SEARCH THEORY AND APPLICATIONS

NATO CONFERENCE SERIES

I Ecology
II Systems Science
III Human Factors
IV Marine Sciences
V Air–Sea Interactions
VI Materials Science

II SYSTEMS SCIENCE

Volume 1 Transportation Planning for a Better Environment
Edited by Peter Stringer and H. Wenzel

Volume 2 Arctic Systems
Edited by P. J. Amaria, A. A. Bruneau, and P. A. Lapp

Volume 3 Environmental Assessment of Socioeconomic Systems
Edited by Dietrich F. Burkhardt and William H. Ittelson

Volume 4 Earth Observation Systems for Resource Management and Environmental Control
Edited by Donald J. Clough and Lawrence W. Morley

Volume 5 Applied General Systems Research: Recent Developments and Trends
Edited by George J. Klir

Volume 6 Evaluating New Telecommunications Services
Edited by Martin C. J. Elton, William A. Lucas, and David W. Conrath

Volume 7 Manpower Planning and Organization Design
Edited by Donald T. Bryant and Richard J. Niehaus

Volume 8 Search Theory and Applications
Edited by K. Brian Haley and Lawrence D. Stone

SEARCH THEORY AND APPLICATIONS

Edited by
K. Brian Haley
The University of Birmingham
Birmingham, United Kingdom

and
Lawrence D. Stone
Daniel H. Wagner, Associates
Paoli, Pennsylvania



Published in cooperation with NATO Scientific Affairs Division
PLENUM PRESS · NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

NATO Advanced Research Institute on Search Theory and Applications, Praia da Rocha, Portugal, 1979.
Search theory and applications.

(NATO conference series: II, Systems science; v. 8)
"Proceedings of a NATO Advanced Research Institute on Search Theory and Applications held March 26–30, 1979, in Praia Da Rocha, Portugal."
Includes index.
1. Search theory—Congresses. I. Haley, Keith Brian. II. Stone, Lawrence D., 1942- III. Title. IV. Series.
T57.97.N37 1979 001.4'24 80-19871
ISBN-13: 978-1-4615-9190-0 e-ISBN-13: 978-1-4615-9188-7
DOI: 10.1007/978-1-4615-9188-7

Proceedings of the NATO Advanced Research Institute on Search Theory and Applications held in Praia Da Rocha, Portugal, March 26–30, 1979.

© 1980 Plenum Press, New York
Softcover reprint of the hardcover 1st edition 1980
A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

PREFACE

The NATO Advanced Research Institute on Search Theory and Applications was held at the Hotel Algarve in Praia Da Rocha, Portugal, from March 26 through March 30, 1979, and was sponsored by the NATO Special Programme Panel on Systems Science. There were forty-one participants representing a wide range of backgrounds and interests.

The purpose of the institute was to bring together people working in search theory and applications with potential users of search techniques to stimulate the increased application of recently developed search technology to civilian problems such as search and rescue, mineral exploration, surveillance, and fishing. Conversely, it was felt that by exposing search analysts to potential applications and new problems, they would be stimulated to develop new techniques for these applications and problems.

The exchange of ideas and problems necessary to accomplish these goals was provided in the meeting workshops. There were three workshops, Search and Rescue, Exploration, and Surveillance and Fishing, each consisting of a small group of search analysts and potential users working together to define areas in which search theory and technology can be applied and to outline plans for implementation. At the end of the conference, each working group submitted a report outlining possible areas of search applications and discussing problems which needed to be solved in order to implement these applications.

These proceedings contain most of the papers presented at the working group and plenary sessions as well as the reports produced by the working groups. While these papers show the diversity of interests and backgrounds represented at the meeting and contain much of the formal technical information presented there, they do not convey the nature of the exciting and interesting informal exchanges which took place and which may, in the end, prove to be the most important aspect of the meeting.

CONTENTS

I. INTRODUCTORY PAPERS

Brief Overview of Search Theory and Applications 1
L.D. Stone

General Operation of Search 9
B.O. Koopman

An Overview of Computer-Assisted Search Information Processing 25
H.R. Richardson and T.L. Corwin

II. APPLICATIONS

1. Search and Rescue

Search and Rescue Working Group Report 45

Search and Rescue in Canada 73
R.P. Hypher

Overland Search for Missing Aircraft and Persons 87
R.J. Mattson

Establishing Search Priorities 93
R.J. Mattson

2. Surveillance

Surveillance Working Group 99

Surveillance and Search Theory 113
S.M. Pollock

Computer Simulation of Search Tactics for Magnetic Anomaly Detection 127
R.H. Kuwahara

On the Optimization of a Search Pattern for Maritime Surveillance 141
M. Reiss

3. Exploration

Exploration Working Group Report 155

Introductory Comments on Exploration and Search Theory 159
E.A. Field

Exploration for Coal in the U.K. 165
J.G. Fergusson

Exploration for Polymetallic Nodules 173
J.P. Lenoble

The Search for Nodules 183
G. Friedrich, J.P. Lenoble, and J.G. Fergusson

Search Theory in the Exploration for and Mining of Polymetallic Nodules on the Ocean Floor 185
D.H. Wagner

4. Medicine

An Application of Search Theory in Medicine: The Detection of Glaucoma 199
P. Kolesar

5. Industry

Industrial Applications of Search Theory 205
J.B. Kadane

6. Clearance

The Search for COSMOS-954 211
R.L. Grasty

III. THEORY

1. Search for Moving Targets

Optimal Search for a Moving Target in Discrete Time and Space With an Exponential Detection Function 221
S.S. Brown

Search for a Moving Target: Upper Bound on Detection Probability 231
A.R. Washburn

Optimization Algorithm for General Target Motion 239
L.D. Stone

Experience With a Branch-and-Bound Algorithm for Constrained Searcher Motion 247
T.J. Stewart

2. Search for Stationary Targets

Optimal Stopping of the Investigation Search 255
T. Kisi

A Solution for the Optimal Multiple Rectangle Search 261
J.H. Discenza

Participants 273

Index 277

BRIEF OVERVIEW OF SEARCH THEORY AND APPLICATIONS

Lawrence D. Stone

Daniel H. Wagner, Associates

Elements of a Search Problem

There are three basic elements which are present in an optimal search problem:

1. Probability Distribution for Target's Location and Motion. Information about the target's location at some initial time (e.g., his last reported position) and his subsequent motion are quantified in terms of probability distributions. The probability distribution for initial location and subsequent motion are combined to produce a probability map for the target's location at any subsequent time.

An example of this might be the last reported position of a submarine which is assumed to be traveling at a known speed, s, but with unknown heading. The reported position has an error which is modeled by a circular normal distribution with standard deviation σ in any direction. A cross section of the density of this distribution taken through the reported position is shown in Figure 1. Assuming the target has no preferred direction (i.e., the distribution of headings is uniform over [0,360] degrees), the distribution at time t later is also shown in Figure 1. Note that at $t = 4\sigma/s$, the reported position is now one of the least likely positions. This example is taken from Koopman [1946].

2. Detection Function. The detection function relates the amount of effort placed in an area to the probability of detecting the target given it is located in that area.

As an example, suppose that we have a sensor with detection

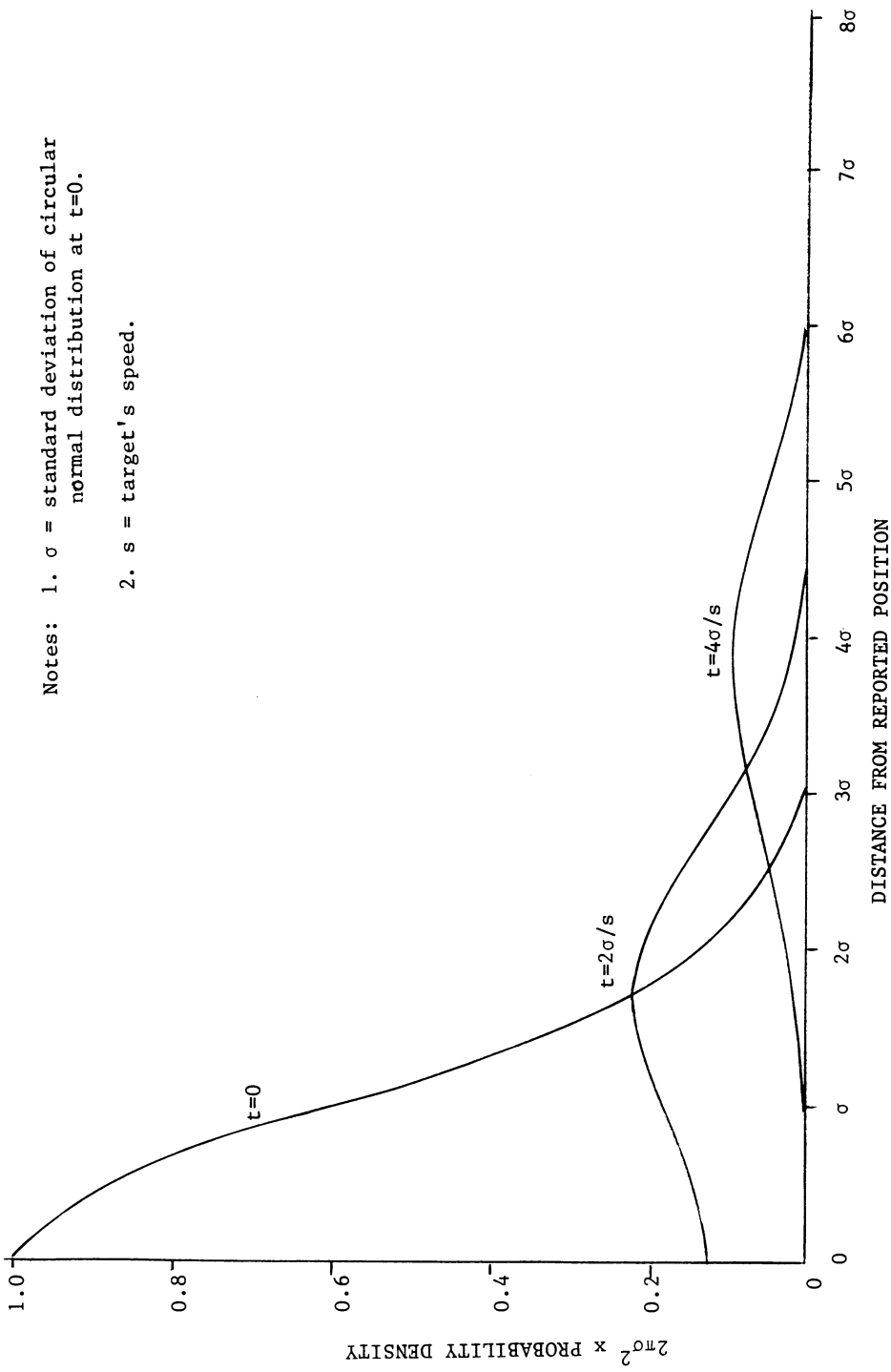


Figure 1. Probability density for a radially fleeing target.

range R, i.e., if we pass within lateral range $r \leq R$ of the target we will detect with probability 1. If $r > R$, then no detection occurs. Define the sweep width W of this sensor to be

$$W = 2R.$$

If one searches in a systematic fashion in a region of area A at speed v for a time t, then

$$b(t) = 1 - \exp(-Wvt/A)$$

is a conservative and often very accurate estimate of the probability of detecting the target given it is in the region. The above equation is the so-called random search formula of B. O. Koopman, and it gives the detection function for this search situation.

3. Constraint on Effort. Typically, the searcher has a limited amount of effort or resource available to conduct the search. If in the above situation the searcher has a limited time T available for search, then we would use time as a measure of search effort and T as the constraint on effort. Often the constraint is not on the total effort available but the rate at which effort is available.

Problem Statement

Using the detection function and the probability distribution for target's location we can compute, for each allocation of total search effort to the various regions of our search space, the probability of detecting the target with that allocation. The optimal search problem is to find an allocation of effort (e.g., search time) in space and time which maximizes the probability of detection by time T.

State of Theoretical Progress in Optimal Search

In order to discuss the present capability for finding optimal search plans, we shall categorize search problems according to target motion and the constraints placed on our ability to allocate search effort. In Figure 2, we categorize targets according to whether they are stationary, or moving in discrete or continuous time. For sensors, we consider short range and long range ones. A short-range sensor is one whose effective search region is confined to an area which is small compared to the spread of the target location distribution. All other sensors are long range. This point is discussed by Professor Koopman in his paper later in this volume on The General Operation of Search.

SEARCH EFFORT		TARGET: STATIONARY	TARGET: MOVING — DISCRETE TIME	TARGET: MOVING — CONTINUOUS TIME
SHORT RANGE	Infinitely divisible	Uniformly solvable for concave detection functions	Solvable with concave detection functions	Necessary and sufficient conditions
SHORT RANGE	Discrete effort	Solvable in principle by enumeration	Necessary but not sufficient conditions	Little or no progress
SHORT RANGE	Constraints on search	Solvable only in special situations	Some progress	Little or no progress
Long Range		Some approximate solutions	Little or no progress	Little or no progress

Figure 2. Summary of present capability to find optimal search plans.

Within the category of short-range sensors we consider infinitely divisible effort, discrete effort, and situations in which there are constraints on searcher motion. By infinitely divisible we mean that effort can be divided as finely in space as desired. This is approximated by a large number of searchers with short-range sensors, e.g., a group of people looking for an avalanche victim by poking holes in the snow with rods. Discrete effort corresponds to a situation in which one searches by making looks in an area, e.g., reconnaissance photographs. The case of constrained searcher motion occurs when someone is searching on foot for a lost person.

The best situations for finding optimal search plans are toward the upper-left part of the figure. For example, we can find uniformly optimal search plans for a stationary target when effort is infinitely divisible and the detection function is concave (Stone [1975]). Moving over one column, we see that optimal plans can be found for targets which move in discrete time provided the detection function is concave (see Stone, et.al. [1978] and the articles by Brown and Stone in this volume). For target motion in continuous time, there are optimal plans available in some special situations but not in general (Stone [1975], Chapter VIII). The necessary and sufficient conditions for continuous time are given in Stone [1979].

We can see from the chart that as we impose more constraints on the search effort, our ability to find optimal search plans decreases (see the article by Stewart in this volume) and that when we are dealing with long-range sensors, we have very limited capability to find optimal search plans. Also, our ability to find optimal search plans decreases as we proceed from a stationary target through motion in discrete time to motion in continuous time.

[In the interim between the meeting in March 1979 and the publication of these proceedings, there has been significant progress in solving optimal whereabouts problems (Stone and Kadane [1979]) and multistate problems (Brown [1979], Discenza and Stone [1980]) for moving targets. Also, the necessary and sufficient conditions of Stone [1979] have been generalized by Stromquist and Stone [1980] to a wider class of problems.]

Tools and Techniques Used to Plan Searches and Find Optimal Search Plans

There are a small number of optimal search problems which yield neat, analytic solutions. Most problems encountered do not have solutions that can easily be written down by hand. Even the calculation of the target location distribution can quickly expand beyond manual capability when one attempts to incorporate all

available information about the target's location and possible motion in a realistic fashion. A number of tools have been developed to assist the search planner in identifying high probability areas for search and in allocating search effort. These tools involve the use of computers, and their use constitutes computer assisted search (CAS). We briefly describe some applications of CAS. A more detailed discussion of CAS is given in the article by Richardson and Corwin in this volume.

U. S. Coast Guard

A series of computer programs called CASP have been developed for use by the U. S. Coast Guard in planning maritime search and rescue (SAR) missions. The programs are designed to produce probability maps for the target's location (e.g., a missing fishing vessel) based on quantifying, in subjective probability terms, all of the information about the target's last known position and possible present and future positions. This is done by making use of the target's planned movements and possible random movements caused by winds and currents. In addition, these programs take into account, in a Bayesian fashion, the negative information one obtains by searching in an area and failing to detect the target. The programs are also capable of suggesting good allocations of search effort (Richardson and Discenza [1980]).

The CASP programs are operational and have been credited with saving a number of lives in difficult SAR operations.

Land Search and Rescue

Although designed for maritime use, the CASP programs have been made available to the U. S. Air Force and the Canadian Forces, for use in planning land searches particularly for downed aircraft.

However, these programs must be modified to account for the distinct nature of the land SAR problems. For example, knowledge of the weather, particularly cloud ceiling and the terrain over which an aircraft planned to fly, can provide valuable clues in planning where to look for a downed aircraft. The articles by Mattson in this volume discuss these points further.

U. S. Navy

The U. S. Navy also makes use of computer programs to plan searches for submarines.

Exploration

There have been some computer tools developed and reported in the open literature. One such tool developed by Surkan [1975] is a computer program which uses prior information from geological surveys quantified into a density map for oil deposits. This prior information is then modified in a Bayesian manner for negative information (i.e., wells drilled which failed to find oil) and positive information (wells which found oil).

Acknowledgement

This paper was prepared with the support of ONR Contract No. N00014-76-C-0696.

References

Brown, S. S. [1979] Defensive Search for a Moving Target in Discrete Time With an Exponential Detection Function (submitted for publication).

Discenza, J. H. & Stone, L. D. [1980] Optimal Survivor Search With Multiple States (submitted for publication).

Koopman, B. O. [1946] Search and Screening, OEG Report No. 56, Washington, D. C. (a modernized edition has been published by Pergamon Press, Inc., New York, 1980).

Richardson, H. R. & Discenza, J. H. [1980] The United States Coast Guard Computer-Assisted Search Planning System (CASP), Naval Res. Logist. Quart. (to appear).

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

Stone, L. D. [1979] Necessary and Sufficient Conditions for Optimal Search Plans for Moving Targets, Math. Oper. Res. 4, pp. 431-440.

Stone, L. D. & Kadane, J. B. [1979] Optimal Whereabouts Search for a Moving Target (submitted for publication).

Stone, L. D., Brown, S. S., Buemi, R. P. & Hopkins, C. R. [1978] Numerical Optimization of Search for a Moving Target, Daniel H. Wagner, Associates Report to Office of Naval Research.

Stromquist, W. R. & Stone, L. D. [1980] Constrained Optimization of Functionals With Search Theory Applications (submitted for publication).

Surkan, A. J. [1975] Sequential Search by Computer in Exploration, Probability Methods in Oil Exploration, Notes of Am. Ass. Petroleum Geologists Res. Symp. Stanford University.

GENERAL OPERATION OF SEARCH

Bernard O. Koopman

General Concepts in Search

Search, in its broadest sense, is an activity not only of modern man, but of man in his most primitive state and of animals capable of motion directed by senses. Looking for food, prey, shelter, mates, and detecting hostile animals or natural dangers, are operations without which they could not survive. During our evolution we have acquired such a set of instincts involving search in this primitive sense that -- as in breathing or walking-- we habitually give it little conscious thought. Only with the advance of modern technology have we learned to examine search in the light of science: as an operation having various structural patterns and obeying laws of its own.

The scientific study of search before World War II was rather sporadic: methodology was applied in the search for minerals or missing things or persons; mass screening was made of substances of desirable physical or medicinal properties. Most navies planned search on the basis of the sure detection of targets within unobstructed horizon distance, and cross-over patrols for air search of surfaced targets were part of many tactical handbooks. In 1924 a paper was read before the Société Mathematique de France on the search for aircraft at night by searchlights and the advantages of certain spiral searches were explained. These and a multitude of other cases could be cited; but each seemed of too individual a nature to be united into a general theory.

During World War II, particularly in detecting invading aircraft (visually or by radar), and in searching for submarines

during the Battle of the Atlantic (by sonar when submerged, visually or by radar when surfaced), it became absolutely necessary to exploit all scientific methods to understand and improve the effectiveness not only of the instruments of search but also the logic of their application. From then on, we could speak of the theory of search as a technology linked to a new branch of applied mathematics.

Admittedly this paper bears the mark and orientation of the author's special experience with military search, particularly with that of anti-submarine warfare during and following World War II, and to the principles that have to be brought to bear upon this rapidly evolving branch of tactics. But from such special and concrete classes of problems general principles have been evolving, increasingly applied to searches of the most diverse sorts -- both civilian and military.

To show how broadly the ideas have spread and found application outside the military field, we briefly note the following: search for lost people; for life-rafts; for celestial objects in astronomy; for minerals and other natural resources; for forest fires or for conditions leading to them; for sources of infection or for individuals or other vectors of contagion; for candidates for various types of medical treatment in a population; for causes of favorable or unfavorable medical conditions in a given individual, as in diagnostics; search for patterns, as in code breaking. Multitudes of other examples could be given. But in closing, we must emphasize that when factors of an unknown kind are sought to explain observed phenomena, we have left the field of search proper and enter that of research.

In every fully developed view of search the four following elements can be distinguished:

A. There must be some knowledge of the whereabouts of the object of search-- the "target;" otherwise most searches would be practically hopeless. In precise terms, this means a more or less rough knowledge of its a priori probability distribution. The sources of such knowledge can be of most varied kind; but whatever they be, they must come from reasoning based on objective evidence. In the search for U-boats in the war, it could combine presumed objectives and tactical methods of the U-boats with the physical characteristics of the latter, as well as the geographical limitations; furthermore, there were often preliminary rough detections, radio transmissions giving a fix, and even information gained from espionage and from code breaking.

B. The characteristics of the detecting equipment must be known to the searcher, as well as its degree of ability to detect various types of targets under the diversity of conditions to be expected. Here again probability enters: the conditional probability of detecting a target, given to be present in a specified position and physical condition. This probability is usually difficult to evaluate: engineering characteristics and tests under controlled conditions are necessary, but seldom sufficient, since the signal from the target, whether visual, electromagnetic, or acoustic, has to pass through a material medium, and at the long ranges usual in military operations slight and unpredictable changes in the physical state of the medium can radically affect the strength of the signal received from the target. Added to these complications are the variables involved in the state of the target itself affecting its detectability: its aspect and reflectivity, its speed and noisiness, electromagnetic emissions, and a host of similar factors, which, if given, might allow the determination of the detectability; but in actual practice are not known to the searcher, and are distributed according to a law which has rarely if ever been objectively established. As in the earlier case, A, we have to deal with a rough approximation.

C. Since B can have no precise meaning unless the amount of searching or the quantity of searching effort is defined in appropriate physical and operational terms, this must be done. In the case of a fixed long range surveillance system, this quantity could be the number of distinct sensor systems and the period of time during which they are activated. With local search by short range sensors on mobile platforms, it could be measured by the track length of the latter or by the amount of time spent in the target region: a more precise definition will be given later. There are as many examples of the evaluation of the amount or search as there are of search situations. Whatever the case, this quantity must be capable of at least a rough numerical expression, definable in principle by objective observations. After this concept has been fixed, the probability described in B becomes precise: it is the detection probability per unit effort.

D. We must realize that in real life any search is a component of a larger operation. In the Battle of the

> Atlantic the search for U-boats was part of the process of protecting over-sea lift--one tactic among many. On the one hand, it was aided by using the information gained from other parts of the operation, leading to improved estimates of target distributions and their tactics; on the other hand the search could cooperate with other searches and do its part in the general defense.

As a peacetime example of the subordination of any particular search to other parts of a global endeavor, we return to the procedures in searching for oil in the ground. First, geological observations are made, followed by studies of their data; next, gravitational anomalies are looked for by airborne sensors over the region indicated by the former studies; then acoustic echo-ranging in promising localities is carried out; only last is the expensive drilling performed in places shown to be useful by the earlier methods. This example illustrates two further notions: the idea of the multiple search and the important concept of the trade-off between two desirable features in a detection system, broad coverage that increases the chance of detecting a target and precision of its localization when detected. In search for submarines, broad area surveillance has the former property, echo-ranging, the latter.

From these broad generalities we pass to a classification of types of search, an idea illustrated by military examples.

First, a search may involve only purely "passive" observations, in which the only thing that is altered is the body of knowledge on the part of the searcher. They have also been called "non-reactive" searches. Measurements in astronomy and in classical physics are observations of this kind. So are photo-reconnaisance and wide-range surveillance. At closer distances from the possible targets, careful consideration must be given to the actual nature of the detection process before assuming that it is of this passive type. Passive observations may be regarded as the "trials" in the theory of probability - their outcomes being the "events ;" these can then be treated, in their various possible combinations, by the classical theory (whether the events are independent or not, mutually exclusive or not).

Second, a search may involve "active" observations, by which the target (if humanly controlled) is alerted, as when the U-boats had effective radar search receivers; or the target, the searcher, or their common environment are so physically altered by the operation of search ("muddying the waters" or covering the lost object by sand or grass, etc.) that a repetition of the same physical conditions in a second trial is excluded. These have also been called "reactive" searches. The best-known example in physics

of this situation is in the quantum mechanics of elementary particles, in which, as Heisenberg first pointed out, the necessary process of measuring the position can change the previously measured value of the conjugate momentum. This is described in modern theoretical physics the incompatibility of observations. The importance in the theory of search of the issue of compatibility of two parts of the operation was first pointed out by Dobbie [1963]. In dealing with the combinations of incompatible observations --the "trials" of probability theory-- one must realize that they do not necessarily combine into physically meaningful events by the classical Boolean algebra of the propositional calculus, as required in the expression of total probability ("at least one event occurs :" the logical sum) or compound probability ("both events occur :" the logical product). This is because the performance of the first trial may make the second trial impossible, as when it destroys or profoundly alters the object needed for the second; e.g., kills the experimental animal etc. More generally, when the very definition of the second trial is based on conditions which are inconsistent with certain possible outcomes of the first, these combinations may lose their meaning. Since probability can only be applied to trials that do have a physical meaning (far more than the "measurable sets" of the "probability space" in the purely mathematical modern view of the subject), it is not surprising that a reformulation of each problem may be necessary before the laws of probability can be applied to the situation. This author has discussed the natural methods for this in the case of quantum theory, and tried to dispel certain misconceptions concerning the "principle of indeterminacy" and the suggestion that the laws of probability cease to hold in the present class of cases (Koopman [1957]).

Third, a search may be by means of long range detection equipment, as in wide area surveillance, satellite monitoring, and astronomical observations. Or it may be by short range types, as in most visual search; in echo-ranging by traditional sorts of equipment, World War II radar, infra-red, and magnetic anomaly detection (MAD). In the former case, the detection systems are usually fixed; or if their geographical positions are changed, this is infrequent and not part of the searching procedure. In the latter case on the other hand, the searcher's platform moves about through the search region. This implies that the dimensions of the latter are considerably larger than the detection ranges, a situation typical of the last war. The motion may be relative, as in the case of a barrier, through which the targets are expected to pass.

Fourth, a search for a particular target may have to be confined to a single try--a "one shot" event, as in a single throw of dice. This would be the case when a ballistic or a cruise missile or invading aircraft is penetrating a screen: it may or

may not be detected, but if not, the screen has no second chance. The same is true for the penetration of a sonar screen by a submarine approaching for an attack on a surface formation. In certain cases the target is only detectable temporarily, as in a missile launch or an underwater explosion.

Fifth, a search for a particular target may be repeatable, possibly very often, until it is found or the search discontinued for other reasons. The search for a lost object, a wreck, a foreign missile site, or a hostile submarine known to be patroling in the vicinity of a particular position and not alertable by the search, would all be cases of repeatable search. This is true in particular when the searcher's platform has a much greater speed than the target, as in air search for a submarine or a satellite search for a surface formation. Although the event of a screen's detecting an individual target in a penetration is not repeatable, when a succession of targets attempt penetration at different times, the expected number detected is found by the formulas of repeatable search.

Sixth, in the case of repeatable search by passive observations--whether the individual searches are made by the same or different but similar searchers-- an issue basic to the mathematical formulation of the problem of the optimal employment of the searching effort is raised by the double question: (i) is the search subdivisible into partial searches of the same kind and with each amount of effort remaining bounded? -- and if thus subdivisible, (ii) are the effects of the separate parts additive in the sense that the probability of detecting the target by their combination (meaning that a least one part detects) is the same as that of a single search using the sum of the individual efforts? To see that this situation is not always true, we note that one of the most familiar and frequently employed searches is by having the searcher's platform trace equally spaced parallel sweeps (the boustrophédon). Whether the definite range law of detection, the inverse cube law, or any other law having the same order of symmetry as the set of paths, the detection formulas do not possess the pair of properties stated. Further examples of their failure are cases in which an appreciable time is taken for the build-up of the target's image on a scope, during which it recognizability increases progressively.

In most cases, however, the regularity and the "fine-structure" of the above counter-examples has to be disregarded as being overridden by the essential inaccuracies in most searches in the real world; then properties (i) and (ii) are usefully assumed as approximations. Actually, the many studies of search as a progressive operation with the minimization of the expected time to reach detection, make assumptions essentially like (i) and (ii), more or less implicitly. For accounts of this

point of view towards search optimization, see Dobbie [1963], Stone [1975] and the references given by the latter.

We conclude this analysis of cases by citing an important mathematical consequence of the triple hypothesis: (B) passive nature of the observations; (i), subdivisibility, and (ii) additivity. It is then shown that the detection function is of exponential form $\{1 - \exp(-wz)\}$, where z is the amount of searching effort and w is a function of the system variables noted in (B). This generalizes the law of random search, introduced in the U.S. Navy during the last war (Koopman[1946]) and applied to an ever wider range of problems in the ensuing decades. It is discussed at length in the proof of formula (1) below.

Later we go into more detail and illustrate these generalities by applications to concrete problems.

Generalities on the Theory of Optimal Search

The cost of search. As noted under C above and indicated by earlier examples, every search entails the expenditure of a certain amount of effort: its "cost." In peacetime, this is usually expressible in financial units: in the search for minerals the various steps have to be payed for in money; the cost of police detection is in salaries and equipment-- to which may be added the interval of time expended in cases when this is a critical element, and in some cases, in political or social liability (non-economic costs). In military cases the financial cost is in procurment of the equipment for planned types of search: the platforms, detecting elements, the trained personnel, etc.

On the other hand, during active war the cost is in what is available, under the limitation of its transportation, etc. When, as in the past, a duration of months or years could be expected, the rate of production of essentials was the crucial factor rather than their cost in money: the difficulty was sometimes described by saying that "we could not spend money fast enough." In a medical emergency, the cost may be in the time spent to get to the patient, and it may be better to shorten this period of time than to delay by seeking the very best medical authority or hospital.

After these generalities regarding what might be called the "strategic" side of cost, we pass to the "tactical" factors it presents in a search at a given time and place: we are then restricted to what is available in men, material and time that are needed. We continue to use the military example to introduce the ideas. We are thus dealing with a problem of the optimal allocation of resources available to us at the time, in view of

the results to be accomplished and in the face of the possible dangers incurred. Now the financial economist as such ceases to have a role in the decisions. (Good military economics are a necessary condition for preparing for a war-- but they are certainly not a sufficient condition for fighting it-- to the apparent surprise of some experts.)

In all but the simplest cases a great diversity of factors enters in the formulation of the problems of optimal search: we must not only detect the target, but be sure that it is one and determine its location with sufficient accuracy for our purpose; there are also factors of timeliness and danger. For example, in an anti-submarine search the density of targets may be much greater near enemy harbors than in mid-ocean-- but so will be the danger to the searchers. Again, if we have obtained "special intelligence" of a concentration of targets in a particular locality, the temptation of intensifying our search there may give evidence to the enemy that we have in fact gained such knowledge-- possibly by breaking his codes. We recall the historical decision regarding the sacrifice of Coventry in World War II: a concentrated defense of this city might have revealed the Allied understanding of the German code ENIGMA--and possibly lost them the War.

Such multiplicity of factors in designing optimal searches evidently poses a dilemma: to take them all into account (which, realistically speaking, is impossible)-- or to give but an incomplete basis for rational action. Science brought to bear upon practical realities is constantly faced with such situations: it can handle them (on the spot, i.e., without waiting for further research to be completed) by separating them into parts that are simple enough to be examined by approximate methods, seeing how these parts show various effects, then recombining them and trying other assumptions: it is, of course, a tentative and constantly readjusted set of conclusions--constantly to be tested by comparing the results with observations. We can say about good mathematical knowledge--as we did about good military economics-- that it is very likely to be a necessary condition for answering the practical questions; but it is not a sufficient condition. The perception of enlightened simplifications is more of an art than a science. In the case of the theory of search we may be guided by a fifth basic element to be added to those denoted above by A, B, C. and D. It is the following two-fold one:

> E. The mathematical formulation of search problems should be carried to a refinement consistent with (i) the degree and accuracy of the searcher's knowledge of the factors and quantities affecting it; and (ii) with his ability to implement the theoretical optimum resulting from the mathematics. On the latter

point, theoreticians have often appeared to ignore that the most carefully traced track that the searcher's platform can be made to follow is but a crude imitation of any theoretical optimal distribution. The latter has shown its usefulness in many cases, since the departures from its standard can often be approximately found; then 70% of the theoretical optimum probability of detection can be regarded as satisfactory, whereas 15% should lead to a revision of the plan. Such examples have occurred.

A Classical Problem and its Generation

In his work for the U.S. Navy during World War II, one of this author's tasks involved the improvement of sonar escort screen and air escort plans in defense of merchant convoys. Those involved in such problems knew that the number of defensive units were tragically insufficient for tight protection; therefore the optimal use of those we had was a matter of life and death. Both observation and reasoning indicated that the attacking U-boats had preferential directions of penetration; from the two foreward sides rather than from dead ahead; and of course submerged approach from the rear was kinematically impossible in those days. Could not this fact serve as a guide to improve the placement of the escorts? Another problem concerned the optimal angles of scan in airborn radar. These and similar cases, such as the improvement of surface search for floating objects, life rafts, etc., were formulated mathematically by Koopman [1946] and their results were used and extended during the war and in the decades following. Since this whole subject has been amply recorded in many places in the published literature (Koopman [1946, 1956, 1979]; Dobbie [1963]; Stone [1975], with full references in the latter), we limit ourselves here to a bare outline, and exemplify the ideas by a simple generalization capable of practical applications. We hope in doing this to give some indication of the rather meticulous analysis of physical and operational assumptions that has to be made before the mathematical formulation can ever begin.

We start by making the four assumptions, the nature of which has been explained earlier: (i) the search is exclusively by passive means; (ii) there are a finite number of variables determining the detectability of the target: the "system variables;" they include the coordinates of position (x) and possibly "internal" variables (y); (iii) the searching is subdivisible and additive; (iv) each part can be regarded as "brief," no extended image build-up time.

Under these conditions it is simple to show by factual reasoning (rather than on the basis of pure mathematics) that the parts of the search combine as independent trials: the probability that the whole search fails is the product of the individual failure

probabilities. This fact will be applied to get the form of the detection probability as a function of the search effort under the very special but important case that the search is localized: our assumption (v). Thus we shall be able to derive a more explicit form of the detection probability than the exponential one cited earlier. The situation is of course that of short range detection; but under rather special conditions, as we shall now see.

We embody their description in our final assumption, (vi), needed to explain precisely what we mean by "search density": the function of position (and such variables) that expresses mathematically our distribution of searching effort. The basic assumption (vi) is that there are three orders of size: the "large" dimensions are those of geographic order, over which the probability density $p(x)$ usually varies widely; "small" dimensions are those of the order of the searching platform and the detection ranges of its sensors, so that searches conducted inside a "small" region are able to detect about as many targets outside it as within it; and between the large and the small dimensions we assume those of "intermediate" size. The latter are characterized by three attributes: the target distribution $p(x)$ is sensibly constant through intermediate regions; the method of search bears on the average about as heavily at the different points of such regions; and the search within a particular one of them has but a negligible chance of detecting outside, as compared to that of detecting targets within.

Under these assumptions, the search density $\phi(x)$ is a "smoothed interpolation function" of the ratios of the expected amount of searching within the regions of intermediate size, described about various points in the target region, to the area of the former. Although it would be physically meaningless to contemplate the limit of any such ratio (as one does in defining the material density of a lamina in classical mechanics), it is clearly possible -- at the level of the approximations permissible in search theory -- to assume whatever regularity is convenient in regard to this function, and to express the total amount of search in any intermediate or large portion of the search region as an integral of the function $\phi(x)$ over it.

We now refine the general factors noted above under B and C as they apply in the present situation --our assumptions (i) through (vi). As is shown in the references cited above, particularly Koopman [1979], the detection probability is a function of the system variables (x, y) and density values $z = \phi(x)$ defined as the conditional probability of detecting the target given to be at (x) and with the internal system variables the given numbers (y)-- and that the search density is $z = \phi(x)$. The very important point-- that without our very special assumptions (i)- (vi) the detection

probability could be a functional and not a function of $\phi(x)$--has been repeatedly emphasized by J. M. Dobbie. By combining our six assumptions, the earlier general exponential form noted at the close of Part I now takes on the more detailed form $D(x,y,z)$ given, as proved, e.g., in Koopman [1979] by the equation

$$(1) \qquad D(x,y,z) = 1 - \exp[-w(x,y)z], \qquad z = \phi(x).$$

Here the function $w(x, y)$ may be regarded as a coefficient of detectability, generalizing the sweep width W in the simpler cases considered orginally by Koopman [1946].

