

THE PURPOSE OF MATHEMATICAL PROGRAMMING IS INSIGHT, NOT NUMBERS*

ARTHUR M. GEOFFRION
Graduate School of Management
University of California
Los Angeles

ABSTRACT. The ostensible purpose of a mathematical programming model is to optimize a stipulated objective function subject to stipulated constraints. But its *true* purpose, at least in strategic applications as every experienced practitioner should know, is to help develop insights into system behavior which in turn can be used to guide the development of effective plans and decisions. Such insights are seldom evident from the output of an optimization run. One must know not only *what* the optimal solution is for a given set of input data, but also *why*. The desired insights usually have more to do with the "why" than the "what." This paper advocates the use of highly simplified analytic models to help explain the "whys" behind the solutions of conventional mathematical programming models. A methodological approach is described which permits the development of richer insights than would otherwise be possible. This approach is illustrated with reference to a facility location study carried out recently for a consumer products manufacturer.

The purpose of computing is insight, not numbers.

R. W. Hamming
*Numerical Methods for Scientists
and Engineers*
McGraw-Hill, 1962

I. A Methodological Proposal for Strategic Applications

Mathematical programming models are fine as far as they go. The trouble is that they seldom go far enough due to a serious inherent limitation: they can deliver an optimal solution for a given set of input data, but they do not explain WHY the solution is what it is. Numerical models involving an iterative solution procedure engender a "numbers in-numbers out" user attitude that tends to inhibit the development of insights due to a false sense of (numerical) security.

Yet it is very important to develop fundamental insights into the reasons *why* an optimal solution is what it is, particularly for strategic applications of mathematical programming. One reason is that few if any applications lead to a single perfect numerical model whose solution is directly translatable into practical action. Rather, there is a *family* of imperfect numerical models reflecting alternative assumptions, objectives, and data estimates; an understanding of the solution behavior of the whole family is needed in order to fully support the development of an appropriate plan of action.

Another reason for the importance of insights into the determinants of an optimal solution is that they help to overcome the serious validation/credibility obstacles so often present in practical applications. How can one be convinced a model is a useful representation of the real system? And how can the end-user of a model — usually a managerial or political

* This paper was partially supported by the National Science Foundation and by the Office of Naval Research. The computational assistance of Mr. Steven Kwan is gratefully acknowledged.

figure rather than a technical person — be persuaded to use the model as a problem-solving aid? The answer to both questions is, in substantial part, that purely numerical results *must* be supplemented by intuitively reasonable explanations as to why these results are as they are. Otherwise the validity of a model can only be taken as an act of faith and the end-user will be inclined to revert to intuition or some other more secure mode of analysis.

Where will the desired insights come from to illuminate numerical mathematical programming models? The approach advocated here is that they should come from simplified auxiliary models that are both intuitively plausible and solvable in closed form or by simple arithmetic. The solution behavior of a well-chosen auxiliary model should be vastly more transparent than that of the full mathematical programming model, yet it should yield fairly good predictions of the general solution characteristics of the full model.

A general methodological approach is as follows:

1. Reduce the level of detail and complexity of the full mathematical programming model until it can be solved in closed form or by simple arithmetic. Call this an *auxiliary model*.

2. Derive from the auxiliary model a set of tentative hypotheses concerning the general behavior of the solution of the full model — the cost tradeoffs determining the optimal solution for a given set of data, the nature of the induced change in the optimal solution as certain data are changed parametrically, and so on.

3. Generate specific predictions from the tentative hypotheses and test these numerically using the full model.

4. To the extent that the numerical tests confirm (i.e., do not contradict) the tentative hypotheses, take these hypotheses as a conceptual framework for understanding and interpreting the numerical results provided by the full model.

