A CHANCE-CONSTRAINED MODEL
FOR REAL-TIME CONTROL IN RESEARCH
AND DEVELOPMENT MANAGEMENT

by

A. Charnes and A. C. Stedry*
Introduction

The problem of decision making and control in response to new information is one which has become increasingly important as developments in electronic computers have made possible the collection of data in "real time." By "real time" data collection we refer to the recording of events as they occur (and access to the record) with sufficiently small time lags that alteration of the events may be made, conceptually, at least as they are occurring. Where electronic computers are involved (and this is the ordinary context in which this term is used) the data are recorded in some form of electronic memory device; the data inputs come increasingly from "on line" remote stations which record information at the source and transmit them without delay or human intervention directly to a centrally located processor and/or electronic memory.

A seemingly trivial statement, namely that the use of these data, rather than their existence, determines the value of "real time" data collection, seems frequently to be ignored. Until the data are translated into management decisions, of course, it is impossible to determine whether or not the existence of more and more "timely" data has actually improved the decision process. Nevertheless pure data collection schemes with no means for translating the data into action are often characterized as "information systems." This may be accounted for, at least in part, by the lack of adequate mechanisms for dealing with data as they appear.

More specifically, response to newly received data requires the revision of previously determined plans between planning periods. When substantial time lags exist between the actual implementation of a plan and the availability of information which could be used to alter the original plan, both the wisdom and practicality of significant alteration between planning periods may be questioned. On the other hand, when new data become available substantially in advance of the end of the initial planning period, the decision as to whether or not to adjust the plan and if so, how, becomes relevant.

At the simplest level, the shortening of response time lags may lead to system instability, as may be seen in elementary treatments of servo-mechanical control systems.
In order to avoid misunderstanding on this point, some clarification may be in order. In the usual optimal planning model developments it is assumed that all information relevant to the actions to be taken prior to the development of the next plan is known, at least in stochastic form, at the start of the planning period. For example, if plans are revised monthly, the strategy for the first month is assumed to be implemented as given, even though a longer—e.g., 12-month—planning horizon may be used in determining the one-month plan. The plan for the second month is determined in similar fashion, updating the model with the most recently available information and (generally) moving the horizon forward. Data gathered on the first month’s operation do not, however, affect the first month’s operation—at least in terms of the planning process—but is explicitly taken into account in formulating the next plan. Even assuming that a valid form of optimization technique is used in the planning process, the emergence of new data in the "operation" phase may (or may not) induce action not in conformity with the plan. If, in fact, adjustments are made as part of the implementation of the plan—i.e., in the "control" of the operation as distinct from "planning"—the effect of the actions taken may be to move the operational phase closer to or further from optimality as compared with following the original plan. Thus the remnant plan may also need modification to improve toward optimality. It is this control process—interim adjustments to both newly received data and to the remnant plan—which is the focus of this paper.

It should be pointed out that the adjustment process envisioned does not require instantaneous receipt of data. Rather, it is necessary only that the new data be received in sufficient time that adjustments are feasible during the remainder of the planning period. The model which we develop in this paper does not depend on instant response or information cognizance but on receipt of new facts which demand attention and can be acted upon before an entirely new plan—including resetting of objectives, policies, etc.—can be formulated. It is expected, however, that these results will have applicability

1/ See Charnes and Cooper [2], and Charnes, Cooper and Symonds [5] for exceptions and further discussion of this point.

2/ See Charnes and Cooper [4], Chapter 1.
in the computerized "real time" system as a step toward developing programmed action rules to respond to new data as they arrive.

**Resource Allocation in Research Program Management**

The specific management problem underlying the current development is that of the planning and control of research task assignments in research management. We assume that a funding organization -- i.e., a sponsor of research activity -- can affect the amount of research done in a particular area at a particular institution by the amount of funds granted for research.