If we were given the total probability distribution $P(x, y)$ in system space, the problem of optimal search could be put into the following form:

Find that distribution $z = \phi(x)$ that maximizes the detection probability $P[\phi]$, given as the integral throughout the space of the variables (x, y)

$$(2) \qquad P[\phi] = \iint_{XY} P(x,y)\{1-\exp[-w(x,y)\,\phi(x)]\}dxdy$$

under the conditions that the total available searching effort is Φ and of course $\phi(x)$ is nonnegative:

$$(3) \qquad \int_X \phi(x)dx = \Phi, \qquad \phi(x) \geqslant 0$$

Under the natural physical assumptions of regularity of the functions involved in this unilateral problem of the calculus of variations, the necessary conditions to be satisfied by the function $\phi(x)$ are easily obtained. In the classical case, the situation consisted in assigning approximately known values to the internal variables (y); and then solving the resulting equation explicitly in terms of p(x)--to which P(x,y) reduced--and Φ. When, on the contrary, the variables (y) were either unknown, or worse, varied during the approach in a manner depending on the detailed nature of the paths taken by both target and searcher, as would the target aspect and hence its reflectivity to echoes, the simplification mentioned does not apply. Every attempt was made during and after the war to find a realistic method of taking such unknown behaviour of (y) into account, but until this day there seems to be no objectively based way to treat this question. It is easy enough to write down the necessary mathematical conditons for the optimization, but they have the form of an integral equation involving the unknown function $\phi(x)$ in the exponent. Even then the difficulty is not mathematical: modern computers could obtain numerical solutions. The problem is that we do not have enough realistic knowledge concerning the joint distribution function $P(x,y)$.

De Guenin [1961] has tried to generalize the problem by replacing the detection function D(x,y,z) in (1) by a function b(x,z), that is not exponential, but has some of the qualitative properties of the exponential expression. Many subsequent authors have built mathematical theories upon this more general function as a basis. Recently we have shown (Koopman [1979]) that under the assumptions (i) - (vi), the above equations lead to the fact that b(x,z) must be unity minus the Laplace transform of a bounded non-decreasing function. In spite of this rather drastic constraint on de Guenin's function, the essential difficulties--lack of reliable knowledge--remain. Accordingly, we shall continue our discussion of the present subject on the old workable--if philosophically imperfect--basis, and leave out the variables (y).

We shall make another generalization of the problem of optimum search by adding to the constraint of economy of searching effort (3) a second one regarding the danger to searchers operating near hostile coasts. We suppose that the danger to the searchers arises from a different sort of enemy unit from the target, having (in general) a different location distribution p'(x), but a detection capability against the searcher that is similar to the searcher's capability against the target. We assume a constant w = W, the sweep width; this implies that ϕ is length per unit area (or speed times time/area) so that $W\phi(x)$ is zero-dimensional. Also, that any encounter of a searcher with one of these hostile units produces a probability of killing the former of k, and that there are N such units in all. Then the expected number of kills per searcher per unit area at (x) is g(x) = Nkp'(x). Since the number of searcher-hours per unit area at (x) has an expected value proportional to $\phi(x)$, we may, in terms of appropriate units, express the danger density at this point as the product $g(x)\ \phi(x)$ giving for the total searchers' danger the integral--where g(x) is (unfortunately) nonnegative--

$$(4) \qquad G[\phi] = \int_X g(x)\phi(x)dx.$$

This constraint could, of course be added to (3) in the earlier formulation of the problem of optimization; but the practical view of the issues facing the decision maker becomes clearer if we state our problem in terms of a "value judgment" by introducing weighting, or trade-off, constants; basically, two in number. Let the ratios of the three disadvantages: the failure to detect $Q[\phi] = 1 - P[\phi]$, expenditure $\Phi[\phi]$, and danger $G[\phi]$ be given by the numbers 1, λ, μ. Then our problem is reduced to the choice of a nonnegative $\phi(x)$ which minimizes the linear combination

$$(5) \qquad L[\phi] = Q[\phi] + \lambda\Phi[\phi] + \mu G[\phi].$$

The fact that the Lagrange multipliers λ and μ enter in this manner into the above modification of the problem has long been familiar in the calculus of variations. In the unilateral case in which one or more inequality constraints are present, many more or less sophisticated methods have been developed. We shall use a very simple one suggested long ago to the author by the late Dr. Glenn Camp, then at OEG; it consists in using the necessary property of $\phi(x)$ being nonnegative, and hence having a real positive square root $f(x)$, so that $\phi(x) = (f(x))^2$. Making this replacement in all the quantities involving $\phi(x)$, we ignore the second condition in (3) and set $\delta L[\phi] = 0$.

The usual reasoning in deriving the Euler necessary conditions to be satisfied by $f(x)$ leads to the equation:

$$(6) \qquad [-p(x)W\exp\{-Wf(x)^2\}+\lambda+\mu g(x)]f(x) = 0,$$

from which we conclude that either $f(x)=0$ or that the square bracket vanishes.

Returning to the original function $\phi(x)$, the condition becomes the following: Either $\phi(x)=0$ or else it must satisfy the equation:

$$(7) \qquad -p(x)W\exp\{-W\phi(x)\}+\lambda+\mu g(x) = 0.$$

At points where $p(x)\neq 0$ we can solve the equation, obtaining:

$$(8) \qquad \phi(x) = (\log[Wp(x)/\{\lambda+\mu g(x)\}])/W.$$

These results show that there will in general be a region that it is appropriate to search: we denote it--- as well as its area -- by S, and its complement (not to be searched) by $\bar{S}$. They are evidently defined by the inequalities:

$$(9) \qquad S = \{x:Wp(x) > \lambda+\mu g(x)\}; \quad \bar{S} = \{x:Wp(x) \leqslant \lambda+\mu g(x)\};$$

whereas their boundary is given by taking the equality in (9).

Evidently the Lagrange multipliers λ and μ are found from (3), (4), (7), and (8):

$$(10) \qquad Q = (\lambda S+\mu K)/W, \quad \Phi = \int_S \log[Wp(x)/\{\lambda+\mu g(x)\}]dx/W;$$

$$G = \int_S g(x)\log[Wp(x)/\{\lambda+\mu g(x)\}]dx/W;$$

where

$$K = \int_S g(x)dx.$$

It can be seen that λ and μ can approach 0 inside S, which would correspond to dropping the respective (constraining) term in (5). By their definition they are nonnegative. On the other hand, they cannot increase indefinitely in S; this would impose excessive constraints, leading into $\bar{S}$.

Indeed, as either λ or μ increase from 0 indefinitely, the region of search S contracts and Q, Φ, and G decrease monotonically: this is as it should be, since either type of change amounts to increasing the constraints upon the search. It also shows mathematically that the parameters are determined by the values of Q and G.

To obtain a simple operational interpretation of these results, we consider a straight length h of a searcher's track (of intermediate size), and assume a definite range law, of range W/2; then he "sweeps clean" the rectangle of area hW, and if (x) is a point in this rectangle, hWp(x) is the probability that he detects a target during this part of his track: denote it by D_h. Thus, multiplying by h, (9) shows that in S, $D_h > \lambda h + \mu h g(x)$. To interpret this, consider first the classical case in which the search involved no danger, so that $\mu = 0$ in all the above equations. Then the condition for searching in S is determined purely by limited effort optimally expended. The condition is that $D_h/h > \lambda$; this explains the meaning of λ as the least detection probability per unit path length permitted if the search at (x) is worthwhile.

At the other extreme, suppose that "cost is no object"--only a search with reasonable safety. We set $\lambda = 0$ and have the condition for searching at this point, $D_h/h > \mu g(x)$; thus μ is the least permitted detection probability per unit track per unit density of danger. Equivalently, it is the greatest allowed danger density per unit "specific track detection probability"--the latter being the verbalization of the ratio D_h/h. This ratio evidently contains the idea of an exchange rate, although not exactly as used in the Battle of the Atlantic, in which it referred to defended units lost versus enemy submarines destroyed.

This example can be generalized, with a wide variety of different cases and interpretations given.

References

DeGuenin, J. [1961] Optimum Distribution of Effort: An Extension of the Koopman Basic Theory, Oper. Res. 9, pp. 1-7.

Dobbie, J. M. [1963] Search Theory: A Sequential Approach, Naval Res. Logist. Quart. 4, pp. 323-334.

Koopman, B. O. [1946] Search and Screening, OEG Report No. 56, Washington, D. C. (a modernized edition has been published by Pergamon Press, Inc., New York, 1980).

Koopman, B. O. [1956] The Theory of Search, Parts I, II, and III, Oper. Res. 4, pp. 324-348 and 503-531; 5 [1957] pp. 613-623.

Koopman, B. O. [1957] Quantum Theory and the Bases of Probability in "Procs. Symp. in App. Maths. VII," McGraw-Hill, New York, pp. 97-102.

Koopman, B. O. [1979] An Operational Critique of Detection Laws, Oper. Res. 27, pp. 115-133.

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

AN OVERVIEW OF COMPUTER ASSISTED SEARCH INFORMATION PROCESSING

Henry R. Richardson
Thomas L. Corwin

Daniel H. Wagner, Associates

This paper provides an overview of computer assisted search (CAS) information processing. The objective is to give the reader an idea of how computers can be used in real time search planning to combine subjective assumptions with search results to provide a better answer to the question "Where is the target?"

The paper is organized into six sections and an appendix. The first section presents the framework of information processing in the context of a "Monte Carlo" type system. While this is not the only type of CAS system in use, it is the easiest type to describe in non-technical language. The first section also describes the "multiple-scenario approach" which is characteristic of most CAS systems in use today. The remaining sections discuss the details of information processing as they pertain to the following five basic functions :

Probability map display
Generating an intial probability distribution
Updating for target motion
Updating for negative search results
Updating for positive search results.

The appendix presents a mathematical description of the target motion models used in CAS systems today.

Computer-Assisted Search Information Processing

This section presents an overview of computer-assisted

search information processing. The discussion is given in terms of a Monte Carlo search system because this type of system is easy to describe and because the U.S. Coast Guard CASP system (Richardson & Discenza[1980]), and several important U.S. Navy systems are of this type.

It should be noted, however, that Monte Carlo type systems usually require large computers to achieve the substantial number (1,000 to 10,000) of replications needed to obtain meaningful results. For small computers, non-Monte Carlo ("analytic") methods have been devised in order to achieve the desired accuracy in less time (Corwin[1979]). These methods are extremely useful within their scope of applicability. However, they tend to be less flexible than the Monte-Carlo methods since they require the search problem to be modeled in terms of tractable mathematical expressions.

Most major CAS systems today make use of a multiple-scenario approach, and this term will appear frequently below. The multiple-scenario approach was suggested by Dr. John P. Craven during the Mediterranean H-bomb search in 1966. It was later employed successfully in the 1968 search for the U.S. nuclear submarine Scorpion (Richardson & Stone[1971]).

As the term is presently used, the multiple-scenario approach means the formulation of several different probabilistic descriptions ("scenarios") for target location at some reference time, including stochastic process assumptions about future target movement when appropriate. Each scenario is given a subjective probability weight in order to express its intuitive plausibility. The distributions for the separate scenarios are combined into a composite distribution by use of these weights. Bayesian methods are then used to update the probability assumptions as information is gained from the search.

Information processing, as defined here, consists of the five basic functions mentioned earlier. The purpose of this section is to indicate how these functions are implemented in a typical Monte Carlo type CAS system.

Figure 1 shows a schematic of a CASP-like computer assisted search system. A computer program for one of the five functions is loaded into the computer's central processing unit. The search informaton processing is done sequentially by taking a record (described below) from an input "target state file," processing it, and then returning it in updated form to an output target state file. These files are usually quite large and for this reason are generally stored on a disc rather than core memory.

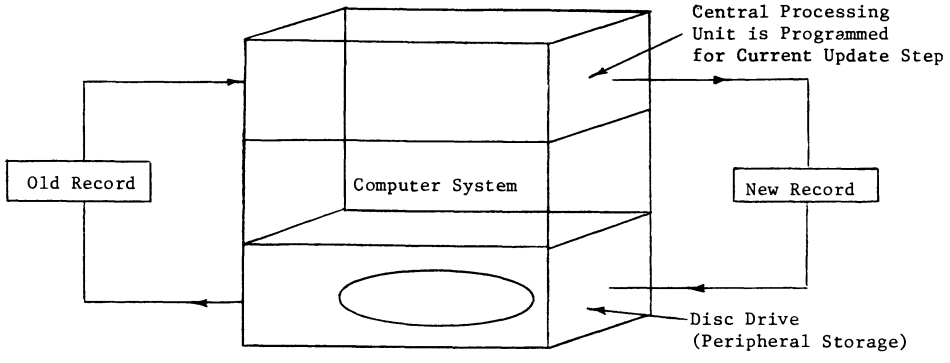


Figure 1. Search Information Processing

Each target state record can be represented by the vector

$$(X, Y, V_X, V_Y, T, K, W),$$

where

(X, Y) = Target's current position

(V_X, V_Y) = Target's current velocity

T = Time of "distress" incident

K = Index of search scenario

W = Probability weight for the target state record

Note: The current CASP system (Richardson & Discenza [1980]) has fewer variables in the target state record than the formulation described here.

Figure 2. Typical search information processing state record.

Each "target state record" (see Figure 2) is an ordered collection of numbers which represents information about the location and motion of one target possibility. Each Monte Carlo replication produces one target state record.

In the example shown in Figure 2, a record consists of seven numbers. The first two of these give the coordinates of the target's current position, and the next two give the components of the target's velocity. These are followed by the simulated time of the distress incident, the scenario index, and the probability weight for the information contained on the record.

The probability weight expresses the likelihood that the target state record is correct. The target state records are created by Monte Carlo sampling at the time that the initial target location probability distribution is generated. Initially all records are assumed equally likely and thus all weights W are equal. The weights are subsequently modified in Bayesian fashion during updates for positive and negative information as described below.

Probability Map Display

The single most important output of a CAS system is the target location probability map display. The probability map allows the search planner to identify the most likely regions for target location and serves as a basis for allocating search resources in a systematic fashion.

As an illustration, Figure 3 (taken from Richardson & Stone[1971]) shows the intial target location probability map for the remains of the submarine Scorpion. The search cells are re about one mile on a side; the number in each cell denotes the number of times that the Monte Carlo simulation placed Scorpion in that cell. The numbers sum to 10,000, the number of replications used. Scorpion (indicated by *) was eventually found within 300 yards of the highest probability cell.

The target location probability maps are generated from the target state files by summing the weights associated with target locations in the region of interest. That is, if S denotes the region of interest (e.g., a search cell) and if

$$(X_j, Y_j, V_X^j, V_Y^j, T_j, K_j, W_j)$$

denotes the target state record for the jth replication, then the probability that the target is located in S is given by

$$\Pr\{\text{target in } S\} = \sum_{(X_j, Y_j) \in S} W_j \Big/ \sum_{j=1}^{J} W_j$$

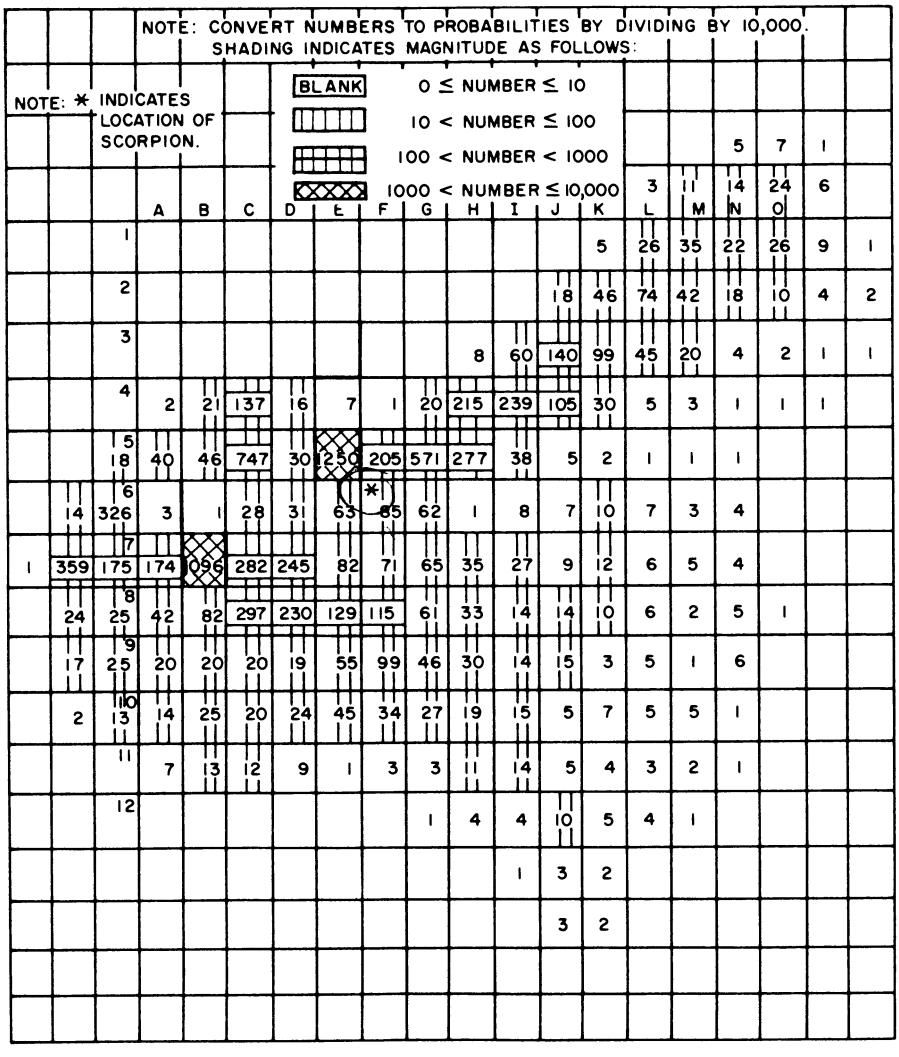


Figure 3. Initial target location probability map for the U.S. nuclear submarine SCORPION.

where J denotes the total number of replications. The summation in the denominator is needed to properly "normalize" the result to be a value between 0 and 1.

Generating an Initial Probability Distribution

In order to illustrate the generation of an intial target location probability distribution, the procedure used in the U.S. Coast Guard CASP system is described (Richardson & Discenza[1980]).

It is not intended to explain the mechanics of Monte Carlo simulation nor is it necessary for the reader to be familiar with the techniques. We wish to indicate the kind of search information which can be used by a CAS system and, in particular, the scenario "building blocks" used by CASP.

The CASP system is intended for use in open ocean search and rescue and, to a large extent, is a crystallization of ideas formulated in the H-bomb and Scorpion searches extended to moving targets. The initial probability map for the target at the time of distress can be constructed from three different types of scenario building blocks. These are referred to as "position," "area," and "trackline."

A "position" scenario (see Figure 4) applies to a situation where the target's location at a specified time can be modeled by a bivariate normal (bell shaped) distribution. A random displacement of the target from the initial point can also be included when appropriate. Such a scenario might be used, for example, to account for the point at which a pilot bailed out of an aircraft and the subsequent displacement based upon parachute drift.

The "area" scenarios permit the intial target location uncertainty to be uniformly spread over user designated regions as shown in Figure 5. Among other things, the area scenarios are intended for use in cases of lost fishermen where the designated regions can describe fishing grounds where the distress may have occurred.

The third and final type of CASP scenario is the "trackline" scenario (see Figure 6). This scenario is used in the case where a base track may be postulated based upon a float plan or other information. Since the exact track followed by the distress unit is

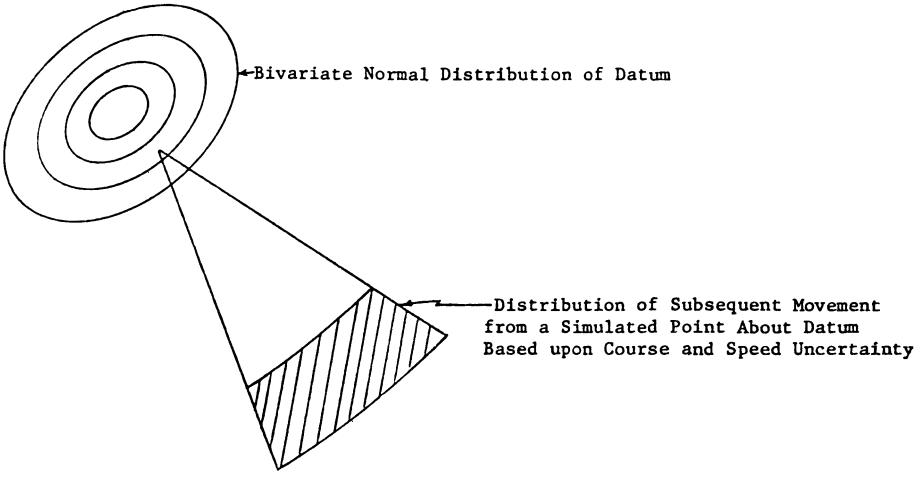
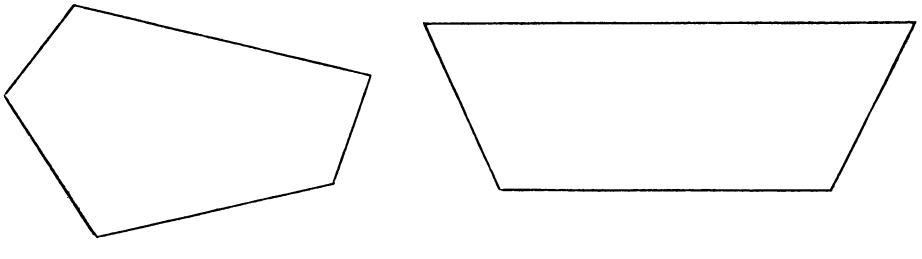


Figure 4. CASP Initial Target Location Probability Distribution (Position Type)



Note: An area scenario specifies a uniform probability density within a convex polygon. The illustration shows two such scenario possibilities.

Figure 5. CASP Initial Target Location Probability Distribution (Area Type)

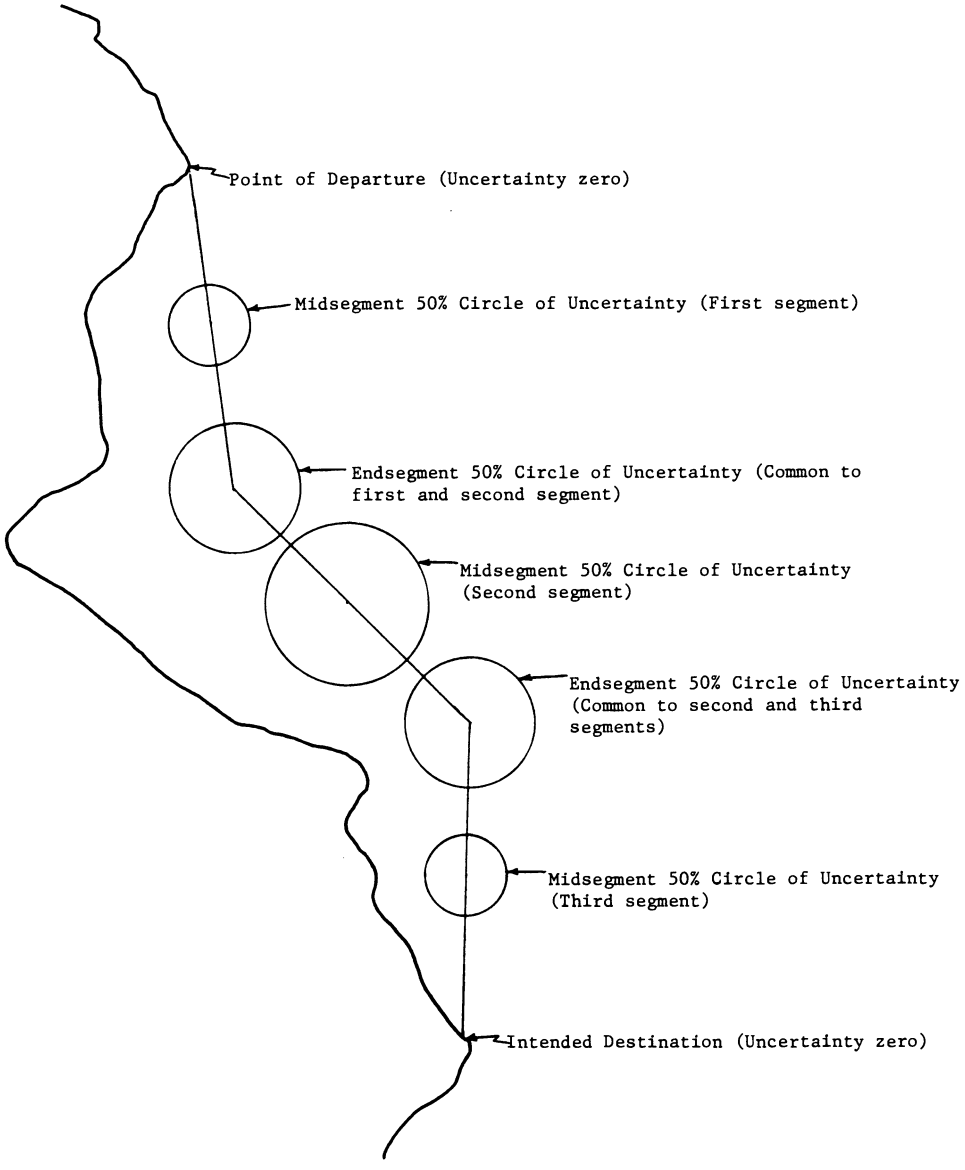


Figure 6. CASP Initial Target Location Probability Distribution (Trackline).

not known, the search planner can generate a large number of simulated tracks about the base track. This is done by user specification of endpoint and midpoint uncertainties for the actual track's location. The search planner also can specify a probability distribution for the location of the distress along track. For example, if bad weather is present in a certain area, the search planner can make this the high probability area for the distress.

The final probability distribution can be a combination of all of the above types if desired. The creation of a target state record begins with a random number draw to select the scenario K. The scenarios will be selected in proportion to their user-input probability weights. Next the computer makes random number draws to determine target position from the probability distribution specified by scenario K. In the present CASP system, the target state record is limited to the three numbers needed to express the target location (X,Y) and weight W. Other systems make use of additional variables as illustrated in Figure 2.

After a target state record is generated, it is placed in the output target state file. Additional records are created until the desired number of replications is achieved.

Updating for Target Motion

Updating for target motion is a sequential operation on the target state file. In order to describe the principle, let Δ denote the target motion update interval and let

$$(X,Y,V_x,V_y,T,K,W)$$

denote the target state record before updating. A typical motion update will modify X,Y,V_x , and V_y but leave T, K, and W unchanged.

In the simplest case, the updated position (X',Y') is given by moving the original position (X, Y) forward a time step Δ using the velocity vector (V_x,V_y), i.e.,

$$(X',Y') = (X+V_x\Delta, Y+V_y\Delta).$$

It is usually assumed that the velocity can change at the end of the update period in a way which depends upon the state of the system. If H denotes the mathematical function which describes this change, then one may write the new velocity (V_x, V_y) as

$$(V_x' ,V_y') = H(X',Y',V_x,V_y,T,K,Z),$$

where Z denotes a vector of random numbers.

In the CASP system, for example, the target motion update takes into account drift due to ocean current, wind generated currents, and "leeway" (i.e., the drift induced by the force of the wind against the side of the distress unit). In certain other CAS systems, these steps are considerably more complicated and are based upon stochastic process considerations described in the appendix.

Updating for Negative Search Results

If search has been carried out for a period of time and the target not found, then certain "negative information" has been gained about the target's location. This negative information is used in CAS systems to modify the weights associated with the target state records. The principal tool involved in updating for negative and positive information is Bayes' theorem, and, hence, this type of updating is often referred to as Bayesian information processing.

In order to explain the calculation, let

J = the number of target records,

Q(n,j) = the probability that the jth record represents the target after n increments of unsuccessful search for $1 \leqslant n \leqslant N$ and $1 \leqslant j \leqslant J$, and

F(n,j) = the cumulative probability that the jth sample would not have been detected by n increments of search.

Further, let A and B denote the following events:

A = "Record j represents the target."

B = "No detection for n search increments."

Bayes' theorem states that

$$\Pr\{A|B\} = \Pr\{B|A\}\Pr\{A\}/\Pr\{B\}.$$

This is a simple consequence of the definition of conditional probability. Accordingly, we have that

$$(1) \quad Q(n,j) = \Pr\{A|B\} = \Pr\{B|A\}\Pr\{A\}/\Pr\{B\}$$
$$= F(n,j)Q(0,j)/\sum_{i=1}^{J} F(n,i)Q(0,i)$$

In order to be able to compute target location probabilities correctly, it is necessary for the target state weights $W_j(n)$ to be proportional to the probabilities $Q(n,j)$. In view of equation (1), this can be accomplished by setting $W_j(n) = F(n,j)\,Q(0,j)$ since the denominator of equation (1) does not depend upon j.

A convenient recursive formula can be used to update the weights when all increments of search are mutually statistically independent. In this case, $F(N,j)$ may be written

$$(2) \quad F(N,j) = f(1,j)f(2,j) \ldots f(N,j) = F(N-1,j)f(N,j),$$

where $f(n,j)$ is the conditional search failure probability for the nth increment $(1 \leqslant n \leqslant N)$, given that the jth record $(1 \leqslant j \leqslant J)$ represents the target.

Since $W_j(n) = F(n,j)Q(0,j)$, equation (2) gives the simple recursion

$$W_j(n) = F(n,j)Q(0,j) = f(n,j)F(n-1,j)Q(0,j)$$

$$= f(n,j)W_j(n-1).$$

Notice that if the position coordinates of a target state record indicate that the target is in a well-searched region (i.e., a low value for failure probability), then the weight of that record will be reduced relative to the other records after the update for negative information. The net result will be to lower the target location probability in this region in accordance with intuition.

Finally, when needed, the probabilities $Q(n,j)$ can be computed in terms of the weights $W_j(n)$ by

$$Q(n,j) = W_j(n) / \sum_{i=1}^{J} W_i(n).$$

The cumulative probability that the search would have detected the target during n increments of search is an important measure of effectiveness used in CAS systems. For brevity, this probability is usually referred to as search effectiveness probability, SEP. It can be computed in terms of the weights W by the formula

$$SEP(n) = 1 - \sum_{j=1}^{J} F(n,j)Q(0,j) = 1 - \sum_{j=1}^{J} W_j(N),$$

where n denotes the number of search increments considered. The first equality above results from the definition of SEP, and the second equality results from the formula for the weights.

Single increment search effectiveness probability (SSEP) is a measure of effectiveness related to SEP. It is defined as the conditional probability of detection for a single increment of search, given that the previous increments have failed. SSEP can also be computed in terms of the weights using the formula

$$SSEP(n) = \{SEP(n)-SEP(n-1)\}/\{1-SEP(n-1)\} =$$

$$= 1 - \{\sum_{j=1}^{J} W_j(n)\}/\{\sum_{j=1}^{J} W_j(n-1)\}$$

Updating for Positive Information

Positive information is the term used to denote the location information associated with a target-related contact. A subjective probability credence c ($0<c<1$) is specified in order to express the likelihood that the contact is associated with the target. The updated weights for the target state records are then computed as an average of the two cases where (1) the contact is assumed true, and (2) the contact is assumed false.

Consider first the case where the contact is assumed true. Let (U,V) denote the contact's "observed" location, and let g(U,V,X,Y) denote the known probability density for a contact at (U,V), given the target is located at (X,Y). The function g is assumed known. Then one applies Bayes' theorem as follows:

$$\Pr\{\text{Sample } j \text{ represents the target} \mid \text{Contact at } (U,V)\}$$

$$= \Pr\{\text{Contact at}(U,V) \mid \text{Sample } j \text{ represents the target}\} \times \Pr\{\text{Sample } j \text{ represents the target}\}/\Pr\{\text{Contact at } (U,V)\}$$

$$= \frac{g(U,V,X_j,Y_j)W_j}{\sum_{i=1}^{J} g(U,V,X_i,Y_i)W_i}$$

where (X_j,Y_j) and W_j are respectively, the target's position and the weight associated with the jth record ($1 \leq j \leq J$).

The second case is easier. If the contact is assumed false, then

$$\Pr\{\text{Sample } j \text{ represents the target} \mid \text{Contact is "false"}\}$$

$$= W_j / \sum_{i=1}^{J} W_i$$

Putting these two cases together we obtain the following formula for the updated weight W_j':

$$W_j' = c \frac{g(U,V,X_j,Y_j)W_j}{\sum_{i=1}^{J} g(U,V,X_i,Y_i)W_i} + (1-c)\frac{W_j}{\sum_{i=1}^{J} W_i}$$

Thus, if the contact's credence c is large, a high weight will be associated with target state records which place the target in the contact's high probability area. If the contact's credence is small, then the relative weights will change very little.

The updating for positive information alters the weights in a way in which the SEP calculation is no longer meaningful. In effect, positive information changes the search problem in a way which could not be anticipated. Usually the weights are renormalized after updating for positive information so that $\Sigma_{j=1}^{J} W_j = 1$. This sets SEP = 0, and SEP will then measure the effectiveness of all search subsequent to the last contact. Alternatively, in some CAS systems, one can also normalize the weights so that $\Sigma_{j=1}^{J} W_j = 1-SEP$. When this is done, SEP will be unaltered by the contact, and continue to increase as additional search is applied.

APPENDIX

Stochastic Process Models for Target Motion in Computer Assisted Search

The target motion models in the CAS programs are constructed using certain types of stochastic processes. These processes, called the "building blocks" of target motion are Markov processes which take values in the state space $\mathbb{R}^n$. The sample paths of these processes have discontinuities of only the first kind. Such sample paths may jump from one continuous segment to another. In general the state space associated with a target describes other attributes of the target than just position, such as velocity or positions at specific times. Following Skorokhod [1965], processes of this sort may be described in terms of an object called a stochastic equation. A stochastic equation is composed of two parts --one part describes the initial distribution of the process on $\mathbb{R}^n$ and the second describes the mechanism by which the target changes position in $\mathbb{R}^n$. The stochastic equation describing the class of models used in CAS appears as follows:

For $t \geq t_0$,

$$(A\text{-}1) \quad dX(t) = [Af(t)\, X(t) + g(t)]\, dt + B(t)\, dW(t) + \int_{\mathbb{R}^m} h(t,X(t),u)\, p_{\lambda 1}(dtxdu)$$

and $X(t_0)$ has initial probability measure μ, where

i) A is an n x n real matrix,

ii) $f:[t_0,\infty) \to \mathbb{R}$ is bounded and continuous where $\mathbb{R}$ denotes the real numbers,

iii) $g:[t_0,\infty) \to \mathbb{R}^n$ is bounded and continuous where $\mathbb{R}^n$ denotes n-dimensional space,

iv) $B:[t_0,\infty) \to \{n \text{ x } n \text{ real matrices}\}$ is bounded and continuous in each entry,

v) $h:[t_0,\infty) \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is bounded and continuous in each argument,

vi) $dW(t)$ represents the increment of an n-dimensional Wiener process defined on $\mathbb{R}^n$,

vii) $p_{\lambda 1}(dtxdu)$ represents the increment of a point process defined on $[t_0,\infty) \times \mathbb{R}^m$ which is such that:

a) for $(\tau_1,\tau_2) \times C \in \mathbf{B}([t_0,\infty) \times \mathbb{R}^m$, $p_{\lambda 1}[(\tau_1,\tau_2) \times C]$ has a Poisson distribution with intensity

$$\lambda(\tau_2-\tau_1) \int_C 1(x)\, dx,$$

b) where $\lambda \geq 0$, 1 is a probability density and $\mathbf{B}([t_0,\infty) \times \mathbb{R}^m)$ are the Borel subsets of $[t_0,\infty) \times \mathbb{R}^m$,

viii) μ is a probability measure defined on $\mathbb{R}^n$.

It should be noted that the representation given by equation (A-1) is a concise representation for a fairly narrowly defined type of target motion. In point of fact, target motion in CAS programs is actually constricted by linking models of this type together. The linking is of two types. For example, a member of the class of models for target motion described by the representation given by equation (A-1) may be adequate to model the motion of a target between two specific times, say t_0 and t_1; target motion for the next time period, say t_1 to t_2 may be better represented using a different member of this class. In such a case the two models are linked together. This type of linking, called linking in series, is done in such a manner that the distribution of the target's position at time t_1, used as a starting distribution for the second leg of the motion, corresponds with the distribution of the target's position (not conditioned upon search results) at time t_1 as predicted under the first model. Links of this sort may be made arbitrarily long if the tactical situation warrants.

A second way of combining the models discussed here may be called linking in parallel. A linking of target motion models of the sort discussed in the previous paragraph is called a target motion scenario. CAS programs typically use many scenarios to describe the motion of the target. Each scenario is given a weight. The scenarios are linked in parallel by computing an average scenario equal to the weighted linear combination of the individual scenarios. Scenarios may be distinguished by different starting times t_0. This dual linking of Markov processes of the type described by the representation (A-1) is the richest description currently implemented in CAS. Consequently, representation (A-1) describes only the building blocks of the models for target motion currently employed in CAS. Presented below is a description of one such building block called the

"distribution-constrained" Brownian bridge.

The distribution-constrained Brownian bridge is used in some CAS programs to model target transits from the vicinity of the point $v \in \mathbb{R}^2$ to the vicinity of the point $w \in \mathbb{R}^2$ in T time units. The model assumes that a start point V and end point W are chosen from a four-dimensional Gaussian distribution centered on $(v,w)^T$. The target then connects V and W with a Brownian path that starts at V at time 0 and ends at W at time T. The model is displayed schematically in Figure A-1.