The use of relatively simple models in conjunction with complex models is not a new idea. “Pilot” or otherwise scaled-down versions of full-scale models have long been used as a practical evolutionary approach to model-building. Since solving them still requires elaborate algorithmic machinery, however, they are generally unsuited to the purposes we have outlined for auxiliary models. So-called “repro-models” [8] have been proposed as simplified input/output approximations to complex models. They are mathematical constructs which lack the explanatory properties essential for an auxiliary model. Much closer to the auxiliary model concept is the use of simple analytical models that are validated and perhaps calibrated with the help of more complex numerical models. Ignall, Kolesar and Walker give several splendid examples of this approach in the context of emergency service deployment [5]. It so happens that their full-scale models are of the simulation rather than mathematical programming variety. The main difference between this and our approach to auxiliary models is apparently one of reversed objectives: they want full-scale models to support the use of simple models, whereas we want simple models to support the use of full-scale models. This reversal of objectives is probably a consequence of the fact that they were primarily concerned with tactical operational applications

where elaborate simulations may be cumbersome, whereas we are primarily concerned with strategic planning applications where the resources needed to support elaborate mathematical programming models may be more readily available. In any case, I definitely intend auxiliary models to supplement full-scale models, *not to replace them*.

The next section summarizes my experience with auxiliary models in the context of a facility location study carried out recently for a consumer products manufacturer. Three different auxiliary models were developed in an effort to understand and explain the solution behavior of the large integer programming model (over 1400 0-1 variables and 250 constraints).

II. Application in the Context of Facility Location

A consumer products manufacturer has a single plant in California making a range of products which, for the purposes of a warehouse location study, can be treated as a single product group. Goods are distributed nationally through full-line warehouses to customers grouped into about 130 customer zones. There are about 65 permissible warehouse locations, the best subset of which is to be selected. Business policy requires that each customer zone must be single-sourced, that is, it may not receive goods from more than one warehouse. The main objective is to decide which warehouses to use and to design their service areas so as to minimize the sum of all relevant costs (freight in, fixed and variable warehouse-related costs, and freight out) subject to the constraints that forecast annual demands will be met, customer service standards will be maintained, and warehouse aggregate throughput limits (both lower and upper) will be honored.

The problem as just described is a familiar one in the literature on facility location (e.g., [3]). It has been simplified for the purpose of this illustration by omitting mention of several complexities treated in the actual study, including: plant direct shipments, economies of scale on inbound freight and warehouse throughput costs, the complexities of customer will-calls and delivery consolidation services, and the need to choose between alternative inventory stockage policies.

I shall describe in turn three of the auxiliary models that have been used in conjunction with the full integer programming model.

A. The First Auxiliary Model

The first auxiliary model is based on one of the simplest related problems to be found in the location economics literature. It makes the following simplifying assumptions:

- A1. Demand is uniformly distributed on the plane with a density of $\rho(CWT/mi^2)$.
- A2. All warehouses are identical and arbitrarily relocatable.
- A3. The supply cost for each warehouse is s (\$/CWT) regardless of its location.
- A4. The fixed cost of each warehouse is f (\$).
- A5. The variable throughput cost of each warehouse is v (\$/CWT).
- A6. The outbound freight rate for each warehouse is t (\$/CWT-mi.).
- A7. There are no throughput limits for the warehouses.

How many warehouses should there be, where, and with what service areas to cover a total area A (mi^2) of unspecified shape at a minimum total

cost? The answer (e.g., [1], [6]) is as follows: there should be

$$(1) \quad n^* = \frac{A}{3.05} (\rho t / f)^{2/3}$$

warehouses¹, each at the center of a circular service area comprising

$$(2) \quad A^* = 3.05(\rho t / f)^{-2/3}$$

square miles, located anywhere on the plane so long as their service areas do not overlap.

It is also known that this optimal solution is remarkably insensitive to the exact geometry of the warehouse service areas. Hexagonal rather than circular service areas change n^* by only about + 0.2%, and even square service areas would change n^* by only about + 1.1%. This insensitivity, along with the arbitrariness of location (one need only avoid overlapping service areas), enables one to position the warehouses and adjust their service area shapes so as to give a fair approximation to the shape of the United States. The degree of suboptimality thereby incurred is likely to be very small if n^* is more than a dozen or so, in which case (1) and (2) give a very nearly optimal solution for the more realistic situation where the demand to be covered is uniform over the continental U.S.

