[44] K. Shanmugam and A. M. Breipohl, "Discriminating between cloud-to-ground and cloud-to-cloud lightning discharges," *J. Geophys. Res.*, (to be published).
[45] R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy*. San Francisco, Calif.: Freeman, 1963.
[46] H. W. Smedes, K. L. Pierce, M. G. Tannguay, and R. M. Hoffer, "Digital computer terrain mapping from multi-spectral data and evaluation of proposed earth resources technology satellite (ERTS) data channels, Yellowstone National Park," in *AIAA Earth Resources Observations and Information System Meeting*, Paper 70-309, Mar. 1970.
[47] J. Spragins, "Learning without a teacher," *IEEE Trans. Inform. Theory*," vol. IT-12, Apr. 1966, pp. 223–230.
[48] D. F. Stanat, "Unsupervised learning of mixtures of probability functions," in *Pattern Recognition*, L. Kanal, Ed. Washington, D.C.: Thompson, 1966, pp. 357–389.
[49] K. Steinbuch and U. A. W. Piske, "Learning matrices and their applications," *IEEE Trans. Electron. Comput.*, vol. EC-12, Dec. 1963, pp. 846–862.
[50] H. Teicher, "Identifiability of mixtures of product measures," *Ann. Math. Statist.*, vol. 38, 1967, pp. 1300–1302.
[51] ——, "Identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 34, 1963, pp. 1265–1269.
[52] ——, "Identifiability of mixtures," *Ann. Math. Statist.*, vol. 32, 1961, pp. 244–248.
[53] R. C. Tryon, *Cluster Analysis*. Ann Arbor, Mich.: Edwards, 1939.
[54] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Ass.*, vol. 58, Mar. 1963, pp. 236–245.
[55] M. Wirth, G. Estabrook, and D. Rogers, "A group theory model for systematic biology, with an example for the oncidunea," *Systematic Zoology*, vol. 15, no. 1, Mar. 1966, pp. 59–69.
[56] Z. Wurtele, "A problem in pattern recognition," *J. Soc. Ind. Appl. Math.*, vol. 13, Mar. 1965, pp. 60–67.
[57] S. J. Yakowitz, "Unsupervised learning and the identification of finite mixtures," *IEEE Trans. Inform. Theory*, vol. IT-16, May 1970, pp. 330–338.
[58] S. J. Yakowitz and J. D. Sprogins, "On the identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 39, 1968, no. 1, pp. 209–214.
[59] S. J. Yakowitz, "A consistent estimator for the identification of finite mixtures," *Ann. Math. Statist.*, vol. 40, 1969, no. 5, pp. 1728–1735.

Correspondence

Nonlinear Programming Model of Crew Assignments for Household Refuse Collection

Abstract—**To determine the manpower requirements for household refuse collection in New York City, a nonlinear programming model has been developed for matching work shifts to curbside refuse demands so as to minimize weekly missed collections subject to union regulations and manpower and truck constraints.**

Introduction

The intent of this paper is to present a nonlinear programming model developed as part of a study of manpower allocation procedures performed by the authors for the Environmental Protection Administration of New York City. This model is used for matching work shifts to curbside refuse demand so as to minimize weekly missed collections subject to union regulations and manpower and truck constraints. In addition it provides insights into the basic interactions and tradeoffs that exist between missed collections, crew allocations, truck availability, wage costs, and personnel requirements for other tasks. While analysis of this model is not explored fully here, typical results are presented to illustrate its usefulness.

Model

In most sanitation districts the amount of refuse which is available at curbside for collection varies according to the day of the week. Since Sunday is not usually a pickup day, a peak load is generated on Mondays and, depending on how the work crews are deployed, some portion of the refuse may remain uncollected at the end of the day. These "missed collections" frustrate efforts to maintain a clean city.

There are two work shifts each day ($i = 1,2$) and the seven days of the week are indexed by the integer j starting with $j = 1$ for Monday.

Manuscript received October 30, 1970; revised January 6, 1971. This work was supported in part by the National Science Foundation under Grant GI-5.

Data available to use is the average productivity p_{ij} (tons collected/crew) on the ith day, jth day, and the average amount of refuse f_j (tons) available at curbside of the jth day. A truck crew consists of three men: a driver and two loaders.

Let n_{ij} be the number of crews assigned to refuse collection and let $\boldsymbol{n}$ denote the vector whose 14 integer valued components are the n_{ij}. Although it is possible to give meaning to fractional crew allocations, we generally want to think of the number of crews as integers.