This process may be described using representation (A-1). Let $(X_1(t),X_2(t))^T$ be the position of the target in $\mathbb{R}^2$ and $(X_3(t),X_4(t))^T$ be the position of the target at time T, for $0 \leqslant t \leqslant T$. Then,

For $0 \leqslant t \leqslant T$,

$$(A\text{-}2) \quad d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = (T-t)^{-1}\begin{bmatrix} X_3(t) - X_1(t) \\ X_4(t) - X_2(t) \end{bmatrix} dt + \sigma \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}$$

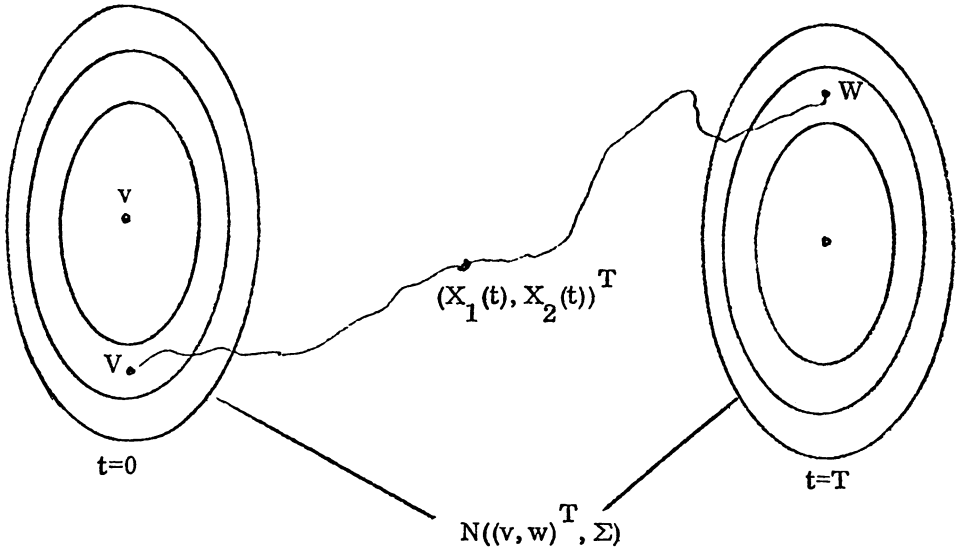
$$d\begin{pmatrix} X_3(t) \\ X_4(t) \end{pmatrix} = 0;$$

where,

$$\begin{pmatrix} X_1(0) \\ X_2(0) \\ X_3(0) \\ X_4(0) \end{pmatrix}$$

has an initial probability distribution which is 4-dimensional normal with mean $(v,w)^T$ and covariance matrix Σ.

Figure A-2 presents examples of the sample paths of a process with representation (A-2).



Note: The target chooses $(V,W)^T \approx N((v,w)^T, \Sigma)$ and connects V and W with a Brownian path which starts a V at time 0 and ends at W at time T. $(X_1(t),X_2(t))^T$ is the position of the target in $\mathbb{R}^2$.

Figure 1A. A Schematic Representation of the Distribution-Constrained Brownian Bridge.

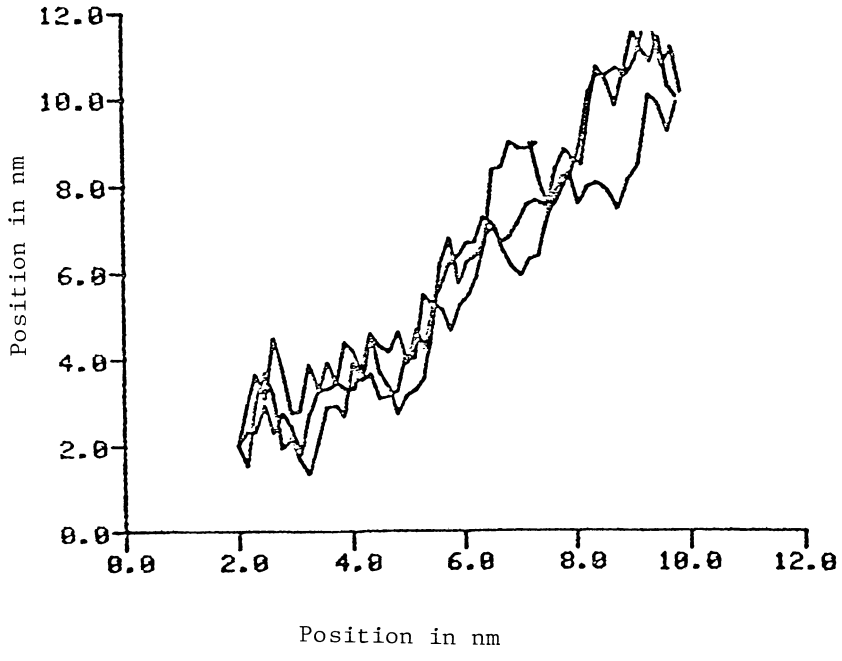


Figure 2A. Sample Paths of the Distribution Constrained Brownian Bridge.

Note: These sample paths were generated under the following assumptions:

i) $\sigma = 1$ nm/$\sqrt{\text{hr}}$

ii) $T = 10$ hrs

iii) $$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 2. \\ 2. \\ 10. \\ 10. \end{pmatrix} \text{ nm}$$

iv) $$\Sigma = \begin{bmatrix} .01 & 0 & 0 & 0 \\ 0 & .01 & 0 & 0 \\ 0 & 0 & .01 & 0 \\ 0 & 0 & 0 & .01 \end{bmatrix} \text{ nm}^2$$

Acknowledgements

We wish to acknowledge the many U. S. activities which have made this work possible. Considerable impetus was given to real-time search planning by the successes achieved in the 1966 Mediterranean H-bomb search and the 1968 search for the remains of the submarine Scorpion. The on-site assistance to both of these searches was provided by the United States Navy Supervisor of Salvage. The Office of Naval Research provided support for theoretical work in information processing and search optimization as well as the implementation of these concepts in various Navy CAS systems. In particular, the work presented in the appendix was supported by ONR Contract No. N00014-78-C-0212. The United States Coast Guard supported the development of the Computer Assisted Search Planning (CASP) system which has served as a prototype for other CAS developments and has been credited with saving more than a dozen lives since it became operational in 1974. The Naval Air Development Center provided support for the first large-scale Navy CAS system in 1972 and expanded the scope of CAS to anti-submarine search.

Finally, we wish to acknowledge our many colleagues at Daniel H. Wagner, Associates who have contributed in numerous ways to CAS. Chief among these is L. D. Stone. Other major contributors (mentioned alphabetically) are W. H. Barker, S. J. Benkoski, S. J. Bloom, S. S. Brown, and F. P. Engel.

References

Richardson, H. R. & Discenza, J. H. [1980] The United States Coast Guard Computer-Assisted Search Planning System (CASP), Naval Res. Logist. Quart. (to appear).

Corwin, T. L. [1979] A Representation for the Posterior Distribution of the Location of a Moving Target, Stochastic Process. Appl. 9, pp. 7-17.

Richardson, H. R. & Stone, L. D. [1971] Operations Analysis During the Underwater Search for Scorpion, Naval Res. Logist. Quart. 18, pp. 141-157.

Skorokhod, A. U. [1965] "Studies in the Theory of Random Processes," Addison-Wesley, Reading, Massachusetts.

SEARCH AND RESCUE WORKING GROUP REPORT

Introduction

Search and Rescue (SAR) is a very demanding activity; its challenges must be met by the practitioners with increased efficiency. Recent developments in search technology are being implemented in computer assisted search systems.

The brainstorming session produced 60 different items which were classified into 4 categories: Management; Search Planning; Technical development and Experimentation; Computer Assisted Search Planning. Recommendations were made concerning operations, implementation of known techniques and development of theoretical work specific to SAR.

The following needs were recognized for enhancing the management aspects of SAR:

i) Interagency and international dissemination of proven methods and doctrine; exchange of technological information on sensor and target characteristics; exchange of training methods.

ii) Development of uniform abstract data base for recording SAR cases.

iii) Public education in preventative SAR in wilderness and marine survival.

In recent years there have been major technological developments. It is important that SAP make maximal use of such advance, e.g., graphical displays, digitizers, microcomputers, etc. Furthermore the interface between the average search planner and the technological devices should be improved so that easily understandable guidelines for decision are rapidly available.

To improve the reliability of the newly developed probabilistic models, it is necessary to obtain better estimates of the parameters needed to specify the prior distribution of the target location. Furthermore, to assess the impact of negative information on the posterior probability distribution, performance of search platforms must be more accurately evaluated.

Finally theoretical investigations need to be carried out in the following areas:

i) Criteria for starting, stopping and restarting a search.

ii) Formulation of single search plan in case of a target which may assume different states.

iii) Application of multi-criteria decision theory to SAR.

iv) Sensitivity of optimal search plans to survivability, motion, detection probability, and discretization.

Management and Human Factors

The following items are discussed in this section: Search Management; Man and Machine Interface; Training and Education and Length of Search.

Search Management. Management of a massive search is a monumental task which can result in utter confusion if not handled correctly. The Searchmaster receives a myriad of information which must be evaluated, weighted and analyzed. He usually has control of numerous resources such as aircrafts, surface vessels and ground teams. All of these assets must be integrated into an effective search plan to find the survivor in the shortest time possible since his life expectancy is time critical. The demanding tasks associated with the Searchmaster duties may require that the function change hands many times during a lengthy search. It is necessary that the new Searchmaster have access to all

information, action and the reason for the action which preceded him so he can apply sound search theory in an effective manner.

Documentation and analysis of mission activities. Detailed documentation of all mission activity is necessary from the onset to ensure continuity and sound analysis ot the effort. This requires the tremendous amount of information received before, during and following each day's activities to be categorized, analyzed and documented. Much of this documentation is presently being accomplished by hand, resulting in incomplete or missing vital information. This in turn results in poor continuity and planning.

A system needs to be devised to apply computer technology in documenting all search mission activities including: information received before during and following each day's activities; assumptions, planning and their reasoning; and search efforts and results. This system needs to interface with search theory and scenario models so inputs can be made quickly and easily by the Searchmaster.

Man and machine interface. Human interface covers many aspects of the search problem. Any search theory or technique must be usable by an operator. Even though the state of the art in search theory is well advanced, it is not being put effectively to general use. This situation is mainly due to the poor interface existing between the average search planner and the computer programs developed to implement the theory. Any procedure or model must be simple enough for the user to understand and use while maintaining sufficient complexity to closely represent the problem. Various aspects of the use of computers in modeling search problems will be discussed.

(i) Intelligent interactive system. The computer system in current use in search planning requires the user to completely formulate the problem in fine detail before obtaining an answer. This requires a very broad knowledge and a well trained operator. There are not a suffient number of highly qualified individuals available to enter the required data. Computer systems could be much more useful and effective if intelligent terminals were programmed to interact with the user. The program could prompt the relatively inexperienced user to supply the required information in an orderly manner. In this way a relatively new Searchmaster could achieve almost the same degree of sophistication as the experienced one.

(ii) Real time systems. Steps must be taken to improve the computer turn around time to allow for more rapid updating of

information. This could allow the Searchmaster to enter new data concerning the search and to receive recommended changes to the search plan while there is still time to take action. Such a capability is particularly needed during searches where clues are rapidly being found. Whenever saving human lives is the reward for a good search, there is need to improve the speed of getting good plans to the searchers.

(iii) Impact of computer technology. Both larger/faster and smaller/cheaper computers are becoming available. The computers currently used for search and rescue are large mainframe systems remotely located from the search site. Communication channels to the computer are generally slow and difficult to use.

Microcomputers and improved input-output devices can significantly improve the interface between user and computer. Intelligent terminals, located at the operating center, could be tied up to a more powerful main computer. The terminal could be equipped with a device which would allow the Searchmaster to "draw" his ideas. A digitizer with a map overlay or a video graphic input system could perform such a function. The data could be initially processed locally and, only when considered correct, sent to the main computer for extensive processing. The output should be easy to understand and to use in the form of a video picture, an overlay for a map, or some type of usable search plan. New technology can allow a large number of Searchmasters to make better use of the development in search theory.

Training and Education

The following items are discussed under this category:

Searchmaster training
Improvement of doctrine
Public education

These items, which are strongly related to one another, are able to make a considerable impact on the success of search on the one hand and reduction of distress on the other.

(i) Searchmaster training. This item is indespensable for the conduct of successful searches and requires constant attention to ensure that the latest techniques for the conduct of a search are readily available.

Certain nations methodically undertake to train personnel to engage in the conduct of live searches. In particular the U.S.A. and Canada have frequent interchange of ideas and personnel in

national programs. To assist in the goal to produce high standards among SAR flyers, exercises are held between the two mations.

(ii) Improvement of doctrine. It is felt that while cooperation does exist in the field of training, there is plenty of scope for methodical plans to be implemented internationally so as to make improvements in the following fields:

(a) Procedures. There is a requirement for timely exchange of information on proven methods and procedures for the set up, diagnosis and conduct of search.

(b) Sensors. Information of the effectiveness and use of sensors is not always passed from one country to another, often because of military classification problems. Nevertheless there are nonmilitary uses of certain commonly available sensors (e.g., the eye, infrared) about which unclassified information would be useful.

(c) Targets. Knowledge of target behavior is not always understood under given environmental conditions (e.g., the problem of overestimation of detectability and poor understanding of a prior probability of target whereabouts).

(d) Methods and aids for search based on theoretical concepts. Although methodological and through training of Searchmasters in the intricacies of the live search problems is undertaken, the teaching of the theoretical foundations of appropriate search theory is less in evidence. There is a lag between the development of useful tools and methods and the indoctrination and understanding by those involved in conducting searches.

Summary

There is an overall need for manuals which are practical, valid and easily available internationally to explain simply and clearly the requirements of Searchmaster training. Such doctrine should emphasize the relevance of known, up-to-date theory and methods to this persuit. Doctrine is one method by which reasonably practical standards can be maintained. Therefore, international exchange of information leading to the accomplishment of these needs is highly desirable.

Implementation

Because such a large number of land and water distress cases in Canada occur in and around the U.S.A., the techniques of computer search methods have been pioneered in that country; it is suggested that information on the framework for Searchmaster training be sought from that country.

(iii) Public education: the problem. One of the more pressing problems connected with SAR is the lack of effective public education into the problems of undertaking searches and the more important aspect of prevention. An extremely large amount of effort is expended because of the lack of user awareness (planes and boats) of the need for defensive behaviors for the use of their equipment especially at sea or in wilderness areas.

The need. There is a need for methodical international dissemination of information to users and regulatory bodies on the hazards inherent in the use of air and sea modes of transportation and the ways in which prevention of distress can be accomplished.

Sources of Assistance

(a) U. S. Coast Guard for Marine SAR.

(b) USAF and Canadian Forces for wilderness flying and survival.

Length of Search

This subcategory addresses those decisions made by the Searchmaster which affect the length of the SAR operation.

The first and most important decision considered is when to stop an unsuccessful SAR operation. Here the Searchmaster must consider the probability of successfully finding the victim with continued search as well as the moral and political issues associated with suspending an operation.

The second problem considered is when to start an SAR operation. Here the biggest problem is the increasing number of false alarms. On one hand, committing search asset too soon would be costly since a large number of SAR alerts are false alarms and could endanger the lives of victims involved in cases which occur later. On the other hand, waiting too long to start a search may result in a prior probability density for victim location which is unacceptably diffuse.

The third problem considered deals with the Searchmaster's allocation of resources when two independent SAR operations must

be planned at the same time. Although this problem is not currently as consequential as the preceding ones, it was felt that in the future it might become important since the number of SAR missions is growing faster than the additional search assets become available.

(i) Stopping rules. One of the most difficult problems which the Searchmaster must address is the decision to suspend or terminate an active SAR operation. This judgment has significant moral and political implications as well as economic impact on future operations. In arriving at a decision the Searchmaster must strike a delicate balance between the value of a human life and the cost of continued search with little probability of success. It is hoped that providing the Searchmaster with search termination guidance will significantly reduce the variability between efforts applied to similar searches and conserve scarce resources without reducing the number of successful SAR resources. Currently decision criteria used by Searchmasters to terminate an SAR operation are extremely subjective and not well documented. To arrive at a final decision the Searchmaster is required to make two judgments. The first judgment is based on analytical consideration of the current posterior probability of success which will be achieved by applying additional effort. If additional search will yield a "reasonable" probability of success, search continues; otherwise the Searchmaster must make a second judgment based on his moral convictions and the politicall climate which he perceives to exist at the time. Only after he decides that these issues warrant search termination will the SAR operation stop. SAR terminations can be made in a more uniform manner and without wasting SAR resources (traditionally SAR operations are known to have been carried on too long) if the Searchmaster is provided with decision aids which give stopping guidance. These aids, although analytical, must consider the probability of successfully finding the distressed person, his survivability, the cost of continued search, the value of human life.

Multiattribute utility theory may be a useful tool in providing this decision guidance. However, it must be kept in mind that specific criteria employed in these aids must be defined by community standards.

(ii) Starting rules. SAR groups are plagued with a large number of distress false alarms. In Canada five Emergency Locator Transmitter (ELT) alerts are received daily, 97% of which are false alarms. A similar situation exists in the United States where the Coast Guard receives 15 false distress inputs each day. The cost of immediately committing search platforms to all alerts is prohibitive. Moreover, the ability of the SAR group to react to

confirmed distress calls would in many cases be significantly degraded, because available assets would be committed to these false alerts. Every Searchmaster is well aware that delaying the search start lowers the probability of finding the victim alive. In SAR operations over sea, the victim may never be recovered. Current operating procedures when an alert is received at the Rescue Coordinating Center all well documented. These procedures vary between land and sea distresses. Typically, the commitment of search resources is delayed between two and five hours while SAR personnel investigate to determine if the alert is a real distress call or not. It must be noted that information collected during this investigation is ultimately used to plan the search; SAR resources are eventually committed only if the alert cannot be shown to be a false alarm. The desire of SAR commands is to initiate search efforts as soon as possible and to locate victims in the shortest possible time. However, in light of the large number of false alarms, search platforms cannot be committed immediately without incurring unacceptable costs and endangering the lives of distressed persons. If the number of false alarms were reduced, search could be started sooner. Furthermore, there is a need for developing analytical procedures for defining starting rules which would take into account the high false alarm rate.

(iii) Allocation of resources. Searchmasters are sometime required to plan operations for two concurrent independent searches. This problem is particularly difficult when the a priori density functions of target location do not significantly overlap. In this case the Searchmaster must allocate the limited resources at his disposal to distinct areas. Lives of distressed persons depend heavily on the selected allocation.

Currently, allocations of effort to multiple search areas are made subjectively. No analytic techniques are known to exist. Factors which will affect the final decision include the number of distressed people in each area and the location and variance of the probability distributions associated with each area. We note that the Searchmaster must make decisions at each step of the search since many of the factors which determine the allocation change with time.

The problem at hand appears to be sufficiently unique that existing results on optimal allocation of resources will not be applicable without some adaptation.

TECHNIQUES FOR SEARCH PLANNING

The planning stage of search and rescue requires techniques which both ensure an efficient use of time and produce the most desirable result in terms of the probable outcome of the search. Further technical developments are deemed necessary in the following three areas: multiple targets, optimal allocations and decision rules.

Multiple Targets

There are two basic kinds of multiple target problems. In the first one, there are actually many homogeneous targets: this class of problem is conveniently handled by current theory and applied probabilistic methods. In the second one, there is really a single target which might be in one of several heterogeneous states, each state having different detectability, survivability and motion characteristics; totally different search objectives may be associated with each state and there is no rational way of combining the behavior of the different states.

Multi-State Target

A single target may assume one of several different states. Each state may have different detectability survivability and motion. This condition is representative of a large portion of actual SAR cases.

Although we can model each state and estimate its probability, there is no theoretical guidance for formulating a single search plan; in fact the search objective is not well defined.

Theoretical investigation is needed to develop some unified approach to planning a search for multiple state targets. The first step in this investigation is to develop some objective which is independent of the target state.

For example in the case of a lost ship which may be
- (i) drifting
- (ii) sunk, leaving survivors in rafts
- (iii) sunk, leaving survivors adrift in lifejackets.

One may set the following objectives:

a) maximize the expected number of survivors which will be rescued.

b) Maximize the probability of rescuing all present survivors.

c) Minimize the expected time to find the first survivor, so as to obtain better information about the disaster which befell the ship.

Once a criterion has been established, the theoretical study might proceed to develop a pure optimal plan. This optimal plan can then be the vehicle by which all the state probabilities and location probabilities are rationally combined and presented to the search planner.

Optimal Allocation of Search and Rescue Resources

Resource allocation is a basic factor in search planning methods, ranging from the the macro-allocation which is dedicating societal resources for the broad function of search and rescue to the micro-allocation problem which reduces to what path a single searcher should take. Even though the most important of these topics involved the strategic real-time allocation of resources and the sensitivity of any allocation plan to the horizon or expected duration of search, the broad problem of resource prepositioning needs more study.

(i) Sensitivity of Search Plans with Respect to Search Deviation

The simplest of search plans is the myopic or incremental plan. This plan is predicated on the assumption that one should look only at the present state and location of the target and plan for the immediate return. This approach is appropriate when the target is stationary and durable. However in most search and rescue problems the target is a person in some kind of distress or in some inhospitable moving environment which is life threatening. In these instances, the most desirable result is to find the missing person(s) alive and well, and the early periods of search become more valuable than the later ones. The optimal plan may be affected by including lifetime and survivability.

Necessary and sufficient conditions for optimal search for a moving target and survivor lifetime searches have been established. But in these searches, no real estimable sensitivity of plans to survivability or motion is known. Present applied search plans do not account for either condition in accordance with theory.

There is a need to develop some easily applied criteria to determine in actual instances the sensitivity of optimal plans to the factors of survivability and motion, so that the search

planner can know with reasonable assurance when to take these complexities into account. In addition, sensitivity analysis may indicate whether more scientific research is needed to improve predictions of both survivability and motion factors.

(ii) Prepositioning of SAR Resources

SAR resources are never permanently situated exactly at a position used in an incident, but some travel time is always necessary. Whatever time is used in transporting the resources must be deducted from the time available for search. Maintenance of SAR equipment at a large number of far flung locations is economically prohibitive.

Current prepositioning schemes seem to be largely unscientific. Many resource distributions for SAR are second-hand positionings, such as co-location with military establishments, airports, etc.

The U. S. Coast Guard has a large-scale simulation model (SARSIM) used to test various prepositioning schemes, which employs no optimization techniques. Studies on this problem tend to be either too complex or too general.

There are numerous schemes for optimal resource location in the literature. A study is needed to determine which approach is most appropriate to the SAR case and to test its application.

(iii) Search Patterns

Deployment of search units is essentially a two stage process: first identify the region which must be searched by the unit; and, second, define the path which must be followed within the area.

Present search methodology uses primarily what is called "parallel sweep search patterns," where a search area is composed of a number of long, straight paths parallel to one another and offset by some fixed distance called track spacing. The search area is then treated as being uniformly swept, with navigation inaccuracies accounted for, if appropriate, in the measure of sweep width. Modern aircraft electronics and microcomputer devices are capable of automatic control of aircraft and can be programmed to execute detailed flight patterns. This capability permits more accurate execution of existing patterns, but more importantly, gives the opportunity to introduce new and unusual search patterns previously disregarded as beyond the capabilities of human pilots.

Are there some types of search paths which might be more

effective for search? Can a path-by-path optimization algorithm be constructed? Would such an optimization be useful for unusual sensor characteristics such as radar, magnetic anomoly detectors or emergency locator searches? Analytical studies should be undertaken and any resulting unusual patterns should be exercised in modern aircraft under automatic control. Certain of these unusual search circumstances may find pathwise optimization more nearly optimal than standard patterns.

Decision Rules

The development of a search plan requires:

i) a clear statement of the goals of the search process and a methodology for combining possibly conflicting goals.

ii) methods for investigation and resolution of uncertainties in the assumptions made at the planning stage.

The first requirement appears to be the most important and is now discussed, and the other is considered in the sections on min-max regret rules; distribution of probability of success; feedback of measures of performance.

(i) Multiple Criteria

Conventionally the goal of an SAR operation is to maximize the probability within a period of time either fixed or random, of detection of the survivor during his lifetime. In fact a multiplicity of goals may be present, including: the detection probability within a given time, the costs of various resources, the satisfaction of political pressures, the differing values of multiple targets. Explicit consideration of each individual's goal is desirable so far as this is practicable, so that compromise between goals does not occur by default.

Up to now there has been no explicit recognition of conflicting goals in the SAR methodology. In other areas, however, there has been intense research into the treatment of multiple criteria in decision making. Approaches to the solution of the problem include: setting goal achievement levels; multi-attribute utility theory; definition of compromise solutions.

It would appear that it is the application of these

approaches to current SAR methodology that requires investigation, rather than new theoretical developments.

(ii) Alternative Decision Rules

In principle a matrix can be constructed in which performance measures are given for each alternative search plan under each combination of assumptions such as scenario, detection performance, etc. Some method of combining these matrix entries into a single performance measure for each plan is required before any selection between the alternative plans is possible.

One approach is to attach subjective probability weights onto each uncertain combination of events and then compute the corresponding expected performance. Currently SAR methodology is exclusively based on ascertaining subjective probability weights; when they are difficult to specify, some alternative summaries of the matrix entries may be required.

Performance assessment in terms of direct minimax approaches is known to be unduly pessimistic: it makes selection purely on the basis of worst case considerations. The "minimax regret" approaches to current SAR methodology that requires investivation, applications. This approach replaces each matrix entry by a discrepancy or "regret" between the entry and the best performance over all alternatives for the associated combinations of event outcomes. Application of the minimax approach to this modified matrix is generally less pessimistic than the direct minimax approach.

It is not entirely clear how this methodology should be applied to SAR planning and this seems a potentially fruitful area of research. If successful, it would make SAR planning less dependent on subjective probability assessments where these are difficult to obtain.

(iii) Distribution of Probability of Success

Evaluation of a probability of success is dependent upon a number of assumptions, and as usually computed should be viewed as an expectation of probability of success.

When the effect of different alternatives is only to shift this expectation without otherwise influencing the distribution of the probability (e.g., the range of possible probabilities on either side of the expectation), then this does not create any problem. It is possible, however, that a modification in search plan may change the whole distribution of probability of success (e.g., reducing expectation by increasing the lower limit of

possible values of probability of success) and in this case selection of a plan on the basis of the expectation only may be invalid.

Problems which need to be investigated are perhaps of a more practical nature. Various possible approaches need to be examined from the point of view of whether the results produced are sufficiently clear and informative to the planner. Some approaches are:

evaluate minimum and maximum probabilities of success over all scenarios;

evaluate quartiles or 10th - 90th percentiles of probability of detection;

compute some weighted sum of the expected probability and some of the above values; and

compute a variance of the probability of detection to be combined with the mean in some way.

(iv) Feedback of Measures of Performance

Cetain input assumptions are common to many search operations. Although not directly observable, these assumptions can perhaps be assessed and updated on the basis of observed performance in the search operations. Assumptions for which this may be relevant include:

winds and ocean currents

sensor peformance.

At present there is no explicit feedback of this type of information. Empirical research is required, aimed at setting up systematic procedures for debriefing after seach operations and analysis of operational performance. Factors which need consideration include:

frequency with which targets are overlooked before being detected;

relationship between actual target positions and the a priori probability distributions generated by CASP to identify any consistent trends (e.g., do prior distributions appear too diffuse or too concentrated? Do searches in certain areas show evidence of unexpected current drifts, etc.?).

TECHNOLOGICAL DEVELOPMENT AND EXPERIMENTATION

The rapid technological advances of the recent years have a direct impact on our ability to conduct search missions more efficiently. The computerized probabilistic model called CASP, which was developed by Daniel H. Wagner, Associates in the early 70's is now in use in the U.S.A. and Canada. A more accurate estimate of the input parameters for such a system, based on past experience and historical data, would result in more efficient and successful searches.

Equipment and Technology

In recent years there have been major technological advances, particularly in the fields of computers and electronics, e.g. image enhancement, infrared detectors, thermal imagers, satellite observation, interim track analysis program (ITAP). It is important that search and rescue make maximal use of such advances.

The current state of the art in this area is advancing rapidly. It is likely that only large SAR organizations will have the capability of keeping abreast of new developments.

To ensure that potential aids to SAR are not overlooked, there needs to be:

i) Exchanges of technical information and experience between SAR organizations

ii) Exchanges of information between practitioners and theoreticians, so that theoreticians are kept informed of the technical developments. This will enable theoretical work to be carried out prior to the introduction of new types of equipments into service in order to optimize their use.

Theoretical work might proceed by modeling possible performance characteristics, irrespective of how such performance might be achieved. In this way:

i) Technical advances can be evaluated to see whether their adoption would lead to increases in search efficiency sufficient to justify the cost.

ii) Desirable performance characteristics can be identified which it may be possible to provide (if not now, then in the future) with advancing technology.

Experimental Data

In probabilistic planning methods, the distribution of the target location is calculated as a weighted combination of the probability distributions of possible scenarios. Models for the most common scenario types are included in the CASP system.

Present probability distribution estimators are based on either historical experience or subjective judgments. More accurate parameter estimation would result in more efficient and successful searches. To provide data on which to base probability distribution parameter estimates, both physical experiments and small-scale digital simulations would be exceedingly useful. In addition, the utilization of historical data for parameter estimation could be improved if all SAR activities could share historical data from other activities. In order for this sharing to be feasible, the data items to be exchanged should be standardized as to units of measurement (distances, dates, geographical location), types of incident, types of search units, etc.

Sweep Width and Detection

Problems related to sweep width and detection have been identified as among the most important in SAR analysis. Probability of detection conditioned upon the target's location determines the impact of negative information on the updated (posterior) probability distribution for target location. Unfortunately the conditional probability of detection or more fundamentally, the sweep width is known only to a rough approximation in most searches. In this section, we will discuss the following topics in an effort to define the problem and suggest direction for further investigation: uncertain sweep width; sensor calibration during search; estimation of sensor performance before search; consideration of survivor lifetime; false targets and multiple sensor search.

(i) Uncertain sweep width. The term uncertain sweep width refers to the whole class of problems in which sensor capability cannot be estimated with high confidence during a search operation. Rather than a rare event, this is a fact of life during

most maritime and land SAR operations. These problems were recognized during World War II, although the pressure of events prevented the development of solutions.

The uncertainty in sensor capabilities can be due to factors related to (a) the performance of the sensor itself, (b) the environment, and (c) the target. The following are brief illustrations of these three factors. Sensor performance might be affected by electronic noise or in the case of visual search by fatigue or lack of attention. The environment may be worse or better in terms of cloud cover or foliage than the average expected. And finally the target may be in an unknown state, e.g., intact or broken up, emitting a signal or not, or a boat, liferaft, or a person in the water.

Evaluation of search effectiveness and search optimization will depend to a large extent on the state of the target.

Present operational search planning systems (e.g., CASP) do not take uncertain sweep width into account. Fortunately, in recent years there has been some theoretical and applied work done on the problem, Richardson and Belkin[1972] and Richardson et al [1971]. This has been from a Bayesian rather than a minimax point of view. Richardson and Belkin[1972] attempts to show the importance of the problem through curves and tables produced by a complete theoretical solution of the search optimization problem for a very special set of assumptions (a normal distribution for a prior target location distribution and a gamma distribution for the prior sweep width distribution). In Richardson et al [1971] algorithms are given for a "cellular" target location probability distribution where the sweep width distribution is described by an arbitrary step function.

It is recommended that new work be carried out in this problem on all fronts, theoretical, applied, and operational. Existing methods especially those in Richardson et al can be applied directly to the land SAR problem as it is currently being treated by Lt. Col. Mattson using small computers. More work is needed to address the moving target problem although the present CASP system can be easily modified to show sensitivity to various sweep width assumptions.

(ii) Sensor calibration during search. Sensor calibration during search is recommended as a way of gauging the accuracy of sweep width estimates on site. This would be done using known target like objects and would resolve some of the uncertainty associated with the sensor.

At present, calibration has been carried out in only a few searches. In the literature there is mention (Richardson et al[1980]) of calibration during the Suez Canal clearance in 1974. Since calibration competes for resources with actual search; its benefits must be established and explained to management before it can take place. This is often a difficult problem.

The major work done in this area is to develop test procedures which can be carried out during the search and which will not interfere unduly with the operation or be too costly. Analysis procedures must also be developed which can rapidly assimilate the data and feed the calibration results back into the search planning process.

(iii) Estimation of sensor performance before search. Tables of sensor characteristics in the form of sweep width and sweep rates are a standard part of publications in most military or quasi-military search organizations. Those tables are indespensable for use in search planning and for the decision process leading to the start of search.

At present many of the tables are out of date and in need of revision. The U.S. Coast Guard is, in fact, carrying out an extensive exercise program to update the tables of visual sweep rate.

It may be useful to present the new sweep width tables with some measure of sample variability. This would give the operational search planner a more realistic impression of the accuracy of the data and also provide the search analyst with a "prior" which could be a starting point for an "on site" update using calibration data.

There is also often the need for the entire lateral range function as well as the sweep width. It might be a good idea to estimate the lateral range function as well as sweep width. Lateral range function estimates were made during the Suez Canal clearance using "monotone regression" (Barlow et al [1972]) and was very important in the sonar search planning process (Richardson et al [1980]).

It was also pointed out that detection experiments might be severely biased by the spotter's knowledge of the existence of a target. Since the vast majority of individual search flights have a very low probability of passing over a target, the testing of human detection capability where a target is present may not be appropriate. The sensitivity of a spotter's detection capability to the probability of target existence should be tested.

(iv) Consideration of survivor lifetime. In many SAR cases the distressed individual is severely injured or placed

unprotected in a hostile environment. Under these conditions survival may only be possible for a few days or hours. Thus the optimal search plan must take survivability into account.

Detailed consideration of this problem belongs in the area of search optimization as well as in sweep width and detection. In the context of optimization the problem has been studied (Stone [1976]) where necessary and sufficient conditions for optimal solution are given. Little or nothing has yet been done, however, on actual modeling of the survival time in stochastic terms as needed in the mathematical formulation of the problem.

As an example of the impact of survival on detection, consider the land SAR situation. In the early phase of search one's detection capability might be enhanced by the expectation of a signal from the target. Later on, however, the target survivor might be too weak to signal, and detection would be governed by direct observation of the crash site. In this case the detection range would be very much reduced.

Work is needed to model the impact of survivability on detection capability and to take these effects into account in both search evaluation and search optimization. This appears to be a wide open area at present with work needed in theory, application and operations.

(v) False targets. False targets arise in at least two distinct ways:

i) sensor noise on transient phenomena, and

ii) actual target-like objects not of interest to the search.

Of the two, ii) presents the most difficult case since they must be investigated by some other sensor before they can be eliminated. The classical theory of signal detection, unfortunately addresses i) using the ROC (receiver-operating characteristics) curve approach. While this theory is adequate as a guide for setting the gain of a sensor it does not allow us to reduce the problems associated with ii).

Several papers and reports have addressed the false target problem from a theoretical, systems analysis and operational point of view. Representatives of these are Stone and Stanshine[1971] and Richardson[1973]. Much more work needs to be done, however, on the problem associated with the relatively new ELT (emergency locater transmitter) devices now mandatory on aircraft.

Considerable benefit would derive from development of methods for reducing the wasted time associated with false targets particularly of type ii). One interesting approach would be to use filtering and tracking algorithms to detect moving ELT signals and hence eliminate them from consideration before search assets are employed. A filtering algorithm was used in the Suez Canal clearance to help discriminate between fish and ordnance. Here the sensor was a side looking sonar (see Richardson et al [1980]).

(vi) Multiple sensor search. This topic considers the problem of measuring the detection capability of multiple sensor search systems, e.g., visual and radar operating simultaneously. The problem of using coordinated units such as a high sweep rate aircraft and a helicopter is more properly covered in search optimization.

It is impractical to develop lateral range functions and sweep width from all possible combinations of sensors which might be simultaneously used. A method is presented by Richardson and Stone[1971] for developing a combined sweep width from individual rectangular lateral range functions. This method was used during the search for the submarine SCORPION and has since been programmed for a hand-held calculator by Dr. Leonardo of NADC. There is little else on this problem described in the search literature.

New theoretical work is needed to develop algorithms to synthesize sweep width from more general types of lateral range functions. This is another reason why experiments should be carried out to provide the search analyst with lateral range functions as well as sweep widths.

COMPUTER ASSISTED SEARCH PLANNING (CASP)

Computational methods for planning real searches require the use of computers because:

a) Search theory as applied to the objective of obtaining maximum yield from limited resources, results in algorithms requiring extensive computations.

b) The situation to be modeled can only be described realistically with extensive data sets characterizing the situation, environment, and search operations.

c) The operational imperative of achieving success quickly from the search (i.e., tc save life or foil an enemy) requires rapid reponse and therefore a short time interval from data acquisition through analysis and planning, to search operations.

While manual rule of thumb procedures do provide guidance in many search planning environments, it has been realized based on operations research advances, that techniques employing new theory and technology could be used to improve both the success rate of search efforts and the efficiency of operations.

Computer methods for search planning have been used in both research and practice by operations analysis. However, the most extensive applications of these methods to be used by operating personnel have been the "CASP" system, developed by the U. S. Coast Guard in conjunction with Daniel H. Wagner, Associates (see Richardson and Discenza [1980]). The "CASP" system has been in use for over six years. To provide a basis of discussing system limitations and areas in need of improvements, the operating environment and system input/output are described below.

(i) U.S. Coast Guard operating environment. The U.S. Coast Guard maintains fifteen major continually manned Operations Centers and numerous (over 230) sub-centers called Groups and Stations for the purpose of managing, planning, coordinating and controlling search and rescue and other response-orientated activities. Coast Guard (CG) personnel, vessels, boats and aircraft conduct most of the operations, but there are significant sharing efforts with units from the military, civilian auxiliary and other nations.

Yearly, over 80,000 search and rescue cases are handled. Most of these are of a direct assistance nature (such as towing) and do not involve problems of locating the distress. About fifteen percent do involve some searching, with the majority being

within twenty miles of shore. Only a relatively few "open ocean" cases occur. Although detailed data about the "profile" of search planning cases are not collected, three types can be identified:

a) Length 1 day. Reported distress requires search of less than 6 hours by boat/helicopter effort.

b) Length 2 to 3 days. Reported overdue requires communications search followed by search by vessels, fixed wing aircraft (and also boats/ helicopters).

c) Length 3 to 6 days. Reported distress or overdue is not resolved by communications search or initial response. Major build up is required involving many units and up to five or six multiple searches over an extensive area.

