U. S. AIR FORCE

PROJECT RAND

RESEARCH MEMORANDUM

A DYNAMIC, SINGLE-ITEM, MULTI-ECHELON INVENTORY MODEL

A. J. Clark

RM-2297

December 8, 1958

ASTIA Document Number AD 209538

Assigned to ______________________

This is a working paper. It may be expanded, modified, or withdrawn at any time. The views, conclusions, and recommendations expressed herein do not necessarily reflect the official views or policies of the United States Air Force.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

This study describes a mathematical model constructed as a computational aid for planners seeking to establish least-cost order and supply policies within a complex and dynamic structure. It is intended only as a useful tool for research in inventory theory; there are no policy implications.

Most past work in inventory theory has sought to derive least-cost ordering and stocking policies for a relatively simple situation: a given facility stocking a single given item. Unlimited resupply has usually been assumed, subject, of course, to ordering costs and delays. Recent work has incorporated demand uncertainty expressed in probability distributions with parameters as functions of time to allow for a changing future. These are dynamic, single-item, single-echelon inventory models which, while useful, have limited utility in complex real-world supply systems -- that of the Air Force, for example. Past attempts to derive an integrated package of decisions for such systems have faltered not for lack of a conceptual approach, but rather because of the numerous dimensions involved.

This study uses a typical Air Force supply system as its frame of reference, in which many bases stock a particular item, a depot replenishes base stocks, and a factory and a repair facility resupply the depot with new and with repaired items. The main

decisions to be made at any given time are how much to ship to each base, how much to repair, and how much to procure from the factory. The dynamic element in the problem is accentuated by bases phasing into and out of operation at different times, and experiencing changing failure patterns while in operation. This is, then, a dynamic, single-item, multi-echelon inventory model; that is, a model which integrates stockage policies for individual activities so as to minimize inventory costs for the logistics system as a whole.

CONTENTS

A DYNAMIC, SINGLE-ITEM, MULTI-ECHELON INVENTORY MODEL

I. INTRODUCTION

Most work in inventory theory has been directed towards the derivation of least-cost ordering and stocking policies for a given facility that stocks a given item. It is usually assumed that unlimited resupply to the facility is available, subject, of course, to ordering costs and delays. Recent work has incorporated demand uncertainty expressed in probability distributions, with parameters as functions of time to allow for a changing future.[1] Models of this kind may be called dynamic, single-item, single-echelon inventory models.[2]

Real-world supply systems are often more complex, of course, since many contain several supply echelons and regenerate failed parts through repair. In the Air Force, for example, there may be many operational bases which stock a particular item, and a

---

[1]The recent book by K.J. Arrow, S. Karlin, and H. Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958, represents a useful collection of work in this field.

[2]The word "echelon" is used rather than "level" to avoid confusion with stock levels, and rather than "stage" -- although the word has been used in this context -- because the term "multi-stage problems" has recently been used to designate problems in which time is divided into discrete decision-intervals or stages.

depot which replenishes base stocks in accordance with base ordering policies. The depot, in turn, is resupplied either from a factory or from a repair facility that repairs those failed items it economically can and scraps the rest. This structure is illustrated in Fig. 1.

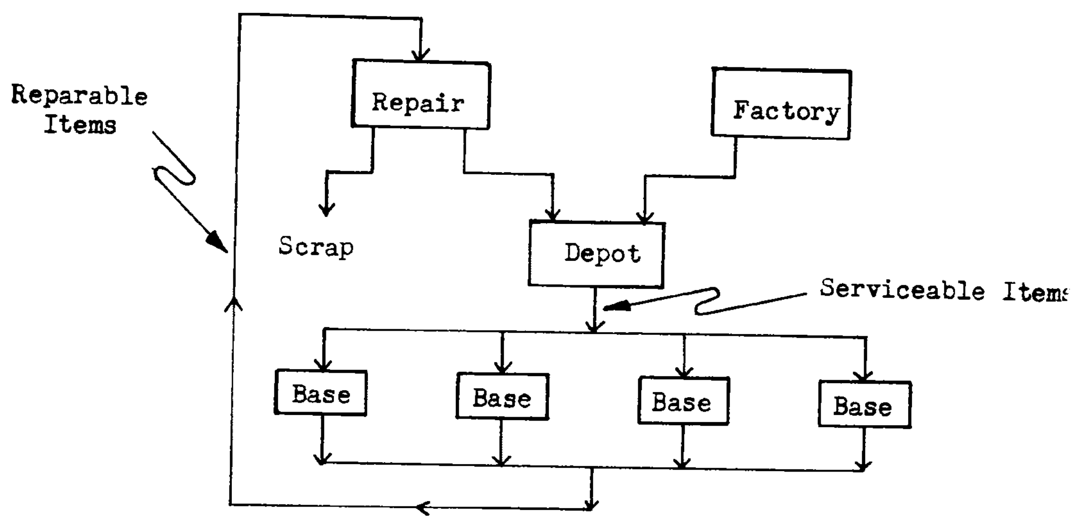

Fig. 1 -- Air Force Supply System

While Fig. 1 is representative of the supply structure for many items in the Air Force inventory, the structure for other items may be much more complex. For example, some failed items may be repaired at the base itself and turned back into the serviceable stock. There also may be several depots, each supporting a complex of bases; more than one depot-level repair facility; and even more than one factory producing the item. Furthermore, there may be a wide variety of bases or "customers" with different usage patterns and maintenance capability.

It is evident that supply decisions in such complex systems should be closely interrelated. Within the structure of Fig. 1, the main decisions at any given time are how much to ship to each base, how much to repair, and how much to procure from the factory. The dynamic element in the problem is accentuated by the various bases phasing into and out of operation at different times, and experiencing changing failure patterns while in operation. In this environment, a particular decision, such as whether or not to resupply a given base, should not be made in a vacuum ignoring the other elements of the system. Instead, the decision should take a large number of relevant factors into account, such as respective transportation and repair delays, asset postures, cost factors, failure forecasts, etc.

The abandonment of many past attempts to derive such an integrated package of decisions for complex systems has been more often due to the numerous dimensions involved, rather than to the lack of a conceptual approach. The purpose of this paper, however, is to show that with certain restrictive assumptions, a practical solution to this problem is possible.

