

ROBUSTNESS ANALYSIS OF PARAMETER SETTINGS IN A GENETIC ALGORITHM FOR THE MULTI- OBJECTIVE SHORTEST PATH PROBLEM

Gerrit K. Janssens

University of Hasselt
Email: Gerrit.Janssens@uhasselt.be
Belgium

José Maria A. Pangilinan

Saint Louis University
Email: joeyp@slu.edu.ph
Philippines

Abstract: This paper presents a sensitivity analysis of a genetic algorithm (GA) as applied to the multiobjective shortest path problem (MSPP). The study explores the robustness of the multi-objective evolutionary algorithm in terms of the sensitivity of its output to changes in its input parameters. Results show that mutation and the recombination operators only slightly affect the exploration and exploitation of solutions in the multiobjective shortest path problem. However, the size of the population and the number of generations of the GA largely influence the output of the genetic algorithm.

Keywords: Multi-objective evolutionary algorithm, genetic algorithm, multi-objective shortest path, sensitivity analysis.

1. INTRODUCTION

Solutions to the multi-objective shortest path problem (MSPP) have been presented by various researchers in the areas of optimization as in route planning for traffic and transport design (Granat and Guerriero, 2003; Lacomme et al., 2003) and information and communications network design (Gen and Lin, 2004; Kumar and Banjerie, 2003). The MSPP is concerned in finding a set of efficient paths with respect to two or more objectives that conflict one another as in the problem of finding efficient routes in transportation planning that simultaneously minimize travel cost, path length, and travel time. The concept of optimization in the MSPP in general is different from the single-objective optimization problem wherein the task is to find one solution that optimizes a single objective function. Whereas, the task in a multi-objective problem is to find an optimal solution that simultaneously optimizes all ob-

jectives and in such a case, there is no single optimal solution, but a set of efficient or nondominated solutions.

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 (Ehr Gott and Gandibleux, 2000). The problem is known to be NP-complete (Garey and Johnson, 1979). It has been shown that a set of problems exists wherein the number of Pareto-optimal solutions is exponential in terms of runtime complexity in the worst-case. But some labeling algorithm studies dispute this exponential behavior. Gandibleux et al. (2006) show that in practice, the number of efficient paths is not exponential. A separate study by Müller-Hannemann and Weihe (2001) also show that the cardinality of efficient paths in a bicriteria shortest path problem is not exponential as long as the instances are bounded by the characteristics as defined in their experiment.

While some researchers focus on exhaustive solutions or on improvements thereof, other researchers are more concerned with better runtime solutions. Hansen (1979) and Warburton (1987) avoid the complexity problem by developing fully polynomial time approximation schemes (FPTAS) for finding paths that are approximately Pareto-optimal. Tsaggouris et al. (2006) present an improved fully polynomial time approximation scheme for the multicriteria shortest path problem and a new generic method for obtaining FPTAS to any multi-objective optimization problem with non-linear objectives.

Interactive procedures similarly avoid the complexity problem by providing user-interfaces that assist the decision-maker to focus on promising paths that identify better solutions according to preference (Coutinho *et al.*, 1999; Granat and Guerriero, 2003).

Evolutionary algorithms (EAs) have been used in single-objective optimization problems but only a few researchers have applied EAs to the multi-objective shortest path problem. Gen and Lin (2004) demonstrate a new multi-objective hybrid genetic algorithm to improve solutions to the bicriteria network design problem (finding shortest paths) with two conflicting objectives of minimizing cost and maximizing flow. Their paper shows how the performance of a multi-objective genetic algorithm can be improved by hybridization with fuzzy logic control and local search. Kumar and Banerjee (2003) describe a method on multi-criteria network design (shortest paths and spanning trees) with two primary objectives of optimizing network delay and cost subject to satisfaction of reliability and flow constraints. They conclude that the primary advantage of EAs to solve multi-objective optimization problems is its generation of diverse solutions and its polynomial runtime. The paper by Chrichigno and Baran (2004) presents similar representations (spanning trees) to Kumar's for a multicast algorithm. The basic difference between both algorithms is the latter adopts the Strength Pareto Evolutionary Algorithm (SPEA) in generating efficient solutions to the multicast routing problem. Pangilinan and Janssens (2007) describe the behavior and performance of an MOEA in terms of diversity of solutions, runtime com-

plexity, and optimality of solutions. Their results, as tested on several network instances, show that a multi-objective evolutionary algorithm (MOEA) is a good polynomial-time method in finding a subset of efficient paths in an MSPP and is particularly beneficial when intractability and memory issues become obstructions to finding efficient solutions to MSPP-related problems.