While the last two categories are not frequent, they consume significant percentages of available resources (in terms of people, search platforms, and funds), so the development of methods for planning efficient operations is essential.

Computer assisted methods are used for this purpose. The objective is to put reference and analysis tools, response rules and decision aids directly in the hands of Operation Center personnel charged with responsibility for controlling response units. The concept of "local control" is fundamental in Coast Guard policy and organization. Responsibility for action is delegated to those in immediate control of operating units. Thus the computer aids must be designed for use by nonmathematicians, nonoperations analysts and those who are primarily interested in an operationally beneficial answer as opposed to the beautiful algorithm that computed it.

(ii) CASP input/output. The maritime search and rescue process consists of data acquisition, situation analysis, search planning and operations coordination/control. Calls for assistance, sighting reports, radio range/bearing information, clues and other intelligence are analyzed firstly to determine if a distress problem exists, and secondly to identify how best to respond to the problem. The following information elements comprise the inputs to the search planning process:

a) Scenario Descriptions (derived from acquired or historical data)
- intentions of the unlocated target, when, where etc.
- suspected casualties, engine failure, etc.
- physical characteristics, to identify drift rate, detectability, survival ability (per the environment)

b) Hazards
- external to the scenarios
- storms, shipping lanes, navigation difficulty, etc.

c) Environment
- on-scene observations
- predictions and historical data
- winds, sea state, currents, temperature, land boundaries

d) Historical Perspective
- case histories, "profiles" of marine activities, etc.
- models of operating methods (fishing, sailing, etc.)

e) Intelligence
- searches (what detected, what not detected), when, where, how (with what sensors), and in what local environment (affecting detection)

f) Resources
- response capability, sensors available, location.

The existing CASP system cannot use the complete set of information listed above and thus does not generate the "complete" search planning answer. The user is required to do much of the analysis and reduction of input data, including the conversion of the many acquired facts and clues to input scenarios. Basic capabilities provided to the search planner are:

a) Enter WIND speed and direction information. "Automatic" reference files of predicted WIND are also available.

b) Enter SEA CURRENT set/drift information. Historical files may be automatically accessed, and some "live" information from satellite sensors provides Gulf Stream current information.

c) Generate distress scenarios in a POSITION (bivariate normal), an AREA (uniform), and a TRACK LINE (linear).

d) Move target scenario distributions by DRIFTING according to user specified leeway and wind current options.

e) Update the target distributions by entering results (nondetection) from a RECTANGULAR parallel path search or a PATH single track search.

f) Display a MAP of the target probability distribution for use in planning the next search, and output probability of success (POS) and search effectiveness percent (SEP).

"CASP" cannot accept as input all information available to the search planner, and provides outputs that require user processing before search units can be allocated. Even so the system is valuable because the outputs that are produced could not be duplicated manually by planners within the required response time.

The limitations of "CASP" must be considered in comparison to the benefits achievable through advancements in planning. As accuracy of data files concerning the environment is improved, motion and detection modeling will produce less uncertain answers. This will result in more return from the same search effort, but note that vague input scenarios are a primary source of planning uncertainty especially during the early period of search. This cause can only be removed by reliable distress information, locating transmitters, general surveillance systems, etc.

A valuable benefit derived from using "CASP" for six years is the accrued understanding of how best to interface computer planning with operating personnel. Computer limitations have precluded expansion of "CASP" capabilities, but with the acquisition of new computers and the reprogramming of the user-interface to allow conversational interaction, a major advance is expected. The primary objectives of this system conversion are identified below:

(iii) Implementation objectives.

(a) The USER must be able to enter data in the form he receives it, and should get outputs giving answers from which he can directly execute operations or make decisions.

(b) The computer model must be self-adjusting to extract the best answer from available data (i.e., accurate current data gives a better distribution, using variable time/position/content file structures).

(c) The computer model must be based on statistical computations (analytical or Monte Carlo) that in all cases consider first and second order estimators (at least). The "law of averages" does not apply to SAR cases where search (sample) size is 1 to 6. Averages, and uncertaintities must be used in decision making. Thus, decision criteria will be based on both expectations and the risks associated with errors of inputs or improper weighting of ideas.

(d) The feedback of operational results from case files, survivor debriefs, on-scene data, etc., must be included as a system function, to provide for the continuing validation and improvement of the model.

(iv) Deficiencies needing resolution.

(a) Outputs should include realistic specifications for searches by available units, based on the optimal allocation of detection efforts to the updated target distributions.

Need:
- theory for optimal allocation
- experiments to determine realistic characteristics data for detection model

Note: Suboptimal allocation is probably satisfactory, because of large model uncertainties and resultant "FLAT" benefit functions.

(b) Relative motion between search unit and target must be considered in both optimal effort allocation and detection modeling.

(c) Model "realism" must be improved by incorporation of land boundaries, and accurate/precise environmental files (currents are now on 1 x 1 grid and winds on 5 x 5 grid).

Need:- Implement available "land" routines
- Gather data and better model air/water motion
- Improve motion model to reflect known uncertainties and variable (space/time) data.

Note: Benefit from environmental experiments does not warrant excessive expenditure for "full" ocean/atmosphere model for SAR use only.

(d) The model must be revised to maintain statistical validity without user intervention. Numbers of replications should

be controlled to maintain an adequate density of sample points which is free from random effects. Regeneration or enhancement techniques should be employed to ensure that search effort allocations are based on true statistical estimates, not model sampling noise.

(e) Planning procedures must be developed to guide users and provide decision criteria for the transformation of case data to scenario inputs.

(f) scenarios must be revised to take available user inputs and transform them to statistical descriptions. This should include simulation of interaction between the search object (or scenario elements) and the environment both before and after the point of distress.

Note: Methods of using capabilities of general simulation methods (e.g., GPSS) should be investigated.

(g) It was felt that a particularly useful tool would be the provision of standard elementary scenarios based on statistical analysis or micro-modeling (e.g., the lateral distribution of accidents from flight paths, correlations of accidents with weather conditions and terrain features). The Searchmaster who must actually conduct the search would then merely enter into the computer possible outline scenarios (e.g., possible flight paths and their order of liklihood). The computer would then generate detailed scenarios based on a mixture of the Searchmaster input, the inbuilt scenario generation programs and the environmental files (containing data, weather conditions, sea currents and terrain features).

References

Barlow, R. E., Bartholomew, J. M., Bremmer, J. M. & Brunk, H. D. [1972] "Statistical Inference Under Order Restrictions," John Wiley, New York.

Richardson, H. R. [1973] ASW Information Processing and Optimal Surveillance in a False Target Environment, Daniel H. Wagner, Associates Report to Office of Naval Research (DDC AD A002254).

Richardson, H. R. & Belkin, B. [1972] Optimal Search With Uncertain Sweep Width, Oper. Res. 20, pp. 764-784.

Richardson, H. R., Corwin, T. L., Stone, L. D., Engel, F. P., Scranton, B. E. & Browning, W. J. [1980] The Contribution of Operations Analysis to the 1974 Clearance of Unexploded Ordnance from the Suez Canal, Oper. Res. (to appear).

Richardson, H. R. & Discenza, J. H. [1980] The United States Coast Guard Computer-Assisted Search Planning System (CASP), Naval Res. Logist. Quart. (to appear).

Richardson, H. R. & Stone, L. D. [1971] Operations Analysis During the Underwater Search for Scorpion, Naval Res. Logist. Quart. 18, pp. 141-157.

Richardson, H. R., Stone, L. D. & Andrews, F. A. [1971] Manual for the Operations Analysis of Deep Ocean Search, prepared for Supervisor of Salvage, Naval Ships Systems Command (NAVSHIPS 0994-010-7010).

Stone, L. D. [1976] Necessary and Sufficient Conditions for Optimal Solutions to a Survivor Search Problem in "Stochastic Systems: Modeling, Identification and Optimization, II," R. J. B. Wets, ed., North Holland, New York.

Stone, L. D. & Stanshine, J. A. [1971] Optimal Search Using Uninterrupted Contact Investigation, SIAM J. Appl. Math. 20, pp. 241-263.

SEARCH AND RESCUE IN CANADA

R. P. Hypher

Air Transport Group Headquarters
Canadian Forces

Introduction

Although Canada has a very large land mass and low population density, she does undertake an appreciable amount of "air" and "marine" activity within her border. The wide use of aircraft for pleasure and business is a common aspect of North American life; additionally, the large area of water--both inland and on the coasts--are the scene of enormous amounts of marine activity comprising every size of vessel. This activity, in its turn, produces cases of distress, forced landings, crashes of aircraft and ship disasters. Table 1 illustrates in summary the characteristics of the Canadian Search and Rescue scene.

Main Features of Air Cases (Where, Who, When, How, and Survival)

(a) Where are Accidents Located? Figure 1 shows the distribution of "air" cases over a five-year period. Although there are a few major concentrations of crashes and forced landings in Canada, there is nevertheless a wide dispersion of cases in large areas of the Arctic and the middle North. This dispersion upsets the balance between useful effort searching and unproductive time in transit to reach the general area of the distress. This poses a readiness problem and pressures our resources.

(b) 68% of all crashes and forced landings involve single engine, fixed wing aircraft under about 2,000 kg in weight, the majority of these are privately owned.

Table 1. Basic characteristics of the Canadian search and rescue scene.

GEOGRAPHICAL	Square Miles Territory: 4 million Water: 1.5 million	Miles Coastline (mainland) 36,000
AIRCRAFT	Number of Planes: over 12,000 Pilots: over 50,000	Major Air Distress Per Year 140 (those involving dedicated SAR = 100)
MARINE	Number of Vessels & Boats	Major Marine Distress Per Year
	Pacific / Gt.Lakes / Atlantic	
Pleasure	200,000 / 1.1 m / 8,000	560
Commercial	6,500 / 2,800 / 42,000	(not all involved dedicated SAR effort)
TOTAL INCIDENTS	Including Medevacs and False Alarms 9,250 (year 1978)	
TOTAL MILITARY SAR HOURS FLOWN PER YEAR	12,500	
NUMBER OF MILITARY PERSONNEL INVOLVED	500	

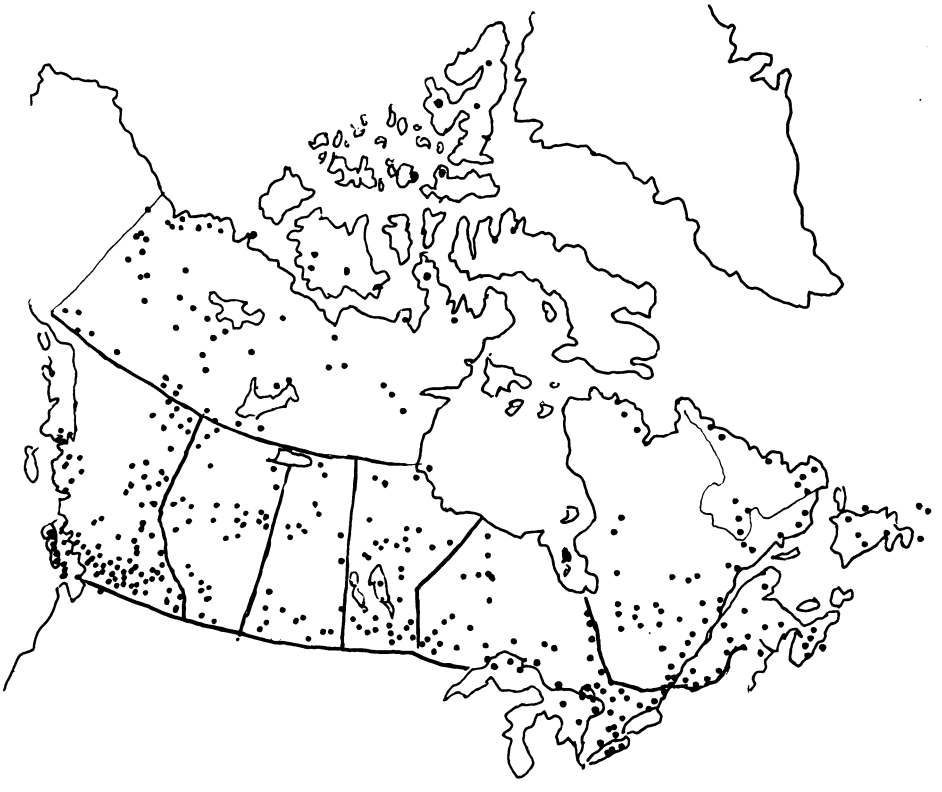


Figure 1. Distribution of "air" cases over a five-year period.

(c) When Do Accidents Occur? Occurrence of accidents is highly correlated to weather and it can be shown that this --in turn-- can be associated with the experience of the pilots, in particular their inability to fly on instruments, or their inexperience in night flying. Although Canada experiences just about every possible type of weather, surprisingly, 91% of light aircraft coming to grief do not have deicing equipment or adequate means of navigating after loss of visual contact. Because most of Canada experiences a very deep winter, it follows that about one-third of the crashess occur in temperatures below freezing, during snow storms or when snow is on the ground. Weather itself is also a serious problem affecting search and, in consequence, it is largely during poor weather that demands made upon the search aircraft (and vessels) are the greatest.

(d) How? Less than 20% of all "air" crashes and forced landings are due to pure mechanical trouble. Of those cases where we have an explanation of what occurred (about 3 out of 4 cases) the pilot could not cope with the weather, or was lost, or out of fuel.

(e) Survival. One study shows that of those who do not survive, (Milligan [1975]) 70% died on impact and the remainder, with a few exceptions, within a few hours afterwards. Of aircraft that crash, 93% have five or fewer on board which gives some indication to us of the rescue requirements. A recent analysis shows that fatalities occur in about half the Air Distress cases and in those involving fatalities, five out of six, of those on board do not survive. Only 7% of all searches are unsuccessful.

The proportion of people killed who are on board the aircraft is higher in cases where weather was associated as a factor, than when the aircraft was lost, out of fuel, or had mechanical problems. Loss of life occurs on occasion to the searcher as well as the searched and this is usually attributable to searching in extremely poor weather.

The survival time is a very important consideration in relation to SAR Readiness and Allocation of Resources. Fortunately, approximately half of the nonfatal crashes are located within 11 hours, and 75% within 20 hours from time of notification. This however hides a very important and troublesome factor and that is the notification time from the time the incident occurred. This element consists of a time distribution

with a median value of 5 hours but also displaying a very long tail, it is sometimes days before it is known an aircraft is missing.

Marine Cases

We have mainly studied Marine cases to determine if we have a requirement for more helicopter assistance and a higher readiness response. Surprisingly enough, when we analysed Marine cases, we found that the problem is not generally as pressing as we first thought. There are quite distinct trends with season and time of day and this gives us a clue in solving readiness problems. Table 2 (drawn from sundry DND staff papers) gives the distribution of serious cases of Marine Distress.

Our studies have shown that considerable differences exist between the features of Marine cases between the Pacific, the Atlantic and Great Lakes regions. For example, the Pacific coast has a far greater incidence of boats foundering, capsizing or cases of man overboard, whereas the Atlantic and Great Lakes suffers more through ships being aground or broken down. Another very important difference is method and time of notification. In the Atlantic 58% are merely overdue, whereas in the other two regions, over half of the cases are alerted by visual observation. As a consequence of this, the rescue time distribution demonstrates a much greater spread of values in the Atlantic. The difference between the West and East Coast is the fact that the former carries out inshore fishing, whereas the latter is largely away from shore. Although the number of cases in which people are rescued from the water is about 1 in 6 of the serious cases of distress, the drownings are few. Our biggest problem however appears to be single major distress cases such as the vessel Gabrielle which foundered and incurred a relatively large loss of life.

Comparisons: Marine and Air Cases

Regardless of the obvious differences between Marine and Air cases of distress, they both involve expenditure of effort on either search or rescue. One of the useful features of the majority of Marine cases, not occurring in Air cases, is self notification. It exists for about one-third of the serious cases and a higher proportion of those of lesser magnitude. Another difference between Air and Marine cases in Canada is the shorter transit time for the latter.

Loss of life on boats often occurs when a rapid capsizing occurs. In such cases, it is impossible to give warning or provide timely rescue services. Herein lies the problem of how to

Table 2. Comparisons of distress cases by region 1976 - 1977.

	PACIFIC	GT.LAKES	ATLANTIC
Percentage of National Total			
Cases	49.9%	33.6%	16.5%
Persons involved	32.5	27.7	39.8
Lives lost	37.9	19.9	42.4
Cases involving loss of life	52.6	21.1	26.3
Average Number of Persons Per Case	2.2	2.8	4.7*
Number of Deaths Per Case Involving Loss of Life	1.2	1.6	2.7
Average Number of Cases Per Month	11.7	9.5**	3.9

* Four cases involving 206 persons not included.

** Based on 20 months, as there were no cases recorded for the months of January and February in either year.

prescribe (within reason) what resources are needed to raise the readiness state.

Electonic Locator Beacons (ELTs)

A dictum has recently been in vogue - "to take the 'search' out of search and rescue." The most popular move in this direction has been to legislate (since revoked) that aircraft (in particular) will carry Electronic Beacons that will automatically transmit on a distress frequency when an aircraft crashes. The basic idea was (and I suppose still is) sensible, however, many problems were not foreseen when the legislation was introduced. The problems can be summarized as follows:

a. there was no mandatory surveillance system giving full coverage of Canada,
b. the false alarm rate was 97% of all signals, and
c. failure of ELTs occurred on a large number of real cases of distress.

Because of these problems, procedures had to be adopted to avoid consuming valuable and scarce resources following false alarms. Studies revealed the fact that 89% of unresolved signals died out in one hour. This permitted the system to play the odds in response to false alarms. Despite the drawbacks involved with ELTs, it is still believed that they are in principle a useful additon to SAR. Since notification is a critical problem we try to keep our military SAR resources at a high degree of readiness, but find much effort on our part is vitiated by poor notificiation on the part of the distressed. The same rule still applies with ELTs. About two-thirds or more of distress cases in Canada are still triggered by the plane or boat being overdue. There is government legislation on "Flight Plans," but there are still problems of enforcement.

Satellites

For some time in Canada we have been carefully examining the use of satellites to monitor all regions of Canada. This has forced us to consider low orbiters rather than geostationary satellites. An international testing program of these will be started soon. At present reading we can foresee that introduction of a satellite system would create a large diminution in the number of cases that are at present triggered by some means of overdue notification. However, the problems of false alarms will obviously not disappear with satellites, so the introduction of a surveillance system will require a corresponding setup to monitor filter and coordinate information.

Problems of Search

The problems we have discussed so far relate to the wider problem of SAR. The problem that bedevils us most of course is search. There are three considerations uppermost, the first being to minimize the elapsed time to succeed, the second is to accomplish this with limited resources, and the third is a criterion for terminating operations. In the case of the latter, we must consider the emotional problems of concerned relatives of the missing. Because search time (like notification times) features a long tail to the distribution (in the Hartwell case 1,400 flying hours were consumed at a cost of 1.5 million dollars and even then the aircraft was discovered by accident), much difficulty is experienced in trying to lay down rules for determining a threshold for termination. Cost is not the only consideration, we have devoted a fair amount of operations research effort to solving other aspects of this problem.

Methods of searching for aircraft in Canada have traditionally been based upon a mixture of estimates of flight route, interpretation of ad hoc reports, effects of weather, terrain and aircraft capability. To assist in limiting the area of interest prior to the commencement of the search, some estimate is made of the area of likelihood. Clearly, to delineate an area on the basis of possibility only would yield little information and would be wasteful of resources. We have devised a system of delineating areas in the absence of more precise data by observing the distribution of location of past cases in respect to origin, (or last known position), destination, and track. It was observed that the heaviest density of crash locations was in the area of three-quarters the distance down the track. A spread of target values was determined for both along track and across track. The system in effect is based upon two basic phenomena:

a) A linear relationship between the radial distance to the accident (track variable) and the track length and,

b) A log-linear relationship between lateral dispersion and track length.

In the case of a, the spread of values around this relationship is also a function of distance to destination. Examining response surfaces of joint probability of the values so derived, permitted us to arrive at optimal areas at given probability levels. The variables of interest are shown on Figures 2, 3 and 4. The system is called OTV (Offset and Track Variables) and is being introduced as a standard procedure and taught to Searchmaster courses. It is based on Canadian experience so we are

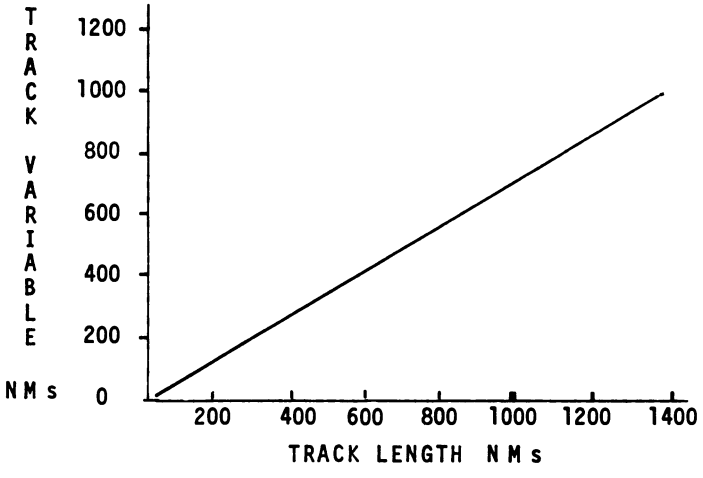


Figure 2. Track variable vs. track length.

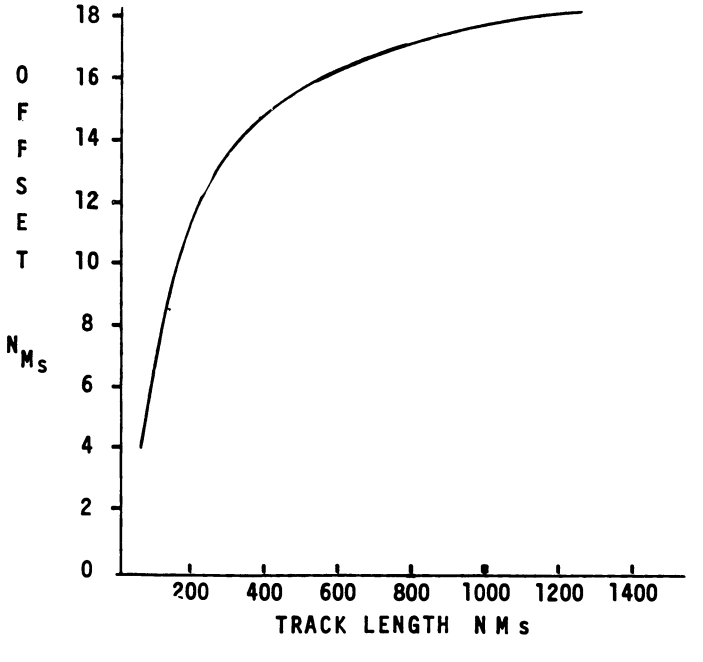


Figure 3. Median value of offset vs track length.

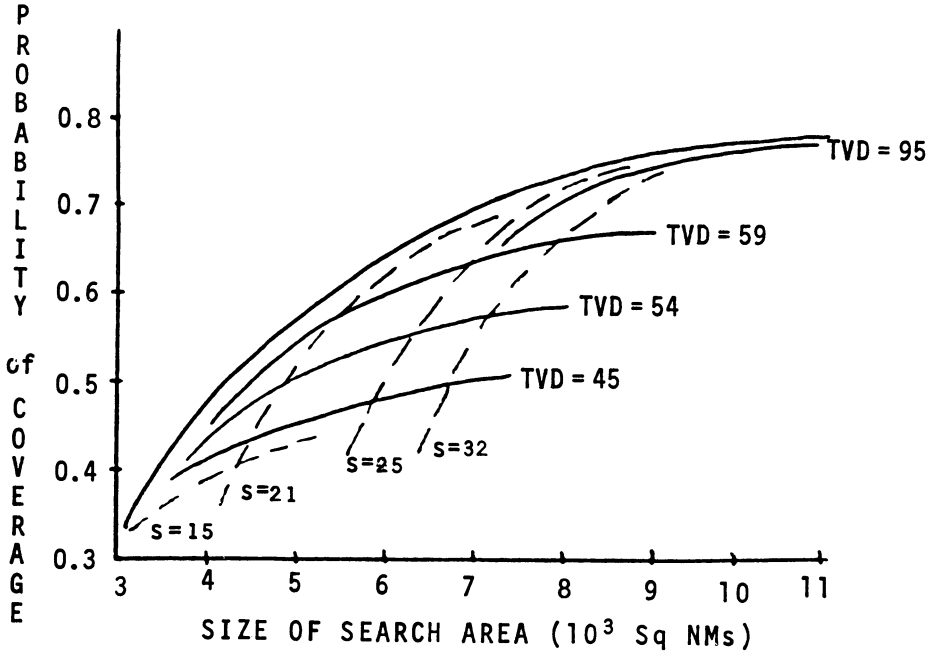


Figure 4. Samples of probability of coverage vs area for track lengths of 100 to 200 nm.

not sure what relevance it has elsewhere. Recent information available from the United States indicates that their crashes tend to be closer to track in that country.

Since our first study we have carried out a rerun on a further sample of cases (over 2 years) and have found excellent correspondence to the original data, so we have in effect conducted a feed-back study. The OTV method has certain limits however. While it indicates an overall probability within a given area, it does not prescribe any sub-areas of high probability within it. It is up to the Searchmaster to combine his a posteriori clues with those obtained a priori. The OTV system also has the advantage of being consistent at varying track lengths which was not the case with the previous method.

The OTV method is only applicable up to a value of approximately 0.8 probability. Beyond that we do not have enough data to make predictions. For track lengths less than 100 NMs, the system still works but is very little different from a previously used rule of thumb. Most track lengths, however, are longer than this in Canada so the OTV method shows greater benefits on longer track lengths.

Probability of Detection

Now one of the most difficult subjects to tackle is that of "Probability of Detection" (PDet). Those of you familiar with anti-submarine warfare will recognize the problem. Although SAR is fraught with all sorts of difficulties, resolving what the PDet is and finding quantitative ways to act upon it is the "achilles heel" of the system. The fact that under a wide spectrum of circumstances the PDet is known to be very low--makes the problem of search stategy very difficult.

The problem of resolving PDet is divisible into two parts PDet depending upon:

a. the platform height, speed, viz., the spotting technique and efficiency, and

b. the terrain, weather, color of aircraft, color of scenery (e.g., red plane, white snow, or a fragmented plane in heavy foliage).

At the moment values of PDet are very subjective and vary from crew to crew. In fact, the difference between the U. S. Civil Air Patrol figures and estimates given by our crews on a search varied by a factor of two. We still tend to search at one or two standard altitudes and assign to them assumed path coverage (e.g., 500 feet altitude and half a mile scan) etc., and we also travel at

given speeds. Little has changed in years in Canada on this subject, except that some excellent research on spotting techniques has been carried out by our DCIEM (Defence & Civil Inst. of Environmental Medicine) studying eye movements (Stager[1973]). It has been shown that spotters do not follow the patterns prescribed and that scan coverage is somewhat inefficient.

As far as the combination of aircraft, terrain and visability is concerned, it is still very much a matter of guess work. It is surprising how aircraft occasionally stumble across an old crash we had been looking for from a few years before and had abandoned after the expenditure of many hundreds of search hours. Sometimes the accidental sighting of the reflection of the sun or a minuscule portion of an aircraft makes all the difference. We have thought of other more sophisticated devices than the human eye (infrared (IR) side-looking airborne radar (SLAR)) to assist search but we have not been given the "green light" because of costs vs pay-off.

The knowledge of PDet is very important, even if only a relative scale of probabilities were to be obtained in order to design efficient search strategies either by manual methods or with the help of a computer. At this time, we are considering that some serious trials must be carried out if we are to improve our search efficiency and thereby help both victims of distress and ourselves. In the meanwhile, we are conducting parametric sensitivity studies on the consequences to be expected of different magnitudes of error in different search strategies. Early results indicate that we are under-searching high probability areas.

C A S P (Computer Assisted Search Planning)

We are associated with the CASP system (discussed elsewhere in this conference) by assisting Professor Michel Jean of College Militaire Royal, St. Jean, Quebec, to carry out the task of evaluating it in the Canadian environment. Suffice it to say that the Military will require some testing on live cases over land if they are eventually to adopt the system. We have, so far, only had the opportunity to participate in a limited number of cases, but so far CASP looks encouraging.

Allocation Problems

The problem of nationwide preallocation of SAR resources is a difficult problem which we have not solved. We did let out a study to a consultant on this subject, but unfortunately the approach was far too complex to be usable.

Conclusion

From the practitioner's point of view, the problem of search and rescue cannot be solved either by simplistic approaches or automatic responses. To a large extent we are dependent upon human intervention and skill in the process of search strategy, search method and in the decision when to terminate a search. Because self-notification systems have been less than fully successful and often fallen below expectations, the need to carry out searches will still be with us for some time. The cost of search is high and, in some cases, running over a million dollars for one case, therefore, any new methods and aids which can be devised to speed up the search process are in great demand.

References

Milligan, S. L. [1975] Causes, Survival Rate and Incidence of Poor Weather Affecting Distress Air Cases, ORAE Report No. R 57, Department of National Defense, Ottawa, Canada.

Stager, P. [1973] Locating Crash Sites in Simulated Air-to-Ground Visual Search, Human Factors. 20, pp. 453-466, Department of National Defense, Ottawa, Canada.

OVERLAND SEARCH FOR MISSING AIRCRAFT AND PERSONS

Robert J. Mattson

Headquarters Civil Air Patrol-USAF(ATC)

This paper gives an overview of the overland search situation in the United States, and will attempt to identify areas where improvement is needed. Both the search for missing aircraft and for missing persons will be covered.

In the United States there are approximately 4000 aircraft accidents each year. More than 3000 of these occur on or near an airport and require no search effort. Almost 800 accidents occur enroute and the crash site is immediately located. The remaining 200+ accidents require some search effort to locate the crash site, and frequently the searches are very involved. In 1978 more than 21,000 hours were flown in civilian and military aircraft on searches for these missing aircraft. In addition to the flying time about 30,000 man-days of effort were expended on the searches.

The magnitude of the missing person search in the United States is difficult to determine. This is because these searches are of a local nature and no national reporting system collects the data. A few states have reported their data and if their experiences are typical, the magnitude of the problem is very large. When looking for a missing aircraft the probability of finding survivors is reduced with each passing hour. Studies indicate that 60% of all survivors of an aircraft crash are injured, and that after 24 hours only 20% of the injured are still alive. The probability of finding even the uninjured alive after 72 hours is very small. One recent study by the Air Force Rescue Coordination Center (AFRCC) showed that of 100 persons involved in aircraft crashes, 35 survive the impact. Of these 21 were injured, and after 24 hours only 4 of these were still alive. Of the 14 uninjured only 7 were found alive after 72 hours. These observations certainly indicate that a quick find is needed to

have any reasonable chance of finding the survivors still alive.

The search for a missing aircraft has a few advantages over many of the other search problems. First, the objective is stationary, once it crashes it does not move. Second, if survivors are able, they will attempt to assist the searchers. Additionally, there are several systems which may aid in locating the crash site.

The first priority when planning a search for a missing aircraft is to reduce the search area from the possible area to the probable area. More than half of the aircraft we search for do not file a flight plan. The possible search area for a light aircraft could easily be over one-half million square miles (4 hrs fuel X 100 knots - 400 nm radius). To reduce this to a more manageable search area some effort is spent determining the probable route of the aircraft, and in estimating the probable crash point. Family, friends, and others must be questioned to gain some insight into the probable routes and actions of the pilot.

One method of narrowing down the visual search area is to review computer processed information received from radar sites along the aircraft's probable routes. If the track of the missing aircraft can be identified, a very accurate crash position (frequently within less than one-half mile) can be determined. A visual search of this area normally finds the aircraft wreckage. In 1978 this process was used on about 25% of the cases and was very effective.

Another method of reducing the possible search area is the use of the emergency locator transmitter (ELT). The ELT is an automatically activated radio transmitter which has been required on most civil aircraft since 1975. The transmitter is activated by the deceleration forces associated with a crash landing, and it transmits a distress signal on international distress frequencies. In 1978, the ELT helped locate 29% of the crashed aircraft for in the U. S. The average flying time expended on these missions was 21 hours. This is less than 20% of the time expended on non ELT aided searches. One negative aspect of the ELT is the high incidence of malfunctions. The ELT frequently is activated during daily operations, and unfortunately does not always work when the aircraft crashes.

Yet another method of reducing the search area is to analyze the weather which may have affected the aircraft. Studies have shown that poor weather is associated with the crash site on up to 80% of the search missions. In addition to standard weather

analysis the AFRCC is beginning to use both infrared and standard satellite photographs to determine the worst weather along the aircraft's probable routes. This analysis is beginning to have some success and several aircraft have been located by searching the areas where the worst weather occurred.

The public is informed of the search and many leads and clues are received. Frequently, the amount of information is overwhelming.

The mission coordinator must use all of the available information when deciding on the best search area. For the most part the mission coordinators are making these decisions subjectively, using their best guess as to which area could be the most probable. Throughout the mission a vast amount of data is being received concerning the probable routes, possible crash locations, and unsuccessful searches. The mission coordinator's decisions become ever more complex.

When searching for a missing aircraft there are basically two types of searches, the trackline or route search, and the grid search. The route search is used for ELT search and to spot visual signals from a survivor. The grid search is used for detailed visual search for the aircraft wreckage.

The search for a missing person can be just as difficult. A missing hunter or hiker could be mobile, or injured or incapacitated. A missing child, or missing patient from a mental institution, or a resident from a home for the aged could be actually hiding from the searchers. When we search for a missing person we not only contend with a very small target, but we may have to deal with a mobile, or uncooperative or even hostile subject. These problems can significantly complicate an already difficult search problem. However, there is one very positive aspect concerning the overland search for a missing person. The person is in physical contact with the environment and must leave physical evidence of his/her passing. To some extent this fact is being exploited with some excellent results.

I will briefly discuss some of the procedures and techniques being used to locate missing persons. First is the need to determine the possibility area and then to reduce this to the high probability areas. Many factors must be considered such as the physical condition of the persons, how far could they have traveled, what probable routes would have been taken, what have similar persons done under similar conditions, what is the terrain and ground cover, what is the weather, etc. Again, using all of these inputs, the mission coordinator makes a subjective decision concerning which are the most promising search areas.

A variety of search units and search techniques are used when looking for missing persons. Containment is a technique used to contine the missing person to a geographical area. Various mobile units can be used for this. Attraction is used to let the subject know that the searchers are present and looking for him/her. Visual and audio signals are used for this purpose. Searchers look for clues as well as for the subject. Since there are more clues than subjects, the clue search is more likely to be successful but the searchers have to be well trained in what to look for. Man tracking by highly experienced personnel has been very effective, but relatively few individuals have the talent to do this effectively. Some dogs have been trained to search for any human within a search area. Visual search of a large grid using large numbers of relatively inexperienced personnel is used to thoroughly search a small area. Aircraft have some value in the missing person search but the type of ground cover dictates their effectiveness. These are only a sample of the techniques and resources being used for ground search for missing persons.

The mission coordinator has a most difficult decision in determining which search technique to use at which time. Some of the techniques are not compatible--for example, the large scale grid search and the use of dogs, or man trackers. What is the optimal search at any given time?

I believe we can make significant improvements in our search planning using even the rough information which is currently available. We are not generally using the statistical information available concerning missing aircraft and missing persons. Preliminary studies show that aircraft crash very close to the route. And some studies concerning the travel patterns of missing people show us where to start looking for them. A few agencies are using this information, but it is not generally being followed. In our selection of search areas and search types we are not considering the fact that every search has many different components. For example, an airborne search for a missing aircraft has at least three components. First, is the search for the ELT. Second, is the search for signals from the survivors. And third, is the visual search for the aircraft wreckage. To effectively use the available resources we must calculate the contribution each unit can make to each type of search. While there may be only a small probability the ELT will work properly, we can conduct a very thorough search for the signal quite rapidly. How much effort should be placed on an ELT search? We can also roughly calculate the probability of detecting the survivors's signals. But, how much effort should be directed towards this search? The amount of effort will probably vary in a direct relationship to the duration of the mission. We have a simple method of quantifying the value of a visual search for the wreckage, but again how much effort

should be spent on this type of search? Some method can be devised to optimize the contribution of each search unit to each of these searches. By optimizing each unit s efforts we should be improving the efficiency of the total search effort.

By keeping track of the three separate but interrelated searches we could also use the data to help terminate the non-productive search. For example, if after 300-400 hours of search we could show:

> For an ELT search 100% of the area covered with 99% POD (probability of detection).
> For a signal, 95% of the area is covered with 90% POD.
> For the wreckage, 80% of the area is covered with 65% POD.
> For an additional 500 hours of search the POD would rise from 65% to 75%.

This information could assist in the difficult decision of when and why to terminate a search effort. This is a real problem and any assistance would be most welcome.

Similar optimization routines could be used to improve the search for missing persons.

Missing aircraft and missing person searches have some differences but they have many similarities. In both searches the search areas are determined primarily by a subjective decision of the mission coordinator. Historical and statistical information is not generally, nor formally, used to help identify high probability areas, and resources are not being used in an optimal manner.