For expository purposes, the model described in this paper will be confined to the particular supply structure portrayed in Fig. 1, although it is not a conceptual requirement to do so. A fully dynamic system will be considered with a number of bases involved, each phasing in and out of operation at different times, and each with different random failure patterns changing over time while operational. Since the model is confined to a single item of supply, it is described as a dynamic, single-item, multi-echelon model.

Although the model uses concepts of the dynamic programming technique, no claim is made for rigour, proofs of any kind, optimality, or even conformity of symbolism. The only claim made is that it incorporates most of the relevant factors, is computable, and seems to make sense for the few cases so far considered.

II. THE SINGLE-ECHELON PROBLEM

Before considering the multi-echelon problem, it is convenient to consider the case of a single echelon and to express the dynamic programming procedure in vector notation.

If a finite future is divided into n equal decision intervals, then certain costs may be defined for the j-th interval. Assuming holding cost as a linear homogeneous function of the stock balance, a holding cost vector, H_j, is defined as follows:

$$H_j: \; h_{j,x} = h_j x \quad (x>0)$$
$$= 0, \quad (x \leq 0),$$

where $h_{j,x}$ is the xth element of the vector H_j, x is the balance at the start of the period, and h_j is the per-unit holding cost. For simplicity, the holding cost is assessed as if the stock balance at the start of the period remains on hand throughout the period.

Assuming depletion cost as a linear homogeneous function of the number of unsatisfied demands, a depletion cost vector, D_j, is defined as follows:

$$D_j: \; d_{j,x} = 0 \quad (x \geq 0)$$
$$= d_j |x| \quad (x<0),$$

where $d_{j,x}$ is the xth element of the vector D_j, d_j is the per-unit depletion cost, and x is the balance at the end of the period, with negative balances representing stockouts during the period. Since the stockout cost is taken as a linear function of the number of stockouts, an equal duration of each stockout is implied which, in turn, implies replacing stockouts by expedite-orders at the time they occur; shortage costs associated with the constant time-delay for the expedited order are reflected in the per-unit cost, d_j. In this case, a negative stock balance has no reality, and is merely a convenient device for reflecting the number of temporary stockouts during the period.

If demands occur during the j-th period according to the probability distribution, $g_j(x)$, then a matrix P_j is defined as follows:

(1) $$P_j: p_{j,x,y} = g_j(x - y) \quad (x \geq y)$$
$$= 0, \qquad (x < y)$$

where $p_{j,x,y}$ is the matrix element in column x and row y.

Let the vector C_j represent, for all possible stock balances, the total expected costs from the start of period j through period n after applying an ordering policy, f, at the start of period j. The dynamic programming recursion formula is then as follows, assuming no delay in receiving amounts ordered:

(2) $$C_j = f(H_j + P_j D_j + \alpha P_j C_{j+1}) = f(T_j),$$

where α is a discount factor.

In multiplying by the matrix P_j, row x of the matrix corresponds to the balance on hand at the end of the period and hence to the x-th element of vectors D_j and C_{j+1}; column y corresponds to possible demands during the period.

If items may be ordered at a per-unit cost of v_j (where ordering costs are assumed to be a linear function of the amount ordered), and there is no fixed ordering cost, then the least-cost ordering policy, f, is given as follows:

$$(3) \qquad \begin{aligned} f(t_{j,x}) &= t_{j,x} \qquad (x \geq S_j) \\ &= t_{j,S_j} + v_j(S_j - x) \qquad (x < S_j) \end{aligned}$$

where $t_{j,x}$ = x-th element of vector, T_j, and

$$S_j = \max_x \left\{ x: \left[t_{j,x} - t_{j,x-1} \right] \geq v_j \right\}.$$

Thus, if there are fewer than S_j on hand at the start of period j, enough are ordered to bring the balance to S_j.

The procedure implied by equations (2) and (3) may be illustrated by plotting the cost vectors T_j and $f(T_j)$ as functions of x, the amount on hand at the start of the period. These functions are shown in Fig. 2 as continuous functions for illustrative purposes; they are actually confined to integral values of x.

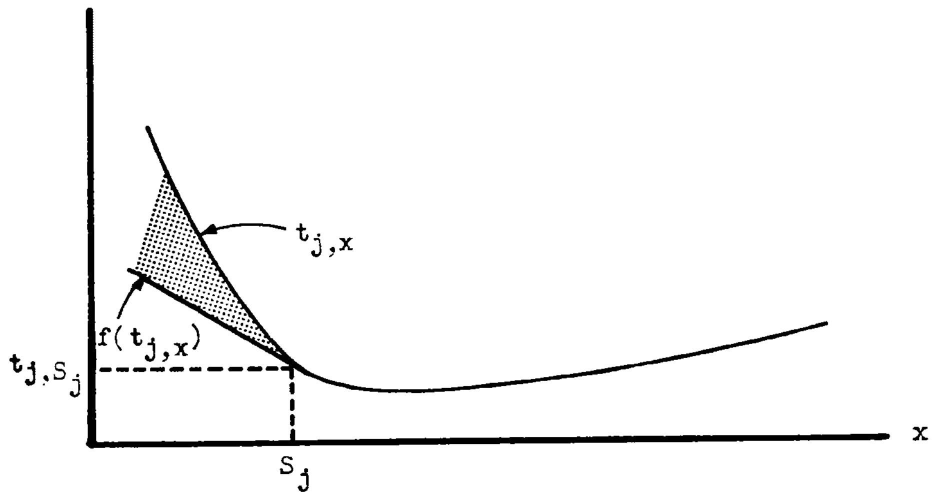

Fig. 2 -- Expected Costs as Functions of Stock Balance -- Without Fixed Order Cost

The function T_j is dominated by the depletion costs for low values of x and by holding costs for higher values of x. The level S_j is the value of x for which the slope of the T_j function is equal to the ordering cost, v_j. The functions T_j and $f(T_j)$ are the same for $x \geq S_j$, since S_j is the amount on hand after ordering. For $x < S_j$, $f(T_j)$ is merely a straight line with a slope of v_j, assuming linear ordering costs. Note that the shaded area in the diagram represents savings derived by ordering, or extra costs if resupply cannot be obtained. These costs will have implications later in the discussion.