One can readily answer almost any "why" question associated with this simple model. The answer at a mathematical level is available if one understands the mathematical derivations supporting (1) and (2). This is in stark contrast to the clumsiness of a large numerical integer programming model. Effective communication of these insights at a managerial level, however, requires that a clear rationale be given for (1) and (2) in terms familiar to management. The key is to interpret A^* as that circular service area size, with the warehouse at the center, which minimizes the average unit cost of satisfying all demand within the area. Supply and variable warehouse throughput costs can be ignored because their cost contributions are independent of warehouse location and service area shape. That leaves fixed costs, which clearly are inversely proportional on a unit basis to service area size, and outbound freight costs, which increase proportionally on a unit basis with the square root of service area size (this follows from purely dimensional considerations). For a service area size less (more) than A^* as given by (2), the unit fixed costs are decreasing faster (slower) than the unit outbound freight costs are increasing. A graph like Figure 1 brings these words alive for manager and management scientist alike. Here we have used average values of ρ , f , t and A obtained from the detailed data prepared originally to support the integer programming model for the consumer products manufacturer.

If a warehouse with a circular service area of size A^* gives the minimum cost per square mile of meeting demand, then A/A^* surely gives the number of warehouses needed to meet demand over a total area A in the most economical way (presuming A/A^* is an integer). Indeed, n^* in (1) can be rewritten equivalently as $n^* = A/A^*$.

Such an explanation gives a quite satisfactory intuitive understanding for (1) and (2). The tradeoff between fixed costs and outbound freight costs is clearly revealed as the factor which controls the optimal solution.

¹ Actually, n^* as given can be fractional; in this case it is necessary to examine the total cost expression for the two integers on either side of n^* . We shall neglect this detail here because n^* is quite large in this particular application.

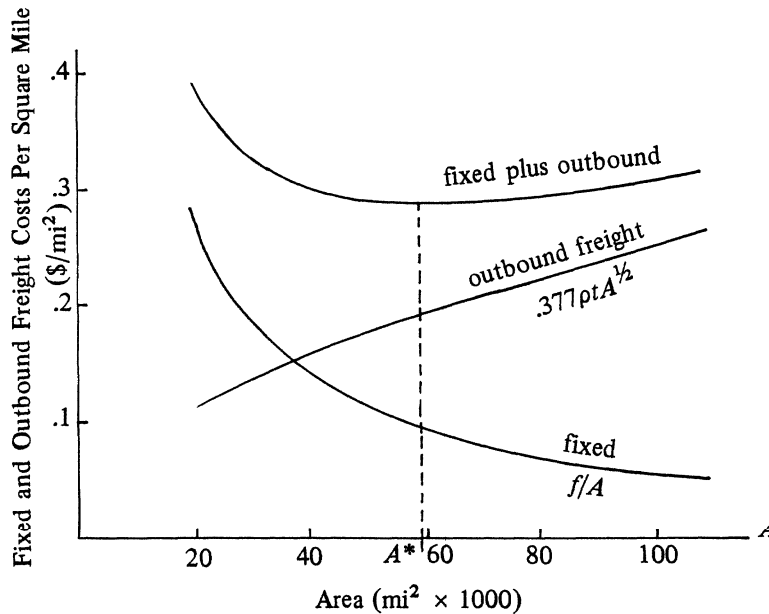


FIGURE 1
 How Fixed and Outbound Freight Costs Determine
 the Best Service Area Size for the First Auxiliary
 Model (data for consumer products manufacturer)

Hopefully this insight, even though gained for but a grossly simplified version of the problem, will contribute to an understanding of the behavior of the real system. One way to test the validity of this reasoning is to generate sensitivity analysis predictions using the auxiliary model and then to test these using the integer programming model (refer to Steps 2-4 of the general methodological approach given at the end of §I.) This will now be done.

Relation (1) implies an easily calculated percentage change in n^* for any given percentage change in ρ , t or f . For instance, increasing ρ or t by 5% implies that n^* will increase by 3.3% (since $1.05^{2/3} = 1.033$). These sensitivity relationships were examined empirically using the integer programming code for several sets of data. The results, given in Table 1, indicate that the sensitivity predictions are surprisingly accurate for small changes in ρ , t and f .

The successes shown in Table 1 add to our confidence in the auxiliary model as a source of insight into the behavior of the real system. They suggest that the explanations for the auxiliary model's sensitivity behavior, which can be spelled out clearly in full detail, may also be operative for the integer programming model and hence (hopefully) for the real system.

