

Multiple criteria performance analysis of non-dominated sets obtained by multi-objective evolutionary algorithms for optimisation

Gerrit K. Janssens¹ and José Maria Pangilinan²

¹ Hasselt University, Transportation Research Institute (IMOB),
Wetenschapspark Building 5
B-3590 Diepenbeek, Belgium
gerrit.janssens@uhasselt.be

² Saint-Louis University, Dept. of Computer Science
Baguio, Philippines
joey.pangilinan@slu.edu.ph

Abstract. The paper shows the importance of a multi-criteria performance analysis in evaluating the quality of non-dominated sets. The sets are generated by the use of evolutionary algorithms, more specifically through SPEA2 or NSGA-II. Problem examples from different problem domains are analyzed on four criteria of quality. These four criteria namely *cardinality* of the non-dominated set, *spread* of the solutions, *hyper-volume*, and *set coverage* do not favour any algorithm along the problem examples. In the Multiple Shortest Path Problem (MSPP) examples, the *spread* of solutions is the decisive factor for the 2S|1M configuration, and the *cardinality* and *set coverage* for the 3S configuration. The differences in *set coverage* values between SPEA2 and NSGA-II in the MSPP are small since both algorithms have almost identical non-dominated solutions. In the Decision Tree examples, the decisive factors are *set coverage* and *hyper-volume*. The computations show that the decisive criterion or criteria vary in all examples except for the *set coverage* criterion. This shows the importance of a binary measure in evaluating the quality of non-dominated sets, as the measure itself tests for dominance. The various criteria are confronted by means of a multi-criteria decision tool.

Keywords: evolutionary algorithms; multi-objective optimization; multi-criteria analysis.

1 Introduction

Many real-world optimization problems are multi-objective by nature and with objectives that are in conflict. Mathematical techniques are available to find best-compromise solutions by aggregating multiple objectives into a single function [9]. They have their drawbacks as they have difficulty dealing with concave and discontinuous Pareto fronts. Stochastic local or global search algorithms or population-based algorithms are often used when exact methods are infeasible to be

applied and are used to solve difficult optimization problems. Among the many types of these algorithms, Evolutionary Algorithms (EA) seem particularly suitable to solve multi-objective optimization problems. An increasing number of research papers report comparative findings of several evolutionary algorithms in terms of computing speed and Pareto optimality as tested on various multi-objective problem instances or applications with known Pareto sets. In most practical or experimental cases however the Pareto sets are unknown.

The objective of the study is to describe the performance of evolutionary algorithms in terms of stability, computational complexity, diversity and optimality of solutions in different multi-objective optimization problems (MOOP). Multi-objective evolutionary algorithm (MOEA) experiments generate a variety of non-dominated sets in each problem domain. Comparisons of the quality of the non-dominated sets have not yet been presented in each of the cases. The stability of an algorithm is concerned with the sensitivity of the results to changes in the MOEA parameters settings. Computational complexity refers to the solution run-time complexity in terms of the size of the problem. Diversity measures the spread of solutions in the non-dominated set in order to provide the decision maker a true picture of trade-off solutions. Optimality measures the proximity of the best non-dominated set to the Pareto-optimal set.

Bosman and Thierens [1] argue that the quest for finding the components that result in the best EAs for multi-objective optimization is not likely to converge to a single, specific MOEA. They stated that the trade-off between the goals of proximity and diversity preservation plays an important role in the exploitation and exploration phases of any MOEA.

A comprehensive discussion of multi-objective evolutionary algorithms (MOEA) can be found in [7]. In addition, Coello [6] gives a summary of current approaches in MOEA and emphasizes the importance of new approaches in exploiting the capabilities of evolutionary algorithms in multi-objective optimization. Zitzler and Thiele [17] performed a comparative analysis of existing evolutionary algorithms in multi-objective optimization by means of well-defined quantitative performance measures. In this research a comparison is made using two well-known techniques: SPEA2 and NSGA-II.