The studies using evolutionary algorithms show several advantages in terms of addressing intractability issues and diversity of solutions over other methods, yet little has been shown regarding its robustness. In this regard, this paper attempts to describe the MOEA's robustness to changes in the input factors such as recombination and mutation probabilities, density of network, and type of network configuration.

2. PROBLEM DEFINITION

2.1 Multi-objective shortest path problem

Given a directed graph $G = (V, E)$, where V is the set of vertices (nodes) and E the set of edges (arcs) with cardinality $|V| = n$ and $|E| = m$ and a d -dimensional function vector $c: E \rightarrow [\mathcal{R}^+]^d$. Each edge e belonging to E is associated with a cost vector $c(e)$. A source vertex s and a sink vertex t are identified. A path p is a sequence of vertices and arcs from s to t . The cost vector $C(p)$ for linear functions of path p is the sum of the cost vectors of its edges, that is $C(p) = \sum_{e \in p} c(e)$ while $C(p) = \min_{e \in p} c(e)$ for min-max functions. Given the two vertices s and t , let $P(s, t)$ denote the set of all s - t paths in G . If all objectives are to be minimized, a path $p \in P(s, t)$ dominates a path $q \in P(s, t)$ iff $C_i(p) \leq C_i(q)$, $i = 1, \dots, d$ and we write $p \preceq q$. A path p is Pareto-optimal if it is not dominated by any other path and the set of nondominated solutions (paths) is called the Pareto-optimal set. The objective of the MSPP is to compute the set of nondominated solutions that is the Pareto-optimal set \mathbf{P} of $P(s, t)$ with respect to c .

The problem of the single-source multi-objective shortest path is to find the set of all paths from s to all other vertices in G , i.e. to find the Pareto-optimal set of $P(s, t)$, $\forall t \in V$

2.2 Sensitivity analysis

The purpose of uncertainty analysis is to determine the uncertainty in estimates for dependent variables of interest (Saltelli 2000, Saltelli 1993). The purpose of sensitivity analysis (SA) is to determine the relationships between the uncertainty in the independent variables and the uncertainty in the dependent variables. Uncertainty analysis typically precedes SA since there is no reason to perform SA when the uncertainty in a dependent variable is under an acceptable bound or range. Sensitivity analysis is the study of how variation in an observed response can be apportioned to

different possible sources or factors X . It tries to determine how the output Y varies according to changes in the inputs.

In the MSPP sensitivity analysis, the problem is concerned with the evaluation of the GA input parameters such as mutation rate and recombination probability, and the operators themselves in terms of their influence on the variation in finding efficient paths.

3. PROBLEM SOLUTION

3.1 Multi-objective evolutionary algorithm framework

Evolutionary algorithms are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Although randomized, EAs are by no means random instead they exploit historical information to direct the search into the region of better performance within the search space. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than their ancestors, just as in natural adaptation.

3.2 Genetic Algorithm for the MSPP

Genetic Representation. A chromosome or an individual consists of integer-identification (ID) nodes that form a path from the source node to a sink node. The length of the chromosome is variable and must not be greater than the number of nodes, n .

Initial Population. A path or a chromosome is generated randomly in an ordered sequence from the source node to the sink node. The ID of the source node s is assigned to the first locus (array index) of the chromosome. The ID of a randomly generated node v_i is assigned to the second locus such that v_i belongs to the set of nodes connected to the source node s . This procedure continues iteratively for the succeeding nodes until a simple path to the sink node t is created.