If there is a fixed ordering cost, u_j, then the optimal policy f, is given as follows:

$$(4)\qquad f(t_{j,x}) = t_{j,x} \qquad (x > s_j)$$

$$= t_{j,S_j} + v_j(S_j - x) + u_j, \qquad (x \leq s_j)$$

where S_j is as previously defined, and

$$s_j = \max_x \left\{ x: \left[t_{j,x} - (t_{j,x} + v_j \; (S_j - x) \right] \geq u_j \right\}.$$

In this case, if the balance is at or below s_j, the reorder point, enough is ordered to raise the balance to S_j.

This case may also be illustrated by plotting T_j and $f(T_j)$ as functions of x.

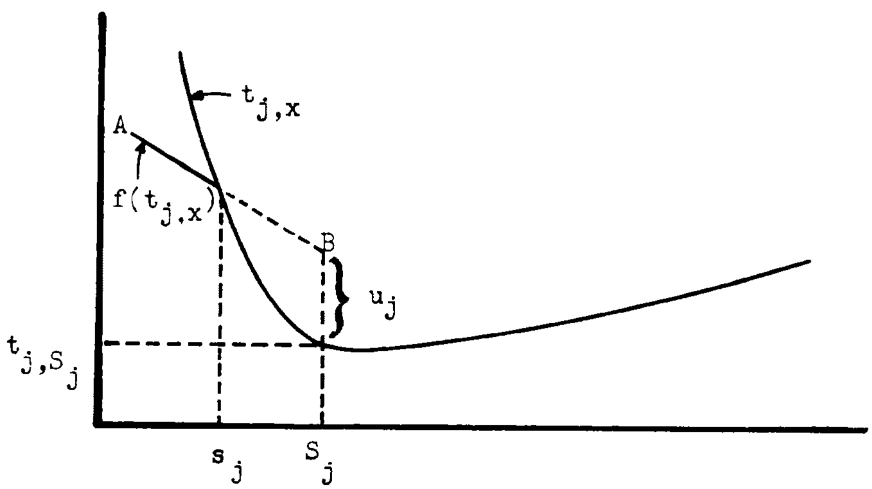

Fig. 3 -- Expected Costs as Functions of Stock Balance -- With Fixed Order Cost

As shown in Fig. 3, the level S_j again is the value of x for which the slope of the T_j function is equal to v_j. The ordering-cost line, AB, with slope of v_j, is raised an amount u_j, the fixed order-cost, above t_{j,S_j}; its intercept with $t_{j,x}$ yields s_j, the reorder point. The functions $f(t_{j,x})$ and $t_{j,x}$ are the same for $x > s_j$, since no orders are placed for balances above the reorder point.

If there is a one-period delay in receiving an order after it is initiated, then equation (1) becomes:

$$(5) \qquad C_j = H_j + P_j D_j + f(\alpha P_j C_{j+1}),$$

where the function f is as previously defined.

The recursion process implied by equation (1) or (2) may be accomplished by setting $C_{n+1} \equiv 0$ under the assumption that the demand for the item ceases after period n. Repeated application of the equation then yields the stock levels S_j and s_j for j = 0 to n.

The above assumptions of linearity in the holding, depletion, and ordering costs are not necessary restrictions for a least-cost solution. Linearity is assumed for expository purposes and also because one is usually degraded to this level when trying to get the costs in an existing supply system. Some of the effects of non-linearity can be readily ascertained from the diagrams in Figs. 2 and 3.[1]

---

[1] See Arrow, Karlin, and Scarf, op. cit., for a complete discussion of this subject.

III. THE MULTI-ECHELON PROBLEM

The multi-echelon problem will be considered as an interconnected set of single-echelon problems. This, however, involves a rather peculiar definition of "echelon" and certain assumptions developed below. Before grappling with the main issue, then, let us take a side excursion into the peculiar definition of "echelon," once more focusing our attention on the Air Force supply organization shown in Fig. 1.

Let us again divide the future into n equal decision-intervals. All routine decisions will be made at the start of these time intervals, and thus all supply lags must be expressed as integral numbers of such intervals. Thus, we define the following:

T_1 = number of intervals in the repair cycle (time between ordering an item into repair and receiving it at the depot).

T_2 = number of intervals in the production lead-time (time between initiating an order on the factory and receiving the goods at the depot).

In these definitions we impose the restriction that $1 \leq T_1 < T_2$. The production lead-time may be viewed as consisting of T_2 discrete stages of production, with items being "ordered" from one stage to the next. Similarly, the repair cycle may be considered as T_1 stages of repair.

Now let us divide all the stock in the system into the following sets:

B = serviceable stock on hand at bases;

S = serviceable stock on hand at the depot;

Q_a = reparable stock that has a stage of repair to be accomplished before becoming serviceable. $(1 \leq a \leq T_1)$;

R_b = in-production stock that has b stages of production yet to be accomplished. $(1 \leq b < T_2)$

It is assumed that this split-up of the stock has occurred at the start of the decision interval, after all deliveries from previous orders have been received but before any new ordering decisions have been made. Thus, there is no "in-between" stock; e.g., stock in transit to bases or stock half-way through a production stage.