Data Change	Integer Programming Model		First Auxiliary Model
	Optimal n before change	Optimal n after change	Predicted n after change ²
Increase ρ by 5%	24	24	24.8
Increase t by 7%	24	25	25.1
Increase t by 13%	26	28	28.2
Decrease f by 5%	24	25	24.9

TABLE 1

Sensitivity Analysis with the First Auxiliary Model
(data for consumer products manufacturer)

Another use to which the auxiliary model can be put is to examine the conventional wisdom quoted so often by practitioners of the art of facility location analysis: that total system costs are quite insensitive to moderate departures from the optimal number of warehouses. The usual way in which this is demonstrated is to calculate the total costs associated with several manually configured systems, each with a different number of warehouses and with service areas drawn as well as manual methods will allow (e.g., Magee [7]). The shortcoming of this approach is that it does not use a least cost system associated with a given number of warehouses; the warehouse locations and their service areas are not necessarily optimal. This shortcoming can be overcome in the context of the auxiliary model. Given that there will be n warehouses, the optimal system under the assumptions of the auxiliary model has total cost

$$(3) \quad TC^*(n) = \rho A(v + s) + nf + .377 \rho t A^{3/2} n^{-1/2}$$

The first term is independent of n and can be ignored. The second (fixed cost) and third (outbound freight cost) terms and their sum are graphed in Figure 2 using the same average values for ρ , A , f and t as were used for Figure 1. The ordinate is given in relative costs, with 1.0 being the fixed plus outbound freight costs associated with n^* . It is evident from Figure 2 that there is indeed considerable latitude to depart from n^* without incurring a substantial cost penalty. For instance, n can range between about 33 and 51 and yet keep fixed plus outbound freight costs within 1% of optimum. The explanation for this relative insensitivity is evident from the graphs of the component cost functions in Figure 2.

² The predictions were computed as:

$$\left(\frac{n^* \text{ from (1) after data changes}}{n^* \text{ from (1) before data change}} \right) \times \left(\frac{n^* \text{ from IP model}}{\text{before data change}} \right)$$

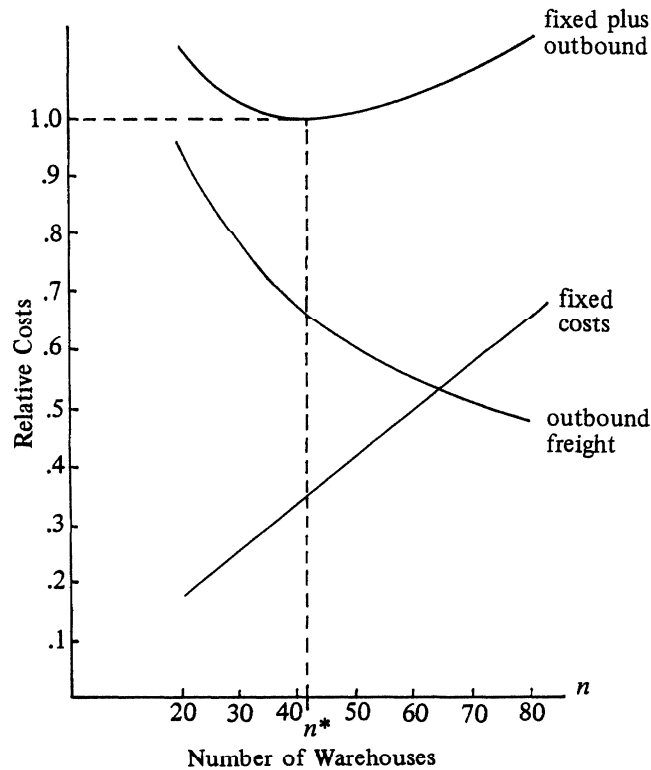


FIGURE 2

Influence of the Number of Warehouses on an Otherwise Optimally Configured System According to the First Auxiliary Model (data for consumer products manufacturer).

An obvious limitation of this first auxiliary model is that assumption A3 is overly stringent. Indeed, A3 virtually ignores the economic impact of plant location and inbound transportation. The next auxiliary model is designed to overcome this limitation.

B. The Second Auxiliary Model

In order to incorporate the influence of inbound transportation economics it seems necessary to replace assumption A3 by a version in which inbound freight costs are given in terms of a $\$/CWT\cdot mi$ rate instead of a flat $\$/CWT$. If all other assumptions remain the same, however, this change introduces an exceedingly messy integral. I have therefore taken the more expedient course of working out a one-dimensional version of the first auxiliary model. The assumptions are:

- A1a. Demand is uniformly distributed on a straight line of length L (mi.) with a density $\rho(CWT/mi)$.
- A3a. There is a single plant at one end of the line, supplying each warehouse at an inbound freight rate $r (\$/CWT\cdot mi.)$, where $r \leq t$.
- A2, A4-A7 as before.

