1668

Cahier No 8020

COMBINATORIAL PROGRAMMING, STATISTICAL OPTIMIZATION AND THE OPTIMAL TRANSPORTATION NETWORK PROBLEM

by

Marc Los and Christian Lardinois

This work was supported by the Research and Development Centre of Transport Canada, by the FCAC program of the Ministry of Education of Québec and the Prêt d'Honneur Inc. of Université de Montréal.

Université de Montréal

Centre de recherche sur les transports - Publication # 171

Mai 1980

ABSTRACT

This paper presents and evaluates a branch and bound algorithm and two heuristic hill-climbing techniques to solve a discrete formulation of the optimal transportation network design problem. For practical applications it is proposed to combine a hill-climbing algorithm with a uniform random generation of the initial solutions, thereby inducing a statistical distribution of local optima. In order to determine when to stop sampling local optima and in order to provide an estimate of the exact optimum based on the whole distribution of local optima, we follow previous work and fit a Weibull distribution to the empirical distribution of local optima. Several extensions are made over previous work in statistical optimization: in particular, a new confidence interval and a new stopping rule are proposed. The numerical application of the statistical optimization methodology to the network design algorithms consolidates the empirical and practical validity of the "Weibull" approach. Numerical experiments with hill-climbing techniques of varying power suggest that the statistical optimization method is best applied with heuristics of *intermediate* quality: such heuristics provide many distinct sample points for statistical estimation while keeping the confidence intervals sufficiently narrow.

RÉSUMÉ

Cet article se propose de présenter et d'évaluer un algorithme de branch and bound et des méthodes heuristiques de hill-climbing pour résoudre une formulation discrète du problème de design d'un réseau de transport optimal. En vue d'applications pratiques on propose de combiner un algorithme de hill-climbing avec une génération uniforme aléatoire des solutions initiales, ce qui induit une distribution statistique d'optima locaux. Conformément à des travaux antérieurs nous ajustons une distribution de Weibull à la distribution empirique des optima locaux, afin de déterminer une règle d'arrêt et afin d'obtenir une estimation de l'optimum exact fondée sur toute la distribution des optima locaux. Plusieurs extensions et améliorations sont apportées aux travaux antérieurs en optimisation statistique: en particulier on propose un nouvel intervalle de confiance et une nouvelle règle d'arrêt. Les applications numériques de cette méthodologie statistique aux algorithmes de design de réseau confirment la validité pratique et empirique de l'approche "Weibull". L'expérimentation avec des techniques de hill-climbing de puissance variable semble suggérer que la méthodologie proposée est plus pertinente avec des heuristiques de qualité *intermédiaire*: de telles heuristiques fournissent de nombreux points distincts pour l'estimation statistique tout en conservant des intervalles de confiance suffisamment étroits.

TABLE OF CONTENTS

I - The optimal network design problem: conceptual and computational problems

A. The optimal transportation network design problem

An important problem in transportation planning is the development of a strategy for allocating investments among different transportation projects. An example of such an allocation problem arises when there are several alternative possible improvements to a given road network in an urban area and when there are budget limitations. There is a tradeoff between the accessibility to the users and the infrastructure cost which has to be explicitly taken into consideration. As in other problems of allocation of investments, mathematical programming has been proposed as a method to generate an optimal investment strategy.[1]

In keeping with previous works in the field we will propose an optimization model which takes into account two criteria: the total cost to the users and the total infrastructure cost. Several modelling strategies are available: we can treat network capacity as a continuous or as a discrete variable and we can treat the infrastructure cost as a constraint or as part of the objective function of the optimization problem. In this work we formulate the optimal network problem as a discrete optimization problem: due to economies of scale in construction this approach is more realistic than a continuous approach. We also include the infrastructure

---

[1] Previous works in the field include, inter alia, Billheimer and Gray (1973), Boyce, Farhi and Weischedel (1973), Dionne and Florian (1979), LeBlanc (1975), Los (1979), Steenbrink (1974a and 1974b), Boyce, ed. (1979), Wong (1978a).

cost in the objective function: this makes the problem easier to solve numerically. By varying the relative weight of the two terms of the objective function we can vary the optimal "expense on the network" thereby adapting the methodology to the case of an actual budget constraint. Besides, it is possible to interpret the weight as a function of a social rate of discount. (See Section III.A below).

B. Objectives of this work and organization of the paper

The proposed mathematical programming formalization of the network problem leads to computational difficulties: in order to solve it exactly we have to use a branch and bound algorithm which becomes very rapidly inefficient for problems of a realistic size. Therefore we have to develop heuristic techniques. Among those, hill-climbing algorithms are found experimentally to be the best, both in terms of solution quality and efficiency (Los, 1975, 1976, 1978; Billheimer and Gray, 1973). A hill-climbing algorithm starts from an initial solution and uses a systematic rule to improve on it until no improvement is possible by the rule. We then have a local optimum. Several local optima can be obtained in this way depending on the initial solution. The global optimum is one of the potential local optima. In order to increase the probability of reaching the global optimum, we can sample initial solutions at random according to a uniform distribution for instance. We would generate as many local optima as we could afford and use the best local optimum obtained as an estimate of the global optimum. This approach was suggested and followed previously (Los, 1975, 1978) in the context of the simultaneous optimization of land use and transportation. It raises two practical problems. How far is the best local optimum obtained from the global optimum? When should we stop sampling initial solutions?

We could answer these two questions using a statistical approach if we could make specific assumptions about the distribution of the local optima. Such an approach was proposed by McRoberts (1971) and applied to the Travelling Salesman Problem by Golden (1977) and Golden and Alt (1979). We will follow and extend this methodology in this paper and apply it to the optimal network problem.

The rest of the paper is organized as follows. Section II describes in detail hill-climbing techniques and presents the statistical optimization approach to combinatorial programming. It presents the state of the art as well as some original extensions concerning in particular a confidence interval for the global optimum and a stopping rule for the generation of local optima. Section III formulates the optimal network problem as a combinatorial programming problem and presents solution algorithms, a branch and bound algorithm as well as two hill-climbing techniques. Section IV applies to the problem and the hill-climbing algorithms of Section III the statistical methodology of Section II and presents in detail the results of the numerical experiments. Section V concludes the paper.

II - Hill-climbing techniques and statistical optimization

A. Hill-climbing techniques

1. Formalization of hill-climbing algorithms

In order to understand the rest of the paper, it is necessary to characterize hill-climbing procedures. This will be done in accordance with definitions due to Reiter and Sherman (1965)[1]. We are concerned with a discrete optimization problem formulated as a minimization problem.

A hill-climbing procedure can be interpreted by means of a *successor structure*. Let M be the set of all possible solutions ρ to a combinatorial programming problem and let $X(\rho)$ be the objective function associated to the solution ρ. A *successor structure* on M is a (π,s) pair where

i) π is a collection of subsets $\{\pi(\rho);\rho\epsilon M\}$ such that

- for each $\rho\epsilon M$ there exists one and only one subset $\pi(\rho)\epsilon\pi$
- for each $\rho\epsilon M$, $\rho\epsilon\pi(\rho)$

ii) s is a function defined on M such that

- $s(\rho)\epsilon\pi(\rho)$
- $X(s(\rho)) \leq X(\rho)$
- $X(s(\rho)) = X(\rho) \Rightarrow s(\rho) = \rho$
- $s(\rho) = \rho \Rightarrow X(\rho') \geq X(\rho)$ for all $\rho'\epsilon\pi(\rho)$.

The collection π is called a neighborhood structure on M,$\pi(\rho)$ is the neighborhood of ρ; the function s is a successor function on M. This

[1] These authors have actually gone as far as one could go in answering questions about the distribution of local optima without the "Fisher-Tippett"analogy used in this work. This fact seems to have been overlooked by later works.

successor structure associates to each solution $\rho\epsilon M$ a solution $s(\rho)$ in its neighborhood where the objective function is at most equal to $X(\rho)$: $s(\rho)$ is the *successor* of ρ.

Other definitions are necessary,

a) If, in $\pi(\rho)$ no solution distinct from ρ has a value strictly less than $X(\rho)$, then ρ is its own successor: $s(\rho) = \rho$. By definition, if $\rho\epsilon M$ is such that $s(\rho) = \rho$, ρ is a *locally optimal* solution and $X(\rho)$ is a *local optimum*.

b) The t^{th} successor of ρ, $s^t(\rho)$ is defined by $s^t(\rho) = s(s^{t-1}(\rho))$ where $t=2,3,\ldots$.

c) A solution $\rho^*\epsilon M$ such that $X(\rho^*) \leq X(\rho)$, for all $\rho\epsilon M$ is *globally optimal* and $X(\rho^*)$ is a *global optimum*.

These definitions formalize the mechanism of a hill-climbing algorithm. Any heuristic algorithm of this type starts from an initial $\rho\epsilon M$ and improves on it by successive systematic but small improvements until it is no longer feasible to improve the current solution. These improvements are obtained by applying systematic rules which proceed as follows:

i) One associates to the current solution the subset of solutions which can be obtained by a minor modification whose type is a priori defined. This subset of solutions, which are *adjacent* to the current solution ρ constitutes its neighborhood, $\pi(\rho)$.

ii) In the neighborhood $\pi(\rho)$ of the current solution ρ, one seeks the adjacent solution with the least value of the objective function and one implements the change corresponding to the "best" adjacent solution. The latter is the successor $s(\rho)$ defined earlier (there must be a rule to break ties). It becomes the new current solution. The procedure is continued until it is no longer possible to improve on the current solution: one then has obtained a local optimum.

Some facts bear emphasis:

a) A solution is locally optimal relative to a given successor structure (π,s), i.e. relative to a given hill-climbing algorithm.

b) For any successor structure (π,s) there exist locally optimal solutions. In particular a globally optimal solution ρ^* is also locally optimal relative to (π,s).

c) For any successor structure (π,s) and for any solution $\rho \in M$, there exists an integer number t (dependent on ρ) such that $s^{t+1}(\rho) = s^{t}(\rho)$; by definition for any $\rho \in M$, $\bar{\rho}$ is called the *ultimate successor* of ρ if there exists an integer t such that $s^{t+r}(\rho) = \bar{\rho}$ for all $r \geq 0$; thus any solution ρ has an ultimate successor $\bar{\rho}$ which is nothing else than a locally optimal solution. This ultimate successor is unique. Distinct initial solutions may lead to distinct local optima.

d) To any solution $\rho \in M$, one can associate the set $T(\rho) = \{q \in M : \bar{q} = \bar{\rho}\}$; containing all the solutions with the same ultimate successor as ρ. Thus, the set M can be partitioned into disjoint subsets, each corresponding to a different locally optimal solution. To a different hill-climbing algorithm (π,s) would correspond in general another partition of M and other local optima.

Reiter and Sherman (1965) also defined a *probability structure*. Let $\bar{\rho}$ be the ultimate successor of an initial solution ρ. If this initial solution is generated randomly according to a uniform distribution, the probability of choosing an initial solution whose ultimate successor is the local optimum $\bar{\rho}$, is:

$$\text{Prob}(\bar{\rho}) = \frac{|T(\bar{\rho})|}{|M|} \tag{1}$$

Thus, the combination of a hill-climbing algorithm and of a uniform distribution to generate initial solutions *induces* a probability distribution on the subset H of the locally optimal solutions associated with the hill-climbing algorithm.

2. Classification of hill-climbing algorithms

The following definitions and theorems will be useful in distinguishing between different hill-climbing algorithms for the same problem and in interpreting the numerical results of the subsequent sections. They aim at formalizing the concepts of power of and *equivalence* between different hill-climbing algorithms for the same problem. The concept of neighborhood $\pi(\rho)$ of a solution ρ was defined as the set of solutions which are *adjacent* to ρ, i.e. which can be obtained from ρ by a small well-defined modification. To this notion of neighborhood, introduced by Reiter and Sherman (1965), we add the concept of *explored neighborhood* $\pi E(\rho)$, defined as the set of adjacent solutions *actually* explored, i.e. for which the objective function is actually computed. For a given hill-climbing algorithm h_i we note the set of local optima H_i, the neighborhood of a solution ρ, $\pi^{h_i}(\rho)$ and its explored neighborhood $\Pi E^{h_i}(\rho)$.

Definition: A hill-climbing algorithm h_1 is *more powerful* than a hill-climbing algorithm h_2 if:

$$\pi E^{h_2}(\rho) \subseteq \pi E^{h_1}(\rho) \quad \text{for all } \rho$$

This will be noted: $h_1 \geq h_2$.

<u>Theorem</u>: If $h_1 \geq h_2$, any local optimum $\bar{\rho}$ for h_1 is also a local optimum for h_2: i.e. $H_1 \subseteq H_2$.

<u>Proof</u>: If $\bar{\rho}$ is a local optimum for h_1, we have:

$$X(\bar{\rho}) \leq X(\rho) \qquad \text{for all } \rho \epsilon \pi E^{h_1}(\bar{\rho})$$

which implies

$$X(\bar{\rho}) \leq X\ (\rho) \qquad \text{for all } \rho \epsilon \pi E^{h_2}(\bar{\rho})$$

<u>Definition</u>: A hill-climbing algorithm h_1 is said to be *locally superior* to a hill-climbing algorithm h_2 if, for any local optimum $\bar{\rho}$ for h_2, h_1 explores exactly the same solutions as h_2, i.e. if:

$$\pi E^{h_1}(\bar{\rho}) \equiv \pi E^{h_2}(\bar{\rho}) \qquad \text{for all } \bar{\rho} \epsilon H_2.$$

We will note "$h_1 >_{\ell} h_2$"

<u>Theorem</u>: If $h_1 >_{\ell} h_2$, any locally optimal solution for h_2 is also a locally optimal solution for h_1, i.e.: $H_2 \subseteq H_1$.

<u>Proof</u>: If $\bar{\rho}$ is a local optimum for h_2, we have:

$$X(\bar{\rho}) \leq X(\rho) \qquad \text{for all } \rho \epsilon \pi E^{h_2}(\bar{\rho})$$

which implies $X(\bar{\rho}) \leq X(\rho) \qquad \text{for all } \rho \epsilon \pi E^{h_1}(\bar{\rho})$

<u>Corollary</u>: If h_1 is more powerful than h_2 and "locally superior" to h_2, then: $H_1 \equiv H_2$. The two hill-climbing algorithms are then said to be *equivalent*, which is noted: $h_1 \sim h_2$. This corollary is in fact counterintuitive; one would not expect a more powerful hill-climbing algorithm to lead to the same set of local optima than a less powerful one.

B Statistical optimization

In this section we will describe a methodology which will allow an answer to the two questions raised in Section I; how far from the global optimum is the best local optimum obtained by a hill-climbing technique? When should we stop sampling initial solutions, i.e. when should we consider that we have generated enough local optima?

1. The state of the art

If we could explicitly enumerate all the possible solutions of a combinatorial programming problem, we could compute the probability for a given solution ρ_i chosen at random in the set M, to have a value less than a given number x. This probability is:

$$\text{Prob } \{X(\rho_i) \leq x\} = \frac{|\{\rho : \rho \in M; X(\rho) \leq x\}|}{|M|} \tag{2}$$

These probabilities would give the cumulative distribution function associated with the set M, provided we considered X as a random variable. This "parent-population" is unfortunately unknown. We only know that it is bounded from below by the global obtimum x^*.

Nevertheless it is possible to estimate the probability that, in a sample of size N, at least one observation is found in the lowest fraction p of the population (McRoberts, 1971). Thus we can determine the size N of the sample necessary to be certain, at a level of 99%, that at least one observation is in the lowest 1 percent of the parent distribution. It turns out that this number is approximately 1,000. That lowest 1 percent still contains too many solutions, which can in fact vary widely in quality.

Another approach consists in combining random search and hill-climbing, as suggested by Reiter and Sherman (1965) and to make use of the empirical distribution of the local optima thus obtained. Given that we do not know much about the parent-population, we will avoid making an assumption on its associated distribution by using non-parametric statistical inference (see Gibbons, 1971). This will allow us to use some well-founded assumptions on the distribution of local optima independently of the unknown distribution of the parent-population.

Dannenbring (1977), Robson and Whitlock (1964) have made use of the theory of order statistics to compute an estimate of the lower bound of any truncated distribution. However they use only random solution values, thereby making no use of the information which would be provided by the empirical distribution of the local optima generated by a hill-climbing algorithm. McRoberts (1971) on the other hand has proposed an approach making use of this empirical distribution which is based on the *theory of extreme values* (Gumbel, 1958; Gnedenko, 1943). His ideas are at the origin of the work reported in this paper and will therefore be presented in detail below.

McRoberts (1971) uses a theorem due to Fisher and Tippett (1928). Consider N independent samples of size m, obtained from a parent-population of density $f(X)$. Let x_i $(i=1,\dots,N)$ be the least value in each sample. We thus have a surpersample $\{x_i,x_2,x_N\}$ of extreme values. Each x_i is the value taken by the first order statistic of a sample of size m. The first order statistic has a distribution which depends on m and f(X). When m becomes very large the distribution of the extreme values x_i tends to a limit distribution which belongs to one of three types (types I, II and III of Gumbel). If the parent-distribution f(X) has a lower bound, the asymptotic distribution is of type III, also called a *Weibull* distribution. It must be emphasized that this limit distribution is independent of the parent-distribution f(X). The cumulative distribution function of this asymptotic distribution is:

$$\Phi(X) = 1 - \exp\left[-\left(\frac{X-a}{b}\right)^{c}\right] \qquad (3)$$

and the density function is:

$$\delta(X) = \frac{c}{b}\left(\frac{X-a}{b}\right)^{c-1} \exp\left[-\left(\frac{X-a}{b}\right)^{c}\right] \qquad (4)$$

$$X \geq a > 0 \ ; \quad c > 0 \ , \quad b > 0.$$

The parameters a, b and c are called respectively the location parameter, the scale parameter and the shape parameter. The parameter a, lower bound of the Weibull distribution is simultaneously the lower bound of the parent-distribution. If F(X) is the cumulative distribution function for the parent-distribution the necessary and sufficient conditions for the existence of a type III asymptotic distribution are the following (see Gnedenko, 1943):

- The domain on which the distribution F(X) is defined must have a lower bound a: $F(X) \equiv 0$, for $X \leq a$.
- F(X) must behave like $b(X-a)^c$ for b, c > 0 when $X \to a$.