Zitzler et al. [18] introduced the Strength Pareto Evolutionary Algorithm 2 (SPEA2), which is an extension and improvement of the original work by Zitzler and Thiele [17]. SPEA2 integrates a fitness assignment strategy, which considers the number of individuals that an individual dominates and the number of its dominators. It uses a nearest-neighbor density estimation technique that guides the search more efficiently and avoids the formation of new solutions in only a few clusters in the search space. SPEA2 has a truncation procedure that preserves the best solutions when the number of non-dominated individuals exceeds the external population size.

Deb et al. [8] introduced an elitist non-dominated genetic algorithm (NSGA-II) that uses not only an elite-preserving strategy but also an explicit-diversity preserving mechanism. Initially, NSGA-II creates a random parent population, sorts the population based on non-domination and assigns each solution a fitness value equal to its non-domination level.

Consider first a case where an optimal algorithm exists to obtain the Pareto set. For example, in the Competitive Facility Location Problem (CFLP) [11], the quality of the non-dominated sets generated by the MOEA can be calculated and compared to the Pareto-optimal set generated by an algorithm by Carrizosa and Plastria [5]. The *error ratio* metric, which measures the closeness of the non-dominated set to the Pareto front in terms of set membership, can be used to measure the quality of solutions in the CFLP. However, if such an algorithm does not exist or is not available, like for two other problems under study here, the Multi-Objective Shortest Path Problem (MSPP) and the Decision Tree (DT) experiments, only approximations to their Pareto-optimal sets are available for performance analysis.

The next sections present some performance metrics that are useful in measuring the quality of non-dominated sets when the Pareto-optimal set is unknown, and utilizes a multi-criteria tool to determine the best non-dominated set for MSPP, and DT problems based on the performance metrics. Several performance metrics exist in literature, and several comparative studies have been conducted that evaluate them.

The studies presented above show a variety of results and no single MOEA performs better in the different performance metrics but most studies compare their algorithms with either NSGA-II or SPEA2 or both. The studies above mostly evaluate the performance of the selection operators of each MOEA without investigating the effect of the parameter settings on its performance. The current study investigates the performance of NSGA-II and SPEA2 in selected multi-objective optimization problems by means of a multi-criteria method in which various performance metrics of the Pareto front are presented to the decision-maker.

2 Performance metrics

Deb [7] states that there are two orthogonal goals for any multi-objective optimisation algorithm: (1) to identify solutions as close as possible to the true Pareto-optimal set and (2) to identify a diverse sets of solutions distributed evenly across the entire Pareto-optimal surface. This has led to several metrics that characterise either closeness, or diversity, or both. Examples of metrics which measure the closeness to the Pareto surface are the Error ratio [14] and the Set coverage [16]. Examples of metrics which measure the diversity across the Pareto surface are the Spacing [13] and the Spread [8]. A measure like the Hypervolume measures both closeness and diversity [17]. Most measures are unary quality measures, i.e. the measure assigns to each Pareto set approximation a number that reflects a certain quality aspect.

Hypervolume

The hypervolume metric [17] calculates the volume in the objective space covered by the members of the non-dominated set \mathbf{Q} . For each solution $i \in \mathbf{Q}$, a hypercube v_i is computed from a reference point and the solution i as the diagonal corners of the hypercube, The reference point can be found by constructing a vector of worst objective function values. The hypervolume (HV) is calculated as:

$$HV = \text{volume} \left(\bigcup_{i=1}^{|\mathbf{Q}|} v_i \right). \quad (1)$$

The hypervolume metric is interesting because it is a single metric which is sensitive both to the overall advancement of the non-dominated set and to the distribution of individual points across the set. The placement of the reference point is critical and determines the sense and the magnitude of the hypervolume. Problems may appear if objectives have dissimilar scales or if some objectives are unbounded.

Spacing

Schott [13] introduced a metric, which is a measure of the relative distances between consecutive solutions in the non-dominated set \mathcal{Q} is calculated as:

$$S = \sqrt{\frac{\sum_{i=1}^{|\mathcal{Q}|} (d_i - \bar{d})^2}{|\mathcal{Q}|}} \quad (2)$$

$$d_i = \min_{k \in \mathcal{Q} \wedge k \neq i} \sum_{m=1}^M |f_m^i - f_m^k| \quad (3)$$

$$\bar{d} = \frac{\sum_{i=1}^{|\mathcal{Q}|} d_i}{|\mathcal{Q}|} \quad (4)$$

Schott's metric measures the diversity of a non-dominated set.

