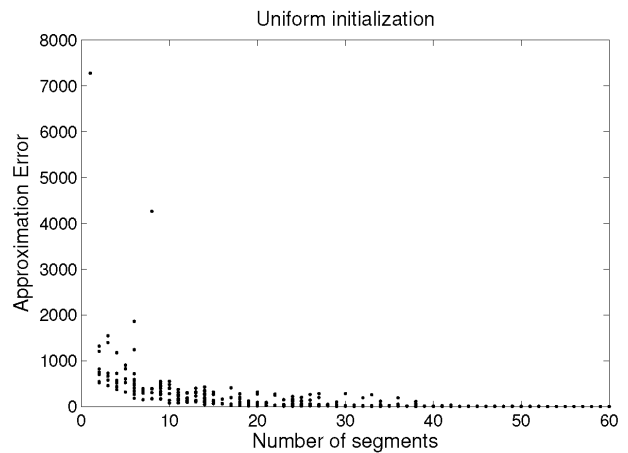

Figure 2.   Initialization example for a MOEA polygonal approximation problem instance

Figure 3.   Initialization example for a MOEA polygonal approximation problem instance

enhanced fitness computation presented in [19]. The specific problem instance used is the chromosome figure, even though similar results are obtained for the different instances considered in the dataset. These instances will be presented in Section V, along with the experimental results.

As Figure 1 clearly shows, the archiving procedure is not only dominating the algorithm running time, but also the enhanced fitness computation (archiving implies more the 99% of the whole running time, including fitness computation). This huge effort to guarantee a well distributed Pareto front seems unacceptable. The following section will present an alternative archiving technique based on the characteristics of the problem.

IV. ALTERNATIVE ARCHIVING PROCEDURE FOR EVOLUTIONARY POLYGONAL APPROXIMATION

The polygonal approximation process has a set of very specific characteristics, mainly its bi-objective nature with a very high degree of conflict between them and the fact that one of these objectives is discrete. This nature of the problem has been used, for instance, to provide computationally efficient initializing methods [21]. Figure 2 shows the result of an initialization process prior to the application of Pareto dominance selection to highlight these characteristics.

Some of the issues related to the costly archiving results exhibited by SPEA2 in this problem, as shown in Figure 1, are related to the multi-objective proposal: the algorithm must be able to store, ideally, one individual per each compression level. This implies that the archive size can get to be really large (and the computational complexity of the environmental selection depends heavily on that archive size Initialization example for a MOEA polygonal approximation problem instance.

A vector $\vec{u} = (u_1,...,u_k)$ is said to ε-dominate another vector $\vec{v} = (v_1,...,v_k)$ for some ε>0 iff $\forall i \in 1,...,k, (1-\varepsilon)u_i \leq v_i$. The idea presented in [11] was, according to ε-dominance, to draw ε-boxes such that at
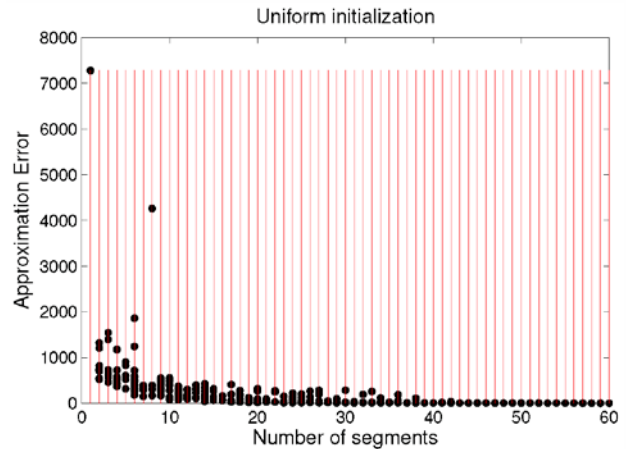
most one element is contained in each box. From the characteristics of the problem presented, the idea for the alternative archiving presented is to use a technique similar to these ε-boxes, considering a box for each of the possible individuals according to the number of segments objective. Figure 3 represents these boxes over the results in Figure 2.