McRoberts' idea consists in interpreting the sequence of solutions which are built during the application of a hill-climbing algorithm to a random initial solution, as one the samples of the Fisher-Tippett theorem. The local optimum x_i^h is interpreted as the minimum value of that sample. When one has N local optima x_i^h (i=1,2,...,N), one is in a situation analogous to the one considered by the Fisher-Tippett theorem and it seems reasonable to assume that the distribution of local optima approaches the Weibull distribution, whatever the parent-distribution may be. By estimating the location parameter of the Weibull distribution, we obtain an estimate of the global optimum. McRoberts did not provide any indication concerning the precision of the estimate.

Golden (1977) and Golden and Alt (1979) improved McRoberts' approach and gave a new interpretation of the application of the Fisher-Tippett theorem to the empirical distribution of local optima. For Golden (1977) the N samples of the Fisher-Tippett theorem as applied to the distribution of local optima are made of the sets of implicit initial solutions which would lead to the different local optima, i.e. the sets T(ρ) (as defined in Section II.A.1). Otherwise the principle of Mc Roberts' method is unchanged: by considering the local optima obtained by a hill-climbing algorithm applied to randomly generated initial solutions, as extreme values of implicit samples, we obtain an estimate of the global optimum by estimating the location parameter of the Weibull distribution fitted to the distribution of local optima.

Several methods have been proposed to estimate the parameters of the Weibull distribution. The procedure suggested by Harter and Moore (1965) is recommended by Mann, Schafer and Singpurwalla (1974) and by Zanakis (1977).

Zanakis compared systematically different estimation methods among which the Harter and Moore method worked very well. This method is based on the principle of maximum likelihood. This method was utilized by Golden and Alt (1979) and has been used in the work reported here. Let $\hat{x}^* = \hat{a}$ be the estimate of the global optimum.

Golden and Alt suggested a confidence interval for the global optimum, in addition to the point estimate. It is obtained as follows. The Weibull distribution always obeys the following equality

$$\phi(a+b) = 1 - e^{-1} \quad . \tag{5}$$

The N observations x_i^h $(i=1,2,\ldots,N)$ are a sample drawn from a Weibull distribution where $x_{(1)}^h$ is the best local optimum.

$$\text{Prob } \{x_{(1)}^h \leq a + b\} = 1 - \text{Prob } \{x_{(1)}^h > a + b\} \tag{6}$$

$$= 1 - \{1 - \Phi_{x_1^h}(a+b)\} \{1 - \Phi_{x_2^h}(a+b)\} \ldots \{1 - \Phi_{x_N^h}(a+b)\} \tag{7}$$

$$= 1 - (e^{-1})^N = 1 - e^{-N} \tag{8}$$

By making use of the estimate $\hat{b}$ of the scale parameter we obtain an approximate $100(1-e^{-N})\%$ confidence interval:

$$\text{Prob } \{x_{(1)}^h - \hat{b} \leq x^* \leq x_{(1)}^h\} \simeq 1-e^{-N} \tag{9}$$

It will be shown below that other confidence intervals can be computed, besides this one.

Golden and Alt (1979) tested the Weibull hypothesis by means of the Kolmogorov-Smirnov test (Gibbons, 1971). They also tested the hypothesis that the N local optima are independent by means of *tests based on runs* (Gibbons, 1971). They applied their statistical methodology to the Travelling Salesman Problem.

2. Discussion of the assumptions implicit in the proposed methodology

In this paragraph we will examine the validity of the analogy which is made in order to apply the theory of extreme values to the distribution of the local optima produced by a hill-climbing algorithm.

The first assumption which is made consists in considering the local optima as extreme values of implicit samples of *statistically independent* observations. This assumption cannot be rigorously justified: all the elements in each implicit sample $T(\bar{\rho})$ have in common the property of leading to the same local optimum $\bar{\rho}$.

Secondly, the N local optima which are characteristic of a particular successor structure cannot be on *a priori grounds* considered as statistically independent, even though this assumption has to be made to apply the Fisher-Tippett theorem. However it is possible to test statistically the independence of the local optima, once they have been generated.

Thirdly, the theory of extreme values is valid for *continuous* distributions, whereas for a combinatorial programming problem the number of possible values is finite. However the combinatorial explosion of the number of solutions of discrete problems, such as the optimal network design problem, makes the approximation of a discrete distribution by a continuous distribution acceptable in practice.

The Fisher-Tippett theorem assumes that all the N samples have the same size m, and that m is very large. In fact the implicit samples corresponding to the local optima do not all have the same size and it cannot be excluded that some of the implicit samples are small.

Thus the validity of the Weibull hypothesis is not well established theoretically. Numerical experiments and in particular tests of goodness-of-fit are extremely important in justifying the hypothesis in a heuristic and empirical sense. In the work reported here, as well as in the work by Golden and Alt, the "empirical validation" of the Weibull hypothesis is made by applying the Kolmogorov-Smirnov test to the empirical distribution of local optima. The role of the test is to determine if we can accept (or have to reject) the hypothesis that the local optima could have been generated by a Weibull distribution whose parameters are the estimated parameters. In addition, the Weibull hypothesis is reinforced in practice if the local optima can be considered as statistically independent, a fact which is ascertained by making a test based on *runs*.

3. Induced distribution and pseudo-distribution

An important distinction has to be made between two kinds of distribution of local optima, which we will call the *induced* distribution and the *pseudo-distribution* of local optima.

Reiter and Sherman (1965) showed that the combination of a hill-climbing algorithm and of a mechanism of random generation of initial solutions *induces* a probability distribution on the subset H of local optima associated with the algorithm. Golden and Alt (1979) following McRoberts (1971) have proposed the hypothesis that the *induced* distribution of local optima should approach the Weibull distribution. However, even if the assumptions stated above (continuity, independence...) are justified, there remains an error in trying to fit a Weibull distribution to the *induced* distribution: at least for small problems, if we generate N local optima, there may be several identical local optima. This is equivalent to repeating several times the same implicit sample, whereas the Fisher-Tippett theorem demands the independence of the N samples. In such a situation we will apply the methodology proposed by McRoberts (1971) or Golden and Alt (1979) to the *pseudo-distribution* of local optima, obtained from the induced distribution by deleting the repetitions of the same values. Thus, if we generate N local optima, the empirical *pseudo-distribution* of local optima is obtained by keeping a sample of R (with $R \leq N$) *distinct* local optima. We then try to fit a Weibull distribution to this pseudo-distribution.

In practice, for combinatorial programming problems of large size, for which the statistical optimization approach is most relevant, there is no difference between the two distributions because the local optima generated tend to be distinct. For problems of small size the distinction is relevant. The fact that we can consider as identical the two distributions for large size problems will be very important for the stopping rule: this rule applies, strictly speaking, to the induced distribution but we will consider as identical the induced distribution and the estimated Weibull pseudo-distribution.

C. Improvements and extensions to Golden and Alt's methodology

In this paragraph we propose some improvements and extensions to the methodology of Golden and Alt (1979) and McRoberts (1971).

1. A new confidence interval

As applied to the pseudo-distribution of local optima, the confidence interval suggested by Golden and Alt is given by:

$$\text{Prob}\ \{x^{hd}_{(1)} - \hat{b} \leq x^* \leq x^{hd}_{(1)}\} \simeq 1 - e^{-R} \tag{10}$$

where $x^{hd}_{(1)} < x^{hd}_{(2)} < \ldots < x^{hd}_{(R)}$ are the ordered distinct local optima generated and where $\hat{b}$ is the estimate of the scale parameter. In the numerical experiments which were made in the course of this work it appeared that this interval was too large: it contained too large a fraction of the range of variation of the R distinct local optima $(x^{hd}_{(R)} - x^{hd}_{(1)})$ and is almost independent of R.

However we can construct a new confidence interval. Let S be any real number. We always have:

$$\Phi(a + \frac{b}{S}) = 1 - \exp\ [-(\frac{1}{S})^c] \tag{11}$$

The distinct local optima x^{hd}_i (i=1,2,...,R) form a sample from the Weibull pseudo-distribution. They are such that

$$\text{Prob}\ \{x^{hd}_{(1)} \leq a + b\} = 1 - \text{Prob}\ \{x^{hd}_{(1)} > a + \frac{b}{S}\} \tag{12}$$

$$= 1 - \{1 - \Phi_{X^{hd}_1}(a + \frac{b}{S})\}\{1 - \Phi_{X^{hd}_2}(a + \frac{b}{S})\}\ldots\{1 - \Phi_{X^{hd}_R}(a + \frac{b}{S})\} \tag{13}$$

$$= 1 - \exp\ (-\frac{R}{S^c}) \tag{14}$$

By introducing the point estimates $\hat{b}$ and $\hat{c}$ of the scale and shape parameters, we obtain an approximate confidence interval whose width is now explicitly dependent on the degree of confidence of the interval:

$$\text{Prob}\ \{x_{(1)}^{hd} - \frac{\hat{b}}{S} \leq x^* \leq x_{(1)}^{hd}\} \simeq 1 - \exp\ (-\frac{R}{S^{\hat{c}}}) \tag{15}$$

Let:

$$1 - \exp\ (-\frac{R}{S^{\hat{c}}}) = 1 - \alpha \tag{16}$$

Then

$$S = (-\frac{R}{\ell n \alpha})^{\frac{1}{\hat{c}}} \tag{17}$$

We can see that, for given values of R, $\hat{b}$ and $\hat{c}$, an increase in the degree of confidence $(1-\alpha)$ is followed, as one would expect, by an increase in the width of the interval. The influence of the shape parameter $\hat{c}$ is also quite explicit. If the Weibull distribution is more spread out (by opposition to steep) on the left, the shape parameter is larger and, ceteris paribus, the confidence interval is also larger.

2. Combination of several heuristic techniques

The $R \leq N$ distinct observations whose distribution is fitted by a Weibull distribution are, in fact, not independent: the "successor structure" characteristic of the particular hill-climbing algorithm used entails that some solutions become local optima rather than other solutions: thus the R observations are "heuristically" related. In fact, the analogy with the Fisher-Tippett theorem only requires that the observations can be interpreted as extreme values of independent samples drawn from the same population. Therefore, instead of using only one heuristic to generate the R extrema, we could use several heuristics. The use of several different algorithms would have the following advantages:

- this method would produce observations some of which would not be "heuristically" dependent on one another.
- it would increase the number of observations used for estimating the Weibull distribution since the *union* of the set of local optima corresponding to each of the heuristics would then be used for estimation purposes.

3. Use of extrema of samples of local optima

As emphasized the elements contained in the implicit samples associated with the R distinct local optima generated by a given heuristic are in fact not independent:they share the property of leading to the same local optimum. The approach proposed here avoids that problem. It consists in explicitly generating samples of local optima and extracting the minimal value in each sample.

We generate E samples G_i of size m, in which each of the m observations in each sample is itself a local optimum obtained by the combination of random generation and of a hill-climbing algorithm:

$$G_i = \{x_1^{h,i}, x_2^{h,i}, \ldots, x_m^{h,i}\} \qquad i=1,\ldots,E \tag{18}$$

From these samples G_i we build a "super-sample":

$$\{x_{(1)}^{h,1}, x_{(1)}^{h,2}, \ldots, x_{(1)}^{h,E}\} \tag{19}$$

with $$x_{(1)}^{h,i} = \text{Min}\ \{x_j^{h,i}, \qquad j=1,\ldots,m\} \tag{20}$$

We can again apply the Fisher-Tippett analogy: the distribution of the E minima approaches asymptotically the Weibull distribution. By estimating its three parameters we obtain again a point estimate and a new confidence interval for the global optimum:

$$\hat{x}^* = \hat{a} \tag{21}$$

$$\text{Prob}\ \{x_{(1)}^{h,(1)} - \frac{\hat{b}}{S} \leq x^* \leq x_{(1)}^{h,(1)}\} \simeq 1-\alpha \tag{22}$$

where

$$x_{(1)}^{h,(1)} = \text{Min}\ \{x_{(1)}^{h,i},\ i=1,2,\ldots,E\} \tag{23}$$

$$S = \left(-\frac{E}{\ell n\alpha}\right)^{\frac{1}{\hat{c}}} \tag{24}$$

Instead of using $R \leq N = Em$ distinct local optima to fit a Weibull distribution we use only E observations $x_{(1)}^{h,i}$ $(i=1,\ldots,E)$. However the E samples of the Fisher-Tippett theorem are now *explicitly* generated and the independence property of the m elements in each sample can be tested. Finally, the observations $x_{(1)}^{h,i}$ $(i=1,2,\ldots,E)$ lead to a Weibull distribution less spread out than the pseudo-distribution estimated from the totality of the Em local optima.

4. Stopping rule

When we have generated N local optima and computed an estimate of the global optimum, we would like to know if it is worthwile to continue the search for better solutions, or if it is better to stop.

a) An elementary stopping rule

Let $x^h = X(\rho^h)$ be the best local optimum among the N local optima generated so far. The point estimate $\hat{x}^* = \hat{a}$ of the global optimum gives immediately a measure of the *potential gain* (or *potential for improvement*)

$PA(\rho^h)$ with respect to the best locally optimal solution "so far" ρ^h:

$$PA(\rho^h) = X(\rho^h) - x^* \simeq x^h - \hat{x}^* \qquad (25)$$

If C_0 is the benefit associated with one unit of value of $X(\rho)$, the quantity $PA(\rho^h).C_0$ represents the benefit of the potential gain. If C_s is the cost of searching for one more local optimum, a first stopping rule is:

If $PA(\rho^h).C_0 \geq C_s$ => continue the search

If $PA(\rho^h).C_0 < C_s$ => stop

b) <u>The probability of gain</u>

The previous rule is not sufficient: even if the potential gain is large, the *probability of gain* (or *probability of improvement*), i.e. the probability that an additional observation would give a better solution than ρ^h can be very small. If we treat as identical the *induced* distribution (in Reiter and Sherman's definition), with the *pseudo-distribution* estimated by means of the R distinct local optima of the sample of size N,[1] the probability of gain is given by:

$$\text{Prob } \{X \leq x^h\} \simeq \hat{\Phi}(x^h) \ , \qquad (26)$$

where $\hat{\Phi}(X)$ is the Weibull cumulative distribution function estimated on the basis of the R distinct local optima so far generated.

[1] This equivalence becomes valid only for problems of large size for which $\frac{R}{N} \rightarrow 1$.

c) The expected value of the gain and McRoberts' rule

A third criterion, introduced by McRoberts (1971) consists in the *expected value of the gain of an additional observation*:

$$E(\text{gain}) \approx \int_{\hat{X}^*}^{x_h} (x^h - X)\ \hat{\delta}(X)\ dX \qquad (27)$$

where $\hat{\delta}(X)$ is the density function of the Weibull distribution, estimated here on the basis of the R distinct local optima. McRoberts' rule is:

If $E(\text{gain})C_o < C_s$ => Stop

If $E(\text{gain})C_o \geq C_s$ => Continue the search

d) General rule combining the three criteria

The probability of gain and the expected value of the gain are relative to only *one* extra solution. However in order to obtain a better local optimum than x^h we may have to generate many additional local optima. Therefore we would like to extend the two criteria to the case of several additional observations. Besides it would be useful to compare the expected value of the gain to the potential for gain.

We will propose below an iterative mechanism determining

- the number of local optima which has to be generated at each iteration to control the probability of improvement on all of the additional observations,
- a stopping rule based on the comparison between the expected value of the gain during the next iteration on one side and the potential for gain and the total computing cost of the next iteration on the other side.

The probability that at least one observation out of K additional observations leads to an improvement is given by:

$$1 - [1 - \hat{\Phi}(x^h)]^K \tag{28}$$

The number of additional local optima necessary to guarantee an improvement with a probability of α is such that:[1]

$$1 - [1 - \hat{\Phi}(x^h)]^K = \alpha \tag{29}$$

$$K = \frac{\ell n(1-\alpha)}{\ell n[1-\hat{\Phi}(x^h)]} \tag{30}$$

The expected gain on K additional observations is the expected value of the random variable $(x^h - X_{(1)})$ where $X_{(1)}$ is the smallest value in a sample of K local optima. The density function of the variable $X_{(1)}$ is the density function of the first order statistic of a sample of size K drawn from a Weibull distribution of estimated density $\hat{\delta}(X)$ (Gibbons, 1971):

$$P_{X_{(1)}}(X) \simeq K\,[1-\hat{\Phi}(X)]^{K-1}\,\hat{\delta}(X) \tag{31}$$

Hence the expected value of gain on K additional observations:

$$E_K(\text{gain}) \simeq K \int_{\hat{X}*}^{x^h} (x^h-X)\,[1-\hat{\Phi}(X)]^{K-1}\,\hat{\delta}(X)dX \tag{32}$$

McRoberts' expression (27) is the particular case of expression (32) where $K=1$.

---

[1] K is in fact the smallest integer number greater than $\dfrac{\ell n(1-\alpha)}{\ell n(1-\hat{\Phi}(x^h))}$.

e) <u>Proposed iterative procedure and stopping rule</u>

1) Generate N_1 local optima and set i=1;

2) i^{th} iteration:

- Let R_i be the number of distinct local optima among the N_i optima generated and let $x^{i,h}$ be the value of the best local optimum so far. Estimate the three parameters of the Weibull distribution associated with the R_i local optima => $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$.

- Choose the probability of gain α_i to be guaranteed on the additional N_{i+1} observations of iteration (i+1). Compute N_{i+1}:

$$N_{i+1} = \frac{\ell n(1-\alpha_i)}{\ell n[1-\hat{\Phi}_i(x^{i,h})]} \qquad (33)$$

where $\hat{\Phi}_i(X)$ is the cumulative distribution function of the Weibull distribution of parameters $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$.

- Compute the expected value of the gain on the additional N_{i+1} observations

$$E_{N_{i+1}}(\text{gain}) = N_{i+1} \int_{\hat{x}^*}^{x^{i,h}} (x^{i,h}-X)[1-\hat{\Phi}_i(X)]^{N_{i+1}-1} \hat{\delta}_i(X)dX \qquad (34)$$

where $\hat{x}^* = \hat{a}_i$

- If either one of the two following conditions is met, stop the search, otherwise, go to step 3.

a) $E_{N_{i+1}}(\text{gain}) < \beta_i \cdot PA(x^{i,h})$

or $E_{N_{i+1}}(\text{gain}) < \beta_i \cdot (x^{i,h}-\hat{x}^*)$

where β_i is a parameter to be chosen.

b) $E_{N_{i+1}}(\text{gain}).C_o < C_s \cdot N_{i+1} + C_E$

where C_E is the cost of estimating the three parameters of the Weibull distribution.