How many warehouses should there be, where, and with what service areas to cover the demand over the entire line at minimum total cost? It can be shown that there should be

$$(4) \quad n^* = \frac{1}{2} L(\rho t/f)^{1/2} \sqrt{1 - \left(\frac{r}{t}\right)^2}$$

warehouses¹ each with a service area of size $l \triangleq L/n^*$ miles, configured as shown in Figure 3. Notice that each warehouse is displaced from the center of its service area in the direction of the plant by an amount proportional to r/t . The reader may find it interesting to note the similarities and dissimilarities between expressions (1) and (4).

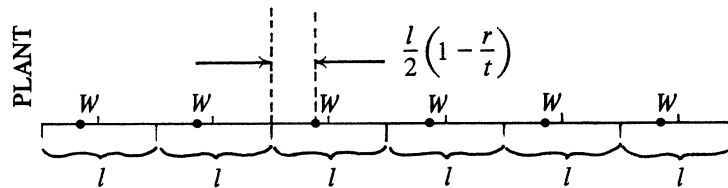


FIGURE 3

Optimal Service Area Configuration for the Second Auxiliary Model

A rationale to explain in simple managerial terms *why* the above constitutes an optimal solution is not given here, but can be worked out along the lines of the rationale presented for the first auxiliary model.

The reader may be curious as to how close n^* from (4) comes to the value obtained from the integer programming model: n^* from (4) is 26, while the integer programming optimum was 24. Surprisingly, this is better than the estimate given by the auxiliary model (for which n^* was $41\frac{1}{2}$).

This model has some remarkable implications concerning the influence of plant location and inbound transportation costs on the optimal warehouse configuration. We see from (4) and Figure 3 that this influence enters only via the ratio r/t . This ratio was .082 for the consumer products manufacturer, a reflection of the fact r was based mostly on rail carload rates but with some truckload shipments, while t was based mostly on much higher less-than-truckload rates. The consequences of such a low r/t ratio are surprising. According to (4), n^* is essentially the same as though r were 0, since $\sqrt{1 - (.082)^2} = .9966 \cong 1$. And each warehouse is displaced from the center of its service area toward the plant in California by a mere 8.2% of the half-width of each service area (less than 5 miles).

In other words, this model indicates that for firms with such a low r/t ratio the influence of inbound transportation economics and plant location on the optimal warehouse configuration is extremely small. In such cases one may as well neglect inbound transportation altogether, as the first auxiliary model essentially did.

An attempt was made to confirm empirically the predicted insensitivity with respect to r . According to (4), the r/t ratio could vary from 0 to more than 0.2 without significantly altering n^* . In fact, however, scaling all inbound freight rates by a factor of $1/2$ (which reduces r/t to .041) causes the number of warehouses in the optimal solution of the integer programming model to increase from 24 to 28. This observation was in clear disagreement with the prediction. It seemed at first to cast serious doubt on the use-

fulness of the second auxiliary model, but further reflection revealed that the empirical test could have been prejudiced by a confounding effect: factoring all inbound freight rates by $1/2$ not only halved the average as desired, but also halved the *deviation* about the average for individual rates. The latter effect, which cannot be reflected in the second auxiliary model, can be eliminated simply by reducing the inbound freight rates in a subtractive rather than multiplicative fashion. More precisely, subtract half the average inbound rate from each individual rate instead of factoring each individual rate by $1/2$. This halves the average rate as before but leaves unchanged the deviation of individual rates about the average. The integer programming model was rerun using the subtractive reductions. As hoped, this integer programming solution exhibited the relative insensitivity predicted by the second auxiliary model: the optimal number of warehouses increased by just one.

Thus we are led by the second auxiliary model to an important *insight* concerning the behavior of the full integer programming model: *the optimal solution is quite insensitive to deviation-preserving changes in the mean inbound freight rate when the r/t ratio is small.* (Technically, of course, this "insight" should more properly be called a tentative hypothesis for which there is some analytical and empirical evidence.)