Echelon k is defined as containing $\mathcal{E}_k$ stock as follows:

$$\mathcal{E}_{-1} = B$$

$$\mathcal{E}_0 = B + S$$

$$\mathcal{E}_k = B + S + \sum_{a=1}^{k} Q_a + \sum_{b=1}^{k} R_b \qquad (1 \leq k \leq T_1)$$

$$= B + S + \sum_{a=1}^{T_1} Q_a + \sum_{b=1}^{k} R_b \qquad (T_1 < k < T_2)$$

Thus, echelon -1 consists of all serviceable stock at bases; echelon 0 contains all serviceable stock in the system; echelon 1 contains all serviceable stock plus stock that has only the last stage of production or repair yet to be accomplished; and so forth. In general, stock in echelon k is augmented by stock from echelon k + 1. Echelon T_1-1 has an unlimited source of supply, since, at a price, any amount may be ordered into the first stage of production. Note that there is no echelon T_1, since items ordered into the first stage of production are delivered into echelon T_1-1 when the stock split-up defining the echelons is made.[1]

Now let us consider, in a general way, costs associated with stock in these various echelons. The routine per-unit ordering cost for echelon k is merely the cost of obtaining an item from echelon k + 1. For echelon -1 (base stock) this is the cost of obtaining an item from the depot or from another base, including

[1]Actually, set notation should be used in these definitions, since we are dealing with sets of stock rather than numerical quantities. We use this notation, however, since no set algebra is contemplated.

transportation and other processing costs. The cost of obtaining an item from another base will be assumed the same as that of obtaining it from the depot; for items repaired at the depot, however, there is usually no necessity for transshipments between bases except possibly to cover stockouts. For echelon 0 (system serviceable stock), augmentation is achieved by moving items through the last stage of production or repair; the associated per-unit ordering cost is then the cost of doing so. Similarly, the ordering cost for echelon 1 is the cost of accomplishing the next-to-last stage of production or repair. In general, the per-unit ordering cost for echelon k ($k \geq 0$) is the cost of completing stage T_2-k of the production process ($k < T_2$) or stage T_1-k of the repair process ($k < T_1$). These costs may be obtained as the marginal values of cumulative production or repair cost functions, as illustrated in Fig. 4.

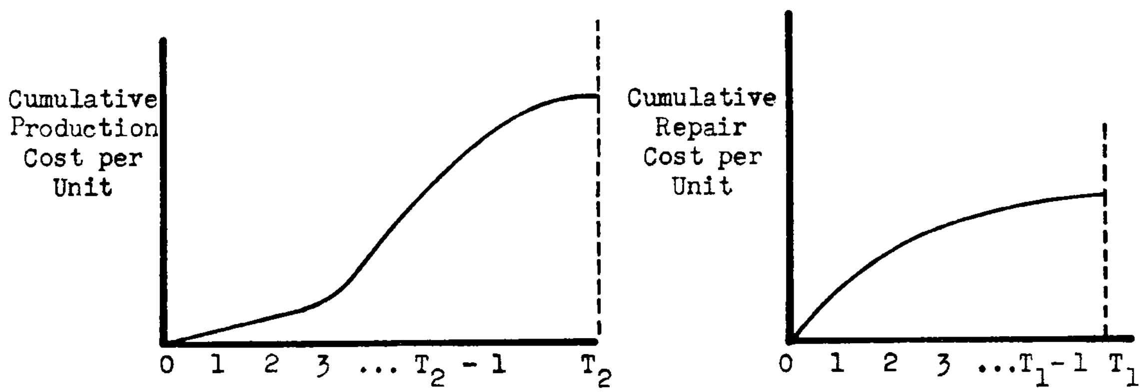

Fig. 4 -- Cumulative Production and Repair Costs as Functions of Processing Times

Costs for the initial periods of the lead time are usually those of paperwork and the acquisition of raw material and components. The main fabrication costs typically occur in the middle periods. Costs at the end periods may be largely those of inspection, packaging, and transportation. (The transportation time from the factory to the depot is included in the total lead time.) The cumulative per-unit cost at T_1 represents the delivered price of the item.

For reasons to be mentioned later, the model must assume the ordering costs from repair to be the same as those from production, for a given echelon. For example, the model assumes that echelon 0 (serviceable stock) may order items through the last stage of repair at the same cost as through the last stage of production. Under this assumption it makes no difference, from a cost point of view, which resupply source is used.

The model also cannot accept fixed ordering costs for any echelon but the highest, T_2-1. For this echelon, the fixed ordering cost includes costs of processing the contract and costs of tooling up for production.

Holding costs are charged to stock in each echelon only to the extent the items have increased in value by being moved from the next higher echelon. They are, then, the extra costs of holding stock in a particular echelon rather than in the one above it. This definition of holding cost stems from the assumption that any given echelon includes all stock in lower echelons; if holding costs for each echelon were assessed to the full value of the stock, they would then be overstated to the extent of the stock's being included in more than one echelon. If, in general, holding costs per unit are stated as a percentage of the unit's value, then the holding cost assigned to each echelon is a percentage of the ordering cost, since this represents the marginal worth of the item at that stage. The percentage charged against this marginal value may vary among echelons because elements of the holding cost, such as costs of engineering obsolescence, depend on what state of completion the item has attained. Other elements, such as interest on the investment, may remain the same for all echelons.

Depletion costs are charged to a particular echelon when demands occur in the system after the echelon's stock has been exhausted. These demands actually represent stockouts at bases, since by definition each echelon includes all stock positioned at bases. If, for example, a stockout occurs for echelon 0, a base demand has occurred after all serviceable stock in the system has been exhausted. The stockout cost for each echelon is determined on the basis of expediting from the next higher echelon, under the assumption that stock is there. It includes the direct costs of the expedite action plus the cost of inconvenience to the base for a further delay. From an over-all point of view, a base stockout will accumulate these individual echelon stockout costs up to the echelon that has assets to satisfy the demand. According to how high upstream one has to go, expedite production or repair costs increase, and more and more delays are encountered in satisfying the demand.

In addition to this natural depletion cost, each echelon other than echelon -1 (base stock) incurs a cost penalty for failure to raise the next lower echelon to its proper level. If, for example, the economical level for echelon k at period j is $S_{j,k}$, then costs are assessed against echelon k + 1 if it has less than $S_{j,k}$ assets. If there are more than $S_{j,k}$ items in echelon k + 1, then echelon k can be raised to its proper level by ordering, regardless of how many items are in echelon k when the decision is made. In this case no penalty is incurred by echelon k + 1. In general, the depletion cost assigned to echelon k + 1 is the difference between costs incurred by echelon k if it can obtain resupply and if it cannot.

Let us now turn our attention to echelon -1, the aggregate of base stocks. If we have m bases, the following inputs are assumed:

$H_{j,i}$ = holding cost vector for period j and base i $(1 \leq j \leq n,\ 1 \leq i \leq m)$

$D_{j,i}$ = depletion cost vector for period j and base i

$v_{j,i}$ = per unit ordering cost for period j and base i

$g_{j,i}(x)$ = failure probability distribution for period j and base i

ρ_i = pipeline time (depot to base) for base i.