3) - Generate N_{i+1} additional local optima.

- Set $N_{i+1} = N_{i+1} + N_i$

  and set $i=i+1$.

  Go back to step 2.

III - The optimal network design problem: Mathematical formulation and solution algorithms

A. Mathematical formulation

Consider a transportation network containing n vertices and Λ potential transportation links. The n vertices are assigned n land use activities which attract or generate traffic and we have $\Lambda \leq \frac{n(n-1)}{2}$. Some of the n activities can be dummy activities: their role is to account for vertices which do not generate or attract traffic. Let

$$L = \{\lambda_1, \lambda_2, \ldots, \lambda_h, \ldots, \lambda_\Lambda\}$$

be the set of potential links between pairs of vertices. Any solution network N contains a subset of L and is characterized by the vector

$$Z = (z_1, z_2, \ldots, z_h, \ldots, z_\Lambda)$$

$$\text{where } z_h = \begin{cases} 1 & \text{if } \lambda_h \in N \\ 0 & \text{otherwise.} \end{cases}$$

There is a one-to-one correspondence between the networks and the elements of the power set of $L, P(L)$. Thus the total number of potential solutions is 2^Λ. However, only *connected* networks are considered as *feasible* solutions.

The demand for transportation is assumed fixed and is expressed by the interaction matrix

$$Q = [q_{ij}] \quad , \quad i,j=1,\ldots,n$$

where q_{ij} is the interaction per unit of time between the activities located at locations i and j. Q is not necessarily symmetrical. Let w_{ij} be the total interaction between i and j:

$$w_{ij} = q_{ij} + q_{ji} \tag{35}$$

If there is a direct link λ_h between i and j, we will sometimes write

$$w_h = w_{ij} \tag{36}$$

The capital cost of building link λ_h is also assumed fixed and known and is noted: $k(\lambda_h)$, $h=1,\ldots,\Lambda$. Each link is assigned a length (or more generally a travel cost). The distances between location pairs are defined as the shortest path lengths between these pairs in the network. We assume therefore an all-or-nothing traffic assignment. The shortest distances depend on the network N and are therefore endogenous. They are noted: c_{ij}^N or $c_{ij}(Z)$, $i,j=1,\ldots,n$ where $c_{ij}^N = c_{ij}(Z)$ is the shortest length between i and j in the network N characterized by the vector Z.

The capital cost of network N is given by:

$$K(N) = \sum_{(h.\lambda_h \in N)} k(\lambda_h) = \sum_{h=1}^{\Lambda} z_h k(\lambda_h) \tag{37}$$

and the total travel cost is given by:

$$C(N) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij}\, c_{ij}^N = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij}\, c_{ij}(Z) \tag{38}$$

The objective function represents the tradeoff between capital costs and travel costs and depends explicitly on a social rate of discount r:

$$\Phi(N) = K(N) + m(r)\ C\ (N) \tag{39}$$

The parameter m(r) is a rate of substitution between K(N) and C(N). The actual computation of m(r) in a specific application depends on the time horizon of the use of the network and on the financing of the capital costs. Let us assume for simplicity that the capital cost K is incurred at the present time while the cost of travel is incurred at time t, $t=0,\ldots,\infty$. The value of the objective function becomes

$$\Phi = K + \sum_{t=0}^{\infty} \frac{C}{(1+r)^t} \tag{40}$$

$$\Phi = K + (1 + \frac{1}{r})\ C \tag{41}$$

In this example[1],

$$m(r) = 1 + \frac{1}{r} \tag{42}$$

In the rest of the paper we will assume that m(r)=1 but that does not change the generality of the approach suggested here. The objective function is therefore:

$$\Phi(N) = K(N) + C(N) \tag{43}$$

The mathematical programming problem we have to solve consists in finding at least one feasible network N such that $\Phi(N)$ as defined by (43) is minimum.

---

[1] Note that if we used a budget constraint and solved the network design problem by a Lagrangean technique, the value of the Lagrangean parameter would give the implicit social rate of discount.

It should be noted that the problem proposed here is a drastic simplification of the real investment problem. For instance only one mode of transportation is taken into account. The assignment of activities to locations is fixed and the demand for travel is also fixed. In reality both demand and land use change as a consequence of changes in the transportation system. No congestion is assumed. In spite of these simplifications the problem stated above is a difficult combinatorial programming problem which can be solved efficiently only by heuristic techniques. Such techniques would be components of solution algorithms for more realistic formulations of the network design problem.

B. A branch and bound algorithm

1. Introduction

The development and improvement of exact algorithms is interesting in its own right, to solve problems of small size and to test the quality of heuristic algorithms. The branch and bound algorithm proposed here uses a lower bound based on a theorem due to Hoang Hai Hoc (1973) and generalized by Dionne and Florian (1979). Both works used this theorem to construct a tight lower bound for the optimal network problem with a budget constraint. The theorem is here used to construct a lower bound for the problem in which the capital cost is in the objective function.

The partitioning rule, the lower bound and the search scheme are each presented in turn.

2. The partitioning rule

The links in L are ordered in the order of decreasing construction costs, i.e. λ_1 is the link with greatest construction cost and λ_Λ is the link with least construction cost. A binary decision tree is built using successively each link λ_s as a pivot. (Figure 1)

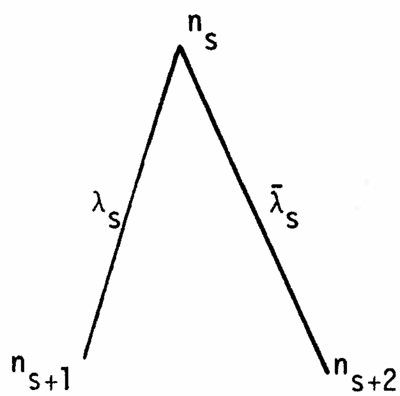


Figure 1: Partitioning rule of the branch and bound algorithm

Let n_s be a vertex of the search tree. It represents a family of networks defined by three subsets of L :

- L_s^I : set of included links
- L_s^E : set of excluded links
- L_s^u : set of unassigned links

A network N belongs to n_s if and only if N includes all the links in L_s^I and does not include any link in L_s^E.

For each n_s: $$L = L_s^I \cup L_s^E \cup L_s^u \qquad (44)$$

The separation of a vertex n_s is done using the first link λ_s on the list of unassigned links L_s^u. The two immediate descendants of n_s, i.e. n_{s+1} and n_{s+2} are characterized as follows:

$$n_{s+1} \begin{cases} L_{s+1}^I = L_s^I \cup \{\lambda_s\} \\ L_{s+1}^E = L_s^E \\ L_{s+1}^u = L_s^u - \{\lambda_s\} \end{cases} \tag{45}$$

$$n_{s+2} \begin{cases} L_{s+2}^I = L_s^I \\ L_{s+2}^E = L_s^E \cup \{\lambda_s\} \\ L_{s+2}^u = L_s^u - \{\lambda_s\} \end{cases} \tag{46}$$

We define N_s^{min} to be the network constructed with the links in L_s^I and N_s^{max} to be the network constructed with the links in $L_s^I \cup L_s^u$.
A vertex of the search tree is terminal if and only if:

$$L_s^u = \phi \tag{47}$$

For a terminal vertex,

$$N_s^{min} = N_s^{max} = N_s \tag{48}$$

i.e. there is only one network N_s in the set associated with the terminal vertex n_s.

3. Lower bound [1]

A lower bound proposed by Los (1975) is

$$v_1 = \mathrm{Max}\ \{K(N^{min}),\ K_{mst}\} + C(N^{max}) \tag{49}$$

where K_{mst} is the construction cost of a minimum construction cost spanning tree. It is easy to check that v_1 is indeed a lower bound: $K(N^{min})$ is indeed a lower bound for $K(N)$ because the construction cost can only increase if we *add* links to N^{min} and since the network has to be connected, K_{mst} is also a lower bound for the construction cost of a feasible network. $C(N^{max})$ is indeed a lower bound for $C(N)$ because, by *deleting* links, the total travel cost between all location pairs can only increase: the lenght of a shortest path between two vertices can only increase if we delete links.[2]

A new lower bound v_2 can be built in the following way. For any network N which contains at least all the links in N^{min}, at most all the links in N^{max}, the following inequality (proved by Hoang Hai Hoc (1973) and generalized by Dionne and Florian (1979)) holds:

$$C(N) \geq C(N^{max}) + \sum_{(h:\lambda_h \in L^u)} \bar{z}_h\, w_h\, f_h \tag{50}$$

where $\bar{z}_h = 1 - z_h$

---

[1] The subscript s is dropped in this paragraph. Since we are concerned with the evaluation of only one vertex n_s, there is no ambiguity.

[2] The monotonicity property of total travel cost does not hold in the presence of congestion.

and $f_h = c_{uv}^{\mathcal{N}^{max} - \{\lambda_h\}} - c_{uv}^{\mathcal{N}^{max}}$

λ_h = link (u,v)

Thus f_h is the increase in the shortest path length between u and v when λ_h is deleted from the network $\mathcal{N}^{max}$.

If we add to the previous inequality the two equalities

$$\Phi(\mathcal{N}) = K(\mathcal{N}) + C(\mathcal{N}) \tag{51}$$

and

$$K(\mathcal{N}) = K(\mathcal{N}^{min}) + \sum_{(h:\lambda_h \in \mathcal{L}^u)} z_h k(\lambda_h) \tag{52}$$

we obtain:

$$\Phi(\mathcal{N}) \geq K(\mathcal{N}^{min}) + \sum_{(h:\lambda_h \in \mathcal{L}^u)} z_h k(\lambda_h) + \sum_{(h:\lambda_h \in \mathcal{L}^u)} \bar{z}_h w_h f_h + C(\mathcal{N}^{max}) \tag{53}$$

A *fortiori*:

$$\Phi(\mathcal{N}) \geq K(\mathcal{N}^{min}) + C(\mathcal{N}^{max}) + \min_{\begin{pmatrix} z_h = (0,1) \\ h:\lambda_h \in \mathcal{L}^u \end{pmatrix}} \{ \sum_{(h:\lambda_h \in \mathcal{L}^u)} z_h k(\lambda_h) + \sum_{(h:\lambda_h \in \mathcal{L}^u)} \bar{z}_h w_h f_h \} \tag{54}$$

The third term of the RHS of this inequality is given by:

$$M(\mathcal{L}^u) = \sum_{(h:\lambda_h \in \mathcal{L}^u)} \text{Min}\ \{k(\lambda_h),\ w_h\ f_h\} \tag{55}$$

Hence a new lower bound:

$$v_2 = K(\mathcal{N}^{min}) + M(\mathcal{L}^u) + C(\mathcal{N}^{max}) \tag{56}$$

Finally, the lower bound implemented is:

$$v = \text{Max}(v_1, v_2) = \text{Max}\{K(\mathcal{N}^{min}) + M(\mathcal{L}^u),\ K_{mst}\} + C(\mathcal{N}^{max}) \tag{57}$$

4. Search scheme

A backtrack programming scheme is used in which, at each iteration, the vertex chosen for separation is the vextex with least lower bound. An upper bound is used to eliminate the descendants of a partitioned vertex, when their lower bounds are greater than the current upper bound. The upper bound is updated whenever a terminal vertex is reached. The algorithm stops when all branches of the search tree have been implicitly examined.

5. Implementation

The branch and bound algorithm has been implemented on the CDC Cyber 74 of the *Université de Montréal* in a program called RESOPT. The program stores the network of included links N^{min} and the maximal network N^{max}, as well as the shortest distances for N^{max}. Shortest paths are computed by the tree labeling procedure of Whiting and Hillier (1960) at the beginning of the algorithm. Whenever a link is excluded from the final potential network, it is deleted from the current maximal network and the shortest distances are updated by means of Murchland's (1967) link deletion algorithm. The minimum construction cost spanning tree is computed by an algorithm due to Prim (1957).

C. Hill-climbing algorithms

1. Introduction

Billheimer (1970) and Billheimer and Gray (1973) proposed a hill-climbing algorithm to solve the optimal network design problem as formulated here. Their algorithm combines link elimination and link insertion. This idea was at the origin of the two algorithms implemented in this work: OPNET1 and OPNET4. The relationship between these algorithms and Billheimer's technique will be described below.

Both algorithms combine a link insertion phase and a link elimination phase: each technique starts with a given network, eliminates one link at a time according to some elimination rule, until no link can be eliminated according to this rule: then it inserts one link at a time according to some insertion rule, until no link can be inserted according to this rule. It then eliminates links again, etc... Each algorithm stops when no improvement of the objective function is possible by elimination of a single link, or insertion of a single link.

2. Link insertion phase

A link is said to be a candidate for insertion if it belongs to the maximal network but is not included in the current network. The consequence for the objective function of adding each candidate link in turn is computed and that link is inserted which brings the greatest decrease to the objective function. The same process starts again with the new network. The link insertion phase stops when no improvement of the objective function is possible by single link insertion. This link insertion method is common to OPNET1 and OPNET4.

3. Link elimination phase

OPNET1 and OPNET4 have different link elimination phases.

Method of OPNET1: The consequence for the objective function of eliminating each link in turn from the current network is computed and that link is eliminated which brings the greatest decrease to the objective function. The same process starts again with the new network. This phase stops when no improvement of the objective function is possible by eliminating single links.

Method of OPNET4: The link elimination method of OPNET1 tests all links which are candidates for elimination. This can be costly in terms of computer time. OPNET4 attempts to avoid the examination of all candidate links by examining only "promising" links. This is done in the following way:

i) Sort the links in ascending order according to the value of the ratio $\frac{x_{ij}}{\ell_{ij}}$, representing the ratio travel cost/capital cost; x_{ij} is the number of trips on the link (i,j) in the current network and ℓ_{ij} is the construction cost of the link (i,j) per unit of travel cost.

ii) Among the first ML links (where ML is a parameter of the program) found by sorting, determine the links whose deletion produces a decrease in the objective function. Store them on a trial list.

iii) If no link appears in the trial list, stop the elimination phase; otherwise, sort the linksby increasing value of the objective function.

iv) Delete the first link of the sorted trial list. Then test the second link of the sorted trial list for deletion. If it leads to a decrease of the objective function delete it and try the next link on the list. If no decrease occurs, try also the next link on the trial list, etc... When the list is finished, go back to (i).

4. Implementation

These two algorithms were implemented in Fortran on the CDC Cyber 173 (and before on the CDC Cyber 74) of the *Université de Montréal*. Their efficiency relies on efficient ways of computing and updating the shortest distances. The shortest path algorithm implemented is due to Whiting and Hillier (1967) and the updating of the shortest distances is done by Murchland's addition and deletion algorithms (see Murchland, 1969).

5. Relationship with Billheimer's algorithm

The improvements provided by OPNET1 and OPNET4 over Billheimer's algorithm are as follows:

i) Billheimer's algorithm has to repeat several times complete traffic assignments, whereas OPNET1 and OPNET4 do a complete traffic assignment only for the initial network.

ii) The evaluation of the elimination of a single link is done differently. OPNET1 and OPNET4 compute an *exact* evaluation (by means of Murchland's link deletion algorithm) whereas Billheimer's approach computes an improvement parameter Δ_{ij} which understates the actual improvement. Some eliminations which could be done are not considered by Billheimer's technique.

iii) OPNET1 does not require the computation of the flows on the links whereas Billheimer's method and OPNET4 require this computation.

6. Comparison of OPNET1 and OPNET4

The parameter ML determines, strictly speaking, distinct heuristic algorithms for distinct values of ML. We will note "OPNET4-ML" this particular algorithm. Following the definitions and theorems of Section II we have OPNET 1 ≥ OPNET4-ML1 ≥ OPNET4-ML2 ≥ ... with ML1 > ML2 > ... This comes from the fact that we have

$$\ldots \subseteq \Pi E(N)^{\text{OPNET4-ML2}} \subseteq \Pi E(N)^{\text{OPNET4-ML1}} \subseteq \Pi(N)^{\text{OPNET1}}$$

for any feasible N.

Thus we must have:

$$H_{\text{OPNET1}} \subseteq H_{\text{OPNET4-ML1}} \subseteq H_{\text{OPNET4-ML2}} \subseteq \ldots$$

IV Numerical experiments

A. Purpose of the experiments

The purpose was two fold:

1) To evaluate the branch and bound algorithm and the hill-climbing algorithms from the point of view of solution quality and efficiency.
2) To implement and evaluate the statistical optimization approach suggested in Section III and apply it to the hill-climbing algorithms developed for solving the optimal network design problem.

B. Evaluation of the algorithms

1. The branch and bound technique

The results of Table 1 show that RESOPT is not a realistic approach for solving problems with more than 40 links: it takes already 207.18 seconds of CDC Cyber 74 to solve a problem of that size. However the knowledge of the exact optimum as computed by RESOPT for some sample problems was useful for the evaluation of the solutions produced by the heuristic techniques.

2. The hill-climbing algorithms

Four options were provided for the choice of the initial network:

Option 1: prespecified network

Option 2: maximal network

Option 3: minimum construction cost spanning tree

Option 4: random network containing all links in the minimum construction cost spanning tree and each additional potential link with a probability of 1/2.

The fourth option is used to increase the probability of reaching the global optimum. The use of option 2 is justified if the weight of the user travel costs is high relative to the link construction costs and the use of option 3 is justified if the reverse is true.

Tables 1 and 2 summarize the results. For the small problems for which the exact optimum is known, it is always obtained by at least one of the two hill-climbing techniques. Both algorithms are very efficient:

OPNET1 (option 3) solves a problem with 30 nodes and 93 links in 8.6 seconds (CDC Cyber 74), while OPNET4 (option 2) solves the same problem in 13.9 seconds (CDC Cyber 74).

As we expect, the quality of the solution obtained by OPNET4 increases with the value of the parameter ML. This is counterbalanced by the fact that the execution time increases with ML. However the running time of OPNET4 increases much less rapidly than the running time of OPNET1 with the size of the problem. For very large size problems ($n > 30$ and $\Lambda > 93$) only OPNET4 can be used efficiently.

C. Numerical experiments with the statistical optimization methodology

1. Procedures used to estimate the global optimum

Three different levels are distinguished.