Set coverage metric

This metric is based on Zitzler [16]. The metric computes the relative spread of solutions between two non-dominated sets \mathcal{A} and \mathcal{B} . The set coverage metric $C(\mathcal{A}, \mathcal{B})$ calculates the proportion of solutions in \mathcal{B} that are weakly dominated by solutions of \mathcal{A} :

$$C(\mathcal{A}, \mathcal{B}) = \frac{|\{b \in \mathcal{B} | \exists a \in \mathcal{A} : a \preceq b\}|}{|\mathcal{B}|} \quad (5)$$

The metric value $C(\mathcal{A}, \mathcal{B}) = 1$ means all members of \mathcal{B} are weakly dominated by \mathcal{A} . On the other hand, $C(\mathcal{A}, \mathcal{B}) = 0$ means that no member of \mathcal{B} is weakly dominated by \mathcal{A} . This operator is not symmetric, thus it is necessary to calculate $C(\mathcal{B}, \mathcal{A})$. The set coverage metric measures convergence based on the concept of dominance relations.

Cardinality

This metric counts the number of solutions in the non-dominated set. It measures neither diversity nor convergence.

In order not to limit the description of the quality of a non-dominated set by using only a single metric, a multi-criteria evaluation seems appropriate. Hence, the computations in this paper evaluate the quality of non-dominated sets according to four criteria, which are mentioned above. The multi-criteria tool employed in the computations is Decision Lab [15]. The Decision Lab software is a multi-criteria

decision making software, which is based on the Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) and the Graphical Analysis for Interactive Assistance (GAIA). The details of the PROMETHEE method are found in Brans et al. [2]. GAIA, which makes use of principal component analysis, is a descriptive complement to the PROMETHEE methods [3].

3 Computations and results

Table 1 shows the computed values for each performance criterion in the different multi-objective optimization problems. There are four criteria. A smaller set cardinality is preferred, especially in the case of a continuous decision space. A spread that has smaller value means that the solutions on the non-dominated front are uniformly spaced therefore this criterion is minimized. Hypervolume and set coverage are maximized. Two non-dominated sets are compared in each MOOP. One set is generated by NSGA-II and the other by SPEA2. Decision Lab can rank more non-dominated sets but since the set coverage is a binary quality measure, only two non-dominated sets can be evaluated each time.

The Multi-objective Shortest Path Problem (MSPP) is an extension of the traditional shortest path problem and is concerned with finding a set of efficient paths with respect to two or more objectives that are usually in conflict. A variety of algorithms and methods such as dynamic programming, label selecting, label correcting, interactive methods, and approximation algorithms have been implemented and investigated with respect to the MSPP [10]. The problem is known to be NP-complete. For fitness and selection, two objective configurations are considered for finding efficient paths: (3-S) and (2-S|1-M). S-type objectives are sum problems that are to be minimised and M-type objectives are max-min problems that are to be minimised. A 50-node of 10% density is the basis for the computed values in both MSPP (Multi-objective Shortest Path Problem) configurations. The *hypervolume* values for the MSPP are blank since they cannot be computed. This reduces the number of MSPP criteria to three.

An evolutionary algorithm is a promising technique to build oblique decision trees (see [4] for a list of advantages). Benchmarking in this field is done by finding oblique partitions on a variety of datasets from the UCI (University of California at Irvine) machine learning repository. The Housing and the Optical Digits datasets are used as the non-dominated sets for the Decision Tree problem.

Decision Lab has a visualization tool that shows the relation between the criteria and the non-dominated set, and shows a preferred solution if it exists. Figure 1 shows a GAIA diagram for the MSPP that shows how each criterion relates to each action. The GAIA plane corresponds to the first principal components of the data, which ensures that a maximum quantity of information is available on the plane. An action in this case refers to a non-dominated set, generated by a specific algorithm. The orientation of the criteria axes indicates which criteria are in agreement with each other. The orientation of the position of an action indicates its strong features. The length of the axis correspond to a criterion's observed deviations between actions, the longer the axis the higher the deviation.