The influence of deviation-altering changes in the inbound rates, as well as other warehouse-specific data sensitivities, can be studied via a third auxiliary model.

C. The Third Auxiliary Model

This auxiliary model is identical with the first except that it drops the assumption that costs are identical for all warehouses. Let k index the candidate warehouses.

A1, A7 as before.

A2a. All warehouses are arbitrarily relocatable.

A3b. The supply cost for warehouse k is s_k (\$/CWT) regardless of its location.

A4a. The fixed cost of warehouse k is f_k (\$).

A5a. The variable throughput cost of warehouse k is v_k (\$/CWT).

A6a. The outbound freight rate for warehouse k is t_k (\$/CWT - mi).

It is not possible in general to derive a closed form optimal solution along the lines of (1)-(4). There is, however, a simple noniterative procedure for calculating how many warehouses there should be and with what size service areas (and even which ones should be selected). It will be convenient to work with the demand D_k (CWT) covered by the circular service area centered on warehouse k , rather than in terms of the actual area A_k covered ($D_k = \rho A_k$). Instead of having to cover a total area A we shall, equivalently, cover a total demand D ($D = \rho A$).

The procedure for solving the problem is as follows. Its validity should be evident from a rationale similar to the one given previously for the first auxiliary model.

Step 1. Calculate for each k the value D_k^* which minimizes

$$(5) \quad ATC_k(D_k) \triangleq \frac{f_k}{D_k} + s_k + v_k + \frac{.377 t_k \rho^{-1/2} D_k^{3/2}}{D_k},$$

where $ATC_k(D_k)$ measures the average total cost in $\$/CWT$ associated with warehouse k when it has a circular service area covering an amount D_k of demand. Denote $ATC_k^*(D_k^*)$ by ATC_k^* .

Step 2. Sort on ATC_k^* and reindex the warehouses so that

$$(6) \quad ATC_1^* \leq ATC_2^* \leq ATC_3^* \leq \dots$$

Step 3. Plot the partial sums $\sum_{k=1}^K D_k^* ATC_k^*$ against the partial sums $\sum_{k=1}^K D_k^*$ for $K = 1, 2, \dots$ and mark D on the abscissa. See Figure 4.

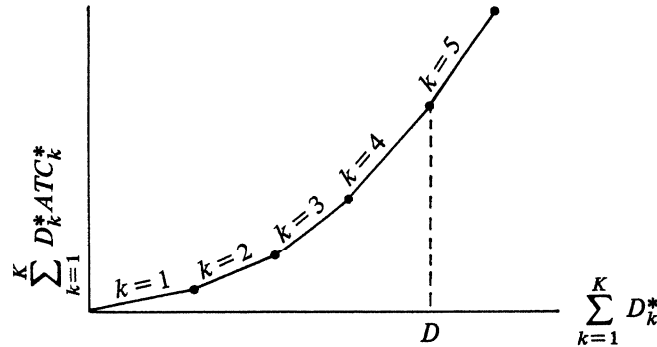


FIGURE 4
Hypothetical Illustration of the Solution Procedure
for the Third Auxiliary Model

Step 4. If D corresponds with one of the partial sums of D_k^* , say the K_0 th, then K_0 is the optimal number of warehouses and the particular ones which should be used are indexed $1 \leq k \leq K_0$. The k th should have a circular service area covering $D_k^* CWT$ of demand, and they can be located anywhere in the plane so long as their service areas do not overlap. In Figure 4, $K_0 = 4$.

If D does not correspond exactly to one of the partial sums of D_k^* , then an optimal solution is no longer immediately obvious. However, unless D is relatively small, it is adequate for our purposes to allow K_0 to be fractional, just as n^* can be fractional in (1) and (4). Another essential comment on Step 4 is that, as explained in the context of the first auxiliary model, only a small degree of suboptimality is incurred if K_0 is larger than a dozen or so and the warehouses are placed as necessary with reshaped service areas so as to cover the continental U.S.

This model is so closely related to the first auxiliary model that a very similar rationale in managerial terms can be given for the "why" behind its solution. For instance, the cost relationships shown in Figure 1 still apply.