Fitness Function. SPEA2 (Zitzler *et al.*, 2002) first assigns a strength value $S(p)$, to each path p from the archive (\bar{N}) and population (N) representing the number of solutions that p dominates. Then the raw fitness $R(p)$ of each path p is calculated which measures the strength of p 's dominators. The raw fitness acts as a niching

mechanism but performs poorly when most paths in $M=N+\bar{N}$ are non-dominated, i.e. the population forms new solutions in only a few clusters, in effect compromising exploration of the search space. This phenomenon is called genetic drift. SPEA2 introduces a fitness sharing mechanism to avoid genetic drift. The density estimator is defined as the inverse of the distance of an individual in objective space to the k -th nearest neighbor. The density value is then added to the raw fitness value to give the final fitness function value.

Selection. SPEA2 offers two selection procedures: environmental and mating selection. The environmental selection is concerned with choosing individuals that will have to move on to the next generation archive from the current archive and population. It is a form of an archive update operation. SPEA2 maintains an archive in each generation composed of the “best” individuals with a fixed size \bar{N} which is equal to the population size N . Two usual situations may occur. First, the number of non-dominated solutions in the next generation archive is less than \bar{N} . This case requires more individuals in the next generation archive to be \bar{N} . SPEA2 resolves this by adding the “best” dominated individuals from the current archive and the population. Second, the number of non-dominated solutions for the next generation is greater than \bar{N} . SPEA2 uses a truncation procedure whereby the individual with the minimum distance to another individual is truncated until the next generation archive equals \bar{N} . On the other hand, mating deals with the selection of parents from the archive population for variation. SPEA2 implements binary tournament selection with replacement to fill in the mating pool. This type of mating selects two solutions at a time in each tournament.

Recombination. The crossover scheme is an adaptation of the one-point crossover. For each pair of paths a locus is randomly selected from one of the chromosomes (the shorter path in terms of number of nodes) and the node ID of the locus is matched with the genes in the other chromosome. If there is a match then crossover is performed otherwise two new paths are selected for crossover until the mating pool is empty. It should be easy to see that the loci of both individuals need not be the same (see Fig. 1).

Mutation. In the mutation operator, a locus is randomly selected from the chromosome. The algorithm proceeds by employing the method in the initialization process (as described previously) to create a new path, but the start node is replaced by the locus.

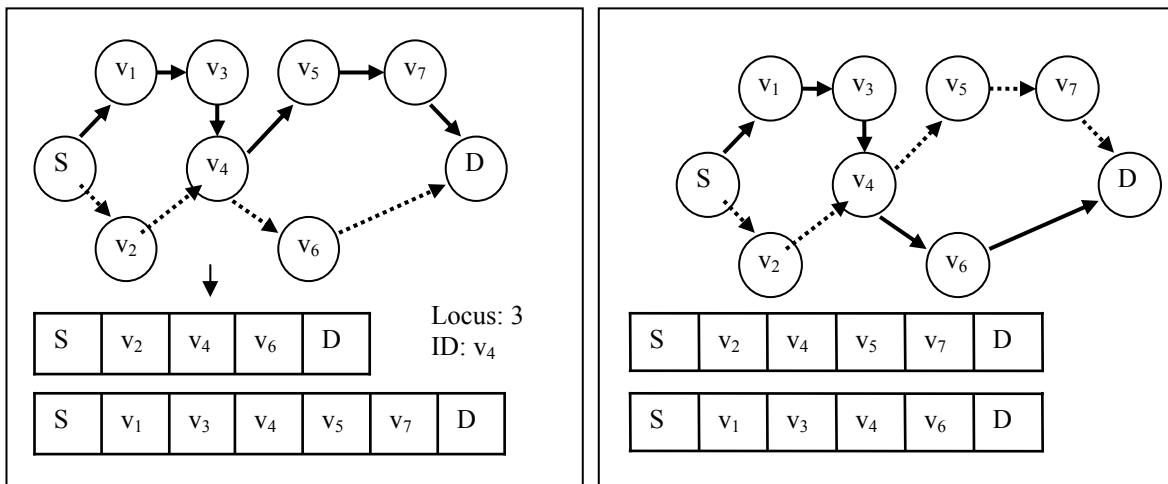


Fig 1. Crossover example. Crossover starts at the locus = position 3. Note the change in chromosome lengths.