There is evidence that relationships exist between expenditures on research and development and inventive output.\(^1\) For so-called fundamental or basic research, measurements of productivity have been related more frequently to organizational factors other than research expenditure.\(^2\) Intuitively, it seems reasonable, however, that research activity levels\(^3\) and the costs of sustaining these levels at particular institutions can be estimated.\(^4\)

We assume further that desired research activity levels to be supported by the granting agency or foundation can be defined,\(^5\) as can the availability of resources of the grantee to provide a certain level of research activity over a class of research areas. Broadly speaking, then, the planning problem of the research sponsor may be described as the allocation of funds so that the desired research levels are maintained at the least cost.\(^6\)

\(^1\) See Mansfield [9] and [10].
\(^3\) Various measurements have been used, including number of papers produced, numbers of research reports, papers weighted by journal quality, citations, etc.
\(^4\) E.g., by recourse to past experience to productivity measurements and funds expenditures by institution or, possibly, class of institution.
\(^5\) While grants are frequently made in response to requests for funds from research individuals or institutions, such requests are undoubtedly influenced by the funds availability and known desires of the potential grantor for research of certain types. Such requests provide data for determination of long-run desired research levels and (as will be discussed below) certain necessary adjustments in the initial desired levels but the funds allocation decision must, in the final analysis, rest with the grantor.

\(^6\) Phrasing the problem in this fashion avoids the spending of funds just because they are available. If funds are too limited to accomplish the desired levels, revision of the latter must be undertaken.
Planning Horizons and Constraints in Research Funding

A distinctive feature of management of research which substantially affects the kinds of model types which can be applied to research management problems is the possibility of the occurrence of "breakthroughs." The occurrence of the unexpected is certainly not confined to research activity so that the planning process described here would be relevant to a class of problems in which the occurrence of events of an "emergency" character is a critical factor. However, we shall outline the planning and adjustment processes in a form specific to the problem of research funding in part because of its intrinsic interest but as well to provide more substance than is possible dealing with a general class of problems.

The research "breakthrough" may be perceived as a substantial advance in knowledge which, albeit possibly the result of years of effort, is suddenly recognized. Furthermore, its occurrence supplies an immediate demand for associated research activity. Older concepts need to be revised; frequently entire sub-fields which have been based on previous theory need to be examined. The questions generated by the breakthrough will presumably lead to research of high (although possibly inestimable) value. Furthering knowledge based on the breakthrough and the immediate increase in research activity in the area of the breakthrough thus becomes of immediate high priority. The granting agency thus would want to adjust its funds to meet this preemptive requirement, cutting back, if necessary, on research in other areas.

To place the breakthrough and attendant adjustments in the framework of control, we distinguish: (1) a short-run plan; and (2) a long-run plan. The short-run plan is formulated for resource allocation for, say, a one-year period. The long run is defined over a much longer horizon--say, five or ten years.

In both the short and long run it is assumed that demands (desired research activity levels), other than those associated with breakthroughs, are known with certainty 1/ for each research area prior to the formulation

---

1/ This is an assumption made for simplicity only.
of the short-run plan. Also, resource availability -- i.e., the ability to sustain a given level of research activity in terms of men, facilities, organizational structures, etc. -- across the set of research areas is assumed to be a random variable whose distribution is known or can be estimated for each research institution which is a candidate for funds. This assumption is predicated on the notion that institutional arrangements -- e.g., departmental separations, institutional reputation in certain fields, history of grants from the subject agency and others -- delimit the amount of funds a research organization can profitably use. On the other hand, it allows variation based on changes in personnel, researcher productivity, etc., which would be expected in the kinds of activities under study.

Thus, both short- and long-run plans are formulated on the assumption that it is possible to allocate funds so as to affect the distribution of research effort among a set of (presumably related) areas at a set of institutions. It would be expected that in the long run greater institutional change is possible so that, in general, the resource availability constraints would be less severe. It should be noted that we are not assuming that the actual research activity level is unaffected by the amount of funds expended but rather that the maximum capability given the existence of funds is constrained.

Adjustments to Meet Emergency Demands

The planning algorithms used for allocation of resources in most management science models do not admit of interim adjustments to meet with initially unforeseen circumstances. While such planning models have been proposed for research management, the omnipresent possibility of breakthroughs (or other emergencies) in fundamental research suggests a more flexible model. In accord with the required interactions among planning, operations, and control discussed above, we propose that the initial short-run funding plan be formulated with the possibility of the occurrence of breakthroughs explicitly included. Further, the adjustments to the initial plan in response to the occurrence of a breakthrough should be made with reference

1/ See Brandenburg and Stedry [1] for a discussion of research distributions.
2/ Cf., e.g., Freeman [7].
to the "posture" after adjustment --- i.e., the capability to carry out the long-run objectives of the funding organization.