How well does this model predict the behavior of the integer programming model for the consumer products manufacturer? Table 2 gives comparative results in terms of n^* and in terms of a cost breakdown. The auxiliary model was applied twice, once with the procedure as stated and once with a simple manual modification which took account of the fact that closely adjacent warehouses cannot be open simultaneously if their service areas would obviously overlap. The agreement in both cases is fairly good.

Moreover, the auxiliary model gives a tolerable prediction as to the particular warehouses which should be selected for use: of the top ranking 12 (24) warehouses according to the model, 11 (14) appeared in the optimal solution of the integer program.

	Integer Programming Model	Third Auxiliary Model ³	
		Unmodified	Modified for Locational Interference
Optimal Number of Warehouses	24	20.6	25.3
Optimal Costs (percentagewise)			
Fixed	6	5	4
Inbound & Var. Thruput	54	53	55
Outbound Freight	40	42	41
	100	100	100

TABLE 2
Results From the Third Auxiliary Model
(data for consumer products manufacturer)

³Actually, a slightly different version of the outbound transportation cost term of (5) was used. Instead of using average values for t_k , which were not conveniently available, a regression was performed to fit an expression of the form αD_k^β in place of $.377 t_k \rho^{-1/2} D_k^{3/2}$. (Note that α is taken to be independent of k .)

Finally, consider again the multiplicative factoring by $1/2$ of all inbound transportation rates, the change which foiled the second auxiliary model by causing an increase of four in the number of warehouses open in the optimal integer programming solution. The third auxiliary model yields the surprisingly accurate prediction that the increase will be 4.4. Since we are seeking insights rather than numbers, it is even more satisfying that an intuitively satisfactory explanation can be given for the increase. The explanation is straightforward but would require going into more numerical detail than seems appropriate here.

III. Conclusion

We have chosen to illustrate the general methodological approach advocated in §I by giving more than one auxiliary model for the same application. This was done to convey the potential diversity of auxiliary models likely to be available, and to stress that developing auxiliary models can be an iterative *design process* guided by the kinds of insights desired. Such models can sometimes be taken from the existing literature on analytical solutions to highly simplified cases. Sometimes available analytic models must be modified or new ones created even to the point of losing a closed form solution. Of course, one should stop short of the point where the loss of auxiliary model tractability due to added complexity destroys its comparative advantage as a source of insights.

A rich source of auxiliary models is available via the notion of Lagrangian relaxation [2]. This technical device is aimed at producing highly tractable simplifications of difficult mathematical programming problems. Although originally conceived for purely algorithmic uses, it turns out that Lagrangian relaxations are frequently amenable to natural managerial interpretation and hence are attractive candidates for adoption as auxiliary models. Initial computational experience along these lines indicates that the idea will be a fruitful one. A very recent paper [4] takes the concept a step further and proposes that auxiliary models based on Lagrangian relaxation be used to help guide the aggregation choices for the full scale model.

It seems not at all unlikely to me that auxiliary models, some based on analytic prototypes and others on Lagrangian relaxation or other formalisms, will eventually be common in strategic applications of mathematical programming. They can be valuable aids for designing full scale models, explaining the reasons behind their numerical results, understanding their parametric sensitivities, and discovering fresh insights into the behavior of the systems being modeled.

REFERENCES

1. Bos, H. C., *Spatial Dispersion of Economic Activity*, Rotterdam University Press, 1965.
2. Geoffrion, A. M., "Lagrangean Relaxation for Integer Programming," *Mathematical Programming Study* 2, 1974.
3. Geoffrion, A. M., "A Guide to Computer-Assisted Methods for Distribution Systems Planning," *Sloan Management Review*, Winter, 1975.
4. ———, "Using an Auxiliary Model to Guide the Design of a Very Large Logistic Planning Model," Discussion Paper No. 60, Graduate School of Management, UCLA, April 1976.
5. Ignall, E. J., P. Kolesar and W. E. Walker, "Using Simulation to Develop and Validate Analytical Emergency Service Deployment Models," P-5463, Rand Corporation, Santa Monica, August 1975.
6. Leamer, E. E., "Locational Equilibria," *Journal of Regional Science*, 8, 2 (1968).
7. Magee, J. F., *Industrial Logistics*, McGraw-Hill, 1968.
8. Meisel, W. S. and D. C. Collins, "Repro-Modeling: An Approach to Efficient Model Utilization and Interpretation," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 4, July 1973.