The challenge is to make better use of the information and resources available when planning a search. We must try to minimize the subjective inputs which currently determine our search area priorites. We can do this by analyzing previous incidents to determine which factors should be given greater weight when establishing search area priorities and probabilities. We could determine a probability for the ELT to be working properly and calculate the effective range of the ELT activated by a crash. This would help us plan better ELT searches and would quantify the value of this search. We should also attempt to quantify the value of other types of searches, such as the search for signals from survivors, and searches by dogs, and searches for clues, etc. When we can quantify the value of each bit of information and the value of the various types of searches, then we can begin to use current search theory to identify the optimal

search. And, unless we can establish some confidence in the values we use, we cannot expect wide acceptance in the use of optimal search theory.

I do not know how many more lives could be saved if we could significantly improve the search effectiveness, but if even one life is saved, it will make our efforts worthwhile.

Acknowledgment

The opinions expressed in this paper are those of the author and do not necessarily represent the position of the United States Air Force.

ESTABLISHING SEARCH PRIORITIES

Robert J. Mattson

Headquarters Civil Air Patrol-USAF(ATC)

Establishing search areas for either a missing aircraft or a missing person generally can be divided into two phases, the initial search phase and the subsequent search phase. Underlying both phases is the unending quest for information which starts with the initial notification. Continually, information which may or may not help in finding the survivors will be received. Every piece of information must be reviewed and placed in perspective; asking the following questions may be of assistance:

Is the information RELEVANT? Will the information be of any assistance in helping to find the survivor?

Is the information TIMELY? Did the lead occur at a time that is within the realm of possibility?

Is the information RELIABLE? Look at the reporting source for sincerity and motivation.

Is the information COMPLETE? If not, do what is necessary to gather complete information and your efforts will be rewarded by a shortened search and an increase in the number of lives saved!!

This process of gathering data is one of the most important in any search effort and should NEVER stop: DIG! DIG! DIG! for more and more information!!!

Initial Search Phase

Early in any mission the Search and Rescue Mission Coordinator (SMC) is faced with the problem of assigning a limited number of resources to search areas. Frequently this decision must be made based on little information which is sometimes contradictory and may be unreliable. But, because of the urgency of the situation he may be compelled to start the search even if he would rather wait for additional information or appropriately trained search staff and equipment (tracking specialists, dogs, helicopters, etc.). At this time the best search areas are probably around the last known position and along the most probable routes. I firmly believe, and statistics will bear me out, that you will find most of your survivors during the initial search effort and the SMC should attempt to cover as much area as rapidly as possible. During this stage of the mission you are looking for the cooperative survivor, one who will signal or respond to your presence, and for obvious crash sites or anything unusual which may lead you to the survivors. Therefore it is appropriate to use a wider track spacing than you would for an unresponsive victim or for a hidden crash site. During this phase, information is changing rapidly and things are happening too fast to really establish formal search priorities. Probably the best tool you have for establishing priorities is "COMMON SENSE " and the experience of those initially on the scene. A review of the studies conducted by the national SAR School concerning aircraft crash locations (True and Mattson [1974] and an analysis of travel patterns of lost persons by Syrotuck [1975, 1977] may be of assistance and guidance in establishing search priorities and area size. But you must remember that every SAR case is a unique one and must be treated as such; look at all the facts and use your "COMMON SENSE!!"

Sometime during the first day or night the SMC is going to think about what must be done next, and it is at this point that the second phase of search area planning is entered.

Subsequent Search Phase

When planning for subsequent searches the SMC must take into account ALL of the facts and consider all of the assumptions that are pertinent to the case. One method of distilling the information into a useful product is to start a list of possibilities and facts as they come to your attention. This list should be started at the time of initial notification. Items such as possible routes, weather, terrain, probable reaction of the individual to various possibilities, historical pattern of

others lost in the area, and anything else that might have a bearing on where the survivors may be found should be added as discovered during the initial phase. These items will be brought out by the many individuals who will be involved, and it is important that nothing is lost or overlooked due to the confusion inherent in the initial response. If the SMC has this list he can be confident that ALL of the information that was available to him will be considered when establishing subsequent search area priorities.

When the SMC is ready to plan for the next search he should consult this list and discuss the information with others who are experienced, knowledgeable, and/or involved in the search effort. Additional items will most likely be added since it is probably the first time that many of the major participants will meet. Next, the possible routes, positions, and areas where the survivors may be found should be listed and thoroughly discussed. Each person should be asked to assign a percentage of probability to each item, and each list should total 100%. To get a consensus, you simply take the average.

Example

Person	"A"	"B"	"X"	"Y"	TOTAL	CONSENSUS
Route I	40	50	50	40	180/4=	45%
Route II	30	30	20	40	120/4=	30%
Pos. A	20	10	20	10	60/4=	15%
Pos. B	10	10	10	10	40/4=	10%
Total	100%	100%	100%	100%		100%

It is best to have each person assign his percentage privately (i.e., secret ballot style) because it will insure that even the meeker individuals will be able to express their opinions without being intimidated by the more vocal members of the group. If the consensus is gathered verbally, you will probably end up with the opinion of only one or two aggressive individuals and may run the risk of polarizing the group toward conflicting ideas which may detract from the efficient operation of the search. Additionally, since everyone is involved in the decision-making process, they will be more inclined to support the final decisions made by the SMC, which will result in a more efficient operation. Most people find that assigning a numerical value to the various possibilities is a very challenging task. This is probably the first time that they have been forced to relate each possibility in comparison to the others, but, hopefully, the process will result in a realistic appraisal of the situation. Two rules go along with this technique:

1. NEVER PLAN A SEARCH BY YOURSELF
2. NEVER DISCARD INFORMATION

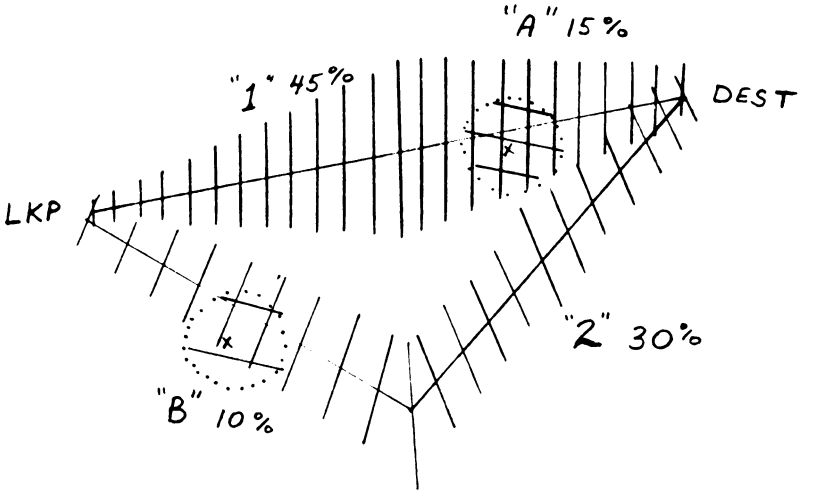


Figure 1. Graphical representation of search information.

The first rule will be obvious once you try this method! No two lists of probabilities will be the same, since we all have some preconceived ideas and no matter how hard we try to be objective we still tend to favor one line of thought over another. We must have help in overcoming our narrow-mindedness; since we do not KNOW where the survivors are, how can we say where they are not (if the information is within the realm of possibility)? Granted, we have had our reasons for discarding information in the past, but these "far out" leads could better serve us if they were included in our search plan with very low probability (i.e., 0.5%, 1%, or 5%); occasionally, one of these leads will be just what is needed to take advantage of new information. Please keep an open mind; don't discard any possible information and keep everything in perspective. Always remember: IF YOU KNOW WHERE THE SURVIVORS ARE, WHY ARE YOU SEARCHING!!!!!!????!!!!???? When the process is complete, which should take 10-30 minutes, the SMC has the consensus of the most informed and experienced personnel available, and he should be able to use this consensus to help him establish search area priorities for the next search. A method of getting a visual picture of the consensus is to draw each possibility on a map or overlay and to shade each possibility according to its probability percentage.

Figure 1 gives an example of a graphic that can be used to establish search area priorities. (The extra width along the track lines represents the navigational errors.) The first priority seems to be in area "A," the second could be where routes "1" and "2" overlap, and the third might be Lead "B."

This consensus may be used to establish search priorities and the SMC can be confident that he is assigning search units to the highest probability areas first, within the limitations of weather, terrain, and capability of the personnel/equipment. This is the beginning of optimal search planning, but many more variables must be considered before the most efficient and effective search can be planned. If you use this procedure early in the search, you will consider all available data, tap the knowledge of the best minds available, and this will give the survivors the best chance of being recovered alive. Good luck!

References

Syrotuck, W. G. [1975] "An Introduction to Land Search Probabilities and Calculations," Arner Publications, Westmoreland, New York, USA.

Syrotuck, W. G. [1977] "Analysis of Lost Person Behavior," Arner Publications, Westmoreland, New York, USA.

True, J. L., Jr. & Mattson, R. J. [1974] A Study of Light Aircraft Crash Locations, Headquarters, Aerospace Rescue and Recovery Service, Scott AFB, Illinois.

SURVEILLANCE WORKING GROUP REPORT

General Conceptual Structure for Surveillance Problems

Surveillance was found to be a difficult concept to structure in the Working Group on Surveillance/Fishing. The work reported here may not reflect the opinions of all the participants. It is nevertheless hoped that the ideas discussed here may prove useful in the analysis of some surveillance-type systems.

Slightly differing approaches or viewpoints are presented by Danskin [1962a, 1962b], Dobbie [1964, 1966], and Richardson [1973] on surveillance in a military environment. Pollock's paper in these proceedings states ideas and approaches that may be more widely applicable.

The reason why surveillance is usually carried out is to be forewarned or made aware of the possible occurrence of some special condition or state of the system ("the state of nature"). Call these states "special states." The information is required because some kind of action has to be taken afterwards. In a sense the system is moving or changing over time, possibly deteriorating. The cost or disutility associated with the action to be taken is usually increasing in time. This cost can be taken as proportional to the "time-late" of determining that the action is required. Even if this is not so, there may be a time-lag between deciding to implement the actions (e.g., corrective measures) and the actual action becoming effective (e.g., corrective measures becoming effective). Another view is that cost or disutility is inversely proportional to the information (neg-entropy) that the user of the surveillance system has about the system being monitored especially if this information is received in an "outdated" mode (i.e., after a time-lag). Pollock mentions that the possible actions may be to "send a message summarizing target location by means of a network that has charges inversely proportional to the (neg-entropy) information gained [by the surveillance system]."

The information, that the system is in a certain state or condition, or that a change of state has occurred, is thus usually used in a predictive manner. Thus, the name "early warning system" can be taken as a descriptive name for a surveillance system.

The user of the surveillance system has to furnish intelligence (in the human sense) in the surveillance system in order to predict which conditions or states may lead to a situation in which some actions have to be taken, i.e., the special states. Some of these states or conditions may not be definable in advance requiring human evaluation of the state of the system in order to find out whether the system is in a possible special condition or abnormal situation. In fact, the consistent aggregation of (state) indicators (see definitions below) of special states is a major problem. The problem of determining the characteristics of the special states in terms of measurable indicators may be the most difficult part of setting up a surveillance system.

The set of all special states or conditions, i.e., the set of all these states of the system which may lead to a situation in which some action must be taken, is the thing that is looked for or searched for by the surveillance system. This set is called the target set, or simply the target, and the elements of this set, the target states. Note that this set may or may not be well defined a priori, in the sense that it may not be evident that the monitored system is in one of these states.

The elements of the target set can be considered as alarm triggers. If the surveillance system detects that the monitored system is in one of the target states, an alarm is given or a warning sounded. Of course the alarm may be false, but the increasing of the probability of detection usually leads to the increase in the probability of false alarms. See Pollock [1971] or his paper in these proceedings for a discussion of how false alarms may be approached. Pollock shows that the overall system performance is a trade-off between desired probability of detection (state discrimination in this case) and the level of false alarms that can be tolerated. This trade-off must be made or guided by the user of the surveillance system (see also discussion under "hierarchical surveillance systems").

The surveillance process is concerned with monitoring of the system in order to detect changes of state(of the monitored system). It is thus possible to associate random variables with the states. The surveillance process is sampling the distributions on the random variables to determing the actual state of the system, to see whether a change occurred. Noise in the sampling process is difficult to eliminate because the distributions on the random variables may themselves be unknown. The process of

surveillance can be thought to consist of

(a) target identification,

(b) gathering target information.

Target Identification

Target identifiers are those indicators which help distinguish between the different states. They, in effect, are tags or labels associated with states in the target set. Target identifiers are very much dependent on the sensor in use in the sense that the sensor may or may not be able to signal the presence of certain, possibly crucial, indicators. In fact determining thresholds for the types of target signals, which serve as target identifiers, is a major practical problem. These thresholds are a trade-off between probability of detection and probability of false alarm, as mentioned above. The user of the surveillance system has to determine these thresholds by considering the cost or disutility of a misclassification of the state of the monitored system versus the cost caused by a false alarm (see also discussion under "hierarchical surveillance systems" where the level of surveillance concept helps to handle this problem).

Once the presence of certain target identifiers has been recorded with a specified level of certainty, the state of the system may be determined (with a certain accuracy). This is simply target classification. The whole process of target classification is taken as the target identification. Thus the target is considered identified when the state of the system has been determined to a certain level of certainty. To avoid confusion it is important to note here that target in normal search theory (e.g.,a ship) is not the target of the surveillance system. The proper classification of the ship (type, origin, etc.) and location in time and space is the "target :" certain combinations type, time, space,..., indicators may represent trigger states--the set of all trigger states is the target set, or simply the target of the surveillance system.

In practice each state will have a probability associated with it that gives the probability that the monitored system is in that state. Thus one can speak about the (probability) distribution over states. A measure of the certainty of target classification, as given by the state probabilities, can be obtained by looking at normal statistical measures of spread of norms. A better way is to consider the information content of the distribution over the states (see Richardson[1973]). This immediately yields an objective function for the surveillance

system, viz., minimize the entropy in this distribution or equivalently, maximize the information content (neg-entropy) of the distribution.

Gathering Target Information

Sometimes there is prior information on the occurrence or distribution of the target identifiers in the search space (the space in which the target identifiers occur). This prior information can be used to determine good places in the search space for allocating monitoring effort. In essence this becomes a search problem in the usual search theory sense, even though the objective function may be to maximize the information in the distribution or the system classification states (see Richardson[1973] for an application). Thus it is only when prior information about the distribution of alarm triggers is available, that one is capable of employing search theory in the classical sense.

Hierarchical Surveillance Systems

Consider the number of differing system states as a measure of surveillance discrimination. At any one level of surveillance the user may accept a low probability of detection as well as a low false alarm rate. The sensor necessary to provide this (very) rough target classification can be reasonably unsophisticated. But as soon as the disutility of not finding the target alarm triggers rise, the user requires another level of surveillance. This may be achieved by the same sensor but set with a lower threshold for target signals, or by a completely different sensor. For example only the CPI may be monitored in an economic system, but as soon as an abnormal pattern is recognized (i.e., economy possible in a trigger state) monetary flow may also be monitored, in order to refine the information on the possible state of the economy. Another example is medical, where temperature may be used as a first indicator and only when this shows abnormality is another symptom checked. Of course most real life systems are so complex that many indicators have to be monitored simultaneously in order to have even a mediocre idea of what the current state of the system is.

The cost trade-off in the user's value space of nondetection of target states and false alarms determines the level of surveillance. As these costs may not be stationary over time (they usually are not) the surveillance system will at various points in time perform at various levels of efficiency, where efficiency is measured in terms of the number of "fineness" of classification and associated certainty.

It is hoped that the reader will find some ideas described here useful in structuring his surveillance problem.

General Formal Structure

We present here a brief outline of a general structure within which a broad class of surveillance problems may be modeled. The interest is not to provide specific solutions, but to hopefully outline the kind of variables, parameters and relationships that might be involved.

The system under study may be considered to be, at time t in some state $S(t) \in S$ where

$$S = \{1,2,\ldots,n\}.$$

We would like to know at any time in which state the system is. However the crux of our problem is that the state of the system changes over time. The surveillance problem then, in principle, becomes one of being able to compute the probability that $s(t) = i$ for all i and t, since most "optimal" or at least "good" decisions can be made by comparing this probability with some threshold (which surely depends on the various costs, and prior information). We also have available a (possibly vector) attribute X which we can measure at various times, and whose value at time t will be represented as X_t. Given this information we need to thus compute

$$pr\ \{S(t)=i \mid X_1, X_2 \ldots X_t\}$$

By Bayes' rule, this is

$$(1) \qquad K pr\{X_1, X_2 \ldots X_t \mid S(t)=i\} pr\{S(t)=i\},$$

where the factor K is a normalizing constant, the second the sampling function, and the third an a priori probability statement about the behavior of the process without any observation information.

Computing (1) is of course straightforward and decision planning problems involving relations of this kind have been attacked with some success. However in most real surveillance problems the sampling function is not well known or specifiable, particularly those conditioned upon states S(t) that have rarely if ever been observed to occur. We can account for this however by allowing the sampling functions to depend upon a parametric vector w which is itself a random variable, and which depends upon the state.

Letting the prior distribution on w, given the state is i, be $f_i(w)$, then we can write

$$(2) \quad pr\{X_1, X_2, \ldots, X_t | S(t) = i\} = \int pr\{X_1, \ldots, X_t | S(t)=i, w\} f_i(w) dw.$$

Putting (2) into (1) allows us in principle to proceed.

Early Warning

The group had some difficulty in defining its task. Gereralities are given about exactly what might be meant by surveillance and early warning.

Two specific areas were mentioned in some depth. In medical surveillance, there are questions of surveillance of a large population to identify persons at risk for various diseases, questions of the timing of the tests during a person's life, questions of identification of disease in persons at risk, and questions of optimal sequencing of tests.

For environmental monitoring, such as air and water pollution monitoring, there are questions of location of monitoring stations, frequency and timing of monitoring, specific variables to measure, and decision rules on when to take what sort of action, taking into account relevant social costs, etc., as well as the physics and chemistry relating pollution to values actually measured.

Brief descriptions of early warning related to detecting business problems and likely student failure were also given.

There are three main problem characteristics:

The state of nature (the "target") is some physical phenomenon that (possibly) changes at some point in time.

The problem is to discover ("detect") the time of change as soon as possible yet without the consequence of false detection.

There may also be an associated desire to make statements about the characteristics of "current" state.

There are five components that must be considered:

The states have associated with them various measurable quantities--variables and identifiers--which can be observed, and serve as the major source of information about the states.

These are in general random variables with some "sampling function" PDF's conditional upon the states of nature.

These sampling functions, however, may not be specified with certainty: Their parameters may in themselves be random variables (in the Bayes statistics sense).

Some a priori information also exists about the times of changes of the states, e.g., in the absence of observed variables one can still try to guess the time of change.

A cost structure must be assumed, or an equivalent subjective function stated. In the former case, explicit costs ascribed to misidentifying states at various times; in the latter other objectives (such as minimizing expected time to detect change, subject to constraints on expected time between false alarms) must be stated.

The following procedure seems to be common.

A number of decisions and questions may be involved:
(a) Should the stated definitions, characterization and associated identifiers be fixed or adaptively changing in time?
(b) How often should the identifiers be measured, which ones and with what precision?
(c) What should the decision rules (the "trigger level") be for deciding when change, and which one, occurs?
(d) How sensitive are the results to all this?

Almost always pertinent decisions will be made by comparing some version of state probability threshold.

In the presentation, an output display of state probabilities at all times in the process may be more useful than decision rules.

Use of semantically loaded terms such as "target," "warning," "trigger," and "false alarm" should be critically examined in the light of the decision maker's reaction.

Pollution Monitoring

The main features of pollution monitoring can be summarized by the following example concerning chlorine pollution by a factory:

State	=	1, factory is dumping pollutant in river 2, factory is not dumping
Time of change	=	time dumping starts
Variables	=	Cl_2 measurements at k sites
Sampling functions	=	river background from experiments given state 1 or state 2
Arrival time	=	past history of production process
Costs	=	a) if time to decide is too long, no "smoking gun" evidence can be taken at the factory b) if missed two costs: decreased deterrence and increased economic/social costs of Cl_2 in the river
Decisions	=	a) where, when, how often to sample on river b) what decision thresholds on Cl_2 levels to use.

Police

There are many possible uses of search theory in police work. The working group concentrated on: missing persons, border patrols and allocation of police patrols. Often in police work the target does not wish to be found and is able intelligently to evade the search.

Missing persons. Resources are devoted to finding, for example, a missing child who has wandered off or has been abducted, a husband or wife who has left home and does not wish to be found, a murder victim, an accident victim who cannot summon help, weapons, stolen goods, etc. These can be divided into evasive and nonevasive categories of which the latter type may be amenable to SAR techniques if the number of alternative "boxes" is not too great. Contact between criminals and their friends also modify the normal approach to this search procedure and hypothetical prior probabilities and distribution functions can be evaluated for petty criminals. The police are usually so experienced that a search approach is unnecessary. Hidden objects

are often sought with metal detectors but control of searched areas is very difficult on the small scale required and the existence of many false alarms adds to the problems.

Border patrols. These are similar to barrier patrols in naval work (Koopman [1946]) but involve very variable terrain and vegetation and the possibility that the patrol units are visible to the target. Although dividing the line into segments and conducting back and forth patrols are possible, intelligent evasion means that existing theory must be extended.

Allocation of resources. This has been studied by Larson at MIT by examining Police car patrol routes which involve the looking for an incident of a very transient nature. See Larson [1972].

Salmon Management

The group discussed the problems of Salmon control which did not seem to offer many opportunities for the application of search methods since the fish are both easy to locate and to catch. The control and catch limits imposed are, however, very critical.

Surveillance in Fisheries Management

The primary objective for management of offshore fish stocks is to develop and utilize the resources for the social and economic well-being of the people and industies of the coastal state. The fish stocks must also be allowed to regenerate compatible with renewable resource management practices so that overfishing is avoided and a future resource base is maintained. In some developing countries, fish are the only significant resource and their extinction would have disastrous consequences.

Fisheries search and surveillance have a major role to play in all the main elements of the management framework:

(i) Resource Assessment.

(ii) Fishing Plans (Domestic and Foreign).

(iii) Licensing Arrangements.

(iv) Surveillance and Enforcement.

(v) Regulations and Penalties (Fines, etc.).

(vi) Information Systems.

(vii) Management Analysis, Evaluation and Administration.

In the case where the fishery resource of a country is the most significant, the fees accruing from licensing arrangements and penalties would contribute a considerable income. Most nations have established a 200 mile exclusive economic zone and this sets the boundary of a coastal state's responsibility for fishery surveillance. However, if fishing an area of open sea becomes a cooperative program, the surveillance problem, in principle, is the same.

There are changes in detail with respect to the species, e.g., tuna, herring, sardines, cod, etc., and also with respect to the fishing method. The basic problem is that there is a bounded area in which licensed and unlicensed fishing vessels operate. The efficient management of the area requires a knowledge of the total fishing effort and also a facility for the detection and arrest of those vessels fishing illegally. The surveillance system would be a combined sea/air system and might have to be financed from the "surveillance revenue."

Fisheries' managers frequently have different motivations from those of individual fishermen. Improved searching techniques, for example, may benefit the fisherman in the short run, but may result in the elimination of fishing licenses in the long run if fish stocks are to be conserved. The question of goals is a particularly delicate one in fisheries.

Other factors which affect the model and objectives in surveillance problems are:

a) Localization after detection.

b) Identification after detection.

c) False targets.

d) Intermittent or continuous contact.

For example in trying to predict the position of several targets where there are no difficulties with identification, the RMS position errors may be a suitable objective. If some targets are misidentified, however, then this criterion would be completely dominated by the occassional large "position error" that appears to result when a target is misidentified. To retain sensitivity to position estimates of correctly identified targets, one might instead use the probability that the error does not exceed one mile (say) as an objective function. Obviously, a great variety of techniques are required to handle the many types of surveillance problems. Methods ranging from game theory to elaborate computer simulations might be appropriate in different circumstances.

Classification of Surveillance Problems

The objective of surveillance and the method of attacking the problem are likely to differ strongly between the two cases of evasive and nonevasive targets.

Surveillance in nonevasive motion might attempt to count the number of fishing boats in a given region to monitor the fishing effort by detecting targets and localize them in a rough manner. In evasive motion (e.g., a boat that fishes illegally) one wishes not only to detect a target but to identify, localize and pursue it.

Surveillance for Enforcement

To illustrate a possible model consider an unlicensed vessel desiring to fish in region i which incurs a cost $(F\lambda(x_i)+c_i)/r_i$ per fish caught, where F is the fine for illegal fishing, x_i is the surveillance effort expended in region i, $\lambda(x_i)$ is the rate of being captured, r_i is number of fish caught, and c_i is the rate of spending on things other than fines in region i. Our hypothetical vessel would choose the region where this ratio is smallest. Thus,

$$\min_i\{F\lambda(x_i)+c_i)/r_i\}$$

would be the objective function to be maximized by the surveillance force by choosing $x_1, x_2, \ldots$, subject to the usual constraints. Specifically the evasive vessels cannot be modeled by assuming a probability distribution for target motion independent of the searchers' motion but need the use of game theoretical methods making use of worst case assumptions.

Note that search theory plays a role in the establishment of the functions $\lambda(x_i)$ but that the surveillance problem requires extra inputs. For example the optimal value of x_i is 0 if r_i is 0, etc. Also x_i cannot be determined without knowing r_i, which means that a separate surveillance of fishing quality may be necessary.

Surveillance for Measuring Effort

The large fishing region is divided into K cells in order to estimate the number of fishing boats in cell i at time t. Surveillance is performed by a quickly moving searcher, so that it is assumed the boats are stationary during the surveillance.

A prior estimate of the number of boats in cell i is given by a set $\{N_i(0), i=1,\ldots K\}$ where $N_i(0)$ is the prior estimate of the number of boats in cell i. Let $p_i(0)$ be defined by

(3) $p_i(0) = N_i(0)/\{\Sigma N_i(0)\}$

so that $p_i(0)$ is the fraction of boats in cell i. This problem is now very similar to a standard search problem except that in the standard problem one would view $p_i(0)$ as the prior probability of the target in cell i. After search, one updates these probabilities in a Bayesian fashion.

After surveillance, one also updates $p_i(0)$ to $p_i(1)$, the estimate of the fraction of boats in cell i after surveillance at time 1. The surveillance does not stop, of course, with detection. Between surveillances the boats may move and so the motion must be included in creating the prior distribution before the next surveillance.

Applications of Search Theory to Finding and Counting Caribou in the Arctic Wastes of Canada

Management of large animals such as caribou frequently involves searching large geographic areas for herds of animals. Caribou herds normally range over areas as large as 100,000 miles during their annual movements. All females and some males will aggregate during June and July for calving in a relatively small area, but during winter the herd will be dispersed into smaller groups, and the groups may be found nearly anywhere on their range. Most of the caribou will winter in a few large groups, but given the large areas to be searched, these groups are not easily found.

Searches are conducted for two purposes, to count the animals and to locate their areas of habitat use. Managers need to know numbers in order to regulate the harvests and generally assess the status of the populations. Location of caribou in winter is important because of current concern about the impacts of highway and pipeline development. Baseline information about distribution is required to assess the impacts of developments after they occur, and potentially to regulate highway use and pipeline access after construction. In addition, caribou vulnerability to hunting is highly dependent upon the location of major aggregations, and managers need information on distribution to understand the impacts of hunting.

The problem of locating herds of animals in large areas appears to be quite suitable for the appication of search theory to 1) generate probability distributions for the location of herds

and 2) determine optimal flight paths for surveillance aircraft.

There are five problem areas:

1. Location of a single herd.
2. Locating several herds.
3. Locating tagged (radio or visual) individuals.
4. Locating herds which are geographically dispersed.
5. Problems of estimating the site of caribou populations.

The solution to these problems involves several steps that fit nicely into the weighted scenario approach of search theory. These steps are:

a. definition of alternative scenarios about the location of target herds or animals;

b. assignment of prior probabilities to each scenario from historical observations and biological knowledge;

c. the assignment of search effort given the distribution probabilities;

d. calculation of new probabilities given the results of the search effort, and the probabilities that the targets will move between searches;

e. iteration of steps c-d to direct subsequent search and maintain updated distribution probabilities.

In the location of a single herd the problem is to determine the optimal allocation of search effort to yield a minimum expected time to detection which is identical to the classic moving target problem developed for submarine search (Brown [1980], Stone et al [1978]).

When there are several herds, whose locations are independent of one another, the optimal solution for multiple herds is an easy generalization of the optimal solution for a single herd (see Strumpfer [1979]).

Individual caribou are frequently tagged with a radio transmitter or visual markers for subsequent location and recognition and since caribou are social the sitings of nontagged individuals do provide information about the probability of location of the tagged individuals.

Although (1) assumes that the herd is a single unit this is rarely the case since many individuals are separated into smaller groups which are indicative of larger herds nearby and this can be modeled as a group of animals with a mean geographic location and a distribution about that mean.

Estimates of the number of caribou are generally obtained by enumerating the animals on the calving grounds by air survey. An alternative is to locate all major aggregations or to fly random transects and estimate numbers.

References

Brown, S. S. [1980] Optimal Search for a Moving Target in Discrete Time and Space, Oper. Res. (to appear).

Danskin, J. M. [1962a] A Theory of Reconnaissance: I, Oper. Res. 10, pp. 285-299.

Danskin, J. M. [1962b] A Theory of Reconnaissance: II, Oper. Res. 10, pp. 300-309.

Dobbie, J. M. [1964] Surveillance of a Region by Detection and Tracking Operations, Oper. Res. 12, pp. 379-394.

Dobbie, J. M. [1966] Solution of Some Surveillance-Evasion Problems by Methods of Differential Games in "Proceedings of IVth IFORS Conf." Boston.

Koopman, B. O. [1946] Search and Screening, OEG Report No. 56, Washington, D. C. (a modernized edition has been published by Pergamon Press, Inc., New York, 1980).

Larson, R. O. [1972] "Urban Police Patrol Analysis," MIT Press, Cambridge, Massachusetts.

Pollock, S. M. [1971] Search Detection and Subsequent Action: Some Problems on the Interfaces, Oper. Res. 19, pp. 559-586.

Richardson, H. R. [1973] ASW Information Processing and Optimal Surveillance in a False Target Environment, Daniel H. Wagner, Associates Report to Office of Naval Research (DDC AD A002254).

Stone, L. D., Brown, S. S., Buemi, R. P. & Hopkins, C. R. [1978] Numerical Optimization of Search for a Moving Target, Daniel H. Wagner, Associates Report to Office of Naval Research.

Strumpfer, J. P. [1979] Search for Multiple Moving Targets, Technical Report, Institute for Maritime Technology, Simonstown, South Africa.

SURVEILLANCE AND SEARCH THEORY

S. M. Pollock

Department of Industrial and Operations Engineering
The University of Michigan

Surveillance in general involves detection devices, search (allocation-of-effort) procedures, and decisions leading to action of some kind. The purpose of this paper is to indicate briefly the role of search theory within the larger context of surveillance problems. I will discuss elements of surveillance other than search theory: detection and related forms of information gathering; appropriate measures of effectiveness; relevant decisions and their consequences; and most important the sequential nature of continuous evaluation of information. Search problems usually concern an entity that is to be "looked for" and associated action to be taken when it is "found." The entity is usually a physical object (although it could be nonphysical, such as the value of the maximum of a mathematical function), and the term "target" is natural to use. The person or organization responsible for the looking is called the "decision maker." The actual expending of effort, with the intent to gain information about the target, is accomplished by the "searcher(s)." By the very nature of search problems, the target's present status is at any time uncertain, and thus describable only probabilistically. Search theory, in this context, has to do with the allocation of searching effort in such a way as to achieve some stated goal, usually presented in terms of producing an extremal value of some objective function.

Surveillance, on the other hand, is usually taken to mean not only the searching for some target, but the larger problem of defining target characteristics, and determining an appropriate time-space window within which target appearance is relevant. This broad definition serves to encompass the everyday usage of the

term surveillance in the following situations:

a) Police: categorized by target type (burglars, known organized crime figures, bunko artists), physical area (diamond district, bank lobbies, red-light districts), and type of search activity (anti-hijacking screening, stop-and-frisk).

b) Medical: the monitoring of communicable diseases in an entire population, in conjunction with anti-epidemic programs; or the monitoring of an individual patient's vital signs in an intensive care unit.

c) Environmental: the continuous measurement of pollution in streams or air; aerial surveys of drought areas; enforcement of regulations (e.g., strip-mining methods, work place health and safety conditions).

d) Military: constant monitoring and reassessment of enemy physical or telecommunication activity; assessment of intent, etc.

e) Library usage: in contrast to search -- when the target (book, article) is known-- there is the activity of "browsing" involving a target not-yet-identified at the onset of the activity.

There are, of course, other forms of surveillance, for example those involving seeking a job or a social partner.

It is convenient to use the general structure of surveillance problems schematically illustrated in Fig.1. At the very start, there is a certain amount of target-location information available to the decision maker. This "prior knowledge" could, of course, be based upon a sizable portion of subjective analysis, and is accomplished by a mechanism labeled "intelligence" (in the military sense).

Using this information in conjunction with a statement of overall objectives (the "criterion"), one may make a decision to allocate effort in such a way as to gain more information about the target location. This allocation, commonly known as a search, influenced by the actual location of the target (the "state of nature"), produces an outcome that serves to change, or "update," the decision maker's information about the target's location. A new search decision may then be made, and the process continues.

At some point, however, the target information is such that, rather than allocating more effort to search, the decision maker may decide to take a nonsearch action: one that has as its

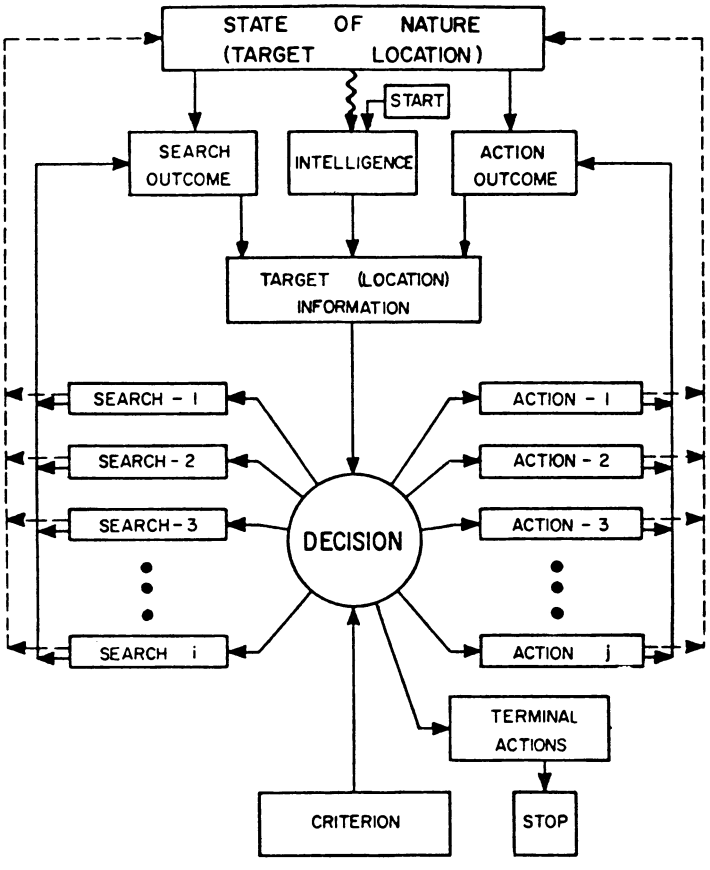


Figure 1. A schematic structuring of a general surveillance problem

objective not gathering target information, but gaining utility. Although some of these actions might be terminal, such that the process stops once the action is taken (the depth charge is dropped, the photo taken, the hideout staked out) other actions produce outcomes that serve to update target information and lead to further decisions.

It is also possible that the decision maker, during the course of the process, receives relevant "free" data about the target, without allocating search effort or taking specific actions. These data are again encoded by the "intelligence" function, a step indicated by the wavy line on the diagram. The figure also shows dotted lines leading from the effort allocation boxes (both search and action) to the state-of-nature box. These are meant to indicate interactions possible when the target can change state (e.g., move) as a direct result of searches or actions.

Such a structuring makes it possible to represent the relations among existing theories and studies by identifying which parts of the figure they analyze. This is shown in the simpler schematic of Fig. 2, in which portions of Fig.1 have been lumped together.

Thus, search theory deals with allocating search effort, given target information and a criterion. Detection theory is the study of the target information (as it turns out, this is invariably probabilistic) obtained when one is given the state of nature and a specified allocation of search effort. When a nonsearch action is taken and the target information is probabilistic, the analysis of the appropriate action to be taken is within the framework of statistical decision theory. When a nonsearch action is taken and the target information is deterministic (i.e., target location is completely known) then the appropriate action either results from deterministic optimization, or continuous game or control theory approaches to pursuit and evasion.

We see that there are six areas of interest (the boxes in Fig.2) that must be analyzed in order to understand a surveillance problem fully.

The state of nature summarizes the actual target status, as well as all "givens" of the problem, and is thus the distillation of the model in terms of constraints, target behavior, physical constants, etc.