Level 1 is concerned with fitting a Weibull distribution to the pseudo-distribution of the local optima generated by a given hill-climbing technique. Let $x_1^h, x_2^h, \ldots, x_N^h$ be N local optima obtained by applying a given hill-climbing algorithm to N initial solutions randomly generated according to a uniform distribution, and let $x_{(1)}^{hd} < x_{(2)}^{hd} < \ldots < x_{(R)}^{hd}$ be the $R \leq N$ distinct and ordered local optima. If we estimate the 3 parameters of the Weibull distribution fitted to these R observations we obtain the following point estimate and confidence interval:

$$\hat{x}^* = \hat{a} \tag{58}$$

$$\text{Prob } \{x_{(1)}^{hd} - \hat{b}/S \leq x^* \leq x_{(1)}^{hd}\} \simeq 1 - \alpha \tag{59}$$

Sample problem and size of the problem	Execution time (seconds of CDC CYBER 74 except when otherwise mentioned)			Value of the solution produced by the algorithm		
	OPNET1	OPNET4	RESOPT	OPNET1	OPNET4	RESOPT
Billheimer's sample problem n=8, Λ=23	.28	.13	5.44	66628*	66820	66628*
test problem n=8, Λ=23	.11	.08	.91	6613770*	6613770*	6613770*
Boyce and Farmi's sample problem n=10, Λ=45	.34 sec. on an IBM 370/168	.22 sec. on an IBM 370/168	1976.73	47170	47170	did not run to completion, best current upper bound = 47170
test problem n=12, Λ=36	.51	.27	45.06	25513100*	25513100*	25513100*
test problem n=12, Λ=40	.74	.61	207.18	25513100*	25513100*	25513100*
test problem n=12, Λ=59	1.47	1.70	477.67	23405100	23405100	did not run to completion, best current upper bound = 23405100
random problem n=30, Λ=93	13.3	17.9	—	258231686	258231686	—

Table 1 : Summary of results[1 and 2]

1 For each algorithm, and each sample problem only one run giving the best solution was chosen for the table, when several initial solutions, or several options had been used.

2 The symbol * means that an optimal solution is obtained.

Problem size: n-number of nodes	Problem size: A-number of links	Value of the initial solution	Value of the local optimum produced by OPNET1	Value of the local optimum produced by OPNET4 (ML=n)	Execution time for OPNET1 (seconds of CDC Cyber 74)	Execution time for OPNET4 (ML=n) (seconds of CDC Cyber 74)	Option
20	64	139857527	129400623	129632416	7.0	7.6	
25	77	194126335	185159977	186198694	14.2	8.3	2
30	93	268042077	258231686	260640032	27.8	13.9	
30	93	369605097	258231686	258231686	8.6	18.6	3
30	93	279724395	258231686	258231686	16.2	23.0	
30	93	281058242	258231686	258231686	13.3	17.9	4
30	93	283382481	258275133		15.2		

Table 2: Results produced by OPNET1 and OPNET4 for problems of different sizes generated at random

with $$S = \left(- \frac{R}{\ell n \alpha} \right)^{1/\hat{c}} \qquad (60)$$

<u>Level 2</u> is distinct from level 1 only insofar as several heuristic algorithms are used to generate the R local optima. This estimation procedure was not followed in the numerical experiments reported in this paper although they were used in experiments reported elsewhere (See Lardinois, 1980b).

<u>Level 3</u> is concerned with the estimates obtained by fitting a Weibull distribution to the distribution of the minima of samples of local optima.

Let $$G_i = \{x_1^{h,i}, x_2^{h,i}, \ldots, x_m^{h,i}\}, \qquad (i=1,2,\ldots,E) \qquad (61)$$

be E samples of m local optima obtained by the combination of a uniform random generation mechanism and of a hill-climbing algorithm. Let

$$x_{(1)}^{h,i} = \text{Min}\ \{x_j^{h,i},\ \ j=1,\ldots,m\}\ (i=1,2,\ldots,E) \qquad (62)$$

be the E minima of the E samples. Let

$$x_{(1)}^{h,(1)} \leq x_{(1)}^{h,(2)} \leq \ldots \leq x_{(1)}^{h,(E)} \qquad (63)$$

be the E *ordered* minima. By estimating the three parameters of the Weibull distribution fitted to these E minima we obtain the following point estimate and confidence interval:

$$\hat{x}^* = \hat{a} \qquad (64)$$

$$\text{Prob}\ \{x_{(1)}^{h,(1)} - \frac{\hat{b}}{S} \leq x^* \leq x_{(1)}^{h,(1)}\} \simeq 1 - \alpha \qquad (65)$$

with $$S = \left(- \frac{E}{\ell n \alpha} \right)^{1/\hat{c}} \qquad (66)$$

All the confidence intervals in the experiments were computed with $1-\alpha = 0.95$. The estimation procedure of Harter and Moore was used.

The test of goodness-of-fit of Kolmogorov-Smirnov was used to test the goodness of fit of the empirical distribution to the theoretical Weibull distribution. The test statistic is noted D. All the tests were done with a level of significance of .05 and the critical values are noted $D_{.05}$. In addition, the goodness of fit was visualized by drawing on the same graphs, the empirical and the theoretical Weibull distributions. On the graphs the value of the location parameter is noted A and the value of the test statistic is noted DN. The goodness-of-fit information is supplemented by two performance measures for the confidence intervals. The first performance measure is defined by:

$$PERF_a(\%) = [(\hat{b}/S)100]/x^* \quad \text{if} \quad x^* \text{ is known} \qquad (67)$$

$$= [(\hat{b}/S)100]/x^h \quad \text{otherwise} \qquad (68)$$

where $$x^h = x^{hd}_{(1)} \text{ at levels 1 and 2} \qquad (69)$$

$$= x^{h,(1)}_{(1)} \text{ at level 3} \qquad (70)$$

The second performance measure is:

$$PERF_b(\%) = [(\hat{b}/S)100]/\Delta \qquad (71)$$

where $$\Delta = x^{hd}_{(R)} - x^{hd}_{(1)} \quad \text{at levels 1 and 2} \qquad (72)$$

$$= x^{h,(E)}_{(1)} - x^{h,(1)}_{(1)} \quad \text{at level 3} \qquad (73)$$

2. Generation of network problems and of initial networks for the hill-climbing algorithms

Data were generated at random as follows: The n^2 cells of the matrix of interactions Q are integers, uniformly distributed between 0 and 10,000. The matrix of direct distances $D = [d_{k\ell}]$ is obtained by generating at random n points in a square of side 100, and computing the $\frac{n(n-1)}{2}$ euclidean distances between the pairs of points. The construction costs are obtained by multiplying the direct distances by 10,000. All n(n-1)/2 possible links are included in the maximal networks which are examined. Two problems were thus generated, one with n=12 and Λ=66, the other with n=15 and Λ=105.

The inital networks for the hill-climbing algorithms are generated by giving a probability 1/2 of being in the network to each potential link. We retain only the connected networks. Each feasible network has an equal probability of being generated as an initial solution. In practice, none of the networks generated was disconnected, for the two problems: this difficulty occurs only for very small problems.

3. Problem with 12 nodes and 66 potential links

The estimation of the global optimum was done at level 1, using three heuristics: OPNET1, OPNET4-1, OPNET4-6: 200 local optima were generated by OPNET1, 150 local optima by both other techniques. Tables 3.A and 3.B sum up the results.

The Weibull hypothesis is not rejected by the Kolmogorov-Smirnov test for the three algorithms. OPNET1 however generated only 16 distinct local optima out of 200. This heuristic appears too powerful for the size of the problem: too few implicit samples exist for the Fisher-Tippett analogy to be credible. These doubts are confirmed by graph #1. For the other two heuristics however, the number of *distinct* local optima is equal to N in one case and almost equal in the other case. This fact makes the Fisher-Tippett analogy more acceptable because of the large number of implicit samples. The graphic curve fittings (graphs #2 and 3) corroborate this fact.

The three density functions are drawn on graph #4 and show, as expected, a clear relationship between the shape of the distributions and the power of the three heuristics. The "weaker" the heuristic, the more spread out on the right the density function is. The performances of the three confidence intervals (Figure 2) reflect these shapes. The best confidence interval corresponds to the heuristic with intermediate power OPNET4-6, whose associated distribution is quasi-exponential ($\hat{c} \simeq 1$). The worst confidence interval corresponds to the weakest heuristic OPNET4-1.

These results suggest that the methodology is best suited to heuristics of *intermediate power*: a heuristic which is too powerful does not produce enough distinct points to fit a Weibull distribution and does not create enough implicit samples to justify the approximation of a discrete distribution by a continuous one and thereby the Fisher-Tippett analogy. On the other hand a weak heuristic produces a Weibull distribution which is too spread out on the right, leading to a trivial confidence interval. Besides, with very weak heuristics the implicit samples contain too few elements for m to be considered sufficiently large for the Fisher-Tippett analogy. A heuristic with intermediate power avoid these two pitfalls.

HILL-CLIMBING METHOD	HARTER-MOORE POINT ESTIMATES $\hat{a}$	$\hat{b}$	$\hat{c}$	GOODNESS OF FIT D	$D_{0.05}$	GRAPHIC NO.
OPNET1(N = 200 ; R = 16)	51758662.36	1031277.66	2.7250	0.1541	0.3273	1
OPNET4 - ML = 1(N = 150 ; R = 150)	50237471.88	15653271.88	6.6089	0.0718	0.1109	2
OPNET4 - ML = 6(N = 150 ; R = 144)	52177155.26	3043801.76	1.0329	0.0934	0.1132	3

TABLE 3.A. Problem of size 12.
Point estimates

HILL-CLIMBING METHOD	$x^{hd}_{(R)}$	95% CONFIDENCE INTERVAL	$PERF_a$ (%)	$PERF_b$ (%)
OPNET1(N = 200 ; R = 16)	53369429	51615637 ; 52173087	1.07	46.59
OPNET4 - ML = 1(N = 150 ; R = 150)	71459251	46285076 ; 54981338	15.82	52.78
OPNET4 - ML = 6(N = 150 ; R = 144)	63335545	52126005 ; 52197691	0.14	0.64

TABLE 3.B. Problem of size 12
Confidence intervals

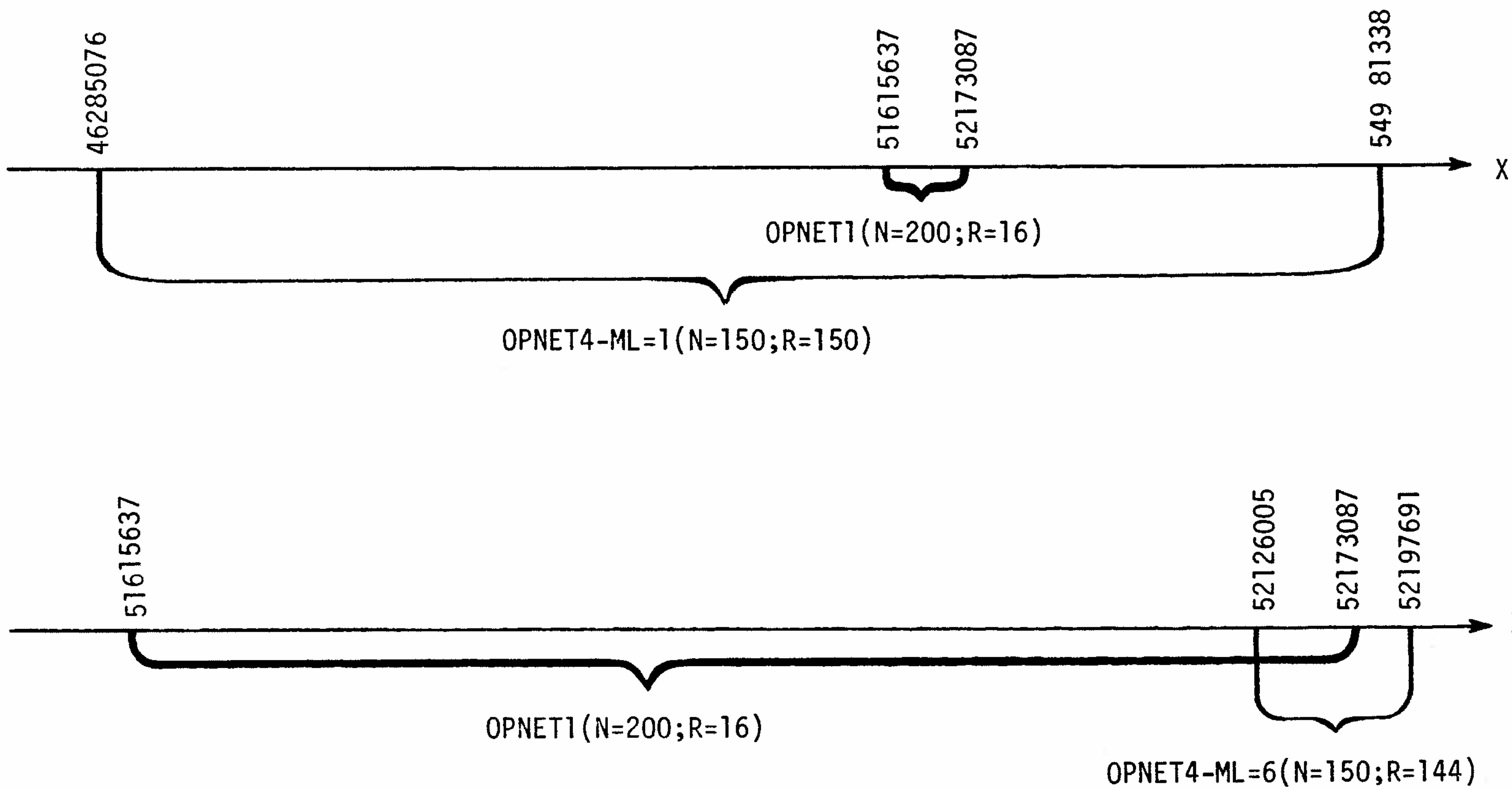
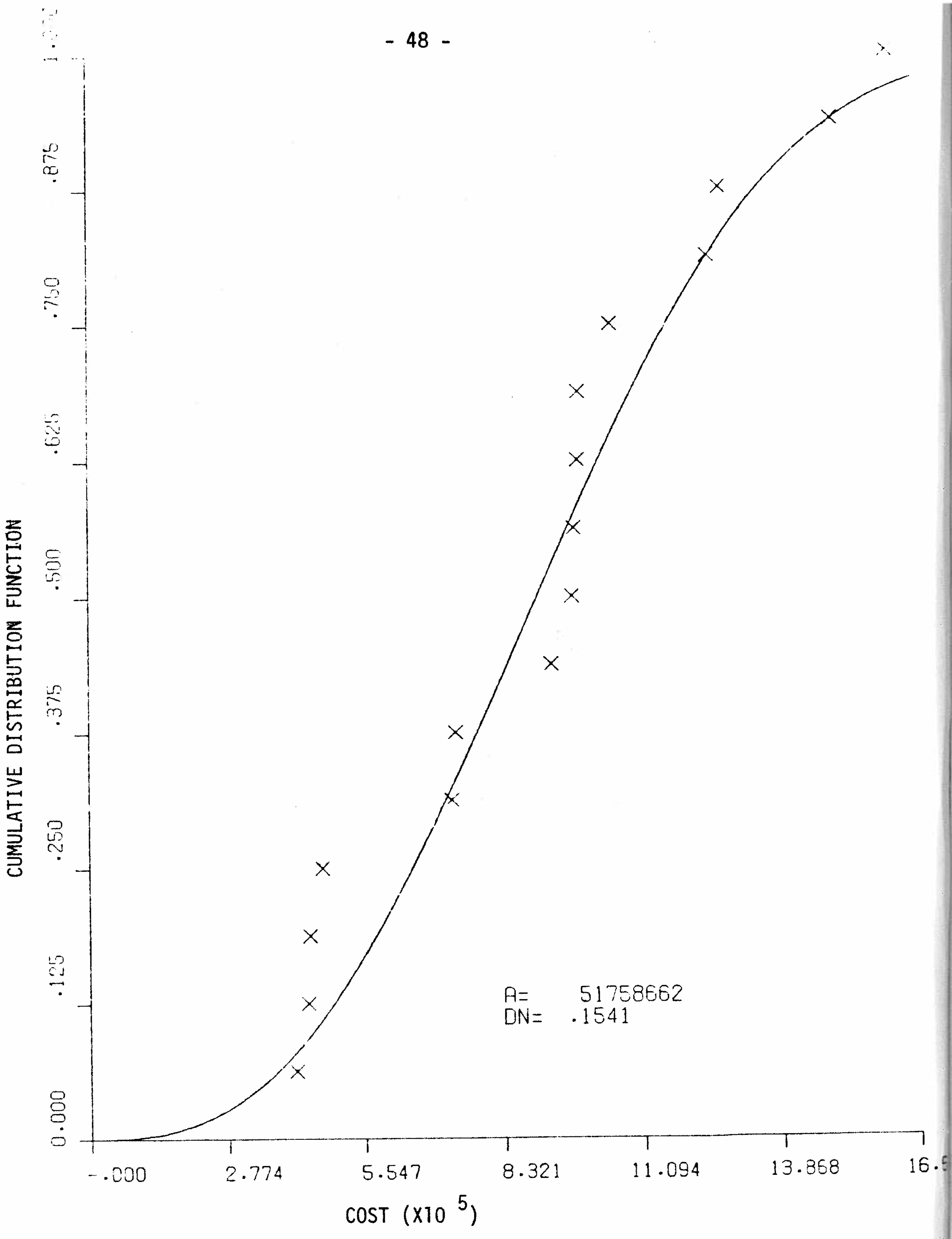


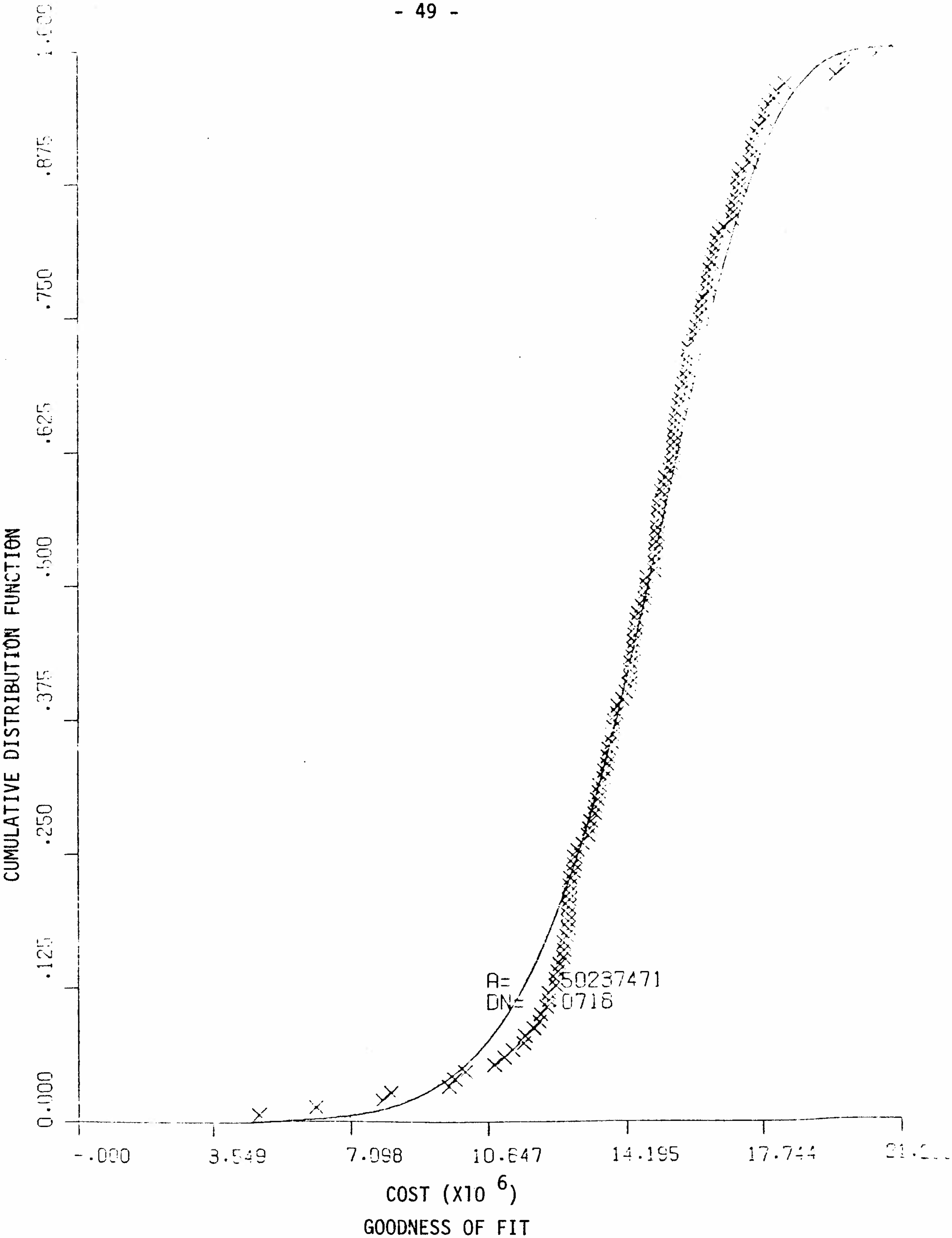
Figure 2. Comparison of the confidence intervals produced by OPNET1, OPNET4 - 6 and OPNET4 - 1.