In summary, the process upon which the model is based involves the explicit consideration of an initial plan, a local modification in operations due to "emergencies," followed by a modification of the remnant plan---these three elements combined in an optimal manner. To clarify this process we now turn to the mathematical formulation of this problem.

Criterion and Constraints

A natural format for a model of this process is that of chance-constrained programming. Let \( b_{ij}^{(1)} \) denote the short-run requirements for activity levels in the \( j^{th} \) research area. Let \( x_{ij}^{(1)} \) be the planned activity level of research area \( j \) at facility \( i \) for the short run. Let the availability at the \( i^{th} \) facility be of the form \( a_{i}^{(1)} + \delta_{i}^{(1)} \) where \( \delta_{i}^{(1)} \) is a random variable with mean zero. Our planned short-run levels, \( x_{ij}^{(1)} \), are then constrained to minimally meet the activity level requirements and with probability at least \( \beta_{i}^{(1)} \), not to exceed the availabilities. These constraints may be written:

\[
(1.1) \quad P(\sum_{j} x_{ij}^{(1)} \leq a_{i}^{(1)} + \delta_{i}^{(1)} ) \geq \beta_{i}^{(1)} , \quad i=1, \ldots, m
\]

\[
(1.2) \quad \sum_{i} x_{ij}^{(1)} \geq b_{j}^{(1)} , \quad j=1, \ldots, n
\]

Next, we suppose an emergency occurs in the short run period. We model an emergency in the \( j^{th} \) area by means of a random variable \( \epsilon_{j} \) which represents the increase (or decrease) in the required research activity level \( j \). The essence of emergency is that \( \epsilon_{j} \) is multimodal---e.g., bimodal---with high probability concentration at 0 and a high enough value at its other peak to cause significant changeover activity (with attendant costs) if the extra demand is to be met. To add further operational realism we assume that the timing of the emergency is random in the short-run period. We model this randomness in terms of its effect on the productivity of the \( i^{th} \) facility by a random variable \( u_{i} \) such that \( u_{i} x_{ij}^{(1)} \) is the amount of research activity up to the occurrence of the (vector) emergency, \( \epsilon_{j} , j=1, \ldots, n \).

---

1/ See Charnes and Cooper [2] and [3], and Charnes, Cooper and Symonds [5].
Now, supposing that the emergency has occurred--i.e., that the sample values of \( u_i \) and \( \epsilon_j \) are known--adjustment process is imminent. We assume that the \( \delta_i^{(1)} \) are now known also. It will be recalled that randomness in maximum availability involved such factors as personnel and institutional changes which, although unknown at the time of formulation of the initial plan, would be quite well specified by the time the operation had commenced.

The interim activity, \( y_{ij} \), is now to be undertaken. We shall specify these in terms of a class of stochastic decision rules involving the (now known) random variables \( u_i \) and the \( \epsilon_k \). We render the availability and emergency conditions on the \( y_{ij} \) via the chance constraints:

\[
\begin{align*}
P \left( \sum_{j} y_{ij} &\leq a_i^{(1)} + \delta_i^{(1)} - u_i \sum_{i} x_{ij}^{(1)} \right) \geq \beta_i^{(12)} , \quad i=1, \ldots, m \\
P \left( \sum_i y_{ij} \geq \epsilon_j - \sum_i u_i x_{ij}^{(1)} \right) &\geq \alpha_j^{(1)} , \quad j=1, \ldots, n .
\end{align*}
\]