3.3 Sobol' method

Sobol' methods (Sobol, 1993; Saltelli et al., 2000) are variance-based global sensitivity analysis methods based upon total sensitivity indices (TSI) that take into account interaction effects. The TSI of an input is defined as the sum of all sensitivity indices involving that input. The TSI includes both the main effect as well as interaction effects. For example, if there are three inputs X_1 , X_2 and X_3 , the TSI of input X_1 is given by $S(X_1) + S(X_1 X_2) + S(X_1 X_2 X_3)$, where $S(X_i)$ is the sensitivity index of X_i . $S(X_1)$ refers to the main effect of X_1 . $S(X_1 X_2)$ refers to the interaction effect between X_1 and X_2 . $S(X_1 X_2 X_3)$ refers to the interaction effect between X_1 , X_2 , and X_3 . Effort has been made to reduce the computational complexity associated with calculation of Sobol' indices (Saltelli 2002a). Sobol' methods can cope with both nonlinear and non-monotonic models, and provide a truly quantitative ranking of inputs and not just a relative qualitative measure (Chan et al., 2000). The types of influence of an input that are captured by Sobol' methods include additive, nonlinear or with interactions. Furthermore, Sobol' methods can be smoothly applied to categorical variables without re-scaling. Sobol (1993) and Saltelli (2002b) describe such an implementation. The Sobol' method can be difficult to apply to models with large number of inputs and complex model structures. The method can be used only with a set of uncorrelated factors.

4. EXPERIMENTS AND RESULTS

Nine random networks have been chosen from Gandibleux et al. (2006) which have the following characteristics: a) number of vertices: 50, 100, 200 and b) network density: 5%, 10%, and 20%. Each network has three objectives to be minimized. The initial population is estimated to be 35 as computed from the general formula $20\sqrt{d}$, where d is the number of dimensions or objectives. The input parameters to be ana-

lyzed are *mr*ate, *xover*, and *maxgen* which are the mutation probability, recombination probability, and number of generations respectively. The configurations of the input parameters are generated using the Sobol method and their distributions are listed in Table 1. There are 128 configurations for the input parameters which are used in all nine networks.

Table 1.

<i>Input factor</i>	<i>Description</i>	<i>Probability Distribution</i>
<i>mr</i> ate	<i>Mutation probability</i>	<i>Uniform(0,1)</i>
<i>xover</i>	<i>Recombination probability</i>	<i>Uniform(0,1)</i>
<i>maxgen</i>	<i>Number of generations</i>	<i>Uniform(1,10)</i>

SIMLAB (2004) is a computer program designed for global uncertainty and sensitivity analysis based on Monte Carlo methods. It offers several techniques for sample generation, sensitivity analysis, and a link to external model execution. The link allows execution of complex models that can hardly be coded as simple mathematical functions such as genetic algorithms. Fig. 2 shows the schema of the external model execution in SIMLAB.