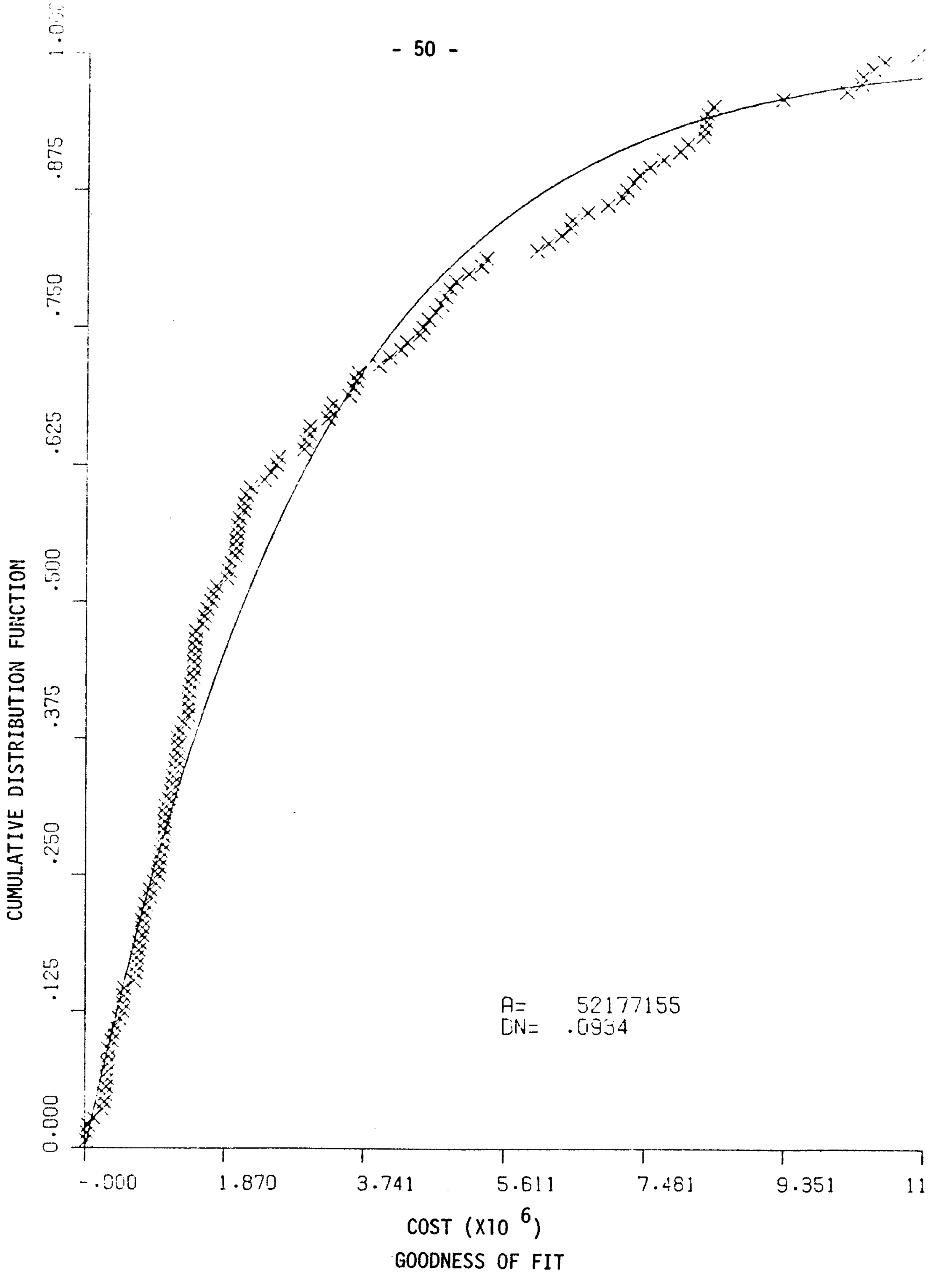
GOODNESS OF FIT

GRAPH 1. Problem of size 12

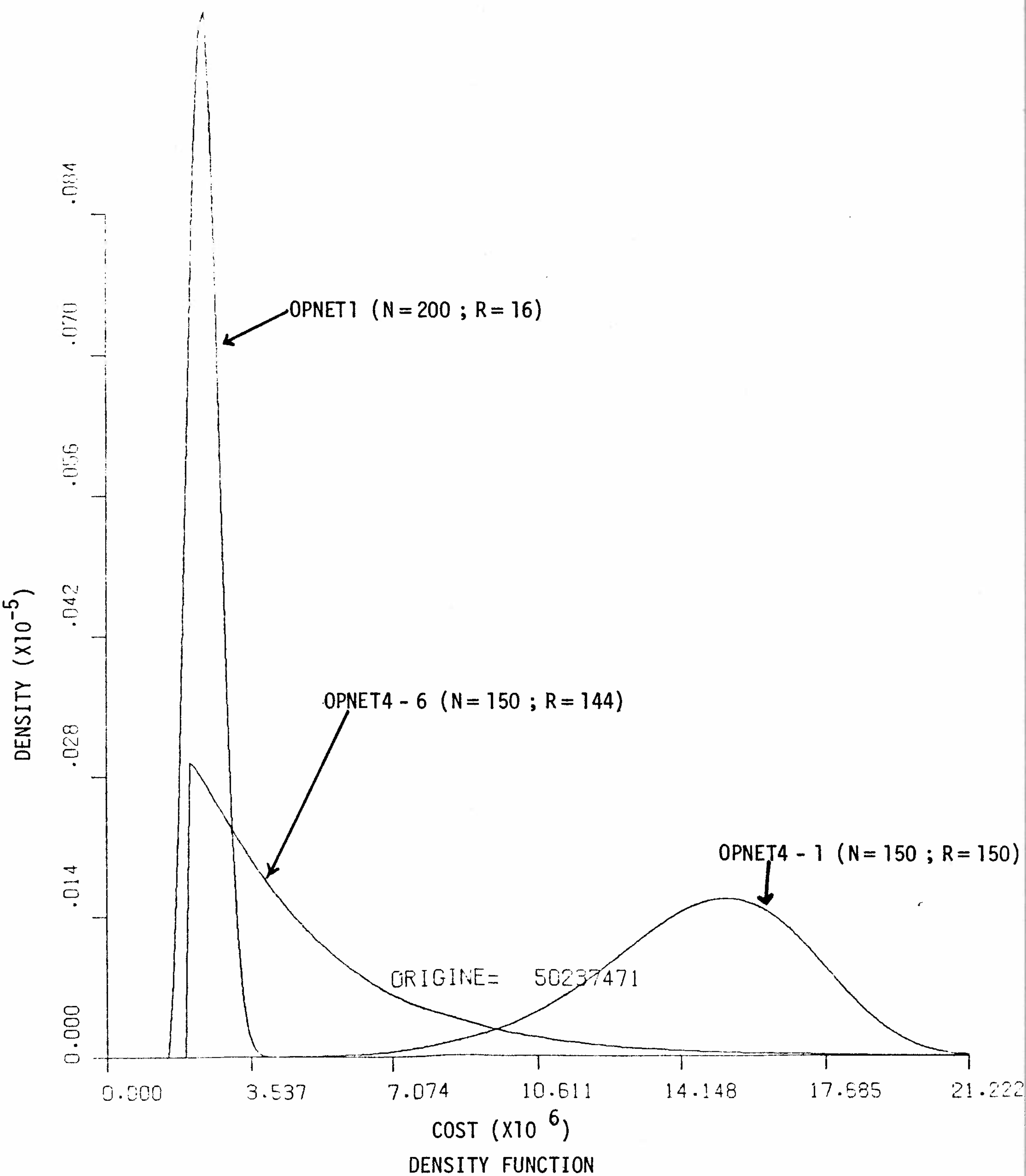
OPNET1(N=200;R=16)



GRAPH 2. Problem of size 12
OPNET4-ML=1(N=150;R=150)



GRAPH 3. Problem of size 12.
OPNET4-ML=6(N=150;R=144)



GRAPH 4. Problem of size 12.
OPNET1(N=200;R=16), OPNET4-ML=1(N=150;R=150)
OPNET4-ML=6(N=150;R=144)

4. Problem with 15 nodes and 105 potential links

Four experiments were done at level 1 with OPNET1 and OPNET4-8. They are characterized as follows:

i) OPNET1 : N=159, R=54 , T=3,150 sec.

ii) OPNET4-8: N=159, R=152, T=2,014 sec.

iii) OPNET4-8: N=56 , R=54 , T=709 sec

iv) OPNET4-8: N=250, R=241, T=3,150 sec.

T represents the execution time in seconds of CDC Cyber 173. Experiments (ii), (iii) and (iv) have one aspect in common with experiment (i): experiment (ii) generated the same number of local optima; experiment (iii) generated the same number of *distinct* local optima; experiment (iv) was made so as to have the same total execution time as the first experiment. Tables 4.A and 4.B summarize the results of the experiments. The test of goodness of fit dit not allow the rejection of the hypothesis of a Weibull distribution in all four experiments of level 1. The graphs #5, 6, 7 and 8 corroborate visually the validity of the Weibull hypothesis. The graphs 9, 10, 11 and 12 compare the density functions of experiments (ii), (iii) and (iv) with that of experiment (i). The most powerful heuristic, i.e. OPNET1, generates the distribution with the least spread: its scale parameter $\hat{b}$ is the smallest of the four experiments. On the other hand the shape parameters $\hat{c}$ are close to 1 for experiments (ii), (iii) and (iv) leading to very performing confidence intervals. Figure 3 compares visually the four confidence intervals. The largest confidence interval corresponds to experiment (i), while the smallest is associated with experiment (iv). For the same computer time of 3,150 seconds, OPNET4-8 leads to a confidence interval which is 73.40% smaller than that produced with OPNET1.

These experiments at level 1 with the network problem of size n=15 and Λ=105 confirm the experiments with the previous network problem (n=12, Λ=66) that a heuristic of *intermediate* power (OPNET4-8 here and OPNET4-6 previously) is better than too powerful a heuristic (OPNET1): it costs less to generate local optima;we thus may obtain more sample points for the estimation procedure. In addition, at least for the numerical experiments reported here, the shape parameter $\hat{c}$ may be such that a quasi-exponential distribution fits the sample points, leading to very performing confidence intervals.

With the same local optima generated during experiment (iv) we can apply the estimation procedure at level 3 with E=25 and m=10. The results are summed up in the last rows of tables 4.A and 4.B. Graph 11 shows the visual fitting of a Weibull distribution to the 25 sample points. The Kolmogorov-Smirnov test was successful. Graph 12 compares the three density functions estimated on the basis of experiments with the same computing cost (3,150 seconds). As we would expect the spread is the lowest for the level 3 experiment. The confidence intervals are compared on figure 4. The two confidence intervals with OPNET4-8 are better than the one obtained with OPNET1. However the estimation procedure at level 3 leads to a worse confidence interval than the estimation at level 1 for experiment (iv): the increase in the shape parameter $\hat{c}$ counterbalances the effect of the decrease in the scale parameter $\hat{b}$, which, by itself, would tend to decrease the confidence interval. It must be emphasized that the evolution of the shape parameter from one experiment to another cannot be predicted.

HILL-CLIMBING METHOD	HARTER-MOORE POINT ESTIMATES $\hat{a}$	$\hat{b}$	$\hat{c}$	GOODNESS OF FIT D	$D_{0.05}$	GRAPH No.	
OPNET1(N = 159 ; R = 54)	67674699.64	851196.81	1.7447	0.0700	0.1814	5	LEVEL 1
OPNET4 - ML = 8(N = 159 ; R = 152)	67710967.34	2612309.87	1.0141	0.0516	0.1102	6	LEVEL 1
OPNET4 - ML = 8(N = 56 ; R = 54)	67711219.00	2404503.33	1.0308	0.0499	0.1814	7	LEVEL 1
OPNET4 - ML = 8(N = 250 ; R = 241)	67710368.88	2662809.98	1.0645	0.0410	0.0875	8	LEVEL 1
OPNET4 - ML = 8(E = 25 ; m = 10)	67711219.00	280951.85	1.4599	0.1863	0.2640	11	LEVEL 3

TABLEAU 4.A. Problem of size 15
Point estimates

HILL-CLIMBING METHOD	GREATEST VALUE[1]	95% CONFIDENCE INTERVAL	$PERF_a$ (%)	$PERF_b$ (%)	
OPNET1(N = 159 ; R = 54)	69535541	67548870 ; 67711219	0.24	8.89	LEVEL 1
OPNET4 - ML = 8(N = 159 ; R = 152)	85176449	67656825 ; 67711219	0.08	0.31	LEVEL 1
OPNET4 - ML = 8(N = 56 ; R = 54)	78383006	67565756 ; 67711219	0.22	1.36	LEVEL 1
OPNET4 - ML = 8(N = 250 ; R = 241)	85176449	67668038 ; 67711219	0.06	0.24	LEVEL 1
OPNET4 - ML = 8(E = 25 ; m = 10)	68475297	67645536 ; 67711219	0.10	8.60	LEVEL 3

TABLE 4.B. Problem of size 15. Confidence intervals

[1] Greatest value = $x^{hd}_{(R)}$ at level 1; $x^{h,(E)}_{(1)}$ at level 3.