The remnant plan must now be modified from the \( x_{ij}^{(1)} \) to values \( x_{ij}^{(2)} \) in accordance with the remnant long-run requirements, \( b_j^{(2)} \), the yet to emerge availabilities, \( a_i^{(2)} + \delta_i^{(2)} \), and the interim activity. Thus we posit similarly to (1.1)

\[
\begin{align*}
P \left( \sum_{j} x_{ij}^{(2)} &\leq a_i^{(2)} + \delta_i^{(2)} \right) \geq \beta_i^{(2)} \\
P \left( \sum_{i} y_{ij} + \sum_{i} x_{ij}^{(2)} \geq b_j^{(1)} + \epsilon_j - \sum_i u_i x_{ij}^{(1)} + b_j^{(2)} \right) &\geq \alpha_j^{(2)}
\end{align*}
\]

Note that the constraint (3.2) represents the effect of the initial plan and the interim adjustment on the posture in which the process is left relative to the attainment of the long-range objectives. We implicitly assume, via (3.1), that the effect of exceeding availability in the initial period, if it should occur, does not carry over into the long run. 1/

1/ It would be difficult to judge whether or not the effect of exceeding availability would be to decrease or increase availability in the subsequent period. Thus this effect, if any, would be included in the random variation already assumed.
We take our optimal control objective as that of minimizing expected cost where cost consists of the following components: (1) realized initial costs \( \sum_{i,j} c^{(1)}_{ij} u_i x^{(1)}_{ij} \); (2) changeover costs \( \sum_{i,j} \frac{\mu_{ij}}{2} \left[ (1 - u_i) x^{(1)}_{ij} - y_{ij} \right]^2 \); (3) interim activity costs \( \sum_{i,j} c^{(12)}_{ij} y_{ij} \); and (4) long-run activity costs \( \sum_{i,j} c^{(2)}_{ij} x^{(2)}_{ij} \). The \( c^{(1)}_{ij}, c^{(12)}_{ij}, \) and \( c^{(2)}_{ij} \) represent the cost of a unit activity level in research area \( j \) at facility \( i \), and the \( \mu_{ij} \) are the marginal costs of a unit change in activity levels from those initially planned for the remainder of the short-run period.

The objective may then be stated as:

\[
\text{(4) Minimize } C = \mathbb{E} \left\{ \sum_{i,j} c^{(1)}_{ij} u_i x^{(1)}_{ij} + \sum_{i,j} \frac{\mu_{ij}}{2} \left[ (1 - u_i) x^{(1)}_{ij} - y_{ij} \right]^2 + \sum_{i,j} c^{(12)}_{ij} y_{ij} + \sum_{i,j} c^{(2)}_{ij} x^{(2)}_{ij} \right\}
\]

**Control Decision Rules**

To complete the statement of the chance-constrained programming problem we must specify the class of stochastic decision rules within which we shall seek an optimal set. For simplicity we shall here use the class of linear decision rules. The character of the \( x^{(1)}_{ij} \) and \( x^{(2)}_{ij} \) as plans leads us to specify these as "zero-order" rules \(1/\) e.g., not explicitly involving the random variables \( u_i, \epsilon_j, \delta^{(1)}_i, \delta^{(2)}_i \). For the \( y_{ij} \) we posit the following class of "operating response" rules:

\[
\text{(5) } y_{ij} = (1 - u_i) x^{(1)}_{ij} + \sum_k \gamma_{ijk} \delta_k
\]

Note that this type of rule is in keeping with the notion of an interim response to an emergency where the coefficients, \( \gamma_{ijk} \), are to be determined by solution of the total chance-constrained problem so as to achieve optimality for this class of operating response rules. Thus, with solution of the mathematical

---

\(1/\) See Charnes and Cooper [3] for discussion and explicit definition.
problem, for every emergency that arises, the \( y_{ij} \) will be specified exactly and not as relative frequencies or mixed strategies.