As previously mentioned, holding costs for bases include only the extra costs involved in holding stock at the bases rather than the depot, and depletion costs are estimated assuming availability at the depot in the event of a base stockout.

The phase-in and phase-out of bases are reflected in the assignment of probability distributions, where $g_{j,i}(0) = 1$ for periods when base i is not in operation. Obtaining the distributions $g_{j,i}$ based on failure estimates, actual failure data, program data, and other factors, is quite a problem in itself; this, however, is considered as outside the scope of this discussion. Given these distributions, matrices $P_{j,i}$ as defined in (1) may be constructed.

The pipeline time, ρ_i, must be either 0 or 1 period for base i. This is an unavoidable consequence of the technique and results from the requirement for a standard decision-interval for all bases, in order to tie in with the higher-echelon decision-intervals.

With the above inputs, levels $S_{j,i}$ may be obtained by using m equations of the type (2) or (5), depending on the value of ρ_k. These levels essentially represent the maximum economical amount to be on hand or due in after ordering, assuming availability of serviceable stock from elsewhere in the system. If all or part of a base's order cannot be satisfied, then the base will incur more costs than assumed by the ordering policy. These extra costs,

$B_{j,i}$, for base i at period j are as follows:

(6) $B_{j,i} = T_{j,i} - f(T_{j,i})$, if equation (1) is used;

$= \alpha P_{j,i}\, C_{j+1,i} - f(\alpha P_{j,i}\, C_{j+1,i})$ if equation (2) is used.

These costs, $B_{j,i}$, represent the difference between base costs incurred if full delivery cannot be made and costs resulting if the balance can be raised to $S_{j,i}$. They therefore reflect what it costs the base to be below its level, or conversely, how much would be saved if an additional unit could be added to the base stock. They are used to obtain depletion costs chargeable to echelon 0 (system serviceable stock) for having insufficient stock to satisfy all base orders.

If there are $\sigma_j = \sum_{i=1}^{m} S_{j,i}$ <u>or</u> more serviceable items in the system at period j, then no penalty should be charged to echelon 0 since all bases can be raised to their levels by redistribution. If there are fewer than σ_i serviceables, then one base or more must be below its level after redistribution. For example, if the system is one unit short of achieving the levels at all bases, then the base that suffers the shortage should be the base for which least additional cost would result; this base is identified as having the least non-zero element in the set of elements comprising the $B_{j,i}$ vectors. If the system is two units short, then the base to suffer the second shortage is the one with the second smallest non-zero element in the set of $B_{j,i}$ elements.

The above argument may be generalized by constructing the vector $\Delta_{j,-1}$ as follows:

$$\text{(7)} \qquad \Delta_{j,-1} : \delta_{j,-1,x} = 0 \qquad (x \geq \sigma_j)$$

$$\delta_{j,-1,x} = b_{j,i,y} \neq 0,$$

$$\text{such that } \delta_{j,-1,x+1} \geq \delta_{j,-1,x} \qquad (x < \sigma_j),$$

where $\delta_{j,-1,x}$ is the x-th element of the vector $\Delta_{j,-1}$ and $b_{j,i,y}$ is the y-th element of vector $B_{j,i}$. (The index -1 indicates that these costs are obtained from echelon -1.) The elements of the vector $\Delta_{j,-1}$ then increase in value as the amount of serviceable stock, x, decreases. For stock positions below σ_j, these values constitute a depletion cost chargeable to echelon 0.

Let us now assume the following factors, applicable to the higher echelons ($0 \leq k < T_2$) as being given:

$H_{j,k}$ = holding-cost vector for period j and echelon k

$D_{j,k}$ = natural depletion-cost vector for period j and echelon k

$v_{j,k}$ = per-unit ordering cost for period j and echelon k

u_j = fixed production-setup cost for period j (applied to echelon T_2-1 only)

$G_{j,k}(x)$ = demand probability distribution for echelon k.

If, for the higher echelons, the holding cost is a linear function of the stock balance and the additional value attained by ordering into the echelon, the holding cost vector $H_{j,k}$ may be defined as follows:

$$H_{j,k}:\ h_{j,k,x} = 0 \qquad (x \leq 0)$$

$$= w_k\ v_{j,k}\ x \qquad (x > 0)$$

where $h_{j,k,x}$ is the x-th element of vector $H_{j,k}$, x is the stock balance, and w_k is the per-unit holding cost expressed as a fraction of the added per-unit worth, $v_{j,k}$.

The probability distribution $G_{j,k}(x)$, for period j and echelon k, represents losses to the echelon during the period. For $0 \leq k < T_1$, losses include all failures at all bases during the period. For $T_1 \leq k < T_2$, losses include only items condemned as not economically reparable . For convenience, these distributions may be obtained from the base distribution as follows: if $g_{j,i}(x)$, with mean = $\bar{g}_{j,i}$, represents demands at base i during period j, then the mean of $G_{j,k}(x)$ for $0 \leq k < T_1$ is given as $\sum_{i=1}^{m} \bar{g}_{j,i} = M_j$. If the system condemnation rate for period j is given by c_j, then the mean of $G_{j,k}(x)$ for $T_2 \leq k < T_1$, is given by $c_j M_j$. The type of distribution assigned to the higher echelons, being in theory a convolution of the base distributions, depends upon the type of distributions used at bases. If Poisson demands are used for bases, then Poisson demands may be assumed for the higher echelons. For other kinds of base distributions, one usually is rather arbitrary in assigning distributions to the higher echelons, since actually computing the convolution is impractical.

With these inputs, and treating each echelon as a single-echelon problem, the dynamic programming recursion formula becomes:

$$(8) \qquad C_{j,k} = H_{j,k} + \Delta_{j,k-1} + P_{j,k} D_{j,k} + f(\delta P_{j,k} C_{j+1,k}),$$

where $0 \leq k < T_2$, f is defined as in the single-echelon problem, and $\Delta_{j,k}$ is given as follows:

$$\Delta_{j,k} = \delta P_{j,k} C_{j+1,k} - f(\delta P_{j,k} C_{j+1,k}); \qquad (0 \leq k < T_2-1)$$

$\Delta_{j,-1}$ is defined by equation (7).