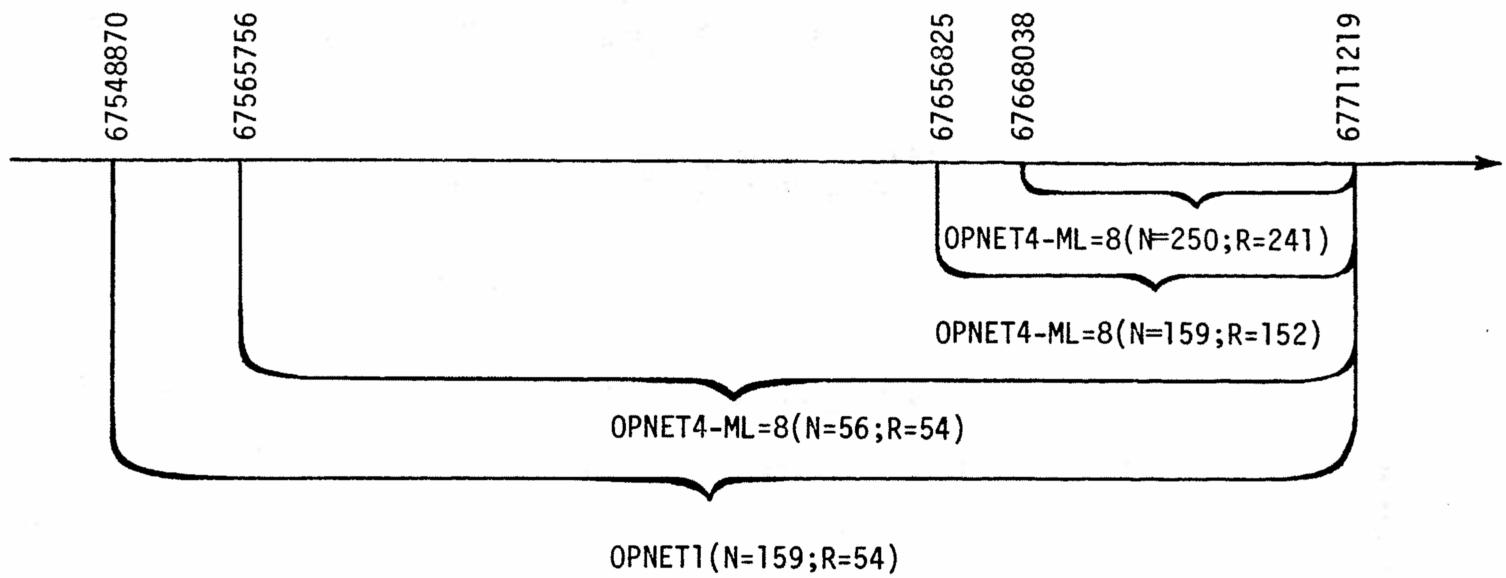
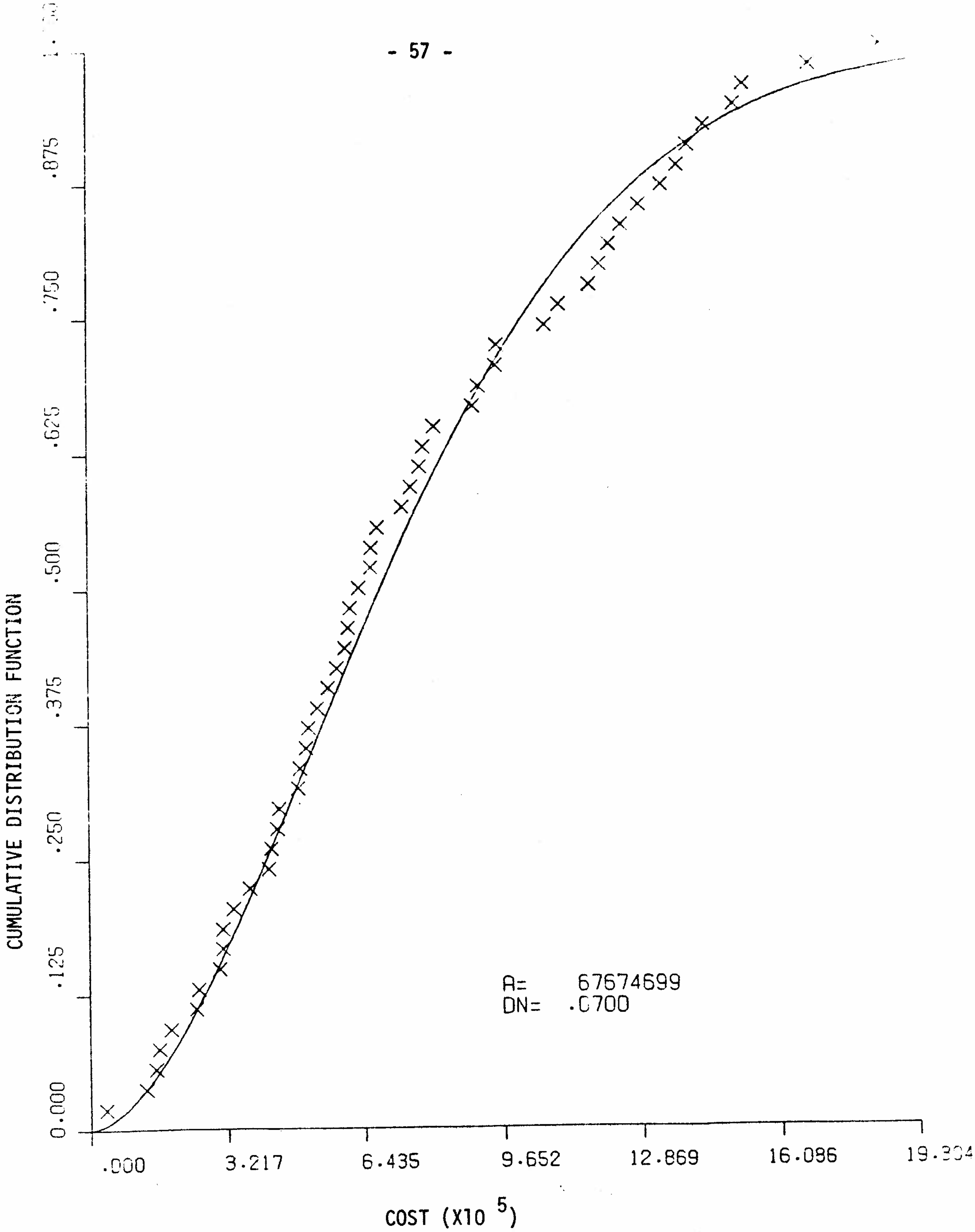
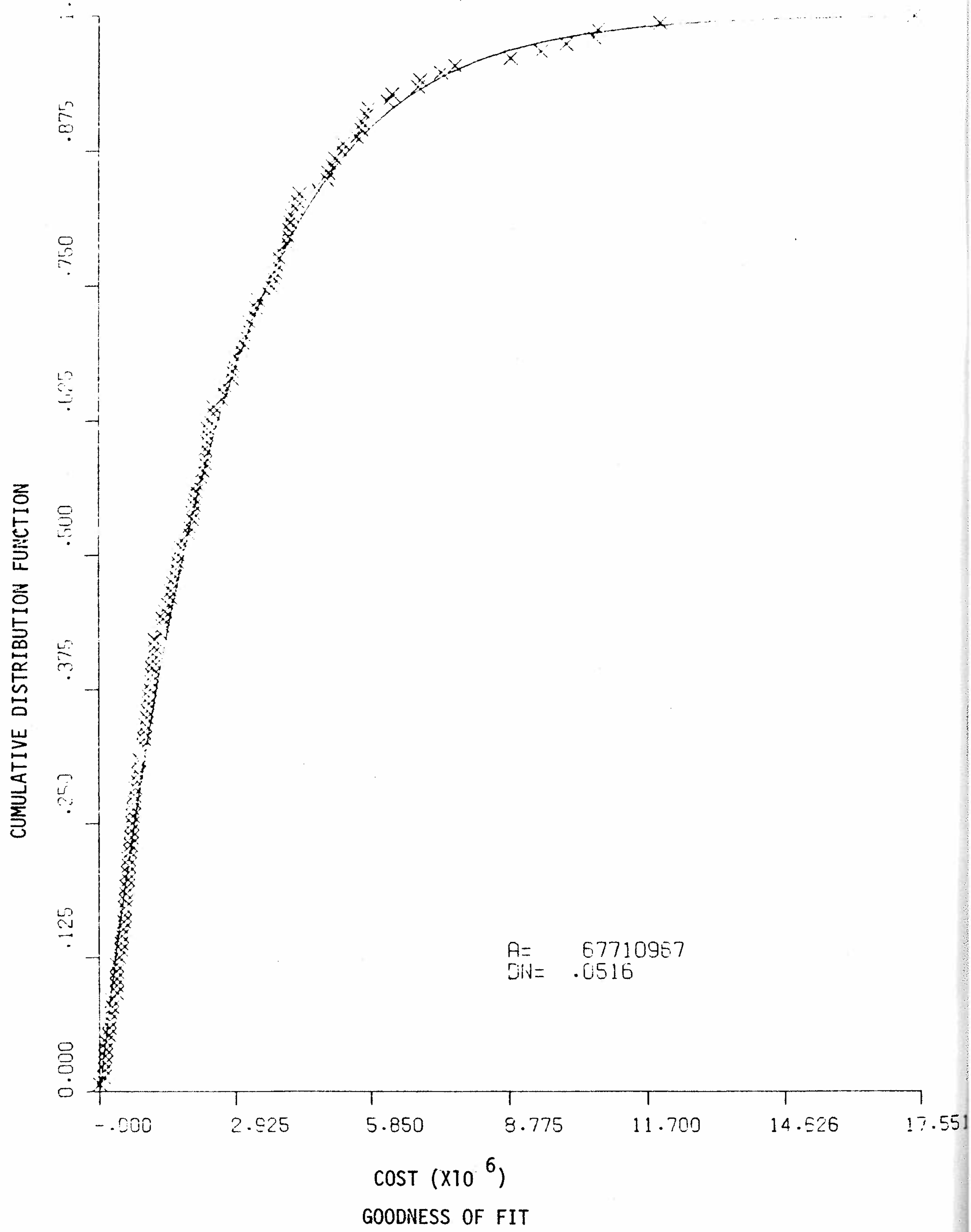


Figure 3. Comparison of confidence intervals produced by OPNET1 and OPNET4-ML=8

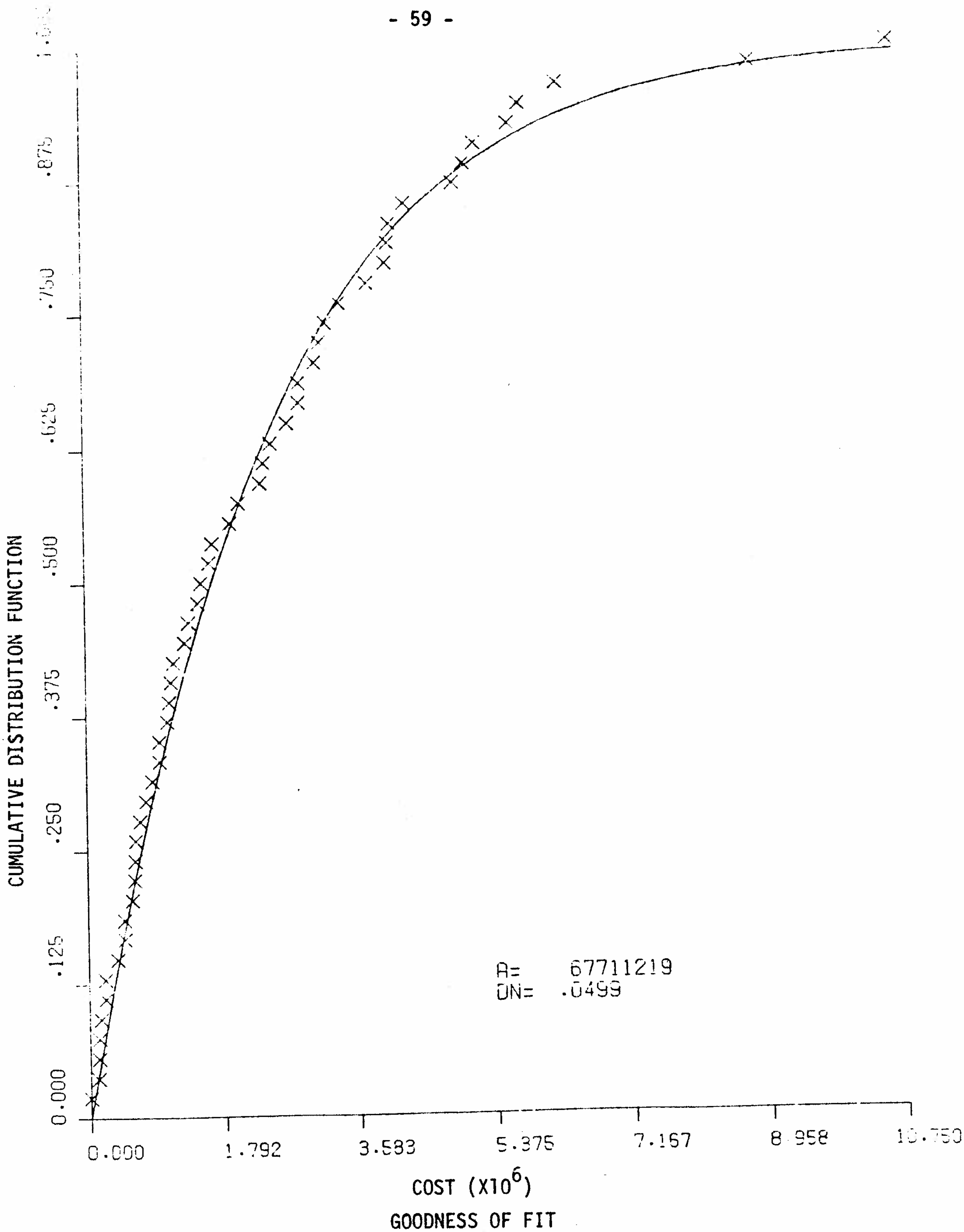


GOODNESS OF FIT

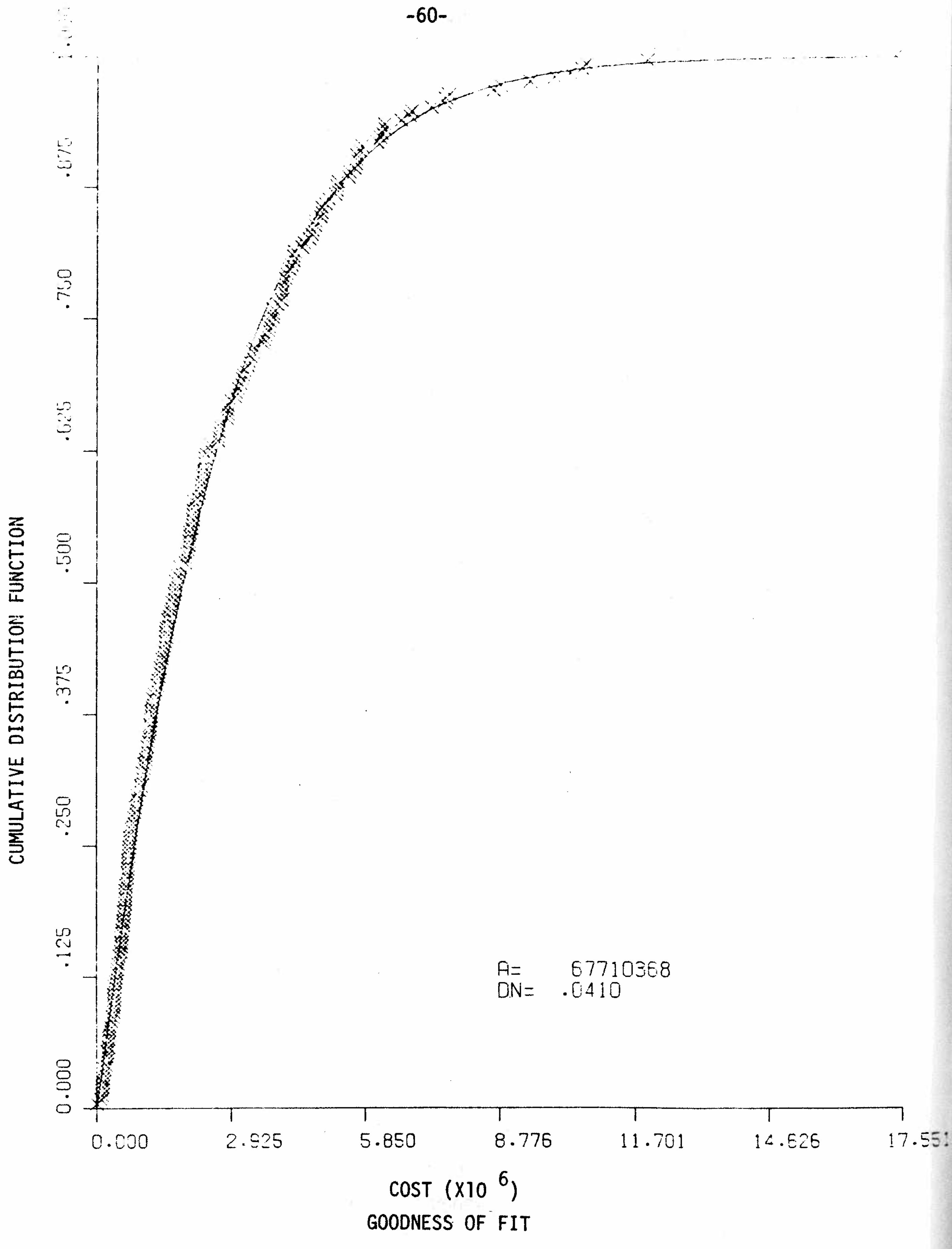
GRAPH 5. Problem of size 15
OPNET1 (N=159;R=54)



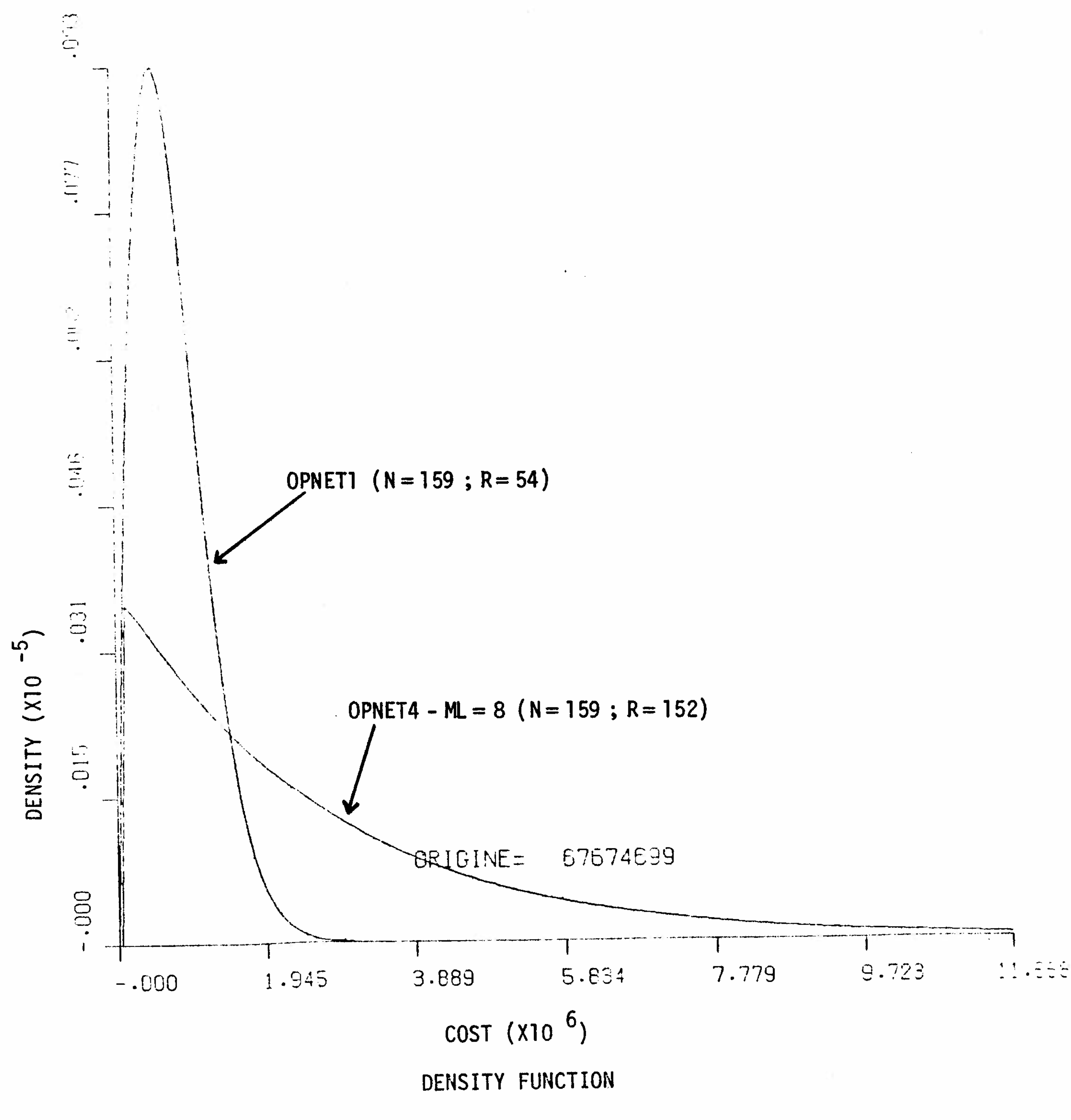
GRAPH 6. Problem of size 15.
OPNET4-ML=8 (N=159;R=152)