Table 1. Computed criteria values of non-dominated sets

Nondominated Set	Cardinality (Minimize)	Criterion		
		Spread (Minimize)	Hyper-volume (Maximize)	Set Coverage (Maximize)
MSPP 2S 1M				
NSGA-II	7	0.8		0
SPEA2	8	0.51		0
MSPP 3S				
NSGA-II	8	0.57		0
SPEA2	5	1.14		0.38
Housing dataset				
NSGA-II	5	0.43	5.47	0.17
SPEA2	6	0.38	6.98	0.6
Optical-digits				
NSGA-II	5	0.36	5.96	0
SPEA2	6	0.63	7.15	0.6

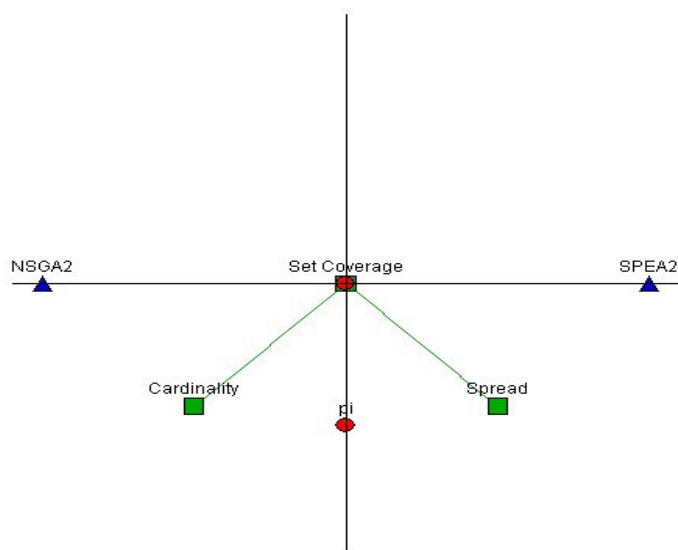


Fig. 1. GAIA diagram for a 2S|1M configuration using the ‘usual criterion’

The orientation of the pi -vector on the figure, which is the decision axis, points to the preferred action or solution considering all the criteria. In this example, pi does not point towards any action, which means that there is no compromise solution. This is due to the condition that NSGA-II has a strong feature on cardinality, SPEA2 has a strong feature on spread, and there is no difference regarding set coverage. All the criteria have the same weight. Adding different weights to each criterion obviously changes the orientation of the pi decision axis.

The criterion weight is independent from the scale of the criterion which means the larger the value the more important the criterion. In order to compare the different criteria independently from their measurement units, the PROMETHEE method provides six preference functions. Five out of six preference functions need some parameters specified by the decision-maker (only the ‘usual criterion’ does not have parameters). The parameters are used to compute the ‘level of preference’ of one action over another.

The preference function translates the deviation d between the values of two actions on a single criterion in terms of a preference degree. The preference degree is an increasing function of the deviation, defined on an interval $[0,1]$. A value 0 has to be interpreted as indifference, while a value of 1 has to be interpreted as strict preference. This research makes use of the ‘linear preference function’, which is defined by Brans et al. [2] as follows:

$$H(d) = \begin{cases} d/p & \text{if } -p \leq d \leq p, \\ 1 & \text{if } d < -p \text{ or } d > p \end{cases} \quad (6)$$

As long as d is less than a threshold value p , the preference of the decision maker increases linearly with d . If d becomes larger than p , a strict preference exists. The linear preference function is used in the following computations and is associated with all criteria. The linear preference function is chosen as the function takes into account even the smallest difference in the scores between two alternatives.

Table 2. The linear preference function threshold values

Criterion	Threshold
<i>Cardinality</i>	75%
<i>Spread</i>	50%
<i>Hypervolume</i>	75%
<i>Set Coverage</i>	50%

Table 2 contains some parameter values to be set by the decision-maker. The threshold values are expressed in a relative way. It should be understood that the choice of these values might influence the final decision. That means the values are to be set either on an objective consensus, or on a subjective but also consensus basis, or – in case these options are not available – are values which should be subject to sensitivity analysis.