The solution of (8) yields levels $S_{j,k}$ for each period and echelon, and reorder points s_j for echelon T_2-1.

The levels obtained from equation (8), together with the base levels, provide a basis for all supply, repair, and production decisions at a designated period j, for which the system asset posture is known. Each echelon is raised to its indicated level, to the extent of stock availability, by ordering from the next higher echelon. The level for the higher echelon is frequently less than that for the lower echelon. This means that the lower echelon cannot be raised to its level since it is uneconomical for the higher echelon to provide the stock.

If more than one base needs resupply and there are not enough assets to raise all bases to their indicated levels, then the available stock may be rationed in accordance with the marginal savings afforded the bases by receiving additional units. By using the cost vectors from equation (7), the first unit of available stock should go to the base for which highest savings are obtained; the second unit to the base for which the next highest savings result; (which may be the same base that receives the first unit) and so forth, to the extent of stock availability. In principle, since the model assumes that shipments from base to base cost the same as from depot to base, base stock should be redistributed in each period so that each base ends up with about the same level of protection. In practice, however, unless the item is almost always repaired at base level, sufficient items become available each period from depot-level repair or production so that base protection can be equalized through the normal echelon flow, without inter-base shipments.

If any of the input factors change significantly, levels should be recomputed to reflect the new data. Normally the system may be allowed to adjust to the new levels through attrition. For drastic changes, decisions outside the model may be indicated.

IV. THE RESTRICTIONS

The multi-echelon problem is conceptually a multi-dimensional dynamic programming problem. Such problems, however, are computationally unfeasible for more than two or three dimensions even with electronic computers. To avoid this difficulty, the problem was converted to a set of interconnected one-dimensional problems by using artifices and imposing certain restrictive assumptions. The main artifice employed was the definition of echelon. With a little experimentation, one can be readily convinced that any other definition leads to multi-dimensional problems. One of the restrictive assumptions was that a fixed ordering cost cannot be permitted for bases without increasing the dimensions of the dynamic programming problems involved. With a fixed ordering cost, reorder points would result, and the cost vector assigned to echelon 0 for failure to meet base orders would depend on the particular stock distribution at the decision time. This may be illustrated by a two-base system: Base 1 with levels s_1, S_1, and Base 2 with levels s_2, S_2. Consider the case in which there are B serviceable items in the system, with $s_1 + s_2 < B < S_1 + S_2$. If Base 1 has an amount $B_1 > s_1$ on hand and Base 2 has an amount $B_2 > s_2$, where $B_1 + B_2 = B$, then no penalty should be charged to echelon 0, since no resupply would be indicated. On the other hand, if $B_1 \leq s_1$ and $B_2 > s_2$, still with $B_1 + B_2 = B$, then a penalty should be applied to echelon 0 since the indicated resupply for Base 1 could not be made. Therefore, the penalty depends on the values of B_1 and B_2 and hence, the ordering policy for echelon 0. If there is no fixed ordering cost, then only the levels S_1 and S_2 obtain, and a single, unambiguous penalty for 0 may be assigned to a balance of B.

The inability to accept a fixed base order cost more or less confines the model to items with a reasonably high value or a low usage rate where one can assert a priori that it pays to order one at a time. In this case, any fixed order cost can be incorporated in the per-unit order cost without incurring too much error in the ordering policy. For essentially the same reasons mentioned above, fixed ordering costs cannot be accepted for any echelon other than the highest. The highest echelon may have this cost because it has an unlimited source of supply. The inability to accept the fixed ordering costs for the other higher echelons is of minor importance, however, since if they exist at all they are probably small compared to the variable ordering costs. Also, the division of the repair cycle and production lead times into discrete stages was artificial; in reality, once a batch of items is scheduled into repair or production, it flows through to completion without halting between processes to allow a different-size batch to accumulate before continuing. This procedure, in effect, implies that a fixed order cost does not normally exist for these echelons. One exception however, is scheduling those items for repair which can be done most economically in accumulated batches. Current Air Force policies require that higher-valued items, which are characterized by low usage rates, be repaired as generated -- usually one at a time on a job-shop basis. This policy again implies that the fixed order cost is relatively small or that it can be subsumed in the variable ordering costs.

Another restrictive assumption is assuming ordering costs from repair to be the same as from production. If different ordering costs were allowed, ordering decisions would hinge on the independent repair and production asset positions -- a two-dimensional problem. If the same ordering costs are assumed, no distinction is made between assets in repair and in production; they may then be lumped together to form a one-dimensional problem. The implications of this assumption are not fully known; however, the few cases so far studied seem to indicate that the results of the model may be relatively insensitive to the shape of the production-cost function (see Fig. 4). If this is true, then the costs of ordering from production may be taken the same as costs of ordering from repair, without incurring too much error. Also, for items that are almost always reparable, such as higher-valued items, results indicate that for a phase-in, phase-out type program, production dominates as a source of supply during phase-in, with repair taking over during the peak program and the phase-out. This dichotomy has relatively little overlap and permits the assignment of actual production costs during phase-in periods with a switch to actual repair costs thereafter.

In general, the assumptions of the model are such that no great errors result for higher-valued items. Also, the rather extensive input requirements, along with the cost of computation, tend to prohibit applying the model to inexpensive items.

V. AN EXAMPLE

A small example will illustrate some of the ways in which the model performs. This example consists of a system with four bases, each with a one-period base-depot pipeline time; a depot; a depot-level repair facility with a two-period repair cycle; and a factory with a five-period production lead time. Although the bases begin and end operations at different times, they are alike in every respect. While in operation, the mean failure rate per period is taken the same for all periods. All cost functions are linear. The base phase-in, phase-out pattern is shown in Fig. 4. The cost and other factors are included in Table 1.