It is clear that the components of $\boldsymbol{n}$ are nonnegative and that if M is the maximum number of trucks available on any shift, then the n_{ij} cannot exceed M. Also no more personnel can be allocated on day j than a preassigned amount N_j which depends on what other tasks must be taken care of in the sanitation district (such as street cleaning or bulk collection).

Denote by m_j the missed collections in tons on day j. We assume as given the amount m_0 which is leftover from the previous week. The amount missed on any day is the sum of the tonnage generated that day plus the backlog from the previous day minus the amount q_j, which is the amount actually collected. Since q_j is given by

$$q_j = \sum_{i=1}^{2} p_{ij}n_{ij} \tag{1}$$

then the amount of missed collection is

$$m_j = \max(f_j + m_{j-1} - q_j, 0) \equiv \lambda(f_j + m_{j-1} - q_j) \tag{2}$$

where $\lambda(x)$ is the ramp function defined by

$$\lambda(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Thus the missed collections must satisfy a nonlinear difference equation.

Our aim is to minimize the nonlinear objective function

$$f(\boldsymbol{n}) = \frac{1}{2}\sum_{j=1}^{7} m_j^2 + \delta C(\boldsymbol{n}) \tag{4}$$

where $C(\boldsymbol{n}) = \sum_i \sum_j c_{ij} n_{ij}$ is the weekly cost in dollars. The quantities c_{ij} are wages computed according to union rules governing Saturday and Sunday work and night shift differential pay. The parameter δ measures the relative significance of *dollar cost* $C(\boldsymbol{n})$ versus the *social cost* of weekly missed collections.

Certain other constraints also need to be satisfied before an allocation $\boldsymbol{n}$ can be considered as a candidate for optimality: no crew can work more than six days a week (union regulation) nor less than five and, finally, crews are to be assigned to the day shift ($i = 1$) before any are put on the night shift ($i = 2$). In effect this last constraint says: if $\sum_i n_{ij} \leq M$, then set n_{2j} equal to zero and replace n_{1j} by the sum; if $\sum_i n_{ij} > M$, then set n_{1j} to be M and replace n_{2j} by the sum minus M. This constraint can be written as

$$n_{2j}(n_{1j} - M) = 0 \tag{5}$$

since it forces n_{2j} to be zero unless n_{1j} equals M. Regarding the five- and six-day workweek constraints, we reason as follows.

Let $r_j = N_j - \sum_i n_{ij}$ be the number of crews on recreation on the jth day.

For simplicity we assume that $N_j \equiv N$, and thus

$$r_j = N - \sum_i n_{ij}. \tag{6}$$

Since each of the N crews gets a day off each week but no more than two, then $2N \geq \sum_{j=1}^{7} r_j \geq N$. This translates into

$$5N \leq \sum\sum n_{ij} \leq 6N. \tag{7}$$

The problem now can be summarized by asking for the resource allocation vector $\boldsymbol{n}$ which will minimize (4) subject to the 44 linear and nonlinear constraints given by (5) and (7) together with the constraints

$$0 \leq n_{ij} \leq M, \qquad \sum_{i=1}^{2} n_{ij} \leq N.$$

Note that the choice of quadratic terms in (4) is dictated by a desire to penalize large missed collections more severely than small violations. A variant of this model in which the objective function is linear in the m_j has been formulated by Ignall *et al.* [1]. They are careful enough to circumvent the positivity constraint inherent in (2), and thus their formulation is in terms of a linear program. In this regard we should also mention some similar work of Heller [2] in which quadratic programming models were developed for the deployment of police forces.

When $\delta = 0$, one is concerned only with minimizing missed collections. There is a risk now that since dollar cost no longer acts to restrict excessive allocations, the resource $\boldsymbol{n}$ may be overallocated. In this case it is useful to redefine m_j by

$$\tilde{m}_j = f_j + m_{j-1} - q_j \tag{8}$$

where m_{j-1} is defined by (2) and $\sim$ indicates that the expression in (8) is no longer restricted to be nonnegative. Using this notation, one can write $m_j = \lambda(\tilde{m}_j)$. The objective function in (2) is now replaced by

$$\tilde{f}(\boldsymbol{n}) = \frac{1}{2} \sum \tilde{m}_j^2 + \delta C(\boldsymbol{n}) \tag{9}$$

and it is clear that (9) allows one to penalize for over as well as under allocations of $\boldsymbol{n}$. In fact if more crews are used than needed, then $\tilde{m}_j$ is negative, and thus (9) cannot be at a minimum.