As shown in Figure 3, these boxes are infinitely thin in one objective (the discrete objective representing the number of segments, where they only cover one value) and infinitely long in the other (the objective representing the representation error). These particular instances of ε-boxes are similar to the glitches from signal processing theory, and so they have been named according to this resemblance. Pareto-dominance is only checked within the ε-glitch which an individual belongs to, not among different glitches. This implies that the complexity of this process is now constant, and the complexity of the whole archiving mechanism is reduced to $O(n)$, where n is the population size. Figure 4
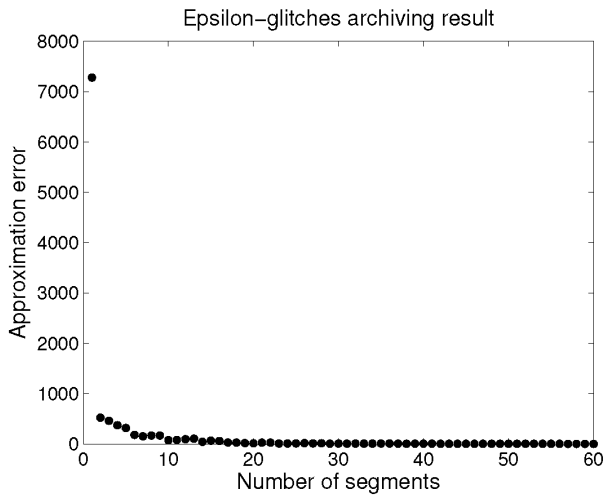
Figure 4.   ε-glitches archiving procedure example

Figure 5.   Chromosome dataset figure

shows the result of the ε-glitches archiving procedure over the population presented in Figure 2.

It must be noted that one of the issues related to the environmental selection process, the fact that its complexity included the archive size, has been overcome, an especially important achievement for this problem, since the required number of individuals in the archive can be very large for some problem instances, as explained in previous sections.

The traditional evolutionary cycle where one full generation is produced at each step is no longer required, since every individual is compared to the correspondent one in its glitch already stored in the archive. With this approach, the evolutionary cycle implies parent selection, children obtaining through the transformation operators and the invocation of the archiving procedure individually

Even though implementation issues are not the focus of this work, the appropriate management of the data structure for the archive is important for the computational cost reduction. Since the archive size is fixed (with an exception detailed in the following paragraph), a fixed size, constant time access data structure is suggested (such as a traditional array). For the initialization of this array, we suggest the generation of an initial random individual for one of the archive boundaries (either of them) and the application of directed mutations to obtain an individual for each of the ε-glitches. This directed mutations (and, thus, the archive initialization process) imply a very low computational cost using the fitness computations detailed in [19].

One final improvement is introduced in the archiving procedure. For MOEA approaches for polygonal segmentation, the initial archive size, as explained in [16], is set to the number of points in the curve. However, very commonly, through the evolutionary process, new individuals with a perfect segmentation (zero error) are found, requiring a lower number of segments than that initial boundary. With the explained archiving mechanism, this would not be taken into account, since it covers Pareto dominance between two individuals which belong to two
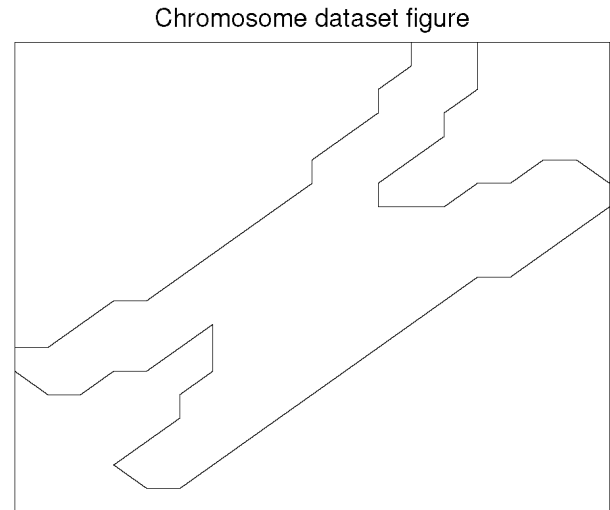
different ε-glitches. To cover this special case, when a new individual is added to the archive with zero representation error, the archive size is reduced to its number of segments.

A final overview of the detailed archiving process can be detailed in the following steps:

- Initialize archive with size *n* and fill with individuals obtained with applied directed mutations from an individual with either, the highest or the lowest possible number of segments, obtaining one individual for each possible number of segments
- While the stopping criterion is not met
  - Select parents
  - Apply transformation operators to parents
  - For each children produced:
    - Compare to archive individual with the same number of segments, and update as required
    - If the updated individual has zero representation error, update archive size if required (the number of segments of the individual is smaller than archive size)
- Output archive results