Figures 1 and 2 are used, for the MSPP, to show which non-dominated set has better quality under two preference functions: the ‘usual criterion’ and the ‘linear preference function’. They show the plane for the 2S|1M configuration. The number of criteria has reduced to three as mentioned previously. The values for *set coverage* in both sets are zero, which means that there are no weakly dominated solutions from each set or that their solutions are similar. Figure 2 shows that SPEA2 is the preferred solution after incorporating the thresholds from Table 2, and the decisive criterion should be interpreted as the ‘spread’ of solutions. The set coverage criterion is not a factor since both sets do not cover any weakly dominated solutions between them.

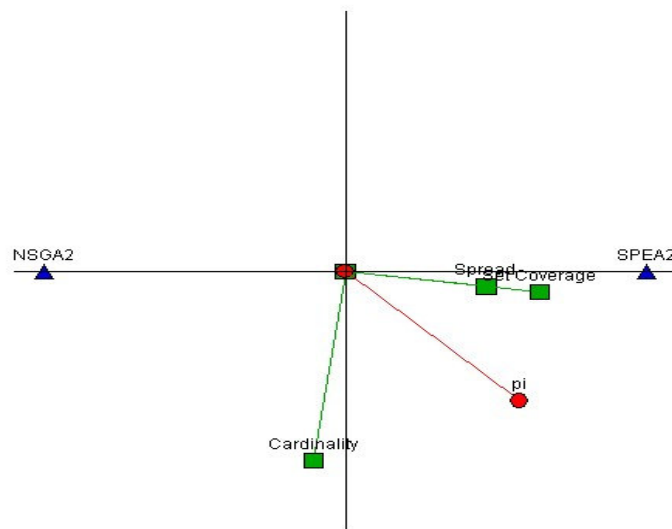


Fig. 2. GAIA diagram for a 2S|1M configuration with preferences

With respect to the application, decision trees or classifiers, the fronts shown in Figure 3 are the non-dominated sets from the ensembles of trees generated for one of the UCI datasets, the Housing Set. The Housing set deals with data regarding housing values in the suburbs of Boston (12 attributes, 506 observations). The OC1 (Oblique Classifier 1 [12]) solutions are dominated by the solutions of either SPEA2, or NSGA-II, or AP (axis-parallel tests). The non-dominated solutions of the AP classifier are dominated by either SPEA2 or NSGA-II. Most of the non-dominated solutions in NSGA are dominated by solutions in SPEA2. It seems that SPEA2 produces the better non-dominated set through the projection of their non-dominated sets but needs to be validated using the PROMETHEE method. The non-dominated sets of AP and OC1 need not be tested for performance quality as their solutions are dominated by both MOEAs solutions.