A delicate question arises here regarding the capacity of the collection system to absorb the requirements which are put on it. We begin a week knowing the previous week's missed collection m_0 and end it with an amount m_7; m_7 now plays the role of m_0 for the next week, and so on. If the system is not to fall behind the demand and if the refuse "inventory" is not to increase from week to week, then one must ask for conditions on the "stability" of the system. By this we mean that for any week one has $m_7 \leq m_0$. Under this condition the system damps down to a steady state. Indeed, the m_7 of subsequent weeks form a monotonely decreasing sequence so that ultimately one reaches a value at which $m_0 = m_7$. It can be shown that a sufficient though not necessary condition for stability is the following.

TABLE I

TYPICAL DATA

	j Values						
	1	2	3	4	5	6	7
c_{1j} ($/crew)	120	120	120	120	120	135	240
c_{2j} ($/crew)	126	126	126	126	126	141.75	252
p_{1j} (tons/crew)	9.35	9.35	9.35	9.35	9.35	9.35	4.93
p_{2j} (tons/crew)	8.93	8.93	8.93	8.93	8.93	8.93	4.29
f_j (tons)	194.7	200.6	142.7	131.2	158.1	162.8	4.46

Value of m_0 was taken to be 0.634 tons.

The collection system is stable in the sense that $m_7 \leq m_0$, for any m_0 and for any given allocation vector $\boldsymbol{n}$, whenever the following conditions hold:

$$\sum_{j=l}^{7} q_j \geq \sum_{j=l}^{7} f_j, \qquad \text{for } l = 1,\cdots,7. \tag{10}$$

One final remark concerning the model. If there are several adjacent sanitation districts which share or borrow personnel and trucks from one another, then one can consider the broader question of how to allocate the resource vector $\boldsymbol{n}$ whose components are now n_{ijk}, where k denotes the district being considered. In this case one would want to minimize the sum of the missed collections over all the districts simultaneously. The details of how to modify the previous model to accommodate this generalization are fairly straightforward and are left to the reader.

NUMERICAL RESULTS

We carried out numerical computations using figures for p_{ij}, c_{ij}, and f_j which are typical of some districts in New York City. The data is shown in Table I. The algorithm used for minimization is a modification of the Fletcher–Powell variable metric gradient scheme. Constraints were handled by augmenting the objective function by penalty terms. A description of the algorithm is given by Kelley *et al.* [3]. The method is basically a descent procedure, and it requires that the gradient of the objective function be known explicitly. Expressions for the partial derivatives are detailed in the following:

$$\frac{\partial f}{\partial n_{kl}} = \sum_{j=l}^{7} m_j \frac{\partial m_j}{\partial n_{kl}} + \delta c_{kl} \tag{11}$$

since $\partial m_j / \partial n_{kl} = 0$, for $j < l$. The partials $\partial m_j / \partial n_{kl}$ are obtained recursively from

$$\frac{\partial m_0}{\partial n_{kl}} = 0 \tag{12a}$$

$$\frac{\partial m_j}{\partial n_{kl}} = \mu(\tilde{m}_j) \left[\frac{\partial m_{j-1}}{\partial n_{kl}} - p_{kl}\delta_{l,j} \right] \mu(j - l)$$

$$= -\mu(j - l) p_{kl} \prod_{i=l}^{j} \mu(\tilde{m}_i), \qquad \text{for } j = 1,\cdots,7 \tag{12b}$$

TABLE II

	$\Sigma\tilde{m}_j^2$	$\Sigma\tilde{m}_j$	Cost
M = 19	747	56.7 tons	$13,874
M = 15	747	56.7 tons	$13,933

$\delta = 0$, $N = 21$ crews

TABLE III

	$\Sigma\tilde{m}_j^2$	$\Sigma\tilde{m}_j$	Cost
M = 19	4,494	212 tons	$13,529
M = 15	4,494	212 tons	$13,674

$\delta = 0$, $N = 19$ crews

TABLE IV

	$\Sigma\tilde{m}_j^2$	$\Sigma\tilde{m}_j$	Cost
δ = 0	747	56.7 tons	$13,874
δ = 1	14,015	68.6 tons	$13,212

$M = 19$ trucks, $N = 21$ crews

TABLE V

	$\Sigma\tilde{m}_j^2$	$\Sigma\tilde{m}_j$	Cost
δ = 0	4,494	212 tons	$13,529
δ = 1	17,678	222 tons	$12,781

$M = 19$ trucks, $N = 19$ crews

where $\mu(x)$ is the step function defined by

$$\mu(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$$

$$\delta_{l,j} = \begin{cases} 1, & \text{if } l = j \\ 0, & \text{otherwise.} \end{cases}$$