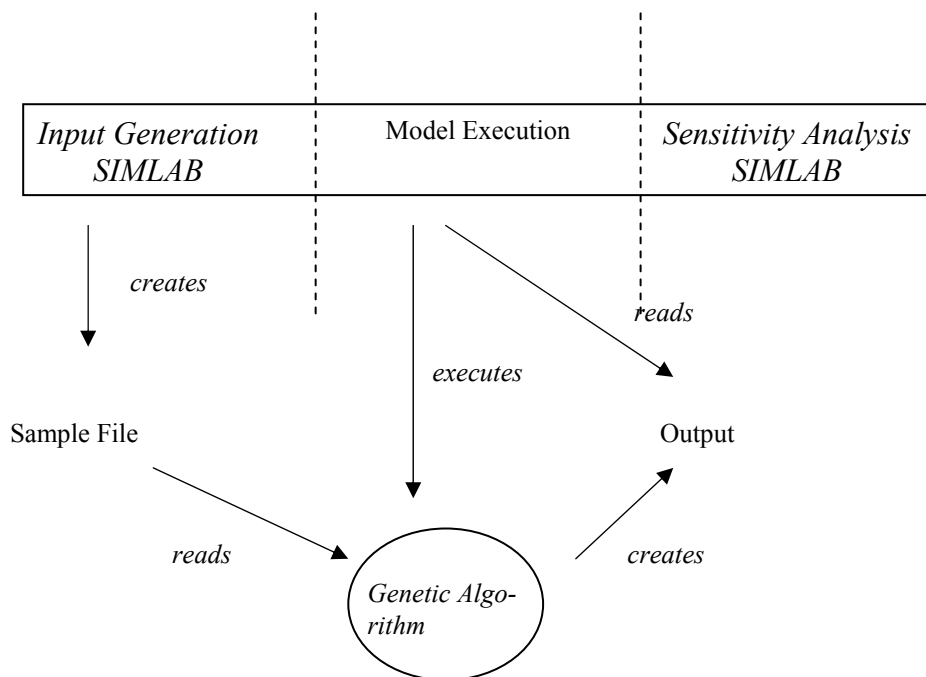


Fig. 2. External model execution.

In general, a Monte Carlo sensitivity analysis involves five steps (Simlab, 2004). In the first step, a range and distribution are selected for each input variable. If the analysis is primarily of an exploratory nature, then rough assumptions of the distribution may be adequate. In the second step, a sample of points is generated from the distribution of inputs specified in the first step. The result is a sequence of samples

(input sample). In the third step, the model is fed with the samples and a set of model outputs is produced. In essence, these model evaluations create a mapping from the input space to the space of the results. This mapping is the basis for subsequent uncertainty and sensitivity analysis. In the fourth step, the results of model evaluations are used as the basis for uncertainty analysis. In the fifth step, the results of model evaluations are used as the basis for sensitivity analysis.

Table 2 shows a sample of a Sobol first-order and total-order indices of a network with 200 vertices and a network density of 20% (output of other networks show similar results). The first-order sensitivity index shows the individual effect of an input factor on the output, i.e. it gives a measure of the direct effect of an input factor on the output variation. An input parameter having a first-order index with the least value means that it has the least influence on the output whereas a factor with the highest first-order value is most important for further investigation. The sum of the Sobol first-order indices does not equal 1.0 which means that the model is nonlinear and implies that some effects on the output are due to interactions among the input factors. Table 2 shows that 39% of the variation is due to the number of generations the GA has executed, and that mutation and recombination have almost no direct influence on the variation in the output. This means that the efficient paths are mainly generated by the *selection operator* of the evolutionary algorithm.

Table 2.

<i>Sobol first-order indices</i>		<i>Efficient paths</i>
	<i>Mutation probability</i>	-0.201
	<i>Recombination probability</i>	-0.207
	<i>Number of generations</i>	0.394582
<i>Sobol total-order indices</i>		
	<i>Mutation probability</i>	0.526636
	<i>Recombination probability</i>	0.557120
	<i>Number of generations</i>	1.122116

The total-order indices on the other hand describe the share of the output variation that is related to each input factor. This includes the direct effect as well as interactions with other factors and only factors with very small total-order indices can be removed to avoid significant changes in the output. Table 2 shows that the total-order indices of mutation and recombination are not small hence, removal of these parameters is not recommended.

Results also show that the MOEA converges to efficient paths after a few generations, in this case a maximum of 10 generations (no improvement after 10 generations). It should be noted that during the sensitivity analysis, the population in each configuration was constant at 35 and that no significant improvements were recorded after the 10th generation. Further runs with the MOEA show that an increase in the population size yields an increase in the number of efficient paths found, but still the

mutation and recombination operators have little influence on the variation in the output.

5. CONCLUSION

The paper explores the robustness of a MOEA to changes in its input factors as applied to finding a set of solutions to the multi-objective shortest path problem. The sensitivity analysis shows that the mutation and recombination does not cause much of the variation in the output. The number of generations on the other hand, has the most influence on the output which means that exploration and exploitation of the search space are mostly influenced by the selection operator. Further experimentation also shows that an increase in population size improves the number of efficient paths found while maintaining the maximum number of generations constant.

While the mutation and recombination operators have almost no direct influence on the output, their total sensitivity indices which include interaction effects, are not negligible. Further investigation is required to know why this is case.

In terms of robustness and as applied to the MSPP, the MOEA is not sensitive to changes in probability values of the mutation and recombination operators and to the operators themselves. However, the MOEA is sensitive to the population size and to the selection operator particularly at the early generations of the MOEA.