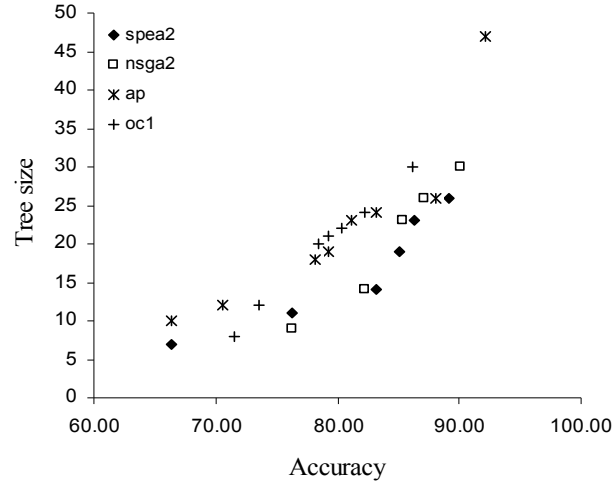


Fig. 3. Non-dominated fronts for the Housing dataset

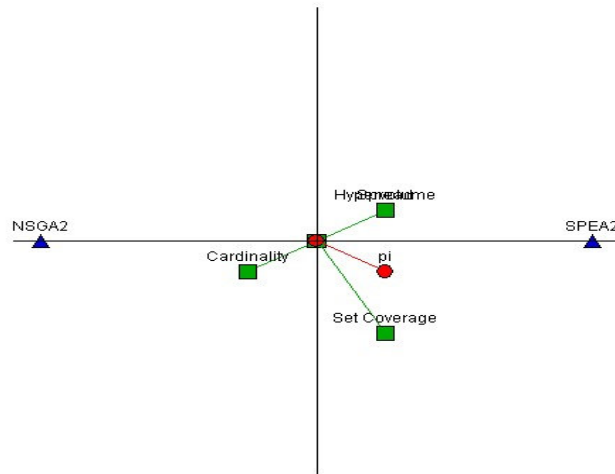


Fig.4. GAIA plane for the Housing dataset using the usual criterion

Figure 4 shows the GAIA plane of the Housing dataset options, and validates that SPEA2 is the preferred solution. The factors that favor SPEA2 are the hypervolume, the spread, and the set coverage. The result does not change when preference thresholds are added. In fact, the pi decision axis leans more to the direction of SPEA2 when preferences are added than it does when without any preferences.

4 Conclusions

Evolutionary algorithms have a great power of generating non-dominated sets in multi-objective optimization problems. Various strategies deliver various non-dominated sets. In order to evaluate the power of such a strategy, several aspects are to be taken into consideration like diversity and convergence to optimality. In literature various metrics have been proposed to measure both performance characteristics. Some metrics measure only one of both, some measure a mix of both. To decide which strategy is better reduces to a multi-criteria problem. In this paper it is illustrated how a plain and clear method, the Promethee method, can be used to decide on the overall quality of an EA strategy. The illustration is made for a combinatorial problem, called the Multi-Objective Shortest Path Problem, and a data mining problem, called the Decision Tree problem.

References

1. Bosman, P. A., Thierens D.: A balance between proximity and diversity in multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 7(2), 174--188 (2003)
2. Brans, J.P., Vincke Ph., Mareschal B.: How to select and how to rank projects: the PROMETHEE method. *European Journal of Operational Research* 24, 228--238 (1986)
3. Brans, J.P., Mareschal B.: The PROMCALC and GAIA Decision Support System for Multicriteria Decision Aid. *Decision Support Systems* 12, 297--310 (1994)
4. Cantu-Paz, E., Kamath C.: Inducint oblique decision trees with evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 7(1), 54--68 (2003)
5. Carrizosa E., Plastria F.: On minquantile and maxcovering optimization. *Mathematical Programming* 71,101--112 (1995)
6. Coello C.: An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys* 32(2), 109--143 (2000)
7. Deb, K.: *MultiObjective Optimization Using Evolutionary Algorithms*. Wiley, Chichester, UK (2001)
8. Deb, K., Agrawal S., Pratap A., and Meyarivan T.: A fast elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2), 182--197 (2002)
9. Ehrgott, M.: *Multicriteria Optimization* (2nd ed.). Springer, Berlin (2005)
10. Ehrgott, M., Gandibleux X.: A survey and annotated bibliography of multiobjective combinatorial optimization. *OR Spectrum* 22(4), pp. 425--460 (2000)

11. Eiselet, H.A., Laporte G., Thisse J.-F.: Competitive location models: a framework and bibliography. *Transportation Science* 27, 44--54 (1993)
12. Murty, S.K., S. Kasif and S. Salzberg (1994). A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2(1):1--32.
13. Schott, J.: Fault tolerant design using single and multicriteria genetic algorithm optimization. Master thesis, Cambridge, MA: Massachusetts Institute of Technology (1995)
14. Van Veldhuizen, D.A.: Multiobjective Evolutionary Algorithms: Classifications, Analyses and New Innovations. Ph.D. thesis, Graduate School of Engineering of the Air Force Institute of Technology, Dayton, Ohio (1999)
15. Visual Decision: Decision Lab and Decision Lab 2000. Montreal, Canada: Visual Decision Inc. (2003)
16. Zitzler, E.: Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications. Ph.D. thesis, Shaker Verlag, Aachen, Germany (1999)
17. Zitzler, E. and Thiele L.: Multiobjective evolutionary algorithms: a comparative study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation* 3(4), 257--271 (1999)
18. Zitzler, E., Laumanns M., and Thiele L.: SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. In: Giannakoglou K., Tsahalis D., Periaux J., Papailou K., Fogarty T. (eds.) *Evolutionary Methods for Design, Optimization and Control*, 19--26, CIMNE, Barcelona, Spain (2002)