Target information involves encoding: how one assimilates and presents (for the use of decision makers) constants of the

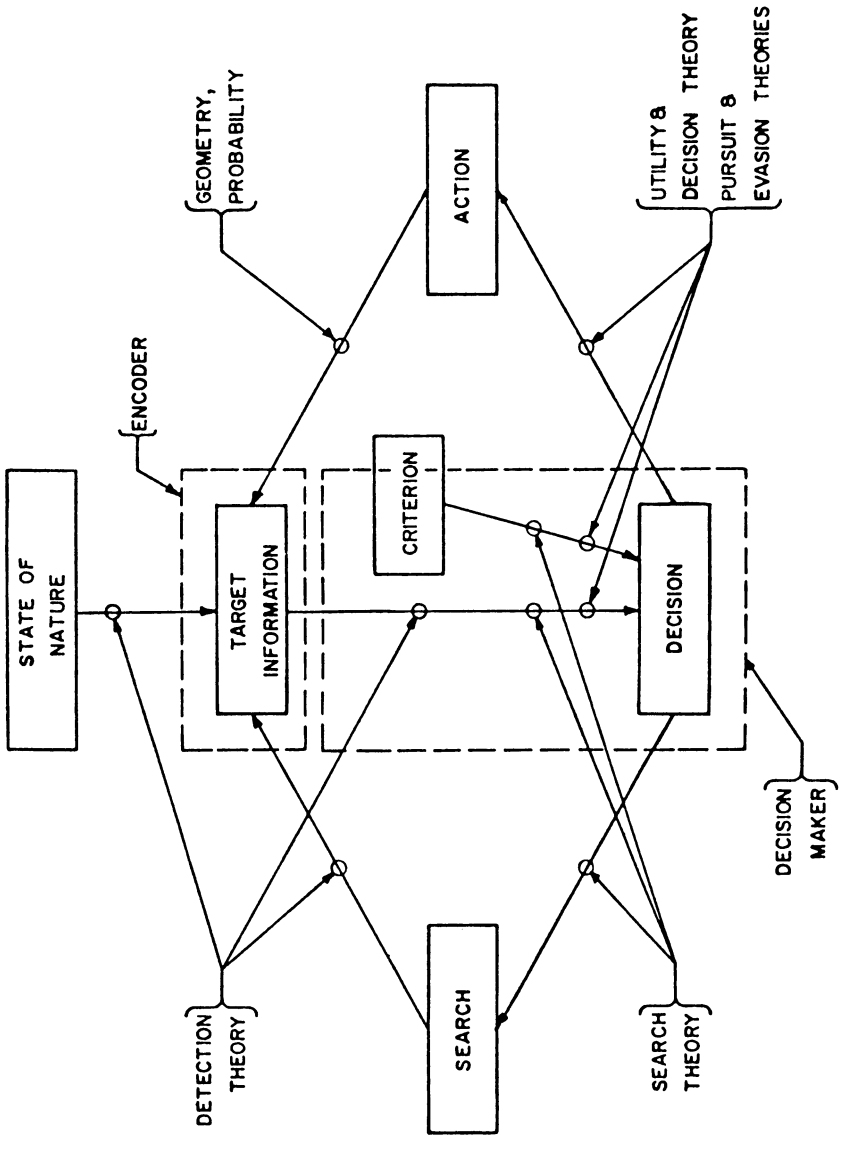


Figure 2. The role of other disciplines in the analysis of surveillance problems.

problem, parameters of the real world, results of decisions, subjective probabilities, etc. The model that is being used to transfer the real world into a theoretical or abstract basis is, of course, the first step in such encoding.

The criterion establishes the primary purpose of the overall process, why the operation is being carried out, what value it is to have various decisions made under possible states of nature.

Search has to do with procedures: the method of allocating resources to provide target information.

Actions are noninformative allocations of resources, made to provide some utility, in satisfaction of the overall criterion.

Decision rules convert information, constraints, and criteria into actions or search.

In what follows, we ignore the difficult mathematical and practical problems of search theory (i.e., the optimal allocation of search effort involved in the left-side loops of the figures), since they will be (or have been) amply covered in other presentations in these proceedings. We first briefly discuss instead one action-oriented aspect of surveillance: dealing with false alarms.

We then present a general approach to the problem of characterizing surveillance for a target whose time of appearance and nature are uncertain.

Dealing with False Alarms

Since nontrivial detection situations involve stochastic elements, the fundamental problem of detection theory is to determine, for a given detection device, environment, target, decision rule, etc., the two related conditional probabilities

$$p_d \equiv \text{detection probability} \equiv \text{prob}(D_A | T_A),$$

$$p_f \equiv \text{false alarm probability} \equiv \text{prob}(D_A | \bar{T}_A)$$

where the events D_A and T_A are defined by "D_A" = "the decision is made that the target is in some time-space region A;" and "T_A" = "the target is in A."

Once these probabilities are calculated for any given detection device, a curve of p_d versus p_f may be plotted as a function of the decision rule, with all other parameters of the problem fixed. This plot is called a "receiver operating characteristic" (ROC) of the detection system. (The use of the word system refers to the combination of detection device and associated decision rule.)

Figure 3 shows the ROC for the special case of detecting a known signal in white Gaussian noise, where a threshold decision rule is used, i.e., if the observed value of a test statistic x is greater than some x*, decision T_A is made. Moving along the curve (by changing the threshold x*), we note that, if p_d goes up, so does p_f--where should one operate on these curves? Note that this question is different from (although clearly related to) the well studied problem of allocation of effort in the presence of false alarms. In the latter case, the false alarm generating structure is assumed to be fixed, and not under the control of the searcher.

As an illustration of how one might decide on a good operating point on the ROC, consider the classical example of random search. The probability of detecting the target by time T is

$$(1) \quad P(T) = 1 - \exp(-\gamma T),$$

where γ is the conditional detection rate per unit time. This rate is related to a system's probability of detection and is often taken to be $p_d/\Delta T$, where ΔT is the time between detection decisions. Whatever the interpretation, however, it is critical to note that for any given detection rate γ, there is an associated false-alarm rate f (which maybe be similarly interpreted to be $p_f/\Delta T$).

Now, the penalty associated with a false alarm is usually the time and effort spent tracking it down (cycling through the non-search action of Fig.1) and discovering that the target information was not representative of the state of nature. Suppose that τ is the average time wasted for each false alarm, and the false alarm rate while searching is f.

In this case, if T_S is the expected time available for searching (the actual time available for searching is, of course, a random variable), the total expected time wasted in tracking down false alarms is $(fT_S)\tau$, since fT_S is the expected number of false alarms that will occur while searching. Since T is the total available time for the mission, we have

$$(2) \quad T = T_S + fT_S\tau$$

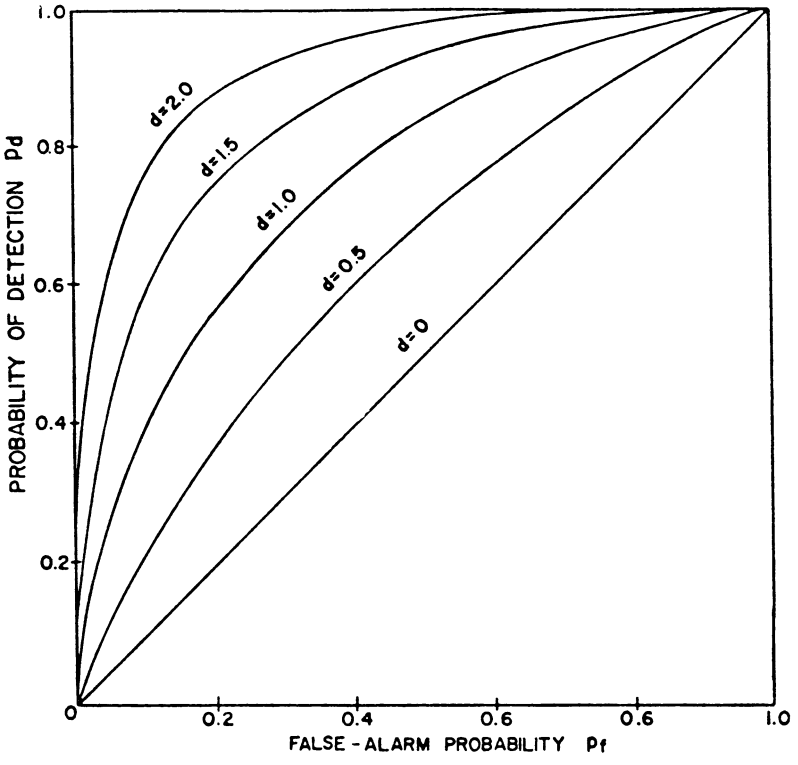


Figure 3. The ROC for the detection of a know signal in white Gaussian noise; d = signal-to-noise ratio.

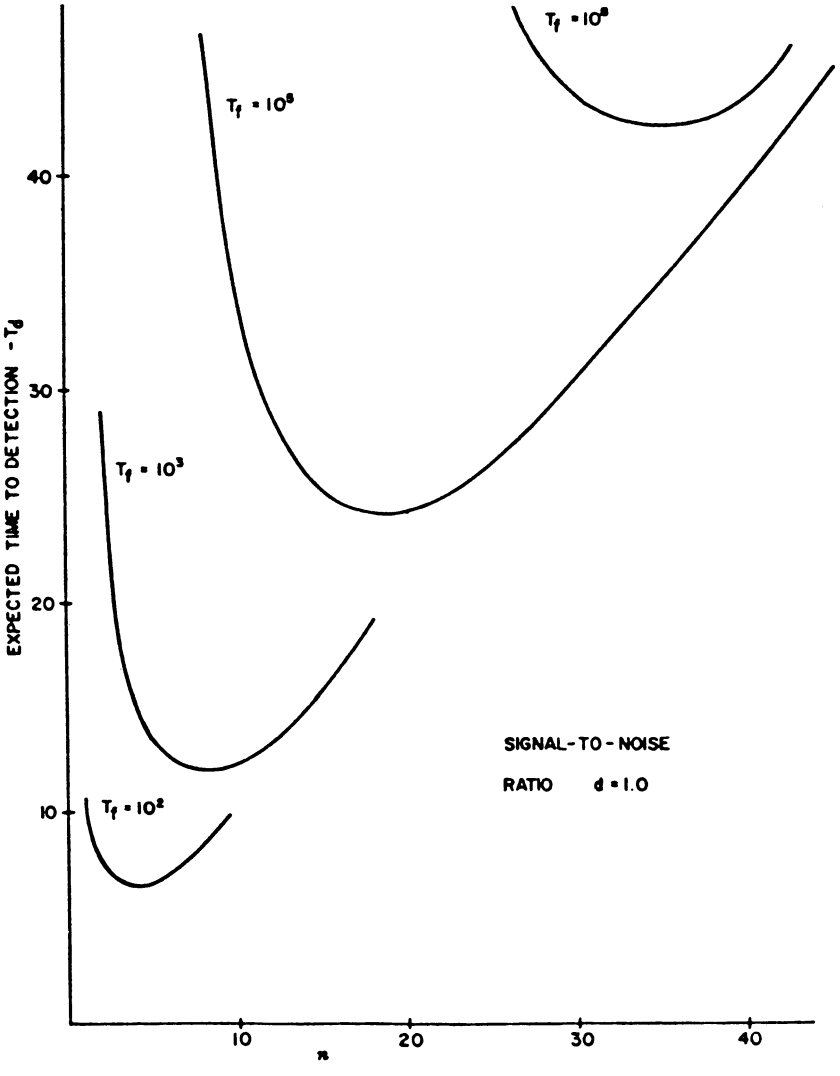


Figure 4. Operating characteristic suitable for continous detection of a know signal in white Gaussian noise. $\Delta T = 1$, $d_0 = 1.0$, and n = number of samples before decision.

or

$$T_s = T/(1+f\tau). \tag{3}$$

As an approximation to finding the probability of detecting the target at or before time T, equation (1) is now used, with T_s replacing T, i.e.,

$$P(T) = 1 - \exp\{-[\gamma T/(1+f\tau)]\}. \tag{4}$$

One way of looking at this result is to say that the rate γ in equation (1) is replaced by an overall effective rate γ' where

$$\gamma' \equiv \gamma/(1+f\tau). \tag{5}$$

We have seen earlier from the ROC that an increase in p_f accompanies an increase in p_d. Thus an increase in γ is accompanied by an increase in f. If it is desired to maximize the detection probability P(T) for a given value of available search time T, for the extremely simple model analyzed using the approximation of replacing T_s by its mean, then the optimum operating point is one that maximizes γ'.

A Time Oriented Operating Characteristic

Characterizing the performance of a detection system by the pair of numbers (p_d, p_f) is useful only if the period of time between decisions is clearly specified and understood. However, suppose system A achieves a higher signal-to-noise ratio than system B at the expense of requiring a larger time to make detection decisions. Which system is "better?"

In such situations, two obviously important operational parameters are T_d= time until detection (length of time to say "target present," given in fact it is present), and T_f= time between false alarms (length of time to first say "target present," given it is not present). In all real operations T_d and T_f are random variables, but their operational aspects may be measured by their respective expectations $E(T_d)$ and $E(T_f)$.

For example, consider the case of detecting a known signal in Gaussian noise (see Fig.3), using n successive indepent samples of the statistic x, taken at intervals of time ΔT.

The expected time until detection is $E(T_d) = n\Delta T/p_d$, and the expected time between false decisions is $E(T_f) = n\Delta T/p_f$, where p_d and p_f are related by the receiver operating characteristic of Fig. 3. Fig.4 shows $E(T_d)$ as a function of n for given values of $E(T_f)$. We see that there is a number of samples n* that minimizes

the expected detection time for a given value of the expected time between false alarms. With such curves, systems can be compared using as a reasonable criterion the minimization of the expected time to detect the target (if it is present), given a constraint on the expected time between false detections. Or, modified ROC curves, plots of $E(T_d)$ versus $E(T_f)$, can be used to compare detection and decision rules.

Surveillance for an Arriving Target

The various types of surveillance activities mentioned at the beginning of this paper have a common aspect: they involve the monitoring of a particular space of activity for the eventual appearance of a target. In such situations, we are again faced with the explicit recognition of the importance of false alarms, i.e., the taking of action appropriate to concluding the target has "arrived," when it in fact has not. Now, however, it is not a matter of seeing how long it takes to find the target, once it arrives. It is rather a problem of deciding whether or not it has arrived. One way of analyzing this kind of monitoring process is suggested by the discussion here. Similar approaches are currently under study for specific cases of surveillance involving pollution monitoring, and automatic machine maintenance (in which case the "target" is a breakdown).

Consider a discrete-time situation, in which:

1. An event E ("target arrival") occurs at time $T(T = 1, 2, \ldots)$ with known probability $p(t) = \text{prob}\{T=t\}$.

2. A surveillance observation is made at time τ $(\tau = 1, 2, \ldots)$, the result being encoded as a random variable $x(\tau)$ which has p.d.f. $p_0(x)$ if $T > \tau$ and $p_1(x)$ if $T \leq \tau$.

3. Immediately after each of these observations, one of the following decisions is made:
   D = "Decide the target has arrived;"
   W = "Wait for another observation."

4. If D is picked and $T > \tau$, then a false-decision cost F is incurred, the knowledge that $T > \tau$ is gained, and the process continues. If D is picked and $T \leq \tau$, then the process is terminated with a cost of $c(T, \tau) = (\tau - T)w$, so that the terminal cost is proportional to the time "late."

We shall also, at first, assume that the occurrence time T

of the event E has the distribution:

$$p(t) = a(1-a)^{t-1} \quad t = 1,2,\ldots .$$

A straightforward dynamic programming argument produces an expression for V(P) = minimum expected cost (achieved by using the optimal strategy) where P is the probability that E has occurred:

$$V(P) = \min\{D(P),H(P)\} = \begin{cases} D(P) & \text{if } P > q \\ H(P) & \text{if } P \leq q, \end{cases} \tag{9}$$

where

$$D(P) = (1-P)(F + V(a)),$$

$$H(P) = Pw + \int_{-\infty}^{\infty} [Pp_1(x) + (1-P)p_0(x)]V[P'(x)]dx.$$

The function

$$P'(x) = \frac{Pp_1(x) + (1-P)p_0(x)a}{Pp_1(x) + (1-P)p_0(x)}$$

is the posterior probability that E has occurred, given an observation x is made. The threshold probability q can be determined numerically. The solution of equation (9) can be obtained by an iterative method. The interpretation, however, is more germane.

From the form of the decision structure of equation (9) we see that P is constantly compared to the decision threshold q. When a D decision occurs, P automatically returns to 0 if the target has not yet, in fact, arrived. With a series of W decisions, however, a series of observations $x_1,x_2,\ldots$ have been made; and P is the appropriately derived posterior probability that the event has occurred.

Specifically, let us consider the process to start with a W decision at $\tau = 1$, and that n successive observations of $x_1,x_2,\ldots,x_n$ are made. We shall also now consider a completely general distribution for T, f(t). Using the definition of conditional probability, where $X = (x_1,x_2,\ldots x_n)$, we define

$$P_n \equiv \text{prob}\{t \leq n | X\} = \text{prob}\{X, t \leq n\}/\text{prob}\{X\}.$$

The unconditional probability of receiving some vector X is

$$\text{prob}\{X\} = \Sigma_{j=1}^{n} f(j)[\Pi_{i=1}^{j-1} p_0(x_i)] \; [\Pi_{k=j}^{n} p_1(x_k)] + \Sigma_{j=n+1}^{\infty} f(j) \; [\Pi_{i=1}^{n} p_0(x_i)]$$

Similarly, we may calculate

$$\text{prob}\{X, t \leq n\} = \Sigma_{j=1}^{n} f(j) \; [\Pi_{i=1}^{j-1} p_0(x_i)] \; [\Pi_{k=j}^{n} p_1(x_k)] .$$

The condition for a D decision, given that X has been observed, is that $P_n \geq q$. Using the above expressions and manipulating terms, this condition becomes

$$(10) \qquad \Sigma_{j=1}^{n} f(j) \left[\Pi_{i=j}^{n} L(x_i)\right] \geq q/(1-q) \Sigma_{j=n+1}^{\infty} f(j),$$

where

$$L(x_i) = p_1(x_i)/p_0(x_i)$$

is the likelihood ratio for each observed x_i.

Equation (10) represents a very complicated process. Not only is the $\Pi_{i=j}^{n} L(x_i)$ term weighted by the $f(j)$, but this weighting is successively compared to a term which gets smaller as n increases. This is a nonMarkovian (because P_n is more than just a function of P_{n-1}) stochastic process with a nonconstant absorbing barrier and is, in general, difficult to analyze.

If we let $f(t) = a(1-a)^{t-1}$, however, an interesting result is shown. By defining $A(x_i) = L(x_i)/(1-a)$, equation (10) becomes

$$(11) \qquad Z_n \equiv \Sigma_{j=1}^{n} \left[\Pi_{i=j}^{n} A(x_i)\right] \geq q/(1-q)a$$

which is a test that compares a variable Z_n to the constant decision threshold, $q/(1-a)a$.

In addition, the sequence Z_n describes a Markov process, in that Z_{n+1} only depends upon Z_n and $A(x_{n+1})$

$$Z_{n+1} = Z_n A(x_{n+1}) + A(x_{n+1})$$

which can be verified by direct substitution into the definition of Z_n.

Equation (11), derived above for the special case of random occurrence, (and equation (10), for the general case) resembles the Wald sequential-probability ratio test (SPRT). In the SPRT, the variable $\Pi_{i=1}^{n} L(x_i)$ is compared to fixed decision thresholds, and all x_i are weighted equally in the decision process. However, equation (11) indicates that more recent observations are more important. It is interesting to note that equation (12) bears great similarity to equations found in exponential smoothing.

Realistic surveillance situations, of course, require much less simplified models than that described above. One particularly important feature is that often, although $p_0(x)$ is a well known sampling function (being estimated from many observations of the "status-quo" condition), the target-present function $p_1(x)$ may be

very uncertain. It is possible to model this as being of some functional form with a distribution $\xi(\theta)$ over parameters θ in some space Θ, so that $p_1(x/\theta)$ is given, and successive observations of x not only update P, but also $\xi(\theta)$. The details of such analysis, for a variety of target appearance models, is currently underway.

COMPUTER SIMULATION OF SEARCH TACTICS FOR MAGNETIC ANOMALY DETECTION

Ronald H. Kuwahara

Research and Development Branch
Department of National Defence, Canada

The use of magnetic anomaly detection (MAD) systems to search for and track submarines is a subject of renewed interest to the Canadian Forces with the recent acquisition of the new Aurora maritime patrol aircraft with its highly sensitive ASQ-502 magnetometer. To enhance the performance of the new magnetometer, to develop signal detection algorithms and to evaluate and refine existing MAD search tactics, the Defence Research Establishment Pacific (DREP) has developed an interactive minicomputer-based MAD search simulator-- the MADSIM program. report summarizes the basic MAD search problem, describes the magnetic signature, outlines the simulation procedure of the program, presents several examples of computer output and concludes with a look at the utility and applications of the search simulator.

Background

Highly sensitive magnetometers which are mounted in tailbooms of maritime patrol aircraft can detect the magnetic field generated by the steel hull of a submarine. Although MAD sensors are relatively short range (much less than a nautical mile), being based on an entirely different physical phenomenon they complement acoustic detection systems. Tactically magnetic detection has been used for small area searches, confirmation of acoustic detection, tracking of submarines and, in wartime situations, for the launching of antisubmarine torpedoes.

A search scenario applicable to maritime operations is the

use of patrol aircraft to conduct a widespread area search for submarines. Acoustic sensors are generally used to survey the area and to isolate possible target positions. Around these localized areas the aircraft uses MAD to reestablish and confirm the contact, to track the submarine and to position itself for optimum weapon launch. The problem is thus a localized search about a datum having a high probability of containing a target. The initial area of the datum is variable, depending on the reliability of the acoustic contact, the submarine's speed and evasiveness and the time late to datum (i.e., the delay between the contact and arrival at datum). As the MAD search progresses the datum area continues to increase as the submarine proceeds away from the initial datum; however, the other aircraft sensors may provide some auxilliary information (such as bearing or speed of the target) which reduces the search area.

In addition to datum search, MAD is used to relocate the submarine after the first contact and thereby track the target. The nature of the magnetic signature and the poor maneuverability of the aircraft at short range precludes continuous tracking of the target; depending on the elusiveness of the submarine, the probability of losing MAD contact may be high.

The search tactics of MAD should be designed to sweep the datum area, while maximizing the detection probability over the expected range of submarine speed and maneuver capabilities. After the initial magnetic detection the tactics should either produce maneuvers which have a high probability of retaining contact and tracking the target, or which position the aircraft at the minimum range and optimum geometry for a torpedo launch. Previously tactics for MAD were developed by paper studies using a definite range law (or"cookie cutter") detection assumption for the MAD sensor. By assuming that the MAD equipped aircraft cleanly sweeps uniform cross sections of water and that the submarine moves along a predictable track, the detection probability of a given maneuver is equal to the probability that the volume of water swept by the aircraft contains the submarine. Unfortunately the peculiarities of the MAD signal puts the cookie cutter assumption in doubt and target evasiveness is not easily considered. Tactics are also evaluted using anti-submarine exercises, which because of their expense and relative infrequency, provide little tactical MAD experience per dollar. In addition the number of uncontrollable variables (sensor malfunction, operator inexperience, spurious noise, etc.) and the low number of encounters for statistical comparison may hinder the proper assessment of the tactics.

To overcome the shortcomings of these methods, a third complementary method of MAD tactics evaluation was developed at DREP-- the simulation of the MAD search problem on a minicomputer. The computer program, MADSIM, was originally conceived to provide

pseudo-realistic data for other MAD development projects. For example, by simulating the aircraft flight characteristics for a sequence of maneuvers, synthetic MAD signals and background noise together with artificial aircraft motion sensor outputs (pitch, roll, yaw, etc.) were generated as input to various computer-based detection algorithms. Examples of these signals are shown in Figure 1. By making the signal generating program interactive with the operator so that the aircraft flight can be actively controlled, the search and tactical aspects of the MAD problem can be studied. Before describing the MADSIM program the behavior of the MAD signal is briefly examined.

The MAD Signal

A magnetic field is produced in the surrounding media by the steel hull of the submarine due to permanent magnetism and magnetism induced by the earth's magnetic field. The resulting lines of force shown in Figure 2 produce a disturbance to the ambient magnetic field of the earth which can be detected by a magnetometer. The submarine's magnetic field is represented by the vectors along the flight path in Figure 2 which are tangential to the lines of force and proportional in length to the strength of the field. The MAD sensor is a magnetometer which measures the components of those vectors which are parallel to the much larger vector representing the earth's field. The continuous variation of this vector component along the sensor path is the detected MAD signal. It is evident that a magnetic anomaly, represented by the lines of force, is carried around by the submarine and the MAD signature is generated by the relative motion of the aircraft through the magnetic anomaly. A signal anomaly surface is a three dimensional graph of the magnetic anomaly variations which is detectable at the aircraft altitude. The example shown in Figure 3 gives a visual picture of the submarine anomaly awaiting detection; the amplitude changes along the search generally have only one of two maxima or minima, the feature of the surface varying with the relative orientation of submarine and earth magnetic field vectors. Several properties of the MAD signal are noteworthy:

1) At a constant aircraft altitude the signal amplitude hence detectability decreases as the inverse Nth power of range at closest point of approach (CPA) where N is greater than 3 and depends on the interfering noise and signal processing.

2) Most signal anomaly surfaces possess an anisotropy whereby for certain search directions a null signature is generated at short range (an example of a null path is shown in Figure 3). Consequently a MAD

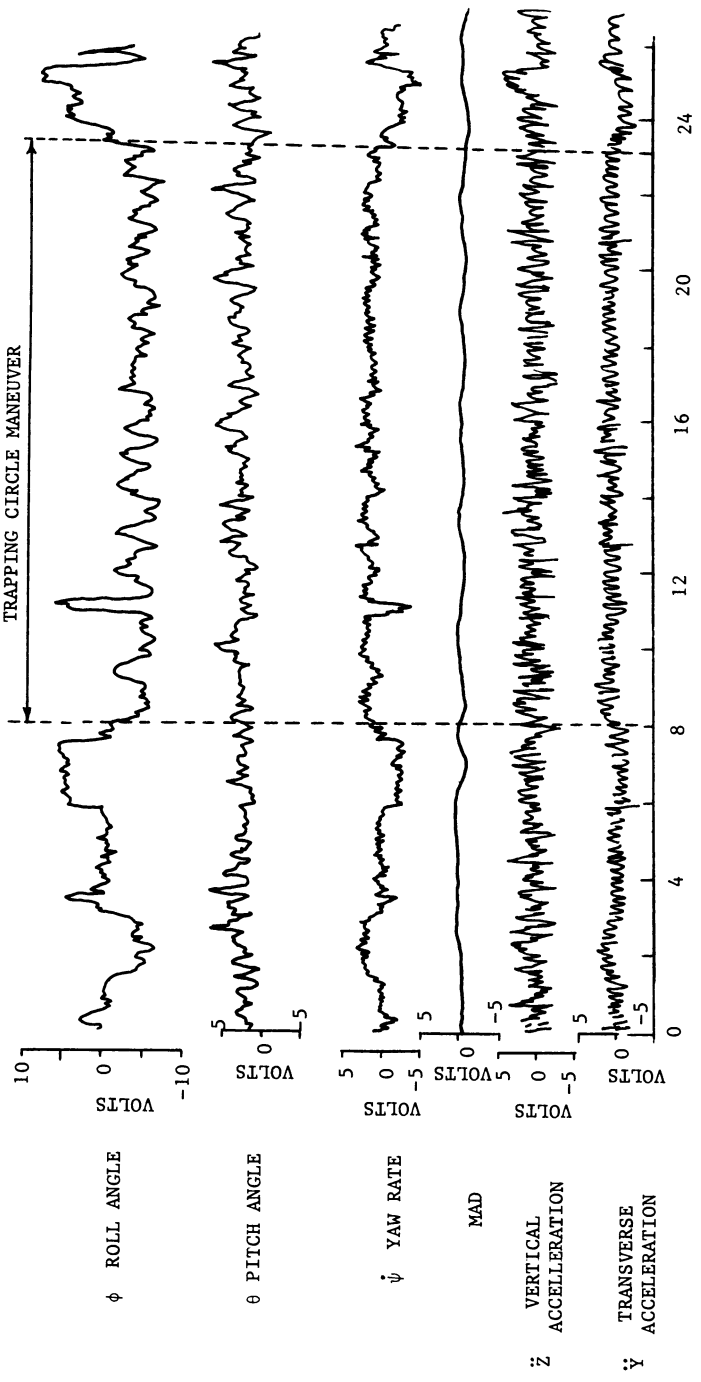


Figure 1. An example of sensor outputs a during a MAD exercise; these signals can be simulated by MADSIM for inputs to signal detection algorithms under development.

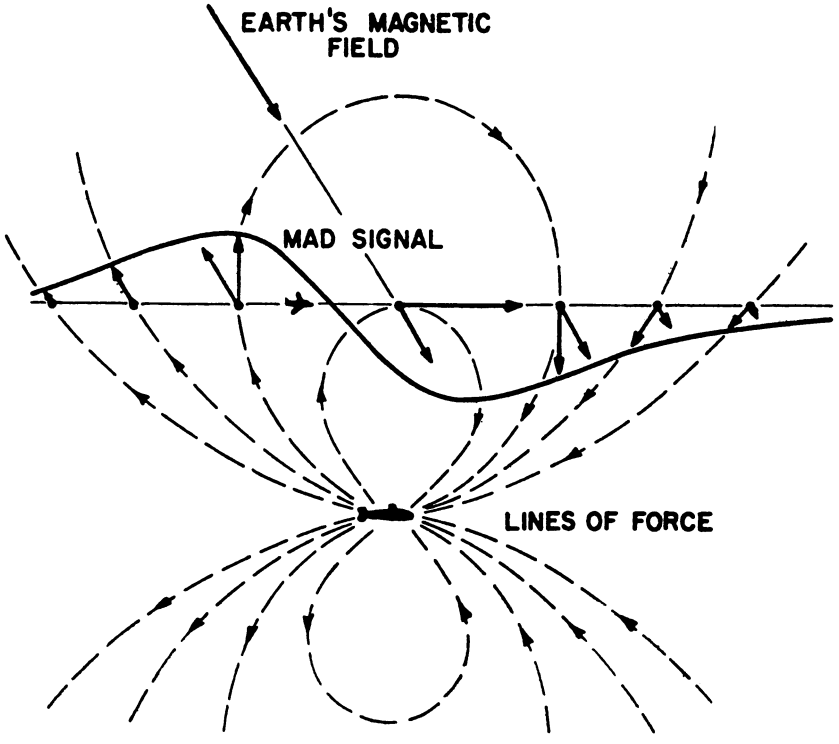


Figure 2. The submarine generates a magnetic anomaly, represented by the lines of force. The magnetometer carried by the aircraft measures the component of the magnetic field vectors which are parallel to the earth's magnetic field direction, and this variation in magnetic amplitude as the aircraft passes the submarine is the MAD signal.

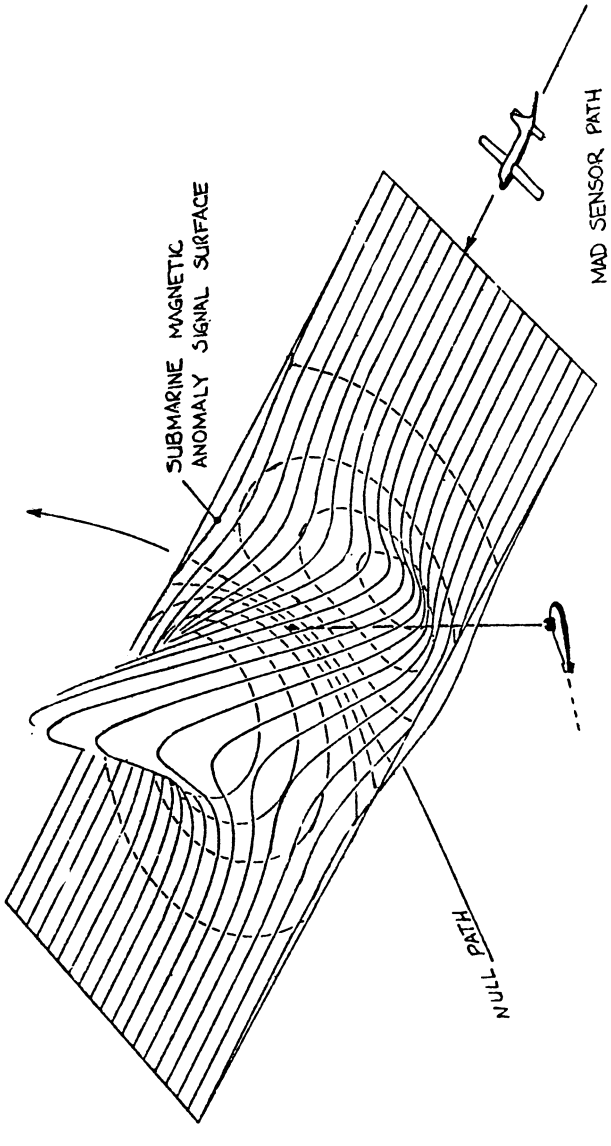


Figure 3. The features of the submarine anomaly are visually displayed in the magnetic anomaly signal surface which is a graph of the magnetic variation plotted over an area above the target at the search altitude. Contours of equal amplitude are shown for clarity. The MAD signal is the variation of the surface along the aircraft track. The presence of a null signal track following the zero contour is indicated.

sweep has a finite probability of a missed detection even at short range.

3) The complete MAD signal is not received until after the CPA has gone by. This delay plus the necessity of moving relative to the target to produce a signal prevents continuous tracking after initial detection. The large turning radius of the aircraft also hampers redetection. The MAD search problem is somewhat analogous to searching for and tracking a snake in tall grass while riding a bicycle and looking into a rear view mirror.

4) Little or no auxilliary information (such as target speed or bearing) is produced by the MAD system.

The MADSIM Program

MADSIM is a Fortran program written for a Texas Instruments 980 series computer equipped with a CRT screen and keyboard. The simulation is controlled by a sequence of subroutines which determine the motion of the aircraft and submarine and is conducted at equally spaced time intervals at which their positions are updated. The flow chart for a simulation interval is shown in Figure 4. The motion of the aircraft and its position are determined by basic aerodynamic equations relating pitch, roll, yaw, speed, altitude, etc. A "PILOT" subroutine controls the evolution of the aircraft altitude to obtain the desired maneuver (such as turn, climb, circle). In turn the aircraft flight sequence is controlled either manually by the operator keyboard instructions or automatically through the use of "TACTIC" subroutines which react to a tactical situation and respond by appropriate commands to "PILOT." The submarine track is controlled in a similar, but simpler fashion. As in a patrol aircraft the aircraft position and the MAD signal are displayed on a CRT (Fig.5) for operator inspection. Depending on the mode, the MAD signal is visually detected by the operator who responds with an appropriate tactic or aircraft maneuver, or the signal is processed by a computer detection algorithm with detections passed onto the tactical subroutines which control the aircraft track. The submarine can either follow a straight and uniform track of random heading and position unknown to the operator, or if put into an evasive mode will make pseudo random course and speed alterations when it passes within the acoustic noise cone of the aircraft. A simulation periodic in time was chosen over an event orientated one to facilitate tactical control; a tactics sequence can be interrupted at any time and modified without difficulty. Also the CRT display of the MAD signal can be updated at equal time intervals.

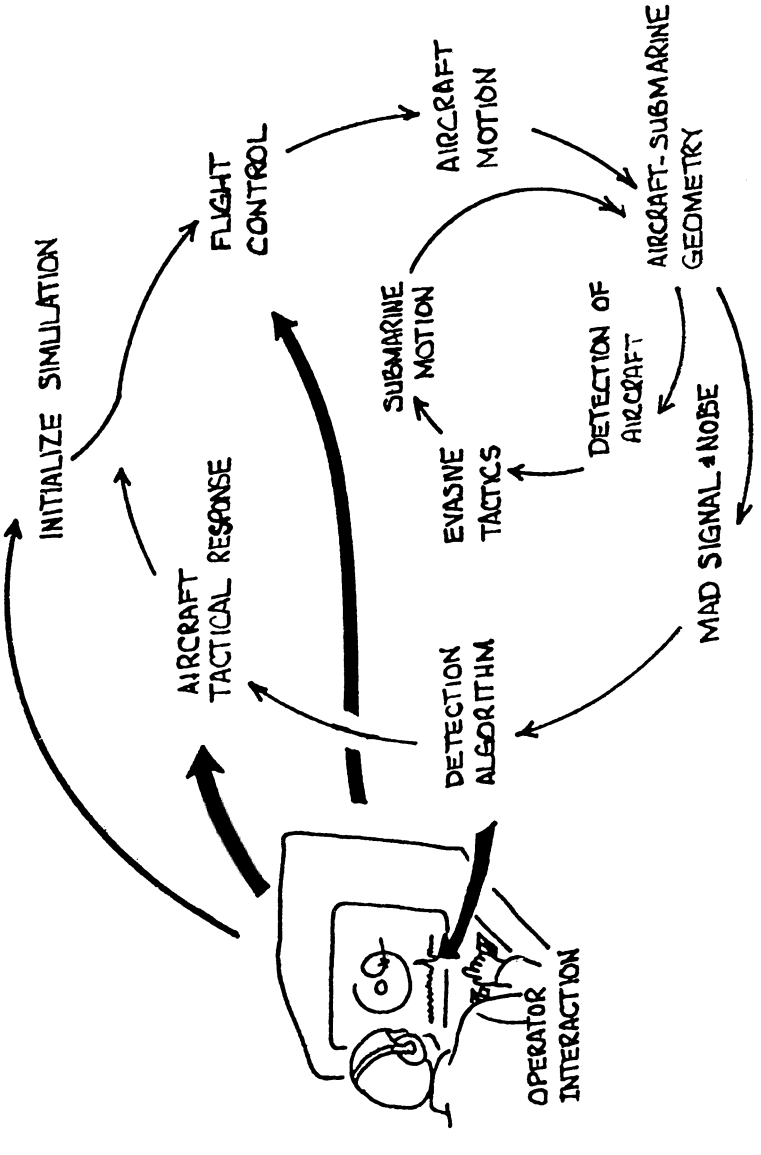


Figure 4. A flow chart of MADSIM for one simulation step showing the basic function of various subroutines and the optional areas of operator interaction.