In case one decides to use (8) in place of (2), then (11) and (12) are replaced by

$$\frac{\partial f}{\partial n_{kl}} = \sum_{j=l}^{7} \tilde{m}_j \frac{\partial \tilde{m}_j}{\partial n_{kl}} + \delta c_{kl} \tag{13}$$

with

$$\frac{\partial \tilde{m}_0}{\partial n_{kl}} = 0 \tag{14a}$$

$$\frac{\partial \tilde{m}_j}{\partial n_{kl}} = \left[\frac{\partial m_{j-1}}{\partial n_{kl}} - p_{kl}\delta_{l,j}\right] \mu(j - l)$$

$$= \begin{cases} -p_{kl}, & \text{for } j = l \\ -\mu(j - l)p_{kl} \prod_{i=l}^{i-1} \mu(\tilde{m}_i), & \text{for } j \neq l. \end{cases} \tag{14b}$$

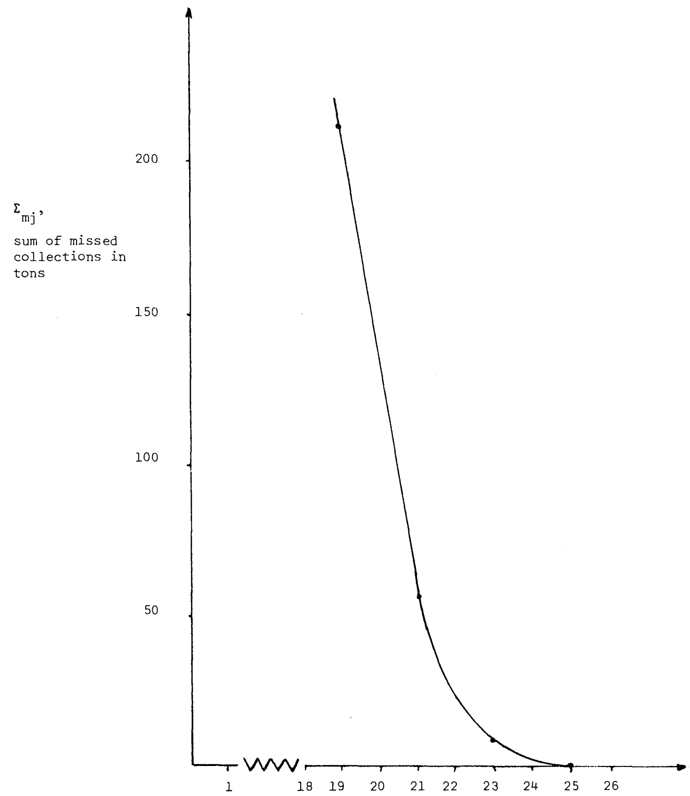

Fig. 1. Sum of missed collections versus N, for $\delta = 0$, $M = 19$.

The computer program generated results over a variety of parameter values. Some typical results are shown in Tables II–V and in Fig. 1. One can draw some general conclusions from these results. First of all, the problem of missed collection is insensitive to the number of trucks available (as long as $N \leq 2M$), although the lack of trucks does force more crews to be run during the night shift at a slight increase in overall cost. Also, as N becomes smaller it costs increasingly more in order to achieve even modest reductions in the missed collections. Indeed, the smaller values of N cause a backlog of uncollected refuse from previous days to develop. This inventory of uncollected refuse grows exponentially as N decreases. What this suggests is that in order to match workload, the emphasis should be put on increasing the availability of manpower rather than increasing the number of trucks.

Acknowledgment

The authors would like to thank Miss Barbara Stevens and J. Mechling of the Environmental Protection Administration of New York City for first bringing this problem to their attention. They also wish to thank J. Indusi of the State University of New York, Stony Brook, for programming the numerical computations and for a number of useful suggestions concerning the model itself.

Stanley M. Altman
E. J. Beltrami
S. S. Rappaport
G. K. Schoepfle
State Univ. of New York
Stony Brook, N.Y.

References

[1] E. Ignall, P. Kolesar, and W. Walker, "Linear programming models of crew assignments for refuse collection," New York City Rand Inst. Doc. D-20497, Aug. 1970.
[2] N. B. Heller, "Proportional rotating schedules," Ph.D. dissertation, Graduate School of Arts and Sci., Univ. Pennsylvania, Philadelphia, 1969.
[3] H. J. Kelley, W. Denham, J. Johnson, and P. Wheatley, "An accelerated gradient method for parameter optimization with nonlinear constraints," *J. Astronautical Sci.*, vol. 13, 1966, pp. 166–169.