**Deterministic Equivalent Problem**

Because of the zero-order character of the \( x^{(1)}_{ij} \) and \( x^{(2)}_{ij} \), conditions (1.1) and (3.1) can be immediately inverted to give:

\[
\begin{align*}
\Sigma x^{(1)}_{ij} & \leq a^{(1)}_i + F^{-1}_{1i}(1 - \beta^{(1)}_i) \\
\Sigma x^{(2)}_{ij} & \leq a^{(2)}_i + F^{-1}_{2i}(1 - \beta^{(2)}_i)
\end{align*}
\]

where the \( F_{1i} \) and \( F_{2i} \) are the marginal distribution functions of \( \delta^{(1)}_i \) and \( \delta^{(2)}_i \), respectively. Inserting the operating response rules (or "certainty equivalent" relations [5]) for the \( y_{ij} \) in (2.1), yields, upon using the spacing variables device of [3]:

\[
\begin{align*}
G_i^{-1}(\beta^{(12)}_i) v_i + \Sigma x^{(1)}_{ij} + \Sigma \gamma_{i,k} y_{ijk} & \leq a^{(1)}_i + \delta^{(1)}_i \\
v_i^2 - \Sigma V(\epsilon_k)(\Sigma \gamma_{i,jk})^2 & \geq V(\delta^{(1)}_i)
\end{align*}
\]

Here we have assumed, for simplicity, that the \( \epsilon_j \) and \( \delta^{(1)}_i \) are independent random variables. The bars over random variables refer to their means (e.g., \( \delta^{(1)}_i = 0 \) by our previous hypothesis), the \( V \) is the variance operator, \( G_i \) is the distribution function for the random variable

\[
\begin{align*}
\Sigma (\epsilon_k - \bar{\epsilon}_k) \Sigma \gamma_{i,jk} - (\delta^{(1)}_i - \bar{\delta}^{(1)}_i) \\
\sqrt{\Sigma (\Sigma \gamma_{i,jk})^2 V(\epsilon_k) + V(\delta^{(1)}_i)}
\end{align*}
\]

which has zero mean and unit variance, and the \( v_i \) are "spacer variables."
Similarly (2.2) may be rendered in the deterministic equivalent form:

(8.1) \[-H_j^{-1}(\alpha_j^{(1)})w_j + \sum_{i} x_{ij}^{(1)} + \sum_{i,k} \xi_{ijk} + \xi_j \geq 0\]

(8.2) \[w_j^2 - \sum_{k} V(\epsilon_k)(\delta_{jk} - \sum_{i} \gamma_{ijk})^2 \geq 0\]

where \(H_j\) is the distribution function for the random variable:

(8.3) \[\frac{(\bar{\epsilon}_j - \bar{\xi}_j) - \sum_{i,k} (\epsilon_k - \bar{\xi}_k) \gamma_{ijk}}{\sqrt{\sum_{k} V(\epsilon_k)(\delta_{jk} - \sum_{i} \gamma_{ijk})^2}}\]

where \(\delta_{jk}\) is the Kronecker delta, the \(w_j\) are "spacer variables" and the other quantities are as defined above. Further, (3.2) may be written as:

(9.1) \[-H_j^{-1}(\alpha_j^{(2)})z_j + \sum_{i} x_{ij}^{(1)} + \sum_{j,k} \xi_{ijk} + \sum_{j} x_{ij}^{(2)} \geq b_j^{(1)} + \xi_j + b_j^{(2)}\]

(9.2) \[z_j^2 - \sum_{k} V(\epsilon_k)(\delta_{jk} - \sum_{i} \gamma_{ijk})^2 \geq 0\]

which introduces only the new spacer variables \(x_j\). Finally, the expression for the function, \(\mathcal{G}\), of equation (4) becomes

(10) \[\mathcal{G} = \sum_{i,j} c_{ij}^{(12)} - (c_{ij}^{(12)} - c_{ij}^{(1)}) \bar{\epsilon}_i \sum_{i,j,k} c_{ij}^{(12)} \xi_{ijk} + \sum_{i,j,k} c_{ij}^{(12)} \bar{\epsilon}_i \gamma_{ijk} + \sum_{i,j,k} \mu_{ij} \sum_{i,j} (\xi_{ij} \gamma_{ijk})^2 + \sum_{i,j,k} \mu_{ij} V(\epsilon_k) \gamma_{ijk}\]

These may be assembled in the form:
which becomes a convex programming problem when the $G_i$ and $H_j$ are independent of the $\gamma_{ijk}$ and the $G_i^{-1}$ and $H_j^{-1}$ values are non-negative. This would be true, for example, if the $\epsilon_j$ and $\delta_i^{(1)}$ have distributions which are mixtures of normal distributions, and the probabilities $\beta_i^{(12)}$, $\alpha_i^{(1)}$, and $\alpha_i^{(2)}$ are sufficiently high.