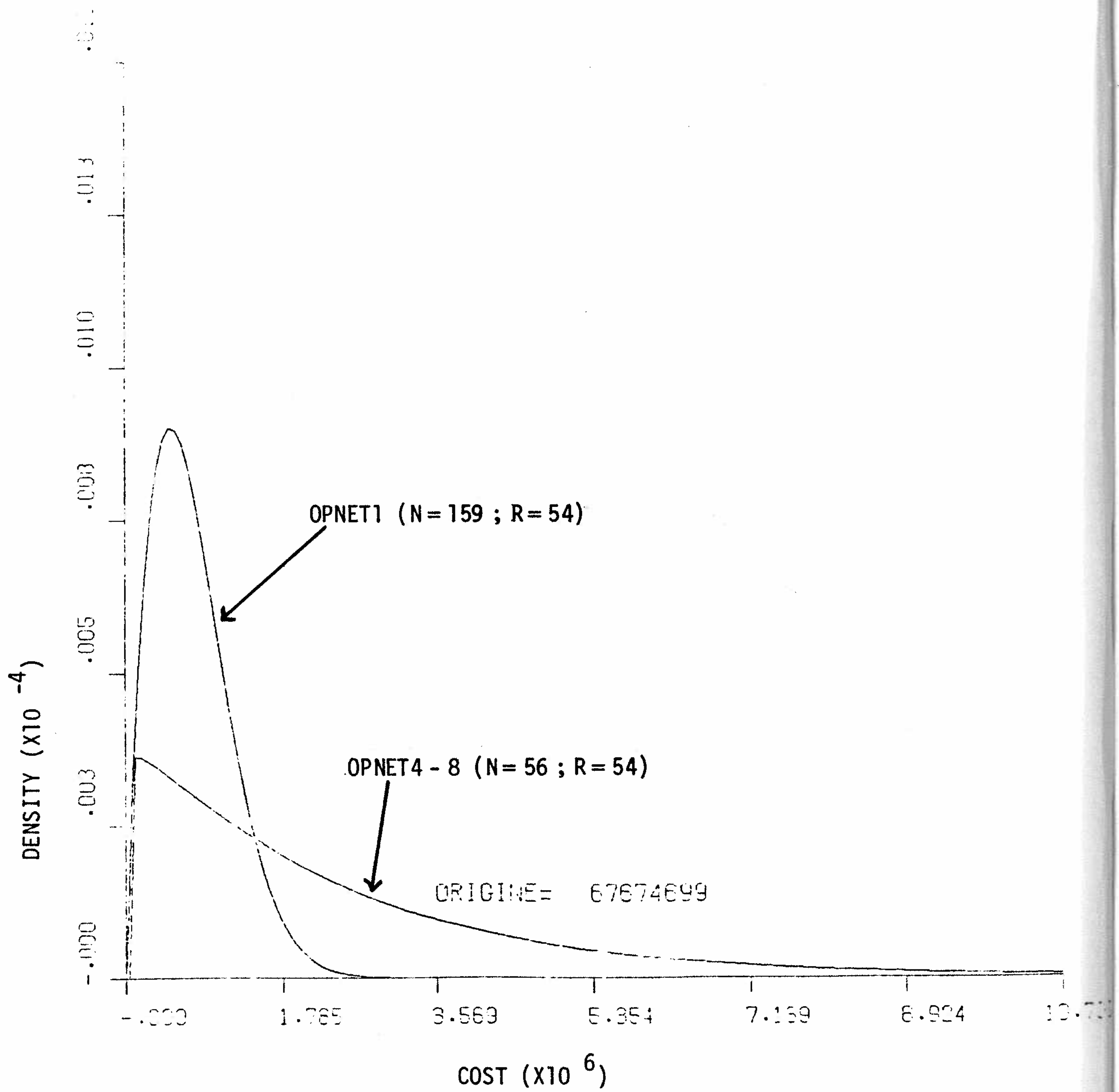
GRAPH 7. Problem of size 15
OPNET4-ML=8 (N=56;R=54)



GRAPH 8. Problem of size 15.
OPNET4-ML=8 (N=250;R=241)

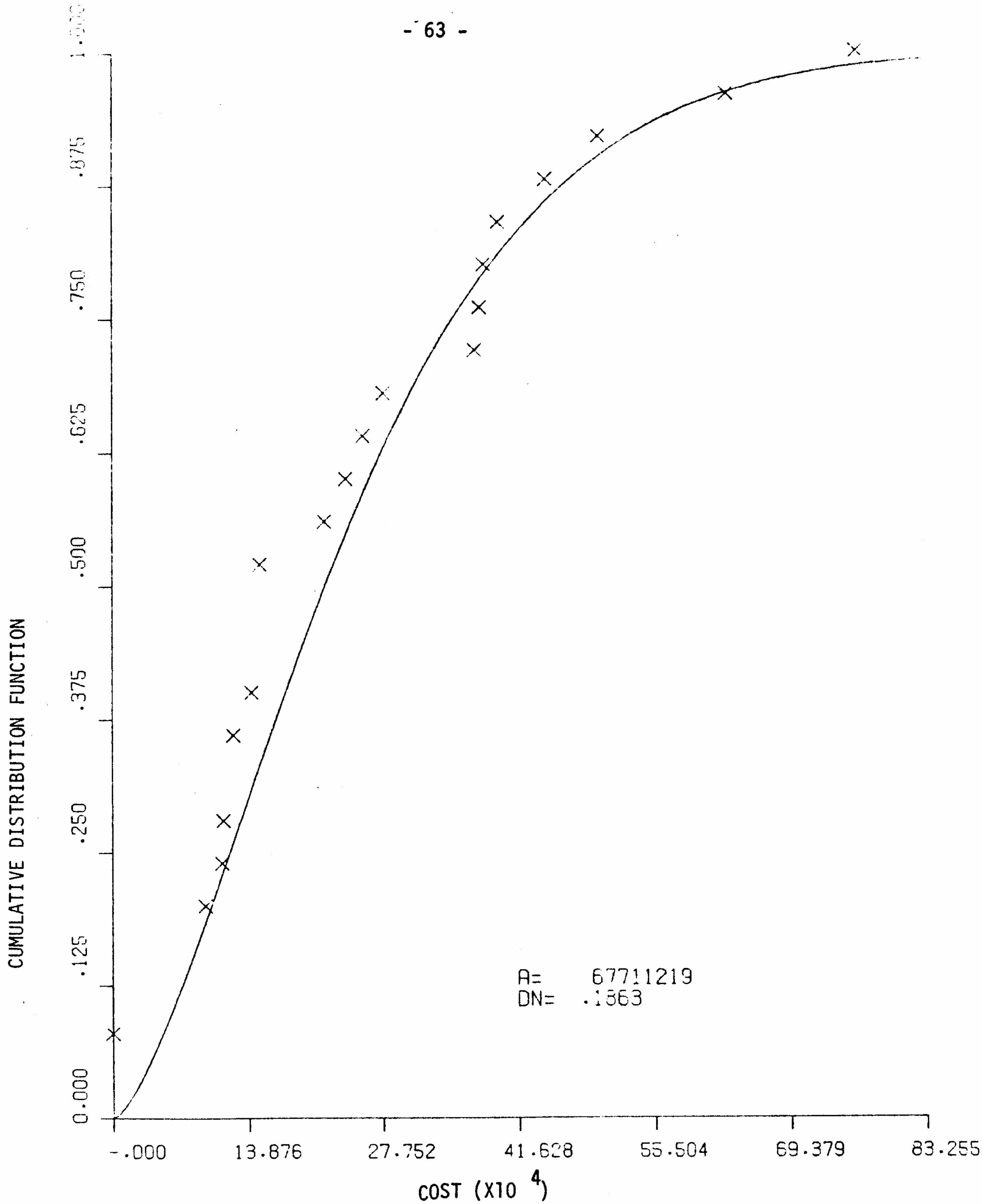


GRAPH 9. Problem of size 15.
OPNET4-ML=8 (N=159;R=152), OPNET1 (N=159;R=54)



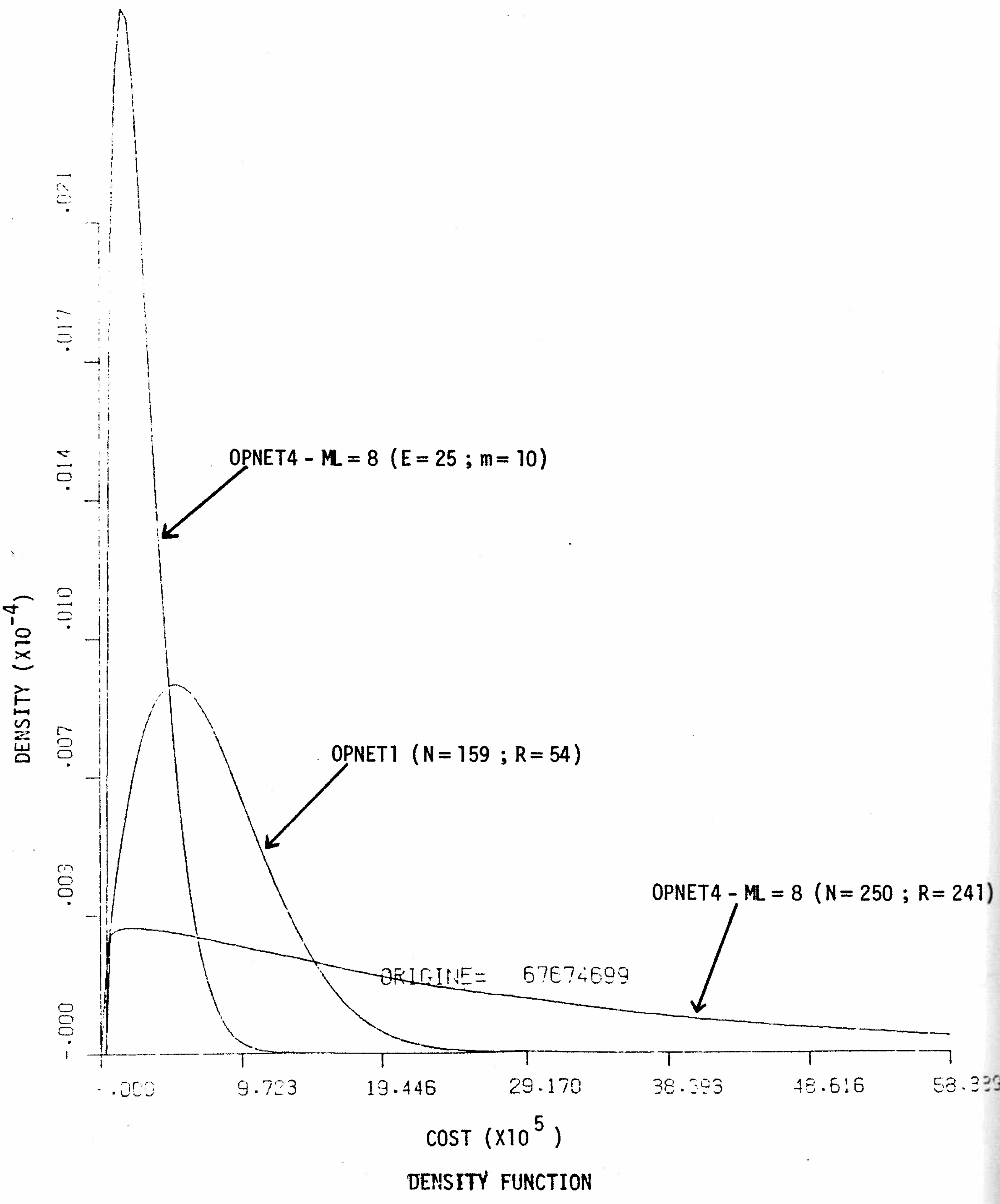
DENSITY FUNCTION

GRAPH 10. Problem of size 15
OPNET4-ML=8 (N=56;R=54), OPNET1 (N=159;R=54)



GOODNESS OF FIT.

GRAPH 11. Problem of size 15.
OPNET4-ML=8 (E=25;m=10)



Graph 12. Problem of size 15.
OPNET1 (N=159;R=54), OPNET4-ML=8(N=250;R=241)
OPNET4-ML=8(E=25;m=10)

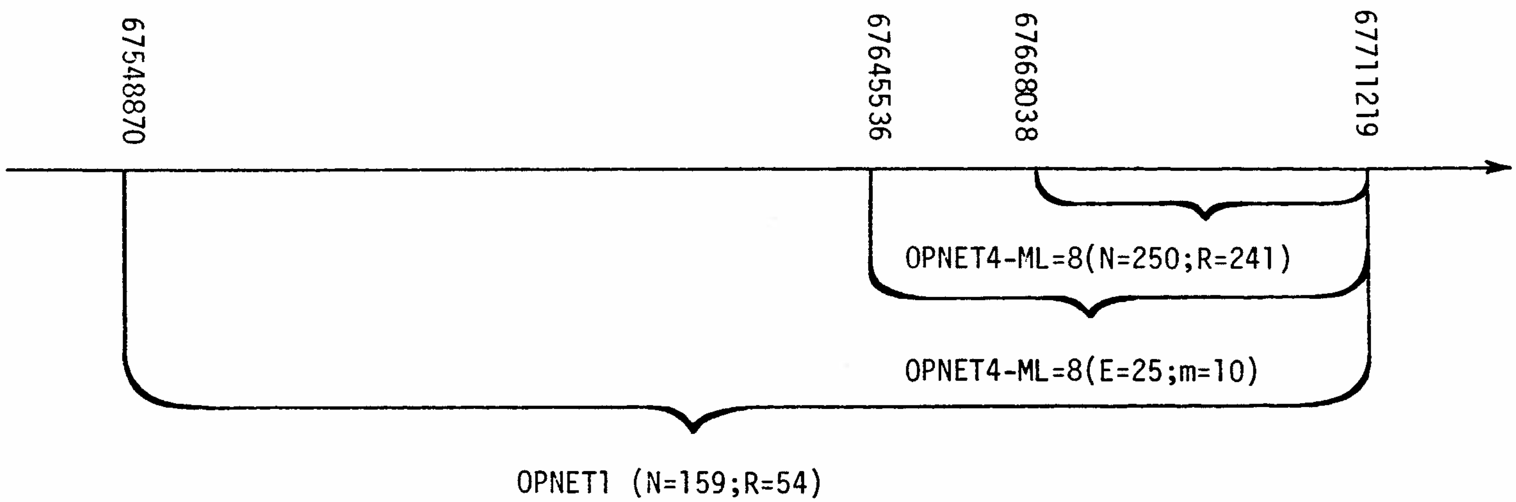


Figure 4. Comparison of confidence intervals produced by OPNET1 and OPNET4-8 at levels 1 and 3.

5. Utilization of the stopping rule

The stopping rule suggested in section II is here applied to the network problem with 15 nodes and 105 potential links, using OPNET1. The parameters, C_E, C_S, C_0, α_i and β_i are used for the stopping rule. The maximal cost of using the Harter-Moore procedure was \$40. We set: C_E=\$40. We set: C_S = \$3.35 [1] for the cost of using OPNET1 with a problem of that size.

First iteration (We start with N_1=50 local optima.)

$R_1=25$; $x^{1,h}=67,711,219$

$\hat{a}_1 = 676\ 902\ 08$

$\hat{b}_1 = 695\ 144$

$\hat{c}_1 = 1.414$

The probability of improvement on one extra observation is: $\hat{\Phi}_1(x^{1,h})=.0071$.
The potential for improvement is: $PA(x^{1,h})=677\ 112\ 19-676\ 902\ 08=21{,}011$
Table 5 indicates the number N_2 of additional observations necessary to guarantee a probability of improvement of α_1=.10, .20, etc., as well as the associated expected values of the gain.

Let us choose α_1 = 30%

Then : $N_2 = 51$

$E_{N_2}(\text{gain}) = 2817.$

First rule. With β_1 = 10%:

$E_{N_2}(\text{gain}) = 2817 > \beta_1 . PA(x^{1,h}) = .10 \times 21{,}011 = 2101.1$

---

[1] These two figures are based on a charge of \$0.17/second of CDC Cyber 173.

α_1	N_2	E_{N_2}(Gain)	E_{N_2}(Gain) / [PA($x^{1,h}$)]
0.10	15	895.16	4.26%
0.20	32	1840.43	8.76
0.30	51	2817.10	13.41
0.40	73	3852.58	18.34
0.50	98	4917.99	23.41
0.60	130	6130.15	29.18
0.70	170	7440.99	35.41
0.80	228	9013.77	42.90
0.90	325	11000.02	52.35
0.95	423	12443.19	59.22

Table 5 - Problem of size 15
Situation at the end of iteration 1.

The first rule does not tell us to stop.

<u>Second rule</u>. Let us assume that 5% of the potential for improvement is equivalent to \$100. Thus

$$C_0 = \frac{100}{.05 \times 21{,}011} = .095$$

$$E_{N_2}(\text{gain}) \times C_0 = 2817 \times 95 \times 10^{-3} = 267.615 > C_s N_2 + C_E = 210.85.$$

The second rule tells us again to continue the search. We therefore generate 51 additional local optima and start the second iteration with $N_2 = 50 + 51 = 101$ local optima.