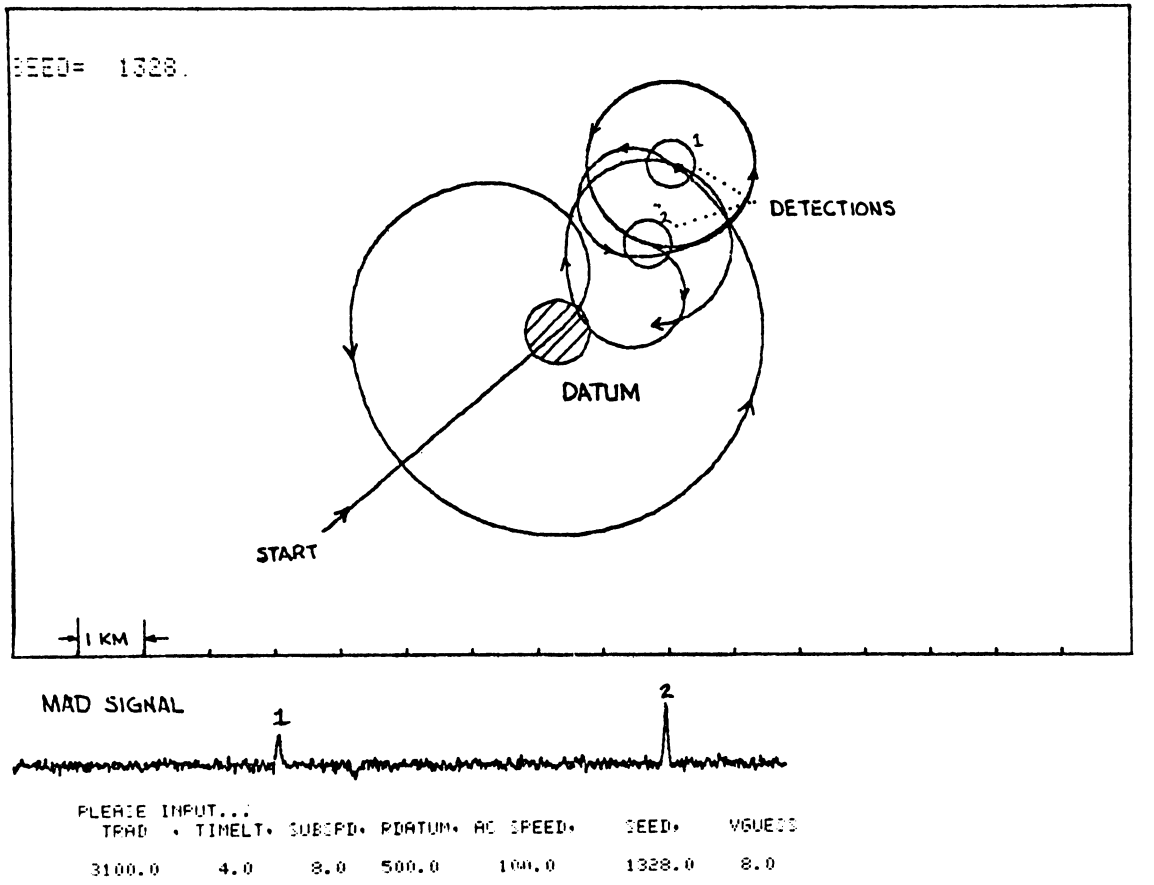


Figure 5. A reproduction of the MADSIM operator's console, showing the aircraft track around the datum and the MAD signal output below. The input parameters for the simulation are listed on the keyboard.

The program operator has a large number of options available at his command through the keyboard. Prior to each simulation the operator chooses the parameters for the initial MAD search; time late to datum, datum error, estimated submarine speed, and noise level. During the search the keyboard is a panel of knobs for adjusting aircraft speed, altitude, wind speed, etc., switches for changing from real to fast time, turning on the auto pilot or the auto detect and buttons for marking a MAD signal or dropping a torpedo.

MADSIM is generally used in three ways. When operator controlled, a simulation run in real time gives the operator a realistic time in which to assess the tactical situation and respond with further search maneuvers. To evaluate a preset tactical sequence the program may be run in fast-time with both aircraft and submarine tracks on display. By observing the search as it unfolds, the operator can quickly see the effectiveness of the tactic. The program has a statistics gathering option in which the starting conditions and tactics for the search are present and several hundred simulations are run for a random set of submarine tracks.

Examples

Pictures of CRT screens in Figures 5, 6 and 7 are examples of MAD search and relocation tactics. Along the bottom of Figure 5 are listed the initial parameters of the search--the time late to datum (TIMELT) = 4 minutes, the trapping circle radius (TRAD) = 3100 metres, the submarine speed (SUBSPD) = 8 metres/second, the aircraft speed (ACSPEED) = 100 metres/second, the datum radius (RDATUM) = 500 metres, a random number (SEED) = 1328 and the operator's estimate of submarine speed (VGUESS). The initial search tactic was to fly over the datum from the aircraft's initial position and into a 3100 metre radius trapping circle in order to contain the submarine traveling in an evasive mode away from the datum. A MAD signal is expected if the submarine exits the circle near the aircraft, and in this example the tactic was successful producing detection #1. The aircraft then initiated a relocation tactic and flew a minimum radius trapping circle about the first detection. This tactic was also successful in producing detection #2, and the maneuver was repeated. The evasive track of the target is now shown.

Figure 6 shows a spiral maneuver in which the radius of the trapping circle is increased to coincide with the estimated exit speed of the target from the datum. Superficial examination

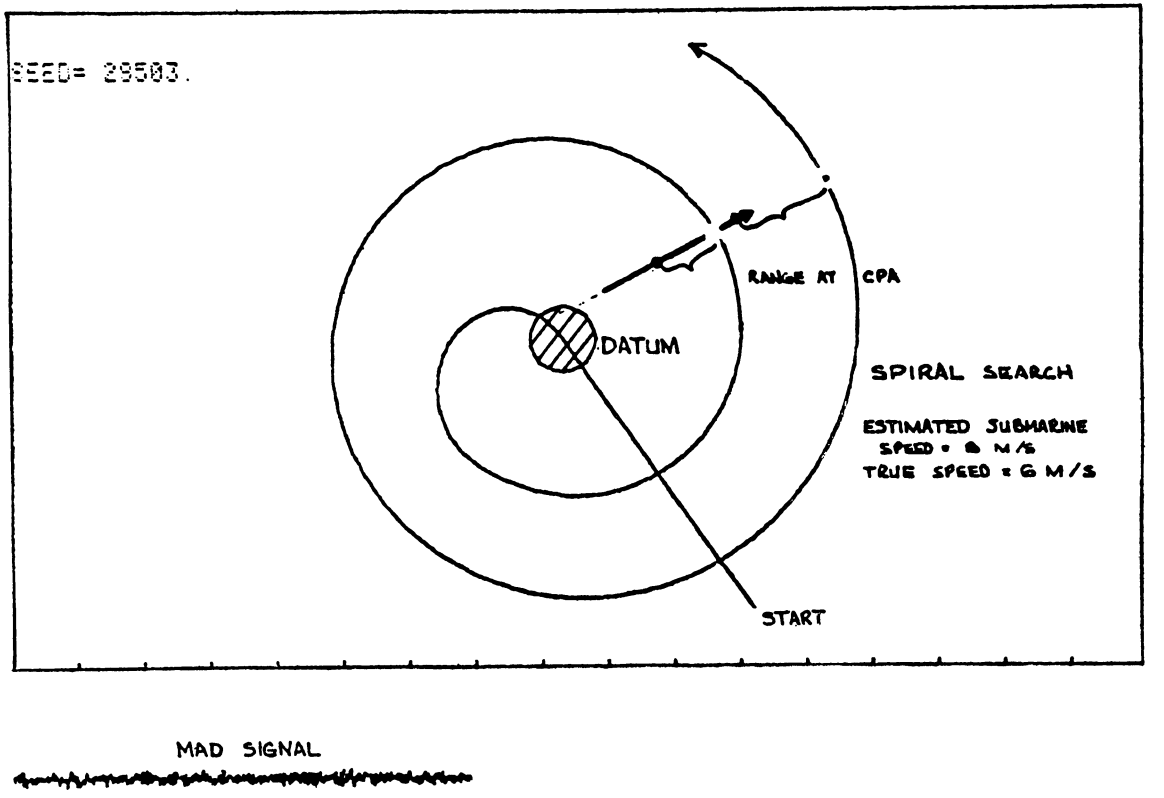


Figure 6. Example of a spiral search tactic around the datum. The aircraft is searching for the target 4 minutes after datum is established and the spiral is based on an estimeted target speed of 8 m/s. The submarine track, for a 6 m/s speed is shown, indicating the failure of the search.

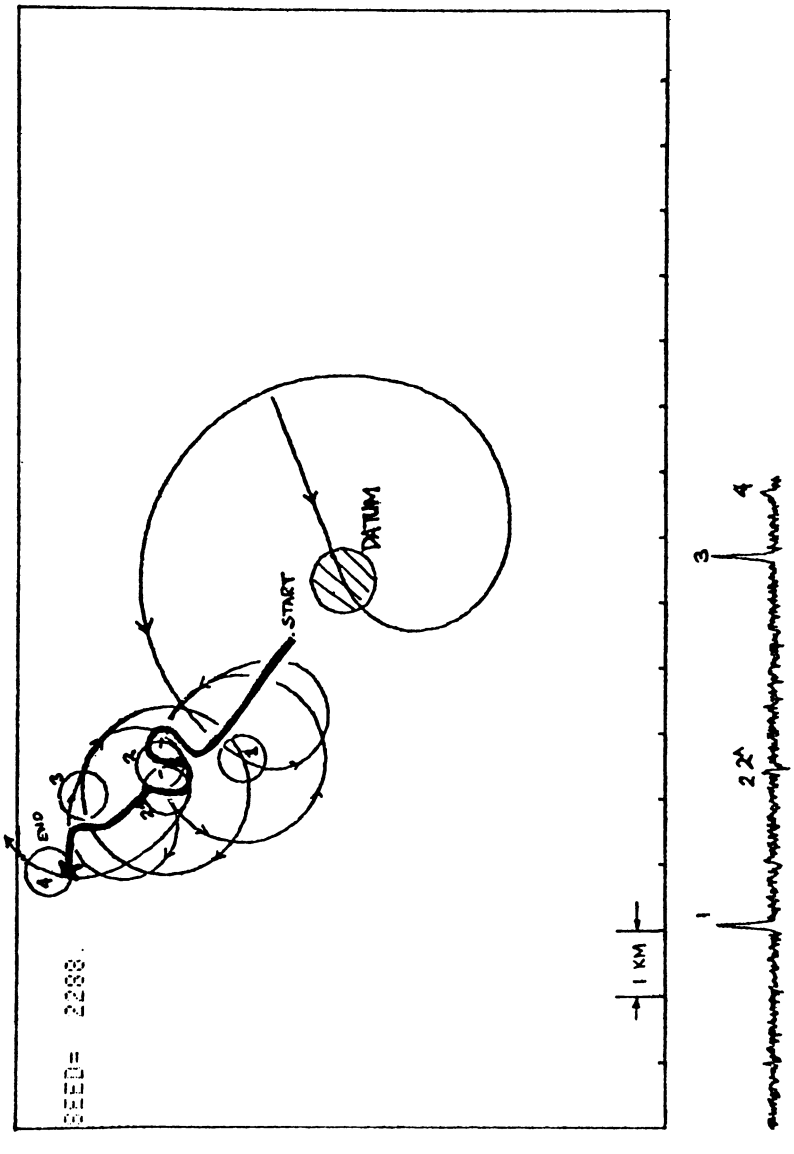


Figure 7. A longer running simulation showing an initial datum search and a sequence of relocation maneuvers. The submarine track clearly shows its evasive track.

suggests this may be a promising search technique but as the search time increases, the probability of a detection rapidly diminishes because the revisit time over such a large perimeter is so infrequent. Figure 7 is an example of a longer running and hence more involved tracking sequence and shows the evasive submarine track.

Utility of MADSIM

MADSIM has proven useful in several applications. To improve MAD search tactics the scenario can be clearly defined and a realistic sensor detection function can be used. Parameters affecting the search can be varied in a controlled manner and statistical comparisons can be generated by repeating a search tactic for a set of simulations having a random distribution of target parameters. For example the distribution of submarine positions within the datum area can be controlled. In this manner the limitations of MAD search can readily be assessed. By changing noise levels, signal strengths, or operating areas the tactics can be assessed to see how they degrade from their performance under ideal conditions. The tactics investigated to date show that the probability of detection of a search tactic varies greatly with the initial conditions and the submarine behavior, suggesting that an all encompassing simple tactic does not exist. However, guidelines for successful search can be set and the MAD operator, by practicing search with the simulator, can learn the MAD game and hence improve his search performance. Complementing this application in tactics development, MADSIM could be used as a stand alone training aid or as part of a tactical trainer.

In addition to improving the search tactics DREP is interested in the effects that changing MAD performance has on tactical effectiveness. Does, for example, the inclusion of a range estimator significantly improve the probability of detection and redetection? Future improvements to MAD such as the utilization of more sensitive sensors can be tactically assessed prior to the development of the new sensors themselves.

ON THE OPTIMIZATION OF A SEARCH PATTERN FOR MARITIME SURVEILLANCE

M. Reiss

Industrieanlagen-Betriebsgesellschaft mbH

The Problem

Consider a rectangular sea area which is crossed by numerous ships. The positions of the ships are random, and they sail on headings parallel to the longer side of the rectangle. The ships are to be detected and identified by a maritime patrol aircraft before they cross a given line at right angles to their headings. This line, i.e., the width of the rectangle, may have a length of 50 to 250 n.m. The problem arises typically in military surveillance tasks but might also occur in the surveillance of fishery protective zones.

The ships will be detected by radar; identification will be possibly visual or with a FLIR (Forward Looking Infra-Red) only. These sensors have a much shorter useful range than the detection equipment. It is obvious, therefore, that it is neither useful nor possible to plan the search pattern with the intention of passing every ship within the identification range. (The identification range is 0.2 to 10 n.m.) The objective must be to develop a search pattern that results in the greatest possible number of ships being within the radar detection range and to divert from the preplanned search pattern only in order to identify a detected ship. The resulting flight path, therefore, consists of a preplanned basic pattern (search pattern) and superimposed stochastic deviations. The deviations are of a stochastic nature for two reasons:

1. The positions of the ships are at any moment random.
2. The radar detection distances can only be described statistically.

Any deviation influences, of course, the detection of other ships. If, for example, a ship in a position abeam to the a/c at the maximum radar range is detected and the a/c heads for that target, then any other ships on the other side of the search pattern may not be detected.

The problem is to find, for the given basic search pattern (Figure 1), values for its parameters (LSP, DMM, SPA), which maximize for given assumptions the searched area per mission (or search time) or minimize the flight time for a given area.

The assumptions are the width of the search area B (WSA), the speed of the aircraft (SAS), the speed of the ships (SAT), the sensor capabilities, the desired probability of detection, and the desired probability of identification (see Figure 2).

Solution

The problem was handled by Monte Carlo simulation. I think it is not necessary to say much here about the technique of this well-known method. The simulation is based on the event-chain method. For a given leg, the coordinates of all possible detections are calculated and stored. Then beginning with the first possible detection, it is determined whether or not this detection causes a change of heading. If a change is necessary, then the new possible detection on this leg is calculated, and so on.

Some important assumptions had to be made about the way the pilot deals with the deviations he has to fly to identify the targets.

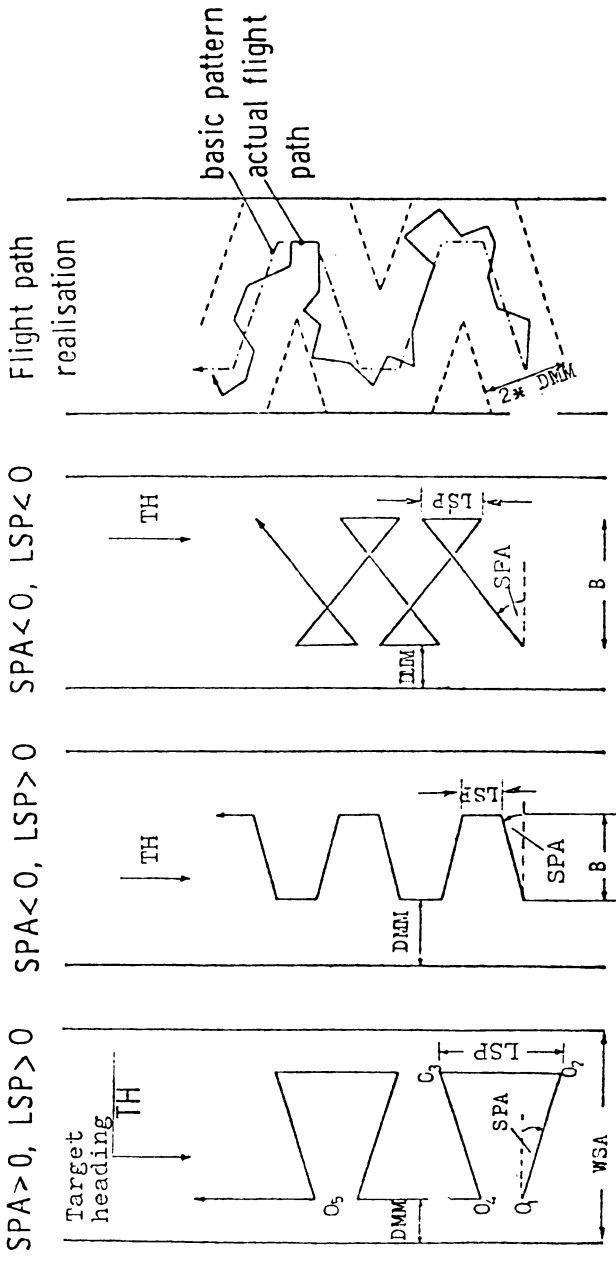
These assumptions are:

1. The a/c heads for the targets, which are less than a distance DMM away from the search pattern.

2. When the a/c is off the search pattern and there is no further ship to identify, then the a/c heads back to the pattern. The heading is chosen so as to make the distance for the return journey constant.

3. If more than one target is detected at the same time, then the a/c heads for the nearest one.

4. Rule 3 is omitted, if the a/c will pass the nearest target within the identification range (VIS) with the present heading or if the target is farther off than EMAX in the direction of the search pattern.



Key: SPA = Angle of cross-leg measured clockwise

LSP = Length of forward-leg (sign convention forward is positive)

Figure 1. Basic search pattern.

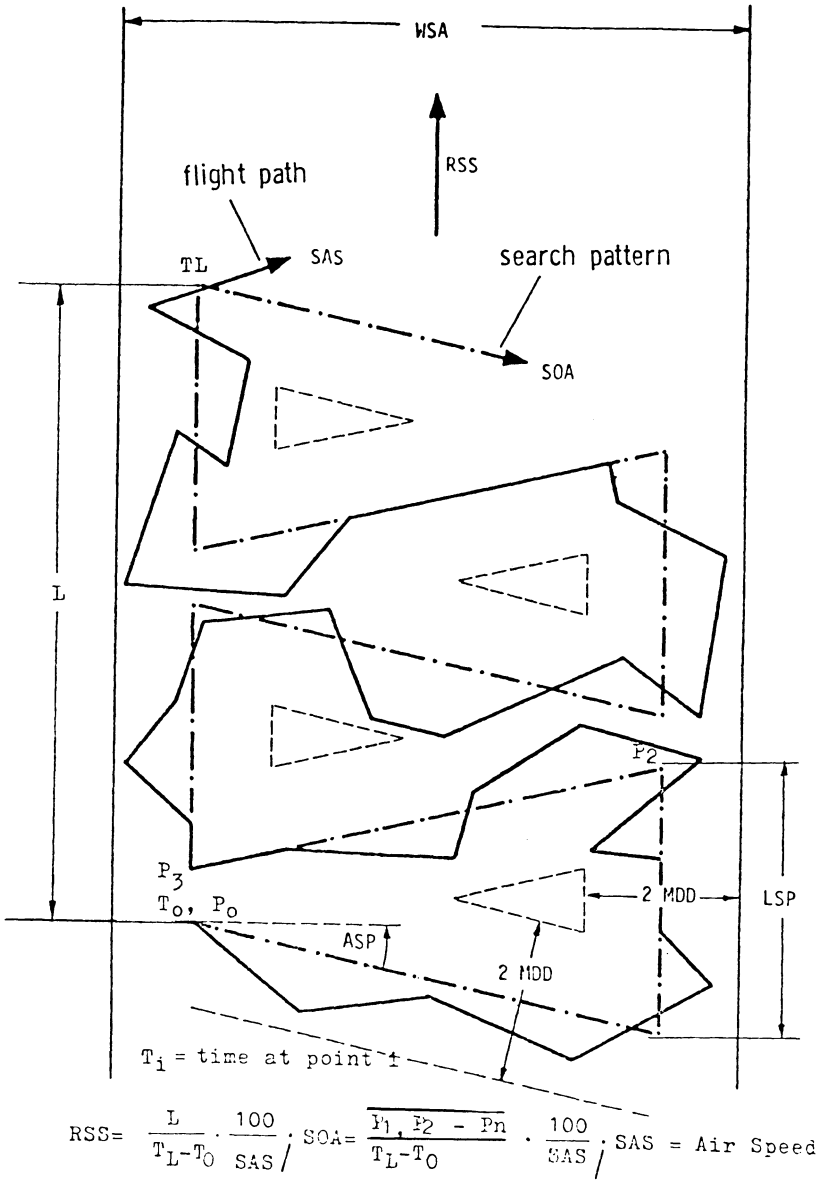


Figure 2. Definitions.

5. If the a/c flies around a target (Figure 3, target 10) that has been detected but was not identified because Rule 3 and 4 prevented it, then the a/c heads back towards that target if there is no nearer target and the distance is less than SCHR. (A good value for SCHR was found to be 1.5 times the maximum detection range.)

These rules are the result discussions with pilots who have flown missions of this type. I think the rules are a compromise between a high speed of advance (of the search), a high probability of identification, and simplicity of application.

These rules do not minimize the total distance for the identification task, but the rules to minimize this distance would be so complicated that it would be unrealistic to assume that they would be applied in practice. Some remarks should be made about the modeling of the radar detection process. The detection is the result of a chain of stochastic events. The model of this process consists of a model of the radar set (resulting in the single glimpse detection probability) and a model of the radar operator's capability to recognize a target on the scope. For both models, the data base is not entirely satisfactory. (The common practice of calling the detection range that distance where the single glimpse detection probability is, under ideal conditions, equal to 0.5 is, in my opinion, very dangerous and misleading.)

The distance for identification (VIS) is assumed to be deterministic and is treated as a parameter.

The surface density of the ships (i.e., the number of ships per unit area) is treated as a parameter also. The simulated positions of the ships are based on a uniform distribution.

Results

For the simple case where the target is not to be identified (i.e., there are not deviations from the search pattern) and a deterministic detection range RS is given, the optimum angle SPA can be obtained from the equation:

$$SPA = \arcsin[TSA/SAS], \qquad (LSP = 2SAS \times SR/\{SAS+TSA\}),$$

where TSA = target speed, SAS = aircraft speed, and SR = sensor range.

The results of the simulations indicate that this simple

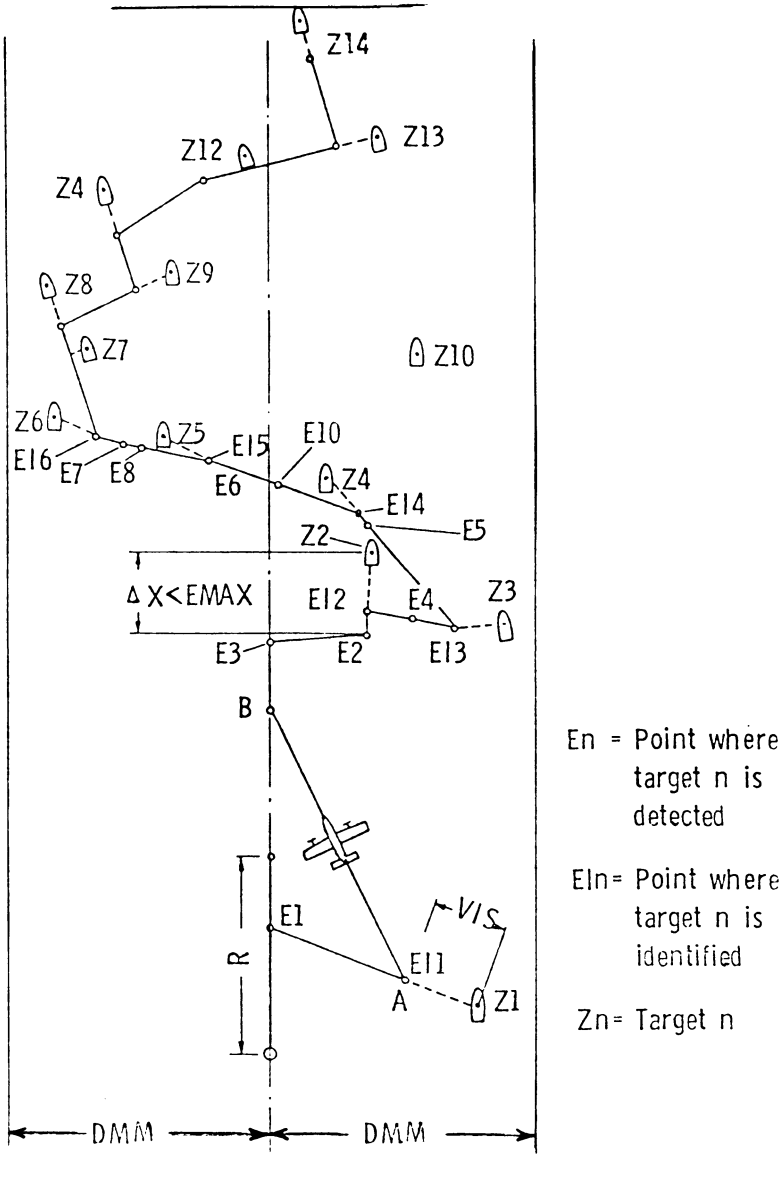


Figure 3. Deviations from the search pattern for the identification of targets.

relation also holds for the complex case if the a/c speed SAS is replaced by the speed SOA. This is the speed of the a/c measured with respect to the search pattern (Figure 2).

The speed SOA is one result of the simulation and is a function of the target density DOT, the maximum deviation DMM, and the sensor capabilities. Figure 4 shows the normalized speed SOA against the maximum deviation DMM. (For the fixed parameters, see Figure 4.) With the above equation and the speed SOA (from the simulation), a graph of the optimum angle SPA against the maximum deviation DMM can be drawn. An example is given in Figure 5. The next figure shows the probability of identification against the search pattern length SLP with the maximum deviation DMM as a parameter (Figure 6).

The figures show that an indefinite number of combinations of SLP and DMM result in the same probability of identification. Figure 7 shows this fact even more clearly and provides an answer to the question: "Which combination of LSP and MDD gives a probability of detection of 0.95?" It does not answer: "Which of these combinations gives the best speed RSS, i.e., the shortest search time?" This information can be obtained from Figure 8, which shows RSS against DMM with the probability of identification as a parameter. (The corresponding value of LSP can be obtained from Figure 6.) The best value for DMM is where RSS is a maximum.

From the graphs we have seen so far, it is possible to choose the optimal search pattern values. However, it is not very convenient to use this kind of presentation because there is a lot of cross referencing necessary. We therefore tried to design graphs which are easier to use. An example is given in Figure 9.

The data required for these diagrams are:

1. The target density DOT(must be estimated) .
2. Radar cross section of the ships RCS.
3. Flight altitude ALT.
4. The ratio SPR of aircraft to target speed.
5. Identification range VIS.
6. Width of search area WSA.
7. Required probability of identification POI.

The appropriate set of graphs can be selected for the first three of the above inputs, and then the optimal search pattern

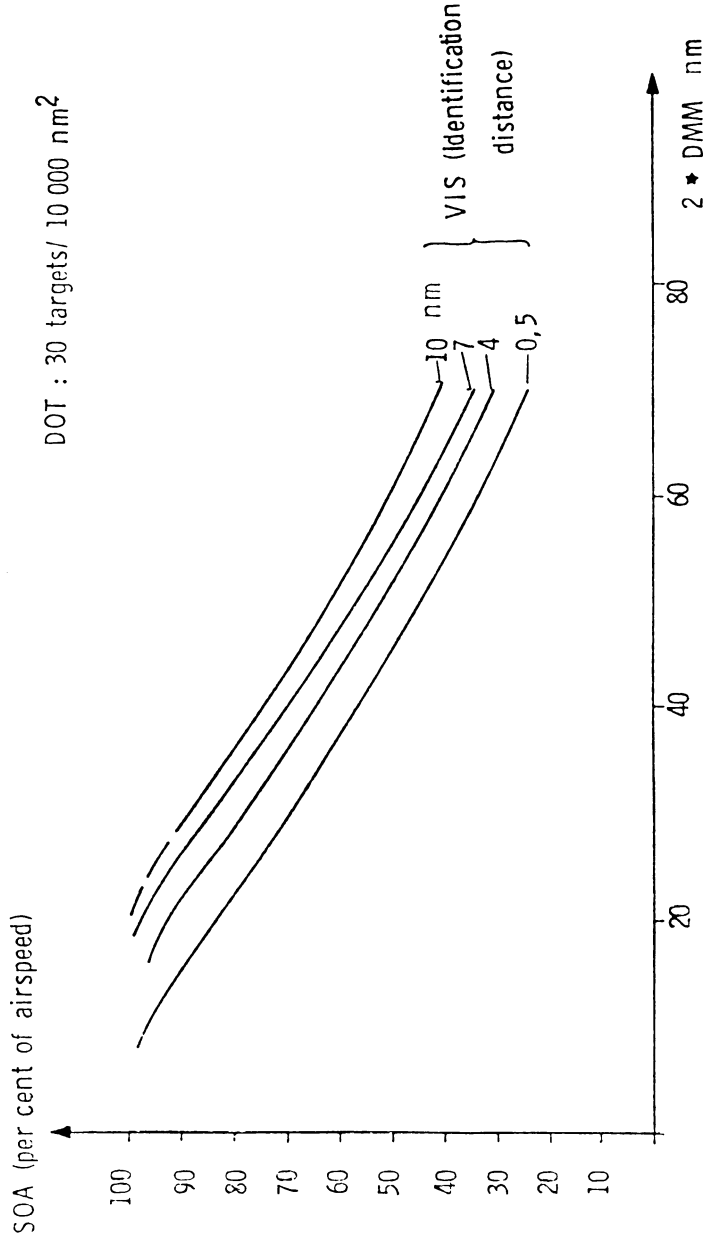


Figure 4. Radar reconaissance, RCS = 500^2, ALT = 1000 ft.

DOT = 15 ships/10 000 nm^2
a/c = type 1
VIS = 10 nm
ALT = 1000 ft
SAS = 180 kn
TAS = 20 kn

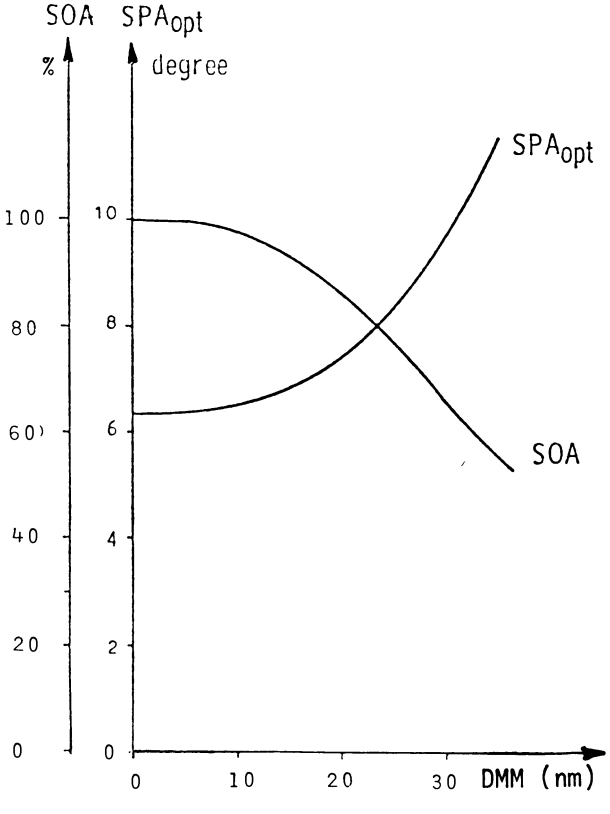
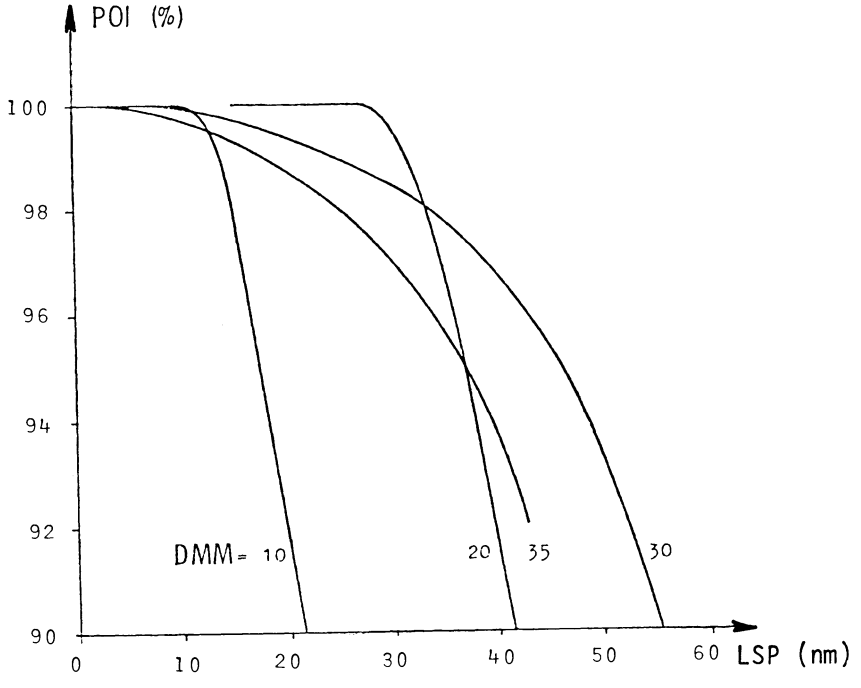


Figure 5. SOA and SPA_{opt} vs. DMM

DOT = 15 targets/10 000 nm^2
WSA = 120 nm
VIS = 10 nm
ALT = 1000 ft
SAS = 180 kn
TAS = 20 kn
SPA = SPA_{opt}
a/c : Type 1



POI: Probability of identification

Figure 6. POI versus LSP.

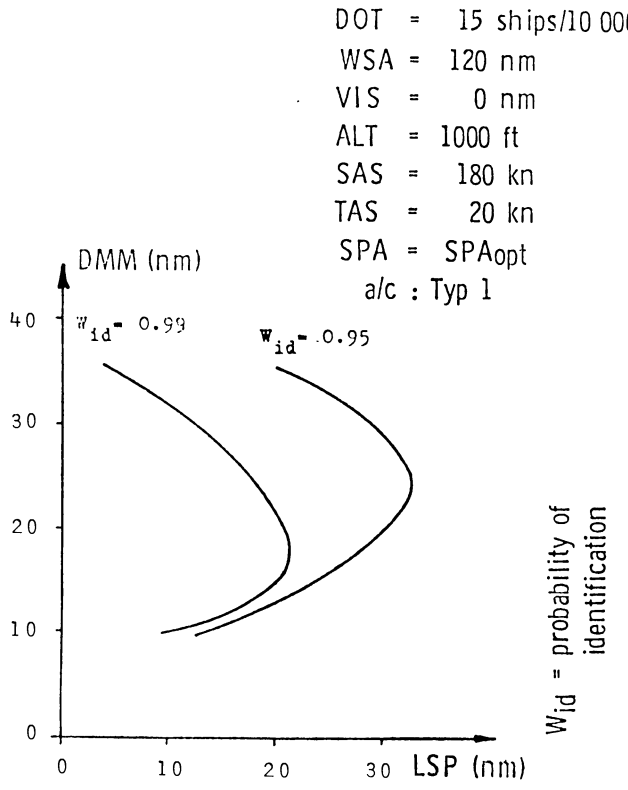


Figure 7. DMM vs. LSP.