From this format it may already be concluded that:

Theorem: In an optimal solution, the $x_{ij}^{(1)}$ and the $x_{ij}^{(2)}$ may be taken as basic (or extreme point) solutions to a linear programming problem of ordinary distribution type.
Proof: If all the variables except the \( x_{ij}^{(1)} \) are specified, the sets of relationships (11.1) and (11.3) reduce to a single set of non-redundant inequalities of type (11.1). Similarly, (11.2), (11.5) and (11.7) reduce to a single set of type (11.2). This, together with the linearity of (10) in the variables \( x_{ij} \), yield our (extreme point) conclusion for optimal \( x_{ij}^{(1)} \).

Similarly, holding all variables fixed but the \( x_{ij}^{(2)} \), we conclude that optimal \( x_{ij}^{(2)} \) may be taken as extreme point solutions to a linear programming problem of distribution type. Q.E.D.

It is also interesting to observe the effect of introducing the possibility of emergencies in terms of the constraint set. For those facilities where the non-redundant constraints are in (11.1) or for those research areas where the constraints (11.2) are binding, no change in the initial plan will result. However, if availability constraints from (11.3) are binding, facilities will have lesser planned activity levels than would be the case in the absence of emergency protection. Similarly, for areas in which (11.5) or (11.7) constraints are non-redundant, activity levels will be increased to "hedge" against an emergency and (in the latter case) against the requirements of the long-run plan.

More specific conclusions are highly dependent on the relative values of the \( \mu_{ij}, c_{ij}, V(t_k) \) and \( V(\delta_1) \). However, the deterministic problem is a convex programming problem of manageable type \(^1\) and specific conclusions for reasonable numerical values of the parameters will be available shortly on the basis of calculations performed using the SUMT method of Fiacco and McCormick [6].

\(^1\) Cf. Charnes and Cooper [3].
Summary

We have postulated a chance-constrained model of a two-stage planning and control process which allows: (1) random availability of facilities in the short and long run; (2) random occurrence of emergency demands at random times during the short run; (3) probabilistic constraints on conformity to availability constraints and emergency demands; and (4) deterministic constraints on desired activity levels.

This model was designed to deal with optimal funding for research support where the possibility of breakthroughs exists, but it also is applicable to a class of problems involving the occurrence of large unforeseen demands. The chance-constrained problem has been reduced to a deterministic equivalent convex programming problem of manageable type, involving at most second degree terms and for which computer routines are available.
REFERENCES


A CHANCE-CONSTRAINED MODEL FOR REAL-TIME CONTROL IN RESEARCH AND DEVELOPMENT MANAGEMENT

Research paper

Charnes, Abraham, and Stedry, Andrew C.

July 1965

Non-1228(10)

NR 047-021

Systems Research Memorandum No. 130

Management Sciences Research Report #47

Releasable without limitation on dissemination.

Funding of research projects is considered as encompassing three stages: (1) an initial short run plan for funding based upon projected regular demands and availability subject to random deviations; (2) adjustment of the initial plan to take into account the actual regular demands and availability and the funding of significant break-throughs which occur at random intervals preempting other demands; and (3) a plan for longer run availability and demands which constitute a "posture" desired subsequent to the funding adjustments of (2). The essence of the distribution of the unexpected demands is multi-modality with low probability of occurrence but high resource demand when they do occur. This approach represents a substantial departure from the usual planning model development which produces only an optimal plan based on forecasted developments without provision for adjustment when the forecasted events actually materialize and additional unexpected demands are placed on resources. The adjustment process explored here -- which provides the mechanism for optimal implementation of the original plan or control of resource allocation -- enables optimal response to information received in "real-time" avoiding the frequently observed over- or under-response to receipt of such information without reference to the impact of the interim decision on future capabilities.
Chance-Constrained Programming
Real Time Control
Optimal Funding
Management of Research
Control of Organizations