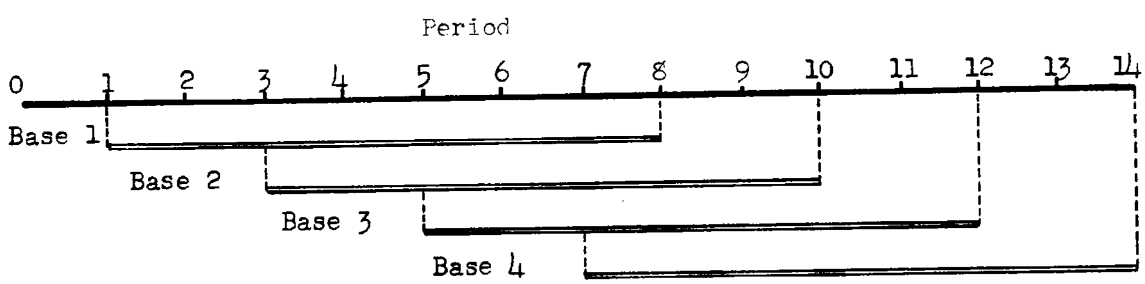

Fig. 4 -- Illustrative Base Phase-in Phase-out Pattern

Table 1

ILLUSTRATIVE COST AND OTHER FACTORS

1. Base-depot pipeline time: 1 period
2. Repair cycle: 2 periods
3. Procurement lead time: 5 periods
4. Base failures per period: Poisson, mean = 1.6
5. Condemnation rate: 50%
6. Interest rate: 0% (discount factor, $\mathcal{L}$, = 1)
7. Base order cost per unit: $50
8. Base stockout cost per unit: $2000
9. Ordering cost per unit for echelons 0-4: $50
10. Holding cost per unit per period for echelons 0-4: $2
11. Stockout cost per unit for echelon 0: $1000
12. Stockout cost per unit for echelons 1-4: $100
13. Fixed procurement cost for echelon 4: $3000

By accumulating the ordering costs shown in Table 1, it is seen that the unit price of the item, delivered to the depot, is $250. The total repair cost is $100 per unit.

With these inputs, the model provides the levels shown in Table 2.

Table 2

LEVELS COMPUTED FROM EXAMPLE INPUTS

Period	Base				Echelon					Reorder Point (Echelon 4)
	1	2	3	4	0	1	2	3	4	
-5									8	
-4								6	9	2
-3							6	9	27	12
-2						6	9	16	27	12
-1					7	9	17	21	28	11
0	9				10	17	22	28	39	30
1	11				18	22	29	34	39	31
2	11	9			21	29	34	37	38	30
3	10	11			29	34	37	36	37	30
4	9	11	9		31	39	36	35	35	28
5	8	10	11		36	37	34	33	32	25
6	7	9	11	9	32	35	31	29	28	22
7		8	10	11	30	30	27	25	24	17
8		7	9	11	23	25	22	21	16	9
9			8	10	20	20	18	14	10	1
10			7	9	14	16	11	6	6	
11				8	12	10	4	4	4	
12				7	6	3	1	1	1	
13										

Blank entries in Table 2 indicate zero-levels. Note that since period 1 is defined as the first period during which failures can occur, levels must be obtained in advance. This is accomplished from the recursion formula by using probability distribution, with the probability of zero demands equal to 1.

To obtain an idea of the effect of these levels in scheduling production, repair, and base distribution, random failures were drawn from a Poisson distribution with mean of 1.6. These failures are listed in Table 3.

Table 3

RANDOM FAILURES

Period	Base 1	Base 2	Base 3	Base 4
1	3			
2	4			
3	3	2		
4	1	1		
5	0	0	3	
6	3	2	0	
7	1	2	1	2
8		1	2	1
9		4	1	2
10			3	0
11			2	1
12				4
13				2

With these failures, the operation of the system was simulated, with all decisions made according to the levels listed in Table 2. The resulting distribution of stock, as a function of time, is shown in Fig. 5.

Fig. 5 shows the initial buildup of production to provide initial stocks at bases. The influence of the high production-setup cost produced a life-of-type buy before the first base became operational. Repair commenced when the first reparables generated