REFERENCES

- Chan., K., S. Tarantola, A. Saltelli, I.M. Sobol (2000). Variance Based Methods, In: *Sensitivity Analysis*, 167-197.
- Coutinho-Rodrigues, J., J. Climaco, J., Current, (1999). An interactive bi-objective shortest path approach: searching for unsupported nondominated solutions. *Computers & Operations Research* 26, 789-798.
- Crichigno, J., B. Baran (2004). A multicast routing algorithm using multi-objective optimization. In: *Proc. Telecommunications and Networking ICT 2004: 11th International Conference on Telecommunications LNCS 3124* (J. Neuman de Souza et al.), 1107-1113. Springer, Berlin.
- Ehrgott, M., X. Gandibleux (2000). A survey and annotated bibliography of multi-objective combinatorial optimization. *OR Spektrum*, 4(22), 425-460.
- Gandibleux, X., F. Beugnies, S. Randriamasy (2006). Martins' algorithm revisited for multi-objective shortest path problems with a MaxMin cost function. *4OR: A Quarterly Journal of Operations Research*, 1(4), 47-59.
- Garey, M. R., D. S. Johnson (1979). *Computers and Intractability: A guide to the theory of NP-completeness*. Freeman, San Francisco CA.
- Gen, M., L. Lin (2004). Multi-objective genetic algorithm for solving network design problem. *20th Fuzzy Systems Symposium, June 2-5, 2004*, Kitakyushu, Japan, 290-291.
- Granat, J., F. Guerriero (2003). The interactive analysis of the multicriteria shortest path problem by the reference point method. *European Journal of Operational Research* (151), 103-111.

- Hansen, P. (1979). Bicriteria path problems. In: *Lecture Notes in Economics and Mathematical Systems* 177 (Fandel G. and Gal, T. eds.), 109-127. Springer, Berlin.
- Kumar, R., N. Banerjee (2003). Multicriteria network design using evolutionary algorithm. In: *Proc. Genetic Evolutionary Computation GECCO 2003, LNCS 2724*, (Cantu-Paz et al.), 2179-2190. Springer, Berlin.
- Lacomme, P., C. Prins, M. Sevaux (2003). Multi-objective capacitated arc routing problem. In: *Proc. Evolutionary Multi-Criterion Optimization: Second International Conference, EMO 2003, LNCS 2632* (Fonseca, C. et al.), 550-564. Springer, Berlin.
- Müller-Hannemann, M., K. Weihe (2001). Pareto shortest paths is often feasible in practice. In: *Proc. Algorithm Engineering: 5th International Workshop, WAE 2001, LNCS 2141* (Brodal, G. et al.), 185-198. Springer, Berlin.
- Pangilinan, J., G. Janssens (2007). Evolutionary algorithms for the multiobjective shortest path problem. *International Journal of Applied Science, Engineering and Technology* (19), 205-210.
- Saltelli, A. T. H. Andres, and T. Homma (1993). Sensitivity of model output: an investigation of new techniques. *Computational Statistics and Data Analysis* 2(15), 211-239.
- Saltelli A., K. Chan, M. Scott (2000). *Sensitivity Analysis*, Probability and Statistics Series. New York, Wiley.
- Saltelli, A. (2002a). Sensitivity analysis for importance assessment. *Risk Analysis* 22(3), 1-12.
- Saltelli, A. (2002b). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 145(2), 280-297.
- Simlab 2.2 (2004). *SIMLAB Reference Manual*. European Commission-ISPC.
- Sobol, I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computation* 1(4), 407-414.
- Tsaggouris, G., C. Zaroliagis (2006). Multi-objective optimization: improved FPTAS for shortest paths and non-linear objectives with applications. *Technical Report No. TR 2006/03/01*, Computer Technology Institute, University of Patras, Greece. Also in: Proc. ISAAC, 389-398.
- Warburton, A. (1987). Approximation of Pareto-optima in multi-objective shortest path problems. *Operations Research* (35), 70-79.
- Zitzler, E., M. Laumanns, L. Thiele (2002). SPEA2: Improving the strength Pareto evolutionary algorithm for multi-objective optimization. In: *Evolutionary Methods for Design, Optimization and Control* (Giannakoglou, K. et al.), 19-26. CIMNE, Barcelona.