V.   EXPERIMENTAL EVALUATION

The dataset used for the experimental evaluation of the proposed technique is based on three traditional figures for the polygonal approximation domain, usually named chromosome, leaf and semicircle [14]. As discussed in [19], these figures, while representative and appropriate for technique comparison, are not very complex. Thus, the procedure presented in that paper will be used to generate three additional figures for the dataset, containing ten linked copies of the previous three, with the names chromosome10, leaf10 and semicircle10. Figure 5 presents the chromosome curve, while Figure 6 shows the chromosome10 curve.
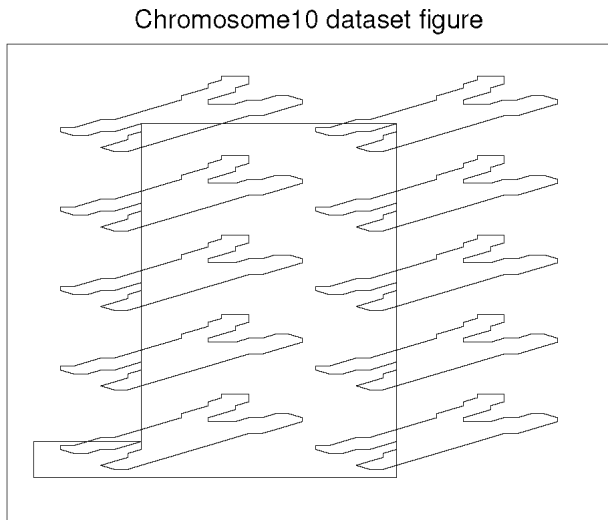
Chromosome10 dataset figure

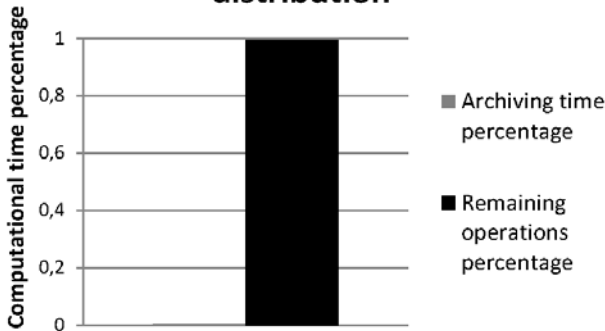Figure 6. Chromosome10 dataset figure, built using ten different instances of the chromosome figure

Figure 7. Computational cost distribution in the proposed algorithm with the ε-glitches archiving technique

The first relevant comparison, according the considerations presented in Section II, is the computational time which is now used by the archiving procedure. The original results for the SPEA2 technique have been presented in Figure 1. The results with the presented technique are shown in Figure 7. These results are general for all the different figures in the dataset. As this figure shows, the percentages have been swapped, spending now more than 99% percent of the available computational time in the evolutionary search instead of the archiving technique.

The experimental configuration is based on thirty different executions of the SPEA2 algorithm as described in [16]. The population size used is 200, and the algorithm is left to run for 200 generations. The running time used for each of this independent executions is measured, and afterwards thirty different runs of the proposed technique are performed, each of them according to an individual

TABLE I.   FINAL POPULATIONS COMPARISON

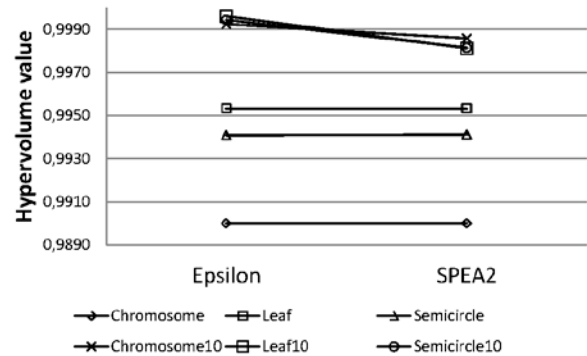| Figure	Epsilon Hyp. results		SPEA2 Hyp. results		Best
	Mean	*Std*	*Mean*	*Std*	
chrom	0,99000	0,00006	0,99001	0,00005	-
leaf	0,99533	0,00001	0,99532	0,00001	-
semi	0,99409	0,00008	0,99412	0,00003	-
chrom10	0,99924	0,00002	0,99858	0,00093	eps.
leaf10	0,99961	0,00001	0,99812	0,00139	eps.
semi10	0,99944	0,00002	0,99814	0,00175	eps.

Figure 8. Computational cost distribution in the proposed algorithm with the ε-glitches archiving technique

previously measured running time as its stopping criterion. The hypervolume indicator values [22] are built afterwards according to these results and the statistical significance of the differences are tested according to Wilcoxon statistical testing [23] with a 95% confidence interval. The results of these procedures are shown in Table I. Figure 8 summarizes these results.

The results for the three initial figures do not show statistically significant differences between the two techniques, while the results for the three harder ones are clearly dominated by the epsilon-glitches based technique. The explanation for this fact is clear: when the final solution can be easily reached, the improved distribution of the solutions provided by the environmental selection technique allows SPEA2 to obtain solutions of similar quality, even though the computational effort spent in the proper search process is smaller (as seen in the comparison of Figures 1 and 7). As the problem instances become harder and the extent of search required to reach a reasonable Pareto front grows, the focus on the search process of the epsilon-glitches technique pays off for the poorer solution distribution, providing it with substantial better results in terms of final hypervolume.