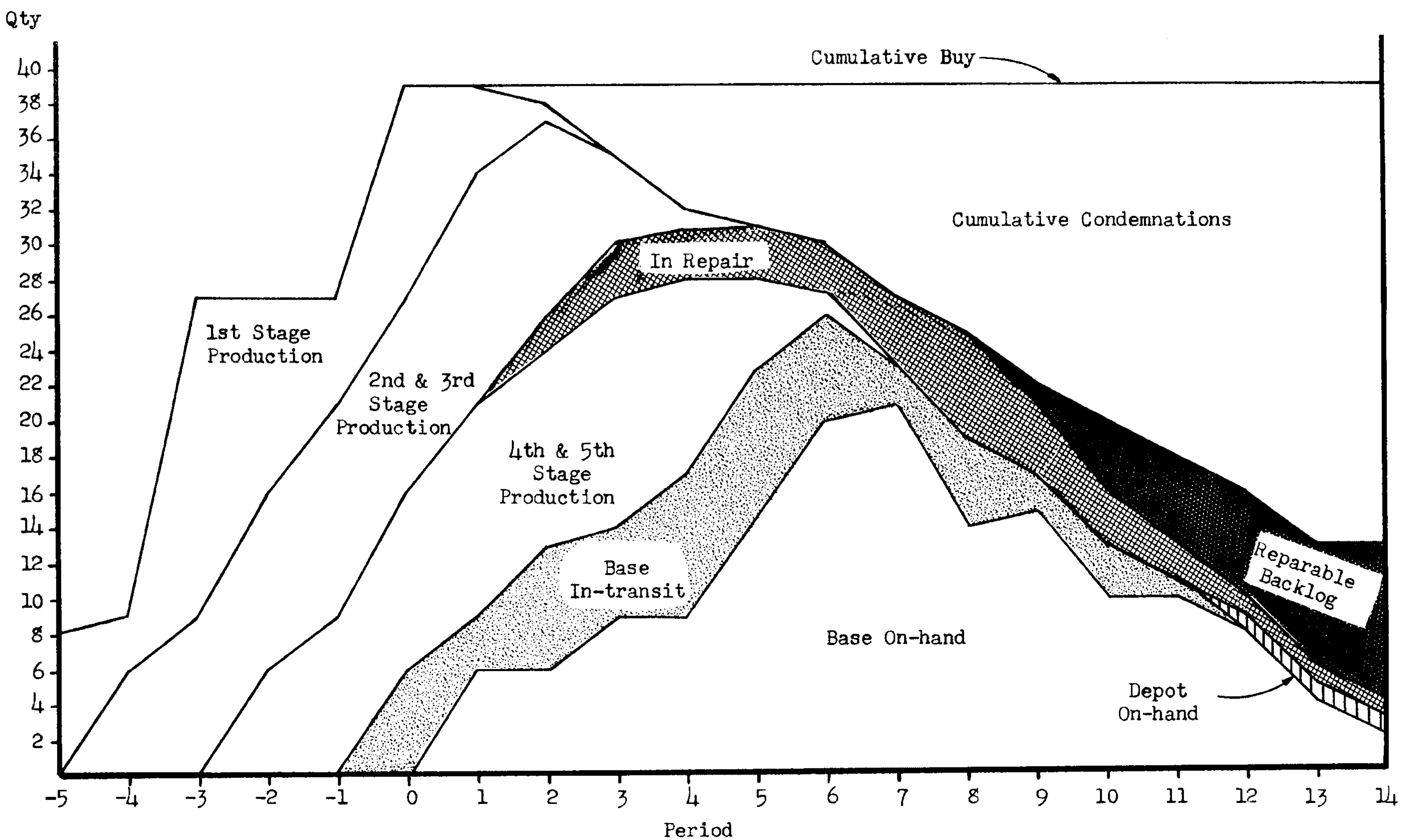

Fig. 5 -- Stock Distribution as a Function of Time

and rapidly took over from production as a source of serviceable items. Shortly after the peak of the program, more reparables generated than were needed, so that a backlog of items needing repair developed. At the end of the program, most of the stock residue was in the form of reparable items. The leftover stock, consisting of 13 units, or 1/3 of the total buy, resulted largely from buys made to afford protection against stockouts at the program peak. This protection was so large that not all the items could be attrited through condemnation before the end of the program. Needless to say, the depletion costs which caused this protection were sufficiently high that no stockouts occurred with the particular failure pattern of Table 3, despite the fact that there were 51 failures and 26 condemnations, against an expectation of 44.8 failures and 22.4 condemnations.

VI. THE MODEL AS PROGRAMMED FOR A COMPUTER

A version of this model has been programmed for the IBM 704 at The RAND Corporation. The programmed model is patterned after the organizational structure of Fig. 1 and permits a maximum of 30 bases, 78 decision intervals, 18 periods in the production lead time, and 6 periods in the repair cycle. Although the phase-in and phase-out dates may be specified for each base separately, the 30 bases are divided into a maximum of 5 groups, with all bases in a group having identical inputs (failure patterns, cost factors, pipeline times, etc.). This provides at most 5 single-echelon problems for the 30 bases, with levels computed for each group and then time-phased in accordance with the individual base phase-in schedule. A maximum of 10 is allowed for the mean failure rate for a base-period.

All probability distributions are internally computed with a choice of Poisson, Negative Binomial, or Normal designated by inputs. The type of distribution, along with the ratio of variance to mean in the case of the negative binomial or normal, may be designated differently for the different base groups, but are assumed the same for all time periods. The means for each base group may be different for the different time intervals. Means for higher echelons are internally accumulated from base means, but the type of distribution and ratio of variance to mean, if applicable, may be separately designated.

All cost factors are assumed linear and remain the same for all time periods, except for the production-setup cost which may be different from a designated period on. This allows for higher setup costs when the end article goes out of production. The cost factors may vary from one echelon to another. Altogether, the programmed model consists of a maximum of 23 interconnected one-dimensional dynamic programming problems.

The length of time required for computation is strongly dependent upon the failure rates, being roughly a function of the square of the maximum rate. This fact suggested the use of scaling factors for higher echelons, where for example, two units were viewed as being one unit with all cost and failure factors correspondingly adjusted. This scaling produces levels that differ by about 5 per cent from levels obtained without scaling. For a scale of two, however, the time required for computation is less than half as much as when no scaling is used. With scaling, a typical problem requires 10 to 15 minutes for computation. A problem involving the full capacity of the model, with maximum allowed failure rates, would take over an hour to compute.

VII. EXTENSIONS OF THE MODEL

Several additional features may be readily inserted in the model. Repair at bases may be easily accounted for if the base repair cycle is assumed the same as the base-depot pipeline time. In this case, probability distributions used in calculating base levels represent gross failures, whereas distributions used for echelons $0 \leq k < T_1$ represent net returns to the repair depot. Distributions for echelons $T_1 \leq k < T_2$ again represent losses to the system through condemnation.

Another useful feature of the model is its adaptability to repair cycles and production lead times of different lengths, for designated time periods in the future. In this way, for example, a shorter repair cycle and/or production lead time may be assigned while the weapon is phasing in or still in production, while longer times may be assigned when the weapon goes out of production. Variations under this feature may be readily inserted, within limits.

Another feature under investigation is the inclusion of disposal policies. At some times it may pay to dispose of surplus items rather than incur holding costs. It is believed that certain kinds of disposal policies may be considered without incurring major changes in the model.

A difficult problem encountered in practice is caused by engineering changes which replace an existing item with one or more new items. Following such changes, several alternative supply decisions are possible; for example, the old item may be used until exhausted, or its use may be restricted in various ways. For some cases, the model may be applied to assets of the new item with assets of the old item included. For other cases, sets of the multi-echelon problem, interconnected via the depletion-cost functions, may result. It is believed that the technique may be adapted to these kinds of item relationships.

Conceptually, the technique may be extended to the general multi-item, multi-echelon problem. This problem, however, is at present computationally unfeasible because of the incurrence of multi-variable and/or conditional probability distributions.

It is believed that the basic technique of this model, as expressed in equation (8) may find application in other problems. One such problem may be in production scheduling, with an end item which may be an assembly of several dozen major components. The sales of the end item and losses of components during fabrication due to spoilage may be subject to uncertainty. The problem here is to determine, as a function of time, how many of each component to schedule for production. It is not known whether or not limitations on resources, such as tools and manpower, can be satisfactorily included in such a model.