Iteration 2

$R_2 = 44;\ x^{2,h} = x^{1,h} = 67711219$

$\hat{a}_2 = 67664211$

$\hat{b}_2 = 846474$

$\hat{c}_2 = 1.7575$

The probability of improvement on one additional observation is:

$$\hat{\Phi}_2(x^{2,h}) = 0.0062$$

The potential for improvement is:

$$PA(x^{2,h}) = 67711219 - 67664211 = 47{,}008$$

α_2	N_3	$E_{N_3}(\text{Gain})$	$E_{N_3}(\text{Gain})/[PA(x^{2,h})]$
0.10	17	1744.87	3.71%
0.20	36	3567.77	7.59
0.30	58	5524.42	11.75
0.40	83	7565.63	16.09
0.50	112	9715.80	20.67
0.60	148	12098.94	25.74
0.70	194	14749.48	31.38
0.80	259	17877.62	38.03
0.90	371	22001.75	46.80
0.95	482	24990.64	53.16

Table 6 - Problem of size 15
Situation at the end of iteration 2.

The number of additional local optima N_3 necessary to guarantee a probability of improvement $\alpha_2 = .10$, $.20$ etc., and the corresponding expected gain are given in Table 6. We examine the two stopping rules with $\alpha_2 = 30\%$, $N_3 = 58$, $E_{N_3}(\text{gain}) = 5524$.

First rule. With $\beta_2 = 10\%$:

$$E_{N_3}(\text{gain}) = 5524 > \beta_2 . PA(x^{2,h}) = .10 \times 47{,}008 = 4700.8$$

We are led to continue the search.

Second rule: We assume that 5% of the potential for improvement is equivalent to \$100.00 i.e.

$$C_0 = \frac{100}{.05 \times 47{,}008} = 42.5 \times 10^{-3}$$

The second rule gives:

$$E_{N_3}(\text{Gain}).C_0 = 5524 \times 42.5 \times 10^{-3} = 234.77 > C_s N_3 + C_E = 3.35 \times 58 + 40 = 234.30$$

We are led to continue the search by this rule. Therefore we generate $N_3 = 58$ additional local optima and we set $N_3 = 101 + 58 = 159$.

Iteration 3:

$R_3 = 54$; $x^{3,h} = x^{2,h}$

$\hat{a}_3 = 67674699$

$\hat{b}_3 = 851196$

$\hat{c}_3 = 1.7447$

The probability of improvement on one additional observation is:

$$\hat{\Phi}_3(x^{3,h}) = .0041$$

The potential for improvement is

$$PA(x^{3,h}) = 67711219 - 67674699 = 36{,}520$$

Table 7 indicates the number of observations N_4 required to guarantee a probability $\alpha_3 = .10$, .20, etc... of improvement, and also the associated expected gain. The stopping rules are used with $\alpha_3 = 30\%$, $\beta_3 = 10\%$ and \$100.00

equivalent to 5% of the potential for improvement

α_3	N_4	E_{N_4}(Gain)	E_{N_4}(Gain) / [PA($x^{3,h}$)]
0.10	26	1377.36	3.77%
0.20	55	2812.26	7.70
0.30	87	4281.63	11.72
0.40	125	5885.37	16.11
0.50	169	7570.67	20.73
0.60	223	9417.56	25.78
0.70	293	11500.93	31.49
0.80	392	13958.02	38.22
0.90	560	17148.39	46.96
0.95	729	19494.47	53.38

Table 7 - Problem of size 15
Situation after iteration 3.

Thus: $C_0 = \dfrac{100}{.05 \times 36{,}520} = \$54.76 \times 10^{-3}$

First rule. $N_4 = 87$, $E_{N_4}(\text{gain}) = 4282$

$$E_{N_4}(\text{gain}) = 4{,}282 > \beta_3.PA(x^{3,h}) = .10 \times 36{,}520 = 3{,}652$$

This rule would suggest to continue the search.

Second rule.

$$E_{N_4}(\text{gain}).C_0 = 4{,}282 \times 54.76 \times 10^{-3}$$

$$= 234.48 < C_s N_4 + C_E = 3.35 \times 87 + 40 = 331.45$$

This rule suggests to stop.

The evolution of the Weibull density function between iteration 1 and iteration 3 is drawn on graph 13. It can be seen that the distribution is very stable between iterations 2 and 3, although it varies considerably between the first two iterations. The estimates obtained at iteration 3 and the confidence intervals were reported in experiment (i) in the previous paragraph.

6. Tests of independence

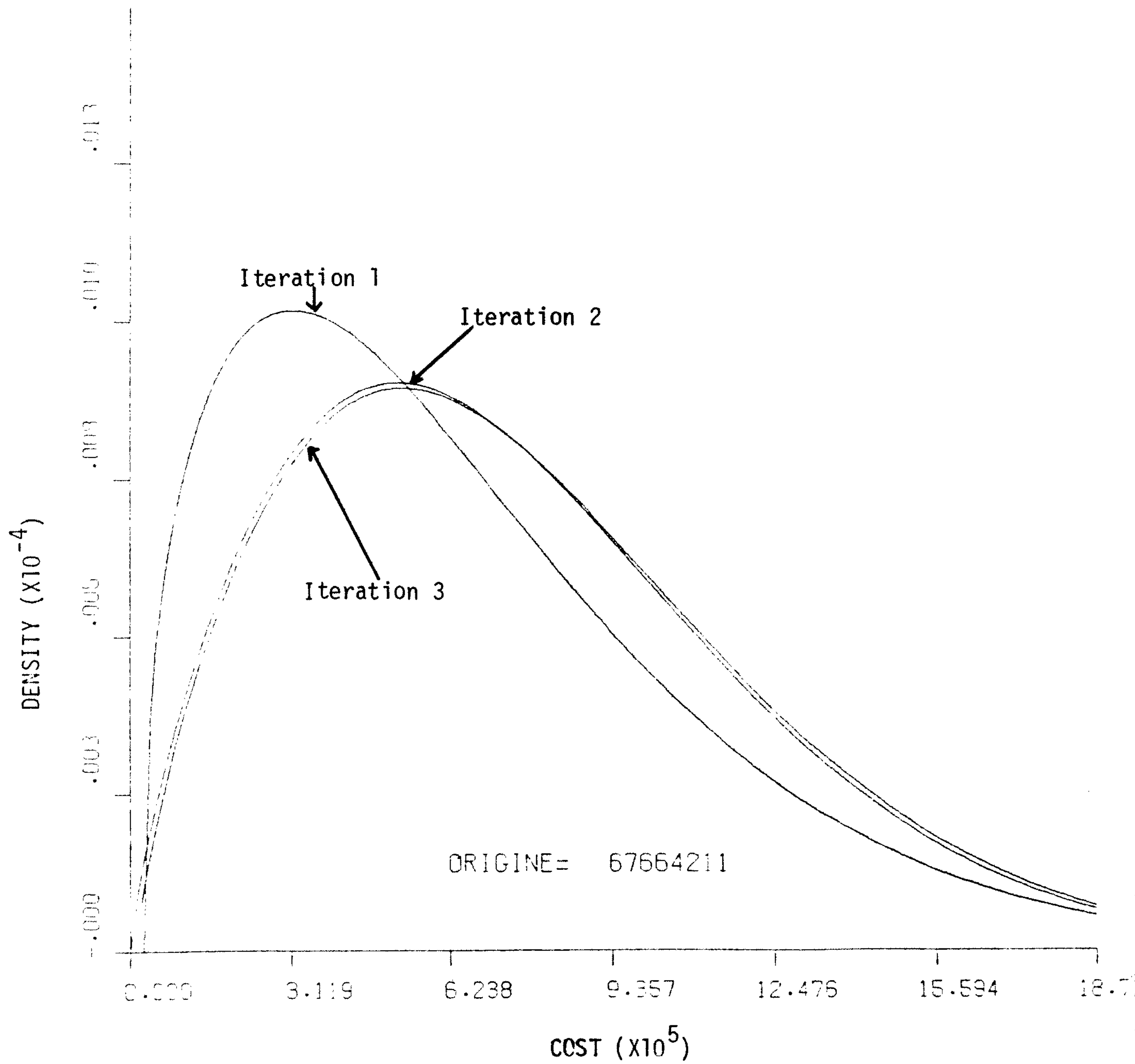
The test of independence was applied to the three following experiments considered previously:

i) Problem ($n=12, \Lambda=66$). Generation of local optima by OPNET1: $N=200$, $R=16$ (see tables 3.A and 3.B).

ii) Problem ($n=15$, $\Lambda=105$). Generation of local optima by OPNET4-8: $N=159$, $R=152$ (see tables 4.A and 4.B)

iii) Problem ($n=15$, $\Lambda=105$). Generation of local optima by OPNET4-8. The estimation was done at level III with $E=25$, $m=10$ (See tables 4.A and 4.B).

In the three experiments the test based on runs did not reject the assumption that the local optima used for estimating the Weibull distribution were statistically independent.



DENSITY FUNCTION

GRAPH 13. Problem of size 15.
OPNET1(N = 50 ; R = 25), OPNET1(N = 101 ; R = 44), OPNET1(N = 159 ; R =

V - CONCLUSION

In this paper we presented and evaluated an exact branch and bound algorithm and two heuristic hill-climbing techniques to solve a specific discrete formulation of the optimal transportation network design problem. Of the two hill-climbing techniques proposed, OPNET1 and OPNET4, only OPNET4 is applicable to large size problems: a parameter (called ML) explicitly controls the tradeoff between efficiency and solution quality in OPNET4. It was proposed to combine in practice a hill-climbing algorithm with random generation of the initial solutions, thus generating a distribution of local optima. In order to determine when to stop sampling local optima and in order to use the whole distribution of local optima to estimate the exact optimum, we followed McRoberts (1971) and Golden and Alt (1979) and fitted a Weibull distribution to the empirical distribution of local optima. Tests of goodness of fit and tests of independence were applied.

Previous work in statistical optimization was extended in several ways. A new confidence interval was introduced which improves over the one proposed by Golden and Alt (1979), in that the tradeoff between the degree of confidence and the width of the interval is now made explicit. Besides, this new interval is dependent on the shape parameter $\hat{c}$. A new stopping rule was introduced which improves over the one proposed by McRoberts (1971), in that we explicitly determine the number of additional local optima to generate in order to obtain a given probability of improvement. The stopping rule weighs the additional computing cost versus the expected benefit and compares the expected gain with the potential for improvement. A new estimation procedure was proposed consisting in using as sample points the minima of explicit samples of local optima: this procedure can be used at very little additional cost once a large number of local optima have been generated. This method presents the advantage of provoking a decrease in the value of the scale parameter $\hat{b}$, which tends to diminish the width of the confidence interval. However the effect of this estimation procedure on the shape parameter $\hat{c}$ is unpredictable and therefore it is not sure that the confidence interval will indeed be reduced.

A distinction was introduced between the *induced* distribution of local optima, obtained directly by the combination of a hill-climbing algorithm and of the uniform random generation of initial solutions, and the *pseudo-distribution* of local optima, obtained from the induced distribution by retaining only distinct local optima. This distinction was introduced because of the repetition of local optima which can occur for problems of small size and very good hill-climbing algorithms such as OPNET1. In the numerical experiments, the Weibull hypothesis was made for the pseudo-distribution of local optima and not for the induced distribution. However the use of the stopping rule assumes that the two distributions are identical, an assumption valid only with large size problems.

The numerical application of the statistical optimization methodology to the network design algorithms reported in this paper consolidates the empirical and practical validity of this approach: the tests of goodness of fit and the graphs showing the empirical and theoretical Weibull curves provide statistical and visual evidence for the empirical validity of this approach. We noted the importance of the quality (or power) of the heuristics in the practicality of the Weibull methodology. The more powerful the heuristic the less spread there is in the distribution: this tends to give better confidence intervals. However a powerful heuristic (such as OPNET1) is too costly and does not generate enough distinct local optima, i.e. does not provide enough sample points for the estimation of the Weibull distribution, at least for small problems (e.g. $n=12$, $\Lambda=66$): there are too few implicit samples and the approximation of a discrete distribution by a continuous distribution becomes questionable. On the other hand, if the heuristic is too weak, the empirical pseudo-distribution is too skewed towards the right, and the confidence intervals are very bad. Besides, with very weak heuristics the implicit samples contain too few elements for m to be considered sufficiently large for the Fisher-Tippett analogy. Thus the experiments with heuristics of varying power suggest that the statistical optimization method is best applied with heuristics of *intermediate* quality: such heuristics provide many distinct sample points for statistical estimation while keeping the confidence intervals sufficiently narrow.

To sum up, the approach proposed in this paper seems promising to make the best possible use of the information provided by reasonably cheap and good hill-climbing algorithms to combinatorial programming problems such as the optimal network design problem. It should be emphasized however, that, if the Weibull hypothesis seems to work in practice it still has no firm theoretical foundation. The Fisher-Tippett theorem cannot, strictly speaking, be applied to discrete distributions. Work remains to be done to ascertain the theoretical validity of the approach.

ACKNOWLEDGEMENTS

The authors are grateful to Michael Florian and Britton Harris for their suggestions at different stages of the development of this work.

REFERENCES

BILLHEIMER, J.W. (1970). Optimal Route Configurations with Fixed Link Construction Costs. Stanford Research Institute, SRI-Project 454531-309.

BILLHEIMER, J.W. and GRAY, P. (1973). Network design with fixed and variable cost elements. Transportation Science 7, No. 1, 49-74.

BIRNBAUM, Z.W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. Journal of the American Statistical Association 47, 425-441.

BOYCE, D.E. editor (1979). Special issue on the optimal network design problem. Transportation Research 13B, #1.

BOYCE, D.E., FARHI, A. and WEISCHEDEL, R. (1973). Optimal network problem: A branch and bound algorithm. Environment and Planning 5, 519-533.

CLOUGH, D. (1969). An asymptotic extreme value sampling theory for the estimation of a global maximum. Canadian Journal of Operations Research 7, 102-115.

DANNENBRING, D. (1977). Estimating optimal solutions for large combinatorial problems. Management Science 23, 1273-1283.

DIONNE, R. and FLORIAN, M. (1979). Exact and approximate algorithms for optimal network design. Networks 9, 37-60.

DREYFUS, S.E. (1969). An appraisal of some shortest path algorithms. Operations Research 17, 3, 395-412.

FISHER, R. and TIPPETT, L. (1928): Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of Cambridge Philosophical Society, 24, 180-191

GIBBONS, J.D. (1971). Non Parametric Statistical Inference. McGraw Hill.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics 44, 3, 423-453.

GOLDEN, B.L. (1977). A statistical approach to the travelling salesman problem. Networks 7, 209-225.

GOLDEN, B.L. and ALT, F.B. (1979). Interval estimation of a global optimum for large combinatorial problems. Naval Research Logistics Quarterly 26, 1, 69-77.

GUMBEL, E. (1958). Statistics of Extremes. Columbia University Press.

HARTER, H. and MOORE, A. (1965). Maximum likelihood estimation of the parameters of the Gamma and Weibull populations from complete and from censored samples. Technometrics 7, 639-643.

HOANG-HAI-HOC (1973). A computational approach to the selection of an optimal network. Management Science 19, 5.

LARDINOIS, C. (1980a). Inférence statistique et problèmes d'optimisation combinatoire de grande taille. Forthcoming in INFOR.

LARDINOIS, C. (1980b). Problèmes d'optimisation combinatoire de grande taille et inférence statistique. Application à la planification urbaine et régionale. Forthcoming Ph.D. thesis in City and Regional Planning, Université de Montréal.

LeBLANC, L.J. (1975). An algorithm for the discrete network design problem. Transportation Science 9, 183-199.

LENSTRA, J.H. and RINNOY KAN, A.H.G. (1978). Computational complexity of discrete optimization problems. In Interfaces between Computer Science and Operations Research, Lenstra, J.K., Rinnoy Kan, A.H.G. and Van Emde Boas, P., editors, Mathematical Centrum, Amsterdam, The Netherlands.

LOS, M. (1975). Simultaneous Optimization of Land Use and Transportation in New Town Design. Ph.D. Thesis in City and Regional Planning, University of Pennsylvania, U.S.A.

LOS, M. (1976). An Optimal Network Problem without Congestion: Some Computational Results. Publication #40, Centre de recherche sur les transports, Université de Montréal.

LOS, M. (1978). Simultaneous optimization of land use and transportation: A synthesis of the quadratic assignment problem and of the optimal network problem. Regional Science and Urban Economics 8, 21-42.

LOS, M. (1979). A discrete-convex programming approach to the simultaneous optimization of land use and transportation. Transportation Research 13B, 33-48.

LUNDQVIST, L. (1973). Integrated location-transportation analysis: A decomposition approach. Regional and Urban Economics 3, 3, 233-262.

MANN, N. SCHAFER, R. and SINGPURWALLA, N. (1974). Methods for Statistical Analysis of Reliability and Life Data. Wiley.

McROBERTS, K. (1971). A search model for evaluating combinatorially explosive problems. Operations Research 19, 1331-1349.

MURCHLAND, J.D. (1967). The Effect of Increasing or Decreasing the Length of a Single Arc on all Shortest Path Distances in a Graph. LBS-TNT-26, London Business School, London, U.K.

OWEN, D.B. (1962). Handbook of Statistical Tables. Addison Wesley. 423-446.

PRIM, R.C. (1957). Shortest connexion network and some generalizations. The Bell System Technological Journal 36, 6.

REITER, S. and SHERMAN, G. (1965). Discrete optimizing. Journal of the SIAM 13, 3, 864-889.

ROBSON, D. and WHITLOCK, J. (1964). Estimation of a truncation point. Biometrika 51, 33-39.

STEENBRINK, P.A. (1974a). Optimization of Transport Networks. Wiley.

STEENBRINK, P.A. (1974b). Transport network optimization in the Dutch integral transportation study. Transportation Research 8, 11-27.

SWED, F.S. and EISENHART, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. Annals of Mathematical Statistics 14, 66-87.

WEIBULL, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanics 18, 293-297.

WEIDE, B. (1977). A survey of analysis techniques for discrete algorithms. Computing Surveys 9, 4, 291-313.

WHITING, P.D. and HILLIER, J.A. (1960). A method for finding the shortest route through a road network. Operational Research Quarterly 11, 37-40.

WONG, R.T. (1978a). Accelerating Benders Decomposition for Network Design. Ph.D thesis in Electrical Engineering and Computer Science. M.I.T., Cambridge, Massachusetts, U.S.A.

WONG, R.T. (1978b). A Survey of Network Design Problems. Working Paper, M.I.T., Operations Research Center, Report # OR 080-78.

ZANAKIS, S. (1977). Computational experience with some nonlinear optimization algorithms in deriving maximum likelihood estimates for the three-parameter Weibull distribution. TIMS Studies in the Management Science 7, 63-77.