VI.   CONCLUSIONS AND FUTURE WORK

This work has presented a study over the archiving procedure considerations for evolutionary polygonal

approximation. Available approaches based on standard, well established multi-objective evolutionary approaches, spend most of their computational time on their archiving procedures (when combined with novel improved fitness computation algorithms). This leads to a waste of computational resources, particularly for hard problem instances. A novel archiving procedure, based on a reduced Pareto-dominance application, has been presented, and its results tested over a set of six different functions with growing difficulty.

The epsilon-glitch technique is applied to the multi-objective polygonal segmentation issue, applying a simplified version of previous epsilon-dominance proposals to this bi-objective problem, focusing on the archive organization and direct comparisons to provide computational cost enhancements. The results show that the techniques, for easier problems, do not yield statistically different results. However, as the difficulty of the problem instances is increased, the novel epsilon-glitch archiving procedure provides clearly better results, since it spends most of its computational effort in the search of the best solutions, instead of the archiving procedure. Future lines include the study of mixed approaches, where full Pareto-dominance can be introduced at certain steps of the presented approach, and the extension of this approach to different sets of problems with similar characteristics.

REFERENCES

[1] C. Coello, G. Lamont, and D. Van Veldhuizen, Evolutionary algorithms for solving multi-objective problems. Springer, 2007

[2] C. Coello and G. Lamont, Applications of multi-objective evolutionary algorithms. World Scientific Publishing Company Incorporated, 2004.

[3] D. Lee, "Multiobjective design of a marine vehicle with aid of design knowledge," International journal for numerical methods in engineering, vol. 40, no. 14, 1997, pp. 2665–2677.

[4] P. Reed, B. Minsker, and D. Goldberg, "A multiobjective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data," Journal of Hydroinformatics, vol. 3, 2001, pp. 71–89.

[5] J. Byrne, M. Fenton, E. Hemberg, J. McDermott, M. ONeill, E. Shotton, and C. Nally, "Combining structural analysis and multi-objective criteria for evolutionary architectural design," Applications of Evolutionary Computation, 2011, pp. 204–213.

[6] E. Talbi, Metaheuristics: from design to implementation. Jonh Wiley and Sons Inc., 2009.

[7] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach", IEEE Transactions on Evolutionary Computation, vol. 3, no. 4, 1999, pp. 257–271.

[8] J. Knowles and D. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," Evolutionary computation, vol. 8, no. 2, 2000, pp. 149–172.

[9] D. Van Veldhuizen and G. Lamont, "Multiobjective optimization with messy genetic algorithms," in Proceedings of the 2000 ACM symposium on Applied computing-Volume 1, 2000, pp. 470–476.

[10] G. Rudolph and A. Agapie, "Convergence properties of some multi-objective evolutionary algorithms," Proceedings of the 2000 Congress on Evolutionary Computation, 2000., vol. 2. IEEE, 2000, pp. 1010–1016.

[11] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," Evolutionary computation, vol. 10, no. 3, 2002, pp. 263–282.

[12] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm," 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: International Center for Numerical Methods in Engineering (CIMNE), 2001, pp. 95–100.

[13] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," Evolutionary Computation, IEEE Transactions on, vol. 6, no. 2, 2002, pp. 182–197.

[14] M. Sarfraz, "Linear capture of digital curves," in Interactive Curve Modeling. Springer London, 2008, pp. 241–265.

[15] A. Kolesnikov, P. Franti, and X. Wu, "Multiresolution polygonal approximation of digital curves,". Proceedings of the 17th International Conference on Pattern Recognition, vol. 2. IEEE, 2004, pp. 855–858.

[16] J. Guerrero, A. Berlanga, and J. Molina, "A multi-objective approach for the segmentation issue," Engineering Optimization, vol. 44, no. 3, 2012, pp. 267–287.

[17] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," Data mining in time series databases, vol 57, 2004, pp. 1–21.

[18] J. Knowles and D. Corne, "Properties of an adaptive archiving algorithm for storing nondominated vectors," IEEE Transactions on Evolutionary Computation, , vol. 7, no. 2, 2003, pp. 100–116.

[19] J. Guerrero, A. Berlanga, and J. Molina, "Fitness-aware operators for evolutionary polygonal approximation," in Applied Computing. IADIS, 2012, pp. 283–290.

[20] J. J. Durillo and A. J. Nebro, "jmetal: A java framework for multi-objective optimization," Advances in Engineering Software, vol. 42, no. 11, 2011, pp. 760–771.

[21] J. Guerrero, A. Berlanga, and J. Molina, "Initialization procedures for multiobjective evolutionary approaches to the segmentation issue," Hybrid Artificial Intelligent Systems, 2012, pp. 452–463.

[22] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, and V. Da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," Evolutionary Computation, IEEE Transactions on, vol. 7, no. 2, 2003, pp. 117–132.

[23] G. Corder and D. Foreman, Nonparametric statistics for nonstatisticians: a step-by-step approach. West Sussex, England: John Wiley & Sons Inc, 2009.