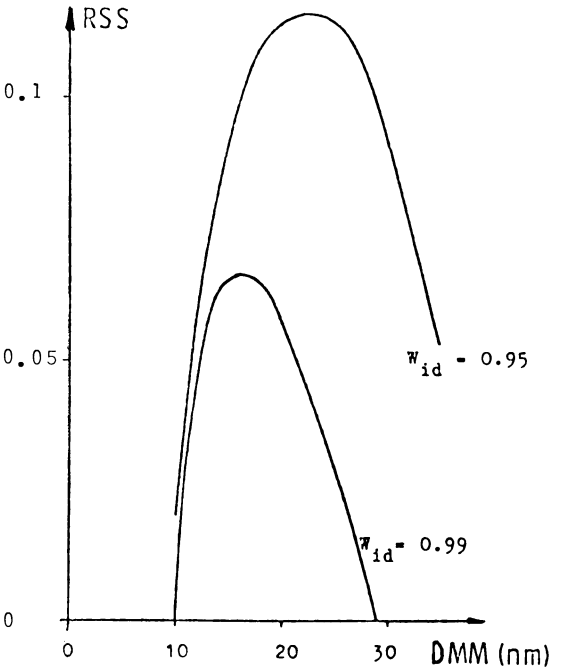
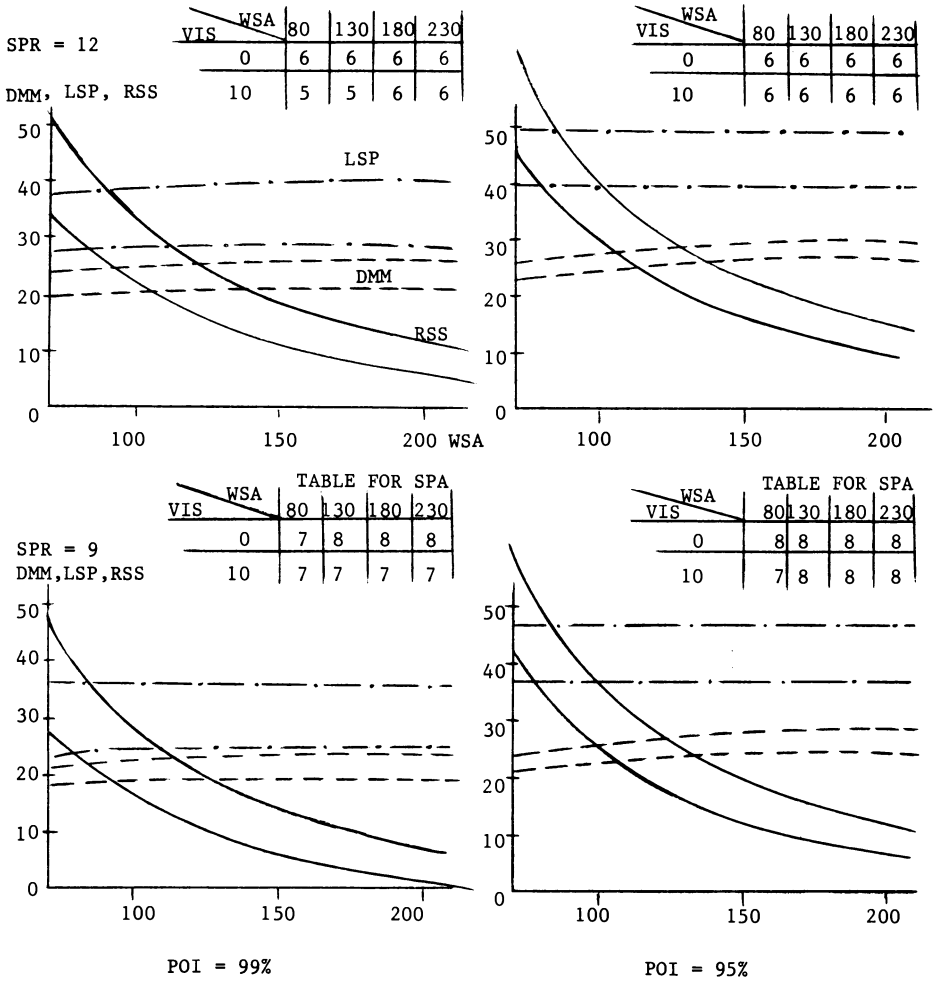


Figure 8. RSS vs. DMM.



Note: DOT = 7.5 ships/10^4 NM^2, RCS = 500 M^2, ALT = 300 ft

Figure 9. Combined Presentation of Results

parameters can be read from the graphs. The angle SPA is found in the table above the graphs. The maximum deviation DMM and the pattern dimension LSP can be read from the graphs. The graphs also contain a curve for the speed RSS (units: percent of the airspeed, see Figure 2). With a search area of length L, the minimum time to search this area is then equal to 100L/{RSSxSAS}.

Conclusion

A satisfactory method has been developed for calculating optimal search pattern measurements. The dependence of the probability of identification on important parameters, e.g., target density, speeds of searcher and target, and width of search area, has been demonstrated. The simulation does not have the same elegance as an analytical approach. However, the method has proved to be a relatively inexpensive tool to use, and it gives the user the freedom to investigate the effect of different search strategies easily.

EXPLORATION WORKING GROUP REPORT

Introduction

The exploration working group received a number of presentations which are summarized later in this report and in some cases appear in a fuller form in other parts of the proceedings. It concentrated on a brainstorming session which was aimed at identifying the potential for the use of search methods in exploration problems. Some details of potential and actual uses are included in papers by Wagner, Field, Fergusson, Lenoble, and Friedrich. A parallel was also seen to the medical work described by Kolesar. A parallel was also seen to the medical work described by Kolesar. This brief report attempts merely to indicate the ideas which seemed to be most important in exploration.

Summary of Papers

The papers, "Search Theory Applied to Exploration," by E.A. Field, "Exploration for Polymetallic Nodules in the Pacific," by J.P. Lenoble, and "Exploration for Coal in the U.K.," by J.G. Fergusson, appear in full elsewhere, but "Theory of Prospecting for Oil," by J. Cozzolino, does not. This latter paper described first the nature of oil exploration and then some features of its exploitation.

Dr. Cozzolino observed that oil exploration is a long-range problem which involves very little information. Experience in geology, exploration history, and some economic information is used. The economic factors are the cost of drilling and the price of oil.

Exploration is normally carried out in large areas such as the Baltimore Canyon. These areas are called basins and contain individual (possibly dry) fields called prospects. An exploratory

well is one drilled away from other producing wells. A company will wish to decide whether it should start or continue to pursue exploration in a particular basin. The decision concerning which specific prospects to consider is best taken using geological knowledge.

The evaluation of a basin's prospects depends on defining the basin and examining its exploration history. A basin is defined so that is encompasses an area of similar geological phenomena and similar exploration costs. A list of wells, which have previously been drilled in a basin, is compiled and their outcomes (strike or dry) and the size of any oil deposit found are recorded. The success rate and average size of a deposit are estimated as a function of the number of exploratory wells drilled in the basin using a maximum likelihood estimator of both the remaining availability of oil (number and size of fields remaining undiscovered) and the efficiency of past exploration. The success rate and average size generally decrease as more wells are drilled because the basin's reserves are depleted.

Finally, the success rate and average size curves are projected forward to predict the value of drilling additional exploratory wells. The goal of the analysis is to evaluate all basins so that exploration effort can be systematically and sequentially allocated among the basins of the world.

Search Theory in Exploration

From a list of 50 items identified by the group as having possible relevance to the exploration activity, it selected the determination of the prior distribution of targets as the most important feature. Certainly, no geologist would disagree with the choice since their main preoccupation is with this problem.

The multistage approach of all exploration studies is designed to contain risk and cost. Thus, with reference to Appendix A on the elements of search, the first stage in exploration is aimed, not at finding oil or minerals directly, but through relatively low cost geophysical surveys at discovering geological indicators that are known to correlate closely with the presence of oil or mineral. The use of detection systems with sweep width characteristics and mobility make them candidates for treatment by the classical search theory methods. These surveys provide the first estimates of target distributions, if not in detail, at least to the extent of indicating favorable and unfavorable areas.

Exploration now proceeds with more detailed surveys within the favorable prospects. The survey maps are extended to include

more definitive data on rock structures and composition, and, through seismic, gravitational, and magnetic measurements, three dimensional contouring begins to identify the rock fault structures that are the targets during this stage. While the elements of search are still present, the multidimensional nature of the detection process is likely to require considerable extensions to classical search methods.

When survey method results have raised confidence in there being a primary target present --oil or mineral-- the decision to drill is taken. Whether the result is positive or negative, the additional geological data obtained from the core and subsequent probing (well logging) will assist in modifying the prior expectations. Positive results will provide data on target quality and size. There are a number of elements present in this problem that are not present in the classical search problem. Thus, the search is not complete when the target is found, and additional drilling is required to establish its size and value. Considerable extensions are required to the classical search theory.

A positive approach to the exploration problem would be to evaluate current methodology in each of the areas identified in Appendix B, "Solution to Search Problem." Just what role, if any, classical search theory would play in any improved methodology remains to be seen.

APPENDIX A

Elements of Search Problem

Physical description, performance data, and cost structure of

targets,

detectors, and

transportation system.

Prior knowledge of

region of search,

distribution of targets (number and size), and

constraints.

Stated objective.

APPENDIX B

Solution to Search Problem

Procedures are required for

determining search budget,

allocating budget between regions and stages of search,

scheduling movement of detectors within search region, and

revising budget and schedules in light of new data.

INTRODUCTORY COMMENTS ON EXPLORATION AND SEARCH THEORY

E. A. Field

Manager Systems Coordination
Esso Petroleum Co.

Introduction

Early search theory, as formulated by Professor B. Koopman, was developed and used primarily for finding lost objects, although in the past two decades the proponents of search theory have made considerable headway in extending its useful areas of application. The purpose of this note is to discuss a number of issues that bear upon the worth of search theory as a means of improving the efficiency of exploration.

Before discussing these issues, I wish to state that all the comments and discussion points made in this note represent my own personal views and are not necessarily those of my employers, the ESSO Petroleum Co. Ltd. However, I am grateful to them for allowing me to take part in this ARI and wish to record the importance they place on the exchange of ideas between educational institutes and industry.

I would also like to record that whilst the literature on the two subjects of search and exploration is vast, only a small selection of it was referred to in preparing these comments.

Elements of Search Problem

We can regard all search problems as systems with the following basic components:

(a) An object that is being looked for called a target.

(b) A location or region in which the target is expected to be present.

(c) A detector that may respond when the target is present.

(d) A transporter for moving the detector over the search region.

(e) A controller for actioning the search plan.

By specifying the practical nature of the components (a) to (e), a wide range of search problems can be modeled. Given a system objective, an efficient plan for transporting the detector over the search region can be determined.

There are two important ways in which the components of the system may modify each other:

(i) As the controller executes the plan, the information produced may be used to modify the prior view of the target's location.

(ii) Information may also be used to change the prior view of the target.

An essential difference between searching for lost objects and exploring for petroleum and minerals is embodied in these two feedback mechanisms.

Classical Search Theory and Extensions

The classical search problem is concerned with "sweeping" or "trawling" a fixed region in order to find a single object. are many variations of the problem (Stone [1975]), including ones in which the prior distribution of the object's position in the search region is modified as the search proceeds using Bayes theorem. Few of the variations deal with the case in which the object (or the searcher's perception of it) changes as the search proceeds. One way of extending the classical theory is to divide

the search problem into several stages. Each stage can be treated as a separate search problem with its own basic components. This multistage model, with the information from one stage used as input to the next stage, provides a likely transition from classical search theory to "exploration and mining theory." Given an overall system objective, an efficient plan for allocating and scheduling search/exploration resources between each stage can be determined.

Contrast of Searching for Lost Objects and Exploration for Oil and Minerals

These comments provide a summary comparison of two activities that may well be at opposite ends of a spectrum of search-like problems. For details of modern petroleum technology, see Hobson [1973].

The classification and cataloging of problems is a prudent first step in understanding how a new problem may be solved. For the purpose of this comparison, we need to consider each basic component in the search system and the system objectives.

(a) Nature of the Target

It is the target that provides the essential differences between search systems. Lost objects must arise in a bewildering variety, but in the significant and practical cases studied by the workers on search theory, the essential variety is on a minor scale compared with that occurring in petroleum exploration. In terms of size alone, an ocean liner in distress may be the maximum-sized object ever searched for. Contrast this with the Brent Oil Field in the North Sea that covers 50 square miles and contains 1,500 million barrels of petroleum.

(b) Region of Search

Given a Mayday signal or a submarine alert, the searcher can quickly delineate a search area and produce a map showing the prior probability of the target's position as a function of time. In contrast, the initiation of exploration is based on a combination of commercial and governmental pressures. Prior information may exist only as interpretation of geological survey data. The choice of an exploration region rests heavily upon the shoulders of the geologist and economic analyst.

(c) & (d)Target Detector and Transporter

The physical characteristics of the target and its environment are the prime determinant of its means of detection. In the classical search theory, a detector device with predefined sweep width is moved over the region of search creating a swept area. The detector is assumed to be imperfect, so that even when the target is in the swept area, it is uncertain of being detected. This uncertainty is called the detection "probability," and the value assigned to it will depend upon the nature of the interaction between detector and target under particular conditions of search.

Exploration and mining take place in well-defined stages in order to contain risks and cost (see Harbaugh, et.al. [1977]). Whilst all stages are ultimately directed at finding commercial quantities of petroleum or minerals, the earlier stages can only be immediately directed at finding secondary targets that are known from experience to be found in some degrees of association with the primary target. As exploration proceeds through its three basic stages of broad geological survey, detailed earth probing, and prospect drilling (see Allais [1957]), so the nature of the search activity changes from one that looks rather similar to a classical search to one that is quite dissimilar.

(e)Control and Actioning of Search Plan

Whilst the level of control and analytical support provided to classical search has advanced in step with computers and telecommunications, at no time has it reached the degree of sophistication accepted as common place by geologists and economic analysts employed in exploration. Perhaps because of the enormous difference in costs between searching for lost objects and exploration, the latter has easily justified the high costs of data processing. Furthermore, because of the much longer response times expected of exploration, there is ample time to undertake detailed resource planning and logging of the activities.

Whilst it may be argued that these are only

differences of degree, it is hard to believe that any practical support systems will not be greatly affected by them.

(f) System Objectives

At its simplest level, all search activity has the same objective, namely, to find the target before the resource allocated to the search runs out and the target loses its value. An efficient procedure for determining both the search budget and the search schedule enables the search controller to start the search. However, as both the location of the target and its detection are uncertain, feedback between the controller, schedule, and budget is required.

Exploration, being a multistage activity, needs a more sophisticated form of budgeting, scheduling, and feedback analysis in order to accommodate the interstage relationships and constraints. The difficulties of meeting exploration objectives are many orders of magnitude more difficult than those of classical search.

(g) Action Subsequent to Finding Target

The objective must also look beyond the finding of the target. In classical search, the simple actions of "render help" or "destroy" might suffice to complete the mission, but in exploration the work has only just begun. For example, in exploration the finding of petroleum or minerals first triggers off further drilling to delineate the reservoir or deposit and determine its commercial worth. Secondly, the engineers take over from the geologists to plan and schedule the production stage. The range, scope, and cost of activities undertaken during production are totally dissimilar to anything that takes place following classical search.

Conclusions and Directions for Further Development

From the review of search theory and exploration, certain general conclusions emerge:

1. Exploration is so dissimilar in character to general search that it is counterproductive to use classical search that it is conterproductive to use classical modeling concepts into any stage of exploration other than the first general survey stage.

2. A theory of exploration is more likely to emerge by using general modeling concepts to help exploration managers reach decisions on a number of specific questions related to the optimum level of exploration activity in a particular area (Allais [1957] and Cozzolino [1979]).

3. In common with many other complex systems, the correct management science approach to helping exploration management is through a general systems' study in order to determine how models and computers can be used to supplement various decision processes.

4. Even without such a study, it is clear that a major problem faced by exploration managers and geologists is the storage and retrieval of vast quantities of data. Such data-handling systems are a prerequisite to any form of quantitative geological modelling (see Grender, et.al. [1974]).

5. Additionally, there is a requirement to keep track of the many case studies developed by geologists and economic analysts. Efficient case-management systems are becoming increasingly necessary in a business environment in which expertise is at a premium and the need to conserve resources is of paramount importance.

References

Allais, M. [1957] Methods of Appraising Economic Prospects of Mining Exploration Over Large Territories, Management Sci. 3, pp. 285-347.

Cozzolino, J. M. [1979] Measurement and Projection of Exploration Search Efficiency, Symposium of Soc. of Petroleum Eng. Dallas.

Grender, G. C., Rapoport, L. A. & Segers, R. G. [1974] An Experiment in Quantitative Geological Modeling, Bull. Am. Assoc. Petroleum Geologists. 58, pp. 448-498.

Harbaugh, J. W., Doverton, J. H. & Davis, J. C. [1977] "Probability Methods in Oil Exploration," John Wiley, New York.

Hobson, G. D. [1973] "Modern Petroleum Technology," Applied Science Publications, Ltd.

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

EXPLORATION FOR COAL IN THE U. K.

J. G. Fergusson

Operational Research Executive
National Coal Board

Introduction

The purpose of this note is to outline the methods and problems of exploration for coal in the U.K. It is written in the hope that it will generate a fruitful discussion on the relevance of the theory of search to exploration for coal. Since in the U.K. the great majority of coal comes from deep mines, it concentrates on that aspect of exploration.

Exploration

The very essence, the raison d'etre of exploration, is to gain knowledge early enough to maintain continuity of production (within the bounds of market needs), that is, to ensure that sufficient workable reserves of coal are available to be mined as existing production units either reach exhaustion or have to be abandoned for other reasons. Until recently, depletion of reserves in the U.K. has been less due to the process of extraction and more due to "write-off" of less valuable, productive reserves to satisfy rising minimum productivity requirements. In 1970-71, it was seen that this, combined with the process of depletion, had continued too far, so even to satisfy the then falling market outlook, it was necessary to renew exploration for reserves for new mines. The momentum of this exploration was, of course, increased in 1973-74 with the change to a long-run, rising market outlook for coal. Different production units, by which I mean a coal face, a colliery, or on a national level the industry, have associated with them different problems of exploration.

Exploration, therefore, is a much wider concept than drilling holes in the ground to see if there is enough coal there to justify a new mine. This, of course, does happen, but it is a very small part of exploration and looking at the industry historically is a very rare event. The vast majority of exploration goes on within existing collieries and is an integral part of the production process (Clarke [1979]).

In this note, I will briefly describe exploration within a colliery because it is so important. I then go on to the national or industry level where, I believe, theory of search may have a role to play.

But first a brief historical review might help to get things into perspective.

Historical Review

Coal was first mined where the coal itself could be seen in outcrops on the surface. As time passed, deep mines became necessary, but reserves have always been in abundance. Each colliery probably lasts on average about 100 years. If a new one was needed, those knowledgeable about the coal field would have a considerable fund of information about a suitable site "up the road" and a bore hole to prove the reserves would almost certainly confirm much more than was needed. Because before nationalization in 1947 the industry was made up of a large number of small companies, the question of looking for coal in "new" areas was virtually unknown. However, nationalization and the current position of reserves means that this rare event is happening now.

Mining methods within a colliery have changed in the last few decades, and this has had a dramatic effect on the problems of exploration within a colliery. Old methods of bord and pillar had a large number of production units, so that if one unit met a geological fault which interrupted production, it was not too serious. Now with highly mechanized methods on just a few production units, the consequences of even one of those units being out of production for a short time has a far greater importance. Not only that, mechanized methods, because of their inflexibility, are much less able to find an easy way of working round the cause of the interruption .

In general terms, the basic method of exploration, drilling a bore holes, remains the same though it is now possible to do it in much less time. But recently, seismic methods developed in the oil industry have been adapted to be of use to the coal industry. The two methods complement each other in a way which I will describe later.

Exploration within a Colliery

Although bore holes are used within a colliery to gain knowledge about coal yet to be mined, by far the most important part of exploration is integrated with the production process. To get the right balance between the exploartion and production content of a colliery plan is a difficult problem and depends on the local circumstances so that no general solution exists. A modern colliery will typically have few (3 to 6) production units, usually longwall mechanised faces working on tow or three production shifts per day.

However, if we look at the way these faces are laid out in space, we can see that different layouts have different exploration contents. For example, two layouts, each operating six machine shifts per day, are illustrated in Figur 1. We can see that one (Layout A) is much more concentrated in space than the other. It is also more efficient since the backup services of transporting coal, men, materials, etc., are reduced to a minimum. However, it is prone to considerable interruption of production if, for example, there is a geological disturbance ahead of faces 1's and 2's big enough to stop them.

In Layout B, 4's face gives the manager of the colliery considerable information about the geological structure of the seam where he is going to have to put his future faces. We can see, therefore, even if it is not easy to quantify, that Layout B has a much higher exploration content. The problem of mine planning is, therefore, to get the best balance between short-term production considerations and exploration.

The geologists in the N.C.B. and members of the O.R. Executive have developed a method of comparing alternative layouts. It is called Geosimplan and is described by Hawes [1973]. Essentially, it uses an operational gaming approach where alternative strategies are compared with an extrapolated or interpolated state of nature represented by a number of configurations within the assumed range of error in extrapolation. Its purpose, apart from serving as a forum for communication between geologists, mining engineers, and other specialists, is to see which strategy is likely to prove the most robust. This approach has been used many times in British coal mines. It is rather cumbersome and time-consuming, but no satisfactory alternative has yet been devised.

Exploration at National or Industry Level

At national or industry level, the problem is to find sufficient suitable reserves to justify sinking a new colliery. At present in the U.K., this involves exploring new coal field areas.

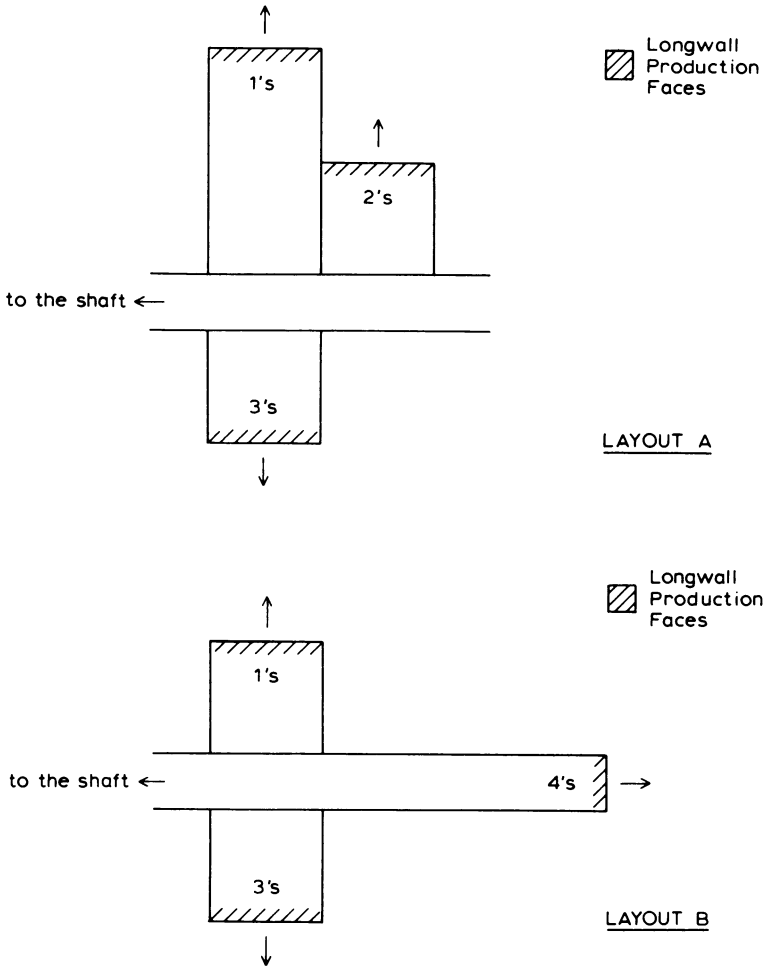


Figure 1. Examples of layouts of faces.

The question of which site to explore next is not straightforward. The geologist's views on where suitable reserves exist is one important aspect. But, of course, one is really to find the next cheapest coal to mine so that financial aspects also come into it. In addition, planning permission must be obtained not only to mine the coal but also to drill bore holes to see if suitable coal is there. What is on the surface can also

have a considerable effect since it will influence the amount of coal that has to be left unmined to avoid subsidence. So to put it into theory of search terms, the whereabouts of the "target" is not the only consideration. Indeed, the decision of which site to explore next is largely a subjective one.

Once it has been decided to explore a new site, four separate phases of exploration occur. The site will be in any area with mining potential but which has had sufficient ambiguity about it for it not to have been mined in earlier phases in the growth of the mining industry.

The first phase is concerned with finding out whether there is any workable coal there. This can be answered by drilling one or two bore holes. This will, of course, provide information about the depth of the target seams and how thick they are and, hence, give an indication of the surface area needed for reserves at least in sufficient bulk to justify access.

The second phase is to find out if the newly discovered coal bearing site is sufficiently extensive to justify making access. This can be found out by sinking three or four bore holes at the limits of the mining area necessary to give sufficient bulk reserves. (If it exceeds this minimum, the same criteria are used to site holes to gain an indication of the best place to start the third phase.)

The next phase is designed to find out if the newly found target seams are of sufficient market quality and sufficiently undisturbed geologically to be likely to give the levels of proceeds and productivity required in a new mine. Here, as well as drilling some more bore holes, a grid of seismic lines is used to identify the orientation, pattern, and intensity of major faults

This third phase will give sufficient data to make an initial selection of a possible access site and a preliminary evaluation of a mine layout in the light of environmental, market, and manpower considerations.

The fourth and final phase is to decide what constitutes enough detailed exploration to justify a particular mine plan, which must, of course, be financially viable. It is, of course, not possible to have complete knowledge of the reserves. This is similar to the third phase. However, the difference is that exploration in this last phase is linked to a specific plan, and its aim is to reduce error in knowledge of the state of nature on points where this is critical to the initial success of the plan. Whereas in the third phase, data is being collected to help select an initial plan.

Given success with each of these phases and, of course, planning permission, arrangements can be made to start sinking the shafts and mining the coal.

Discussion

Having outlined the methods and problems of exploration for coal in the U.K., it would seem appropriate to condlude with some preliminary thoughts on the relevance of the theory of search. I do this by describing three problems and analyzing them on the lines suggested to me by Koopman's [1956] paper, where he describes the four elements that can be distinguished in a search problem. The four elements are:

i) We must have some knowledge of the whereabouts of the "targets."

ii) The characteristics of the detecting equipment must be known.

iii) The quantity of searching effort must be defined.

iv) Any search must be regarded as a component of a larger operation.

The first problem is very much connected with this fourth element. As you will appreciate, exploration, in practive, is not such a discrete process as I have described. Extra effort in one phase can reduce the need for effort at a later stage. Getting the balance right is the problem.

The fourth phase of exploration at national level consists of obtaining sufficient data to justify a particular mine plan. Exploration at this stage is nonproductive in the sense that it does not produce a ton of coal, but the characteristics of the methods used are well-understood and costs are known. However, complete knowledge is impossible. What is more, the value of extra knowledge is not easily quantified.

It is conceptually possible to use the money saved at this stage to increase exploration content at the later stage, i.e., when access has been made, which I have described as exploration within a colliery. This later-stage exploration is productive as it is integrated with the production process. However, it is not easy to quantify the total cost of this exploration, which must include an estimate of the benefits to be derived from the increased knowledge.

Thus, if we define the problem as a two-stage search problem, the target can never be completely "found," The characteristics of

the detecting equipment are different in the two stages, and the cost and benefit in the second stage are difficult to quantify. It seems to me unlikely that the theory of search can help in this situation.

Let us move on to the next problem. Koopman's second element says that the characteristics of the detecting equipment must be known. Presumably, it is possible in certain circumstances to turn things round and use theory of search techniques to assess the benefit of a new piece of detecting equipment or to specify how good it has to be and at what cost for it to be beneficial. Research is currently going on to develop seismic methods which can be used within a seam of coal. This will help identify geological distubances ahead of an advancing face. Clearly, a number of parameters will arise, e.g., how far ahead of the face can it identify a disturbance, with what accuracy can it identify the size of a fault, and its direction. The real difficulty in assessing this new detecting equipment is that even if you specified precisely how well it can perform and how much it will cost, it is not easy to quantify the benefits of the additional information. Perhaps, this is one of the fundamental differences between theory of search applications and exploration for coal. If, in the latter, we describe a geological disturbance as the "target," finding the "target" is not of paramount importance, though it is known that prior knowledge can avoid heavy costs. Perhaps, theory of search has an application in the exploration for coal at national or industry level, particularly at the preliminary stages of deciding which site to explore next. Here, the "target" is to add sufficient to the existing reserves of workable coal to justify creating access to maintain supply within the bounds of market needs. In the search for each individual mine site, the characteristics of the detecting equipment are known and the quantity of effort available in a given time period could be defined. However, the knowledge of the whereabouts of the target is not the only factor to influence the decision. Other factors, such as the likelihood of obtaining planning permission and the amount of coal which will have to be left unmined, are also important.

To summarize, therefore, the vast majority of exploration in the coal industry in the U.K. is concerned with gaining advanced knowledge of geological dip, disturbances, and of coal thickness distribution and quality. The coal itself as a "deposit" is relatively easy to find. In theory of search terms, therefore, if the target is geological disturbance, etc., the benefit of finding these aspects of the target is not easily quantified, though it is known that advance knowledge in some instances can avoid considerable cost. In addition, the "target" cannot be completely found, i.e., complete knowledge of coal seams, their quality and their strata characteristics, as a configuration in space, cannot

be obtained until the coal has been extracted. Perhaps, these two factors make the theory of search, at least in its present state of advancement, inappropriate. We should certainly not forget Koopman's fifth, and perhaps most important, element, that the mathematical formulation of search problems should be carried to a refinement consistent with (i) the degree and accuracy of the searcher's knowledge and (ii) his ability to implement the theoretical optimum results.

Acknowledgements

I would particularly like to record my thanks to A. M. Clarke, Chief Geologist of the N.C.B., for his help in producing this note.

The views expressed are those of the author and not necessarily those of the National Coal Board.

References

Clarke, A. M. [1976] Why Modern Exploration Has Little To Do With Geology and Much More To Do With Mining, Colliery Guardian Annual Review. August.

Dunn, R. B. & Clarke, A. M. [1975] Recent Developments in Coal Exploration in Great Britain, World Coal. April.

Hawes, D. M. [1973] Operational Gaming in the Planning of the Geologically Troubled Colliery in "Operational Research 1972," M. Ross, ed., North Holland/American Elsevier.

Rees, P. B. [1975] Exploitation of Deposits, Mining Engineer. April.

Sloman, M. [1977] How Operational Gaming Can Help With the Introduction of a New Technology, Operational Research Quart. 28, No. 4i.

EXPLORATION FOR POLYMETALLIC NODULES

J. P. Lenoble

Centre National pour l'Exploitation des Océans

This paper is an introduction to the field of polymetallic nodule exploration, most of which lies at the bottom of the ocean below 15,000 ft. Among the minerals that are expected to be found under the ocean, polymetallic nodules are probably the most unusual. Other minerals, which include:

(1) oil and gas;
(2) accumulation of heavy minerals called placers bearing tin, titanium, chromium, and iron;
(3) sulphides of zinc, copper, lead, and iron which are found in red sea brines; and
(4) phosphate nodules;

are already known and mined on land. Polymetallic nodules are very rare on land and do not seem to have ever been mined.

"Polymetallic" refers to a chemical composition of oxides of metals which are in order of decreasing importance:

-- manganese up to 50% in weight;

-- iron but in such quantities that cannot give the nodule the name iron ore;

-- silicon which is not a metal and appears here as a waste material;

-- aluminum which is generally tied to silicon as aluminosilicate and is usually considered unrecoverable;

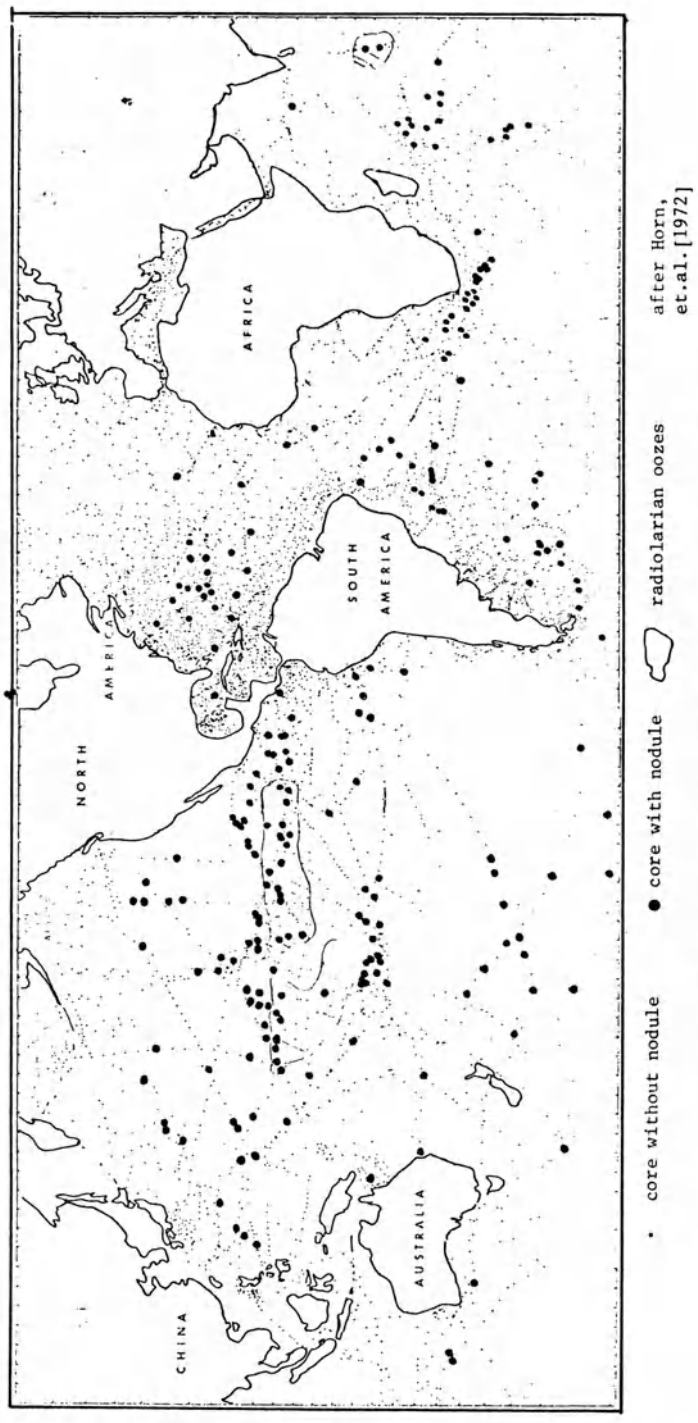


Figure 1. World nodule distribution from cores.

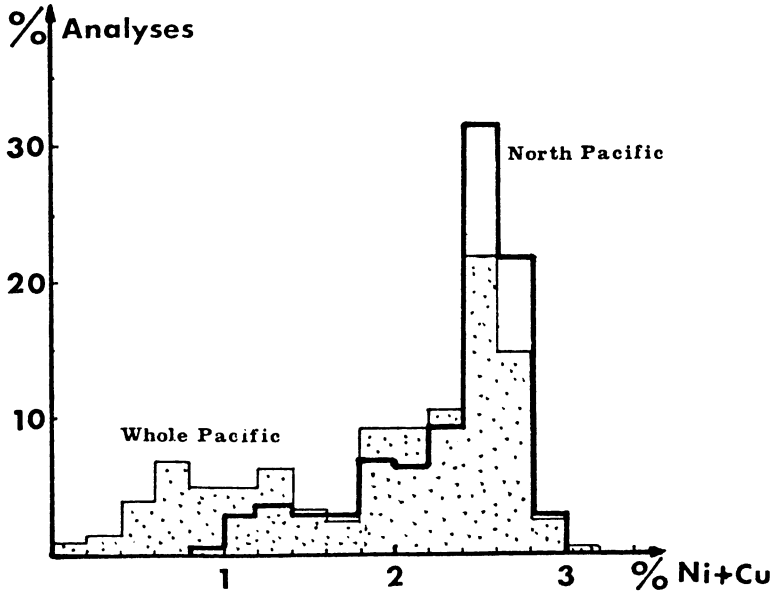


Figure 2. Histograms comparing the nickel and copper content of nodules in the north Pacific and whole Pacific (dotted).

manganese and abundant on the bottom appeared to be correlated with a special underlying sediment which contained remains of planktonic animals called radiolaria.

These radiolarian oozes are mainly distributed in the North equatorial Pacific with some patches lying in the central Indian Ocean (see Figures 1 and 2). Most of the work concerning polymetallic nodules has been concentrated in the central Pacific area.

Statistical theory was applied to geological features to determine a probability of finding nodules of economic interest on the sea bottom at the current state of knowledge. As a result, the search area was reduced to less than 10 million sq.km., i.e., by more than 70 times. However, knowledge is still insufficient to localize areas of at least 50,000 sq.km. where nodules are present on the bottom with an average of 10 kg/sq.m. concentration and a Ni + Cu content above 2.5% which was estimated to be the mining target. Actually, the sampling distribution was less than 6 sample locations per million sq.km. with a sample taken every 400 km.

Evidently, sampling had to be improved. No indirect method using geophysical en route measurement was available. The fastest way to explore is still the launching of free-fall equipment, e.g., on arrival on the bottom, shoot a strobe light picture, take a sample, and release their dead weight in order to return to the surface 3 to 4 hours later. Cost is a function of time depending itself on distance between station, ship speed, station time, transit time, etc. A one-month cruise, including analysis and data computation, costs $500,000. This cost is optimized using geostatic techniques.

Variograms of the main economic parameters are computed: metal content and concentration. These variograms express the variation of the quadratic deviation of the variable between two locations a distance h apart where h is increasing. Let γ be the variogram function; then

$$\gamma(h) = 0.5E\{[f(x) - f(x+h)]^2\},$$

where x is a point of the space V, $E\{\cdot\}$ is the mathematical expectation, and $f(x)$ is the value of the variable which is actually a stochastic function of space.

A bounded variogram means that above a certain distance between two locations, the value taken by the variables must be considered as independent. Below this distance, the values are related by an unknown probability law, and the variable is said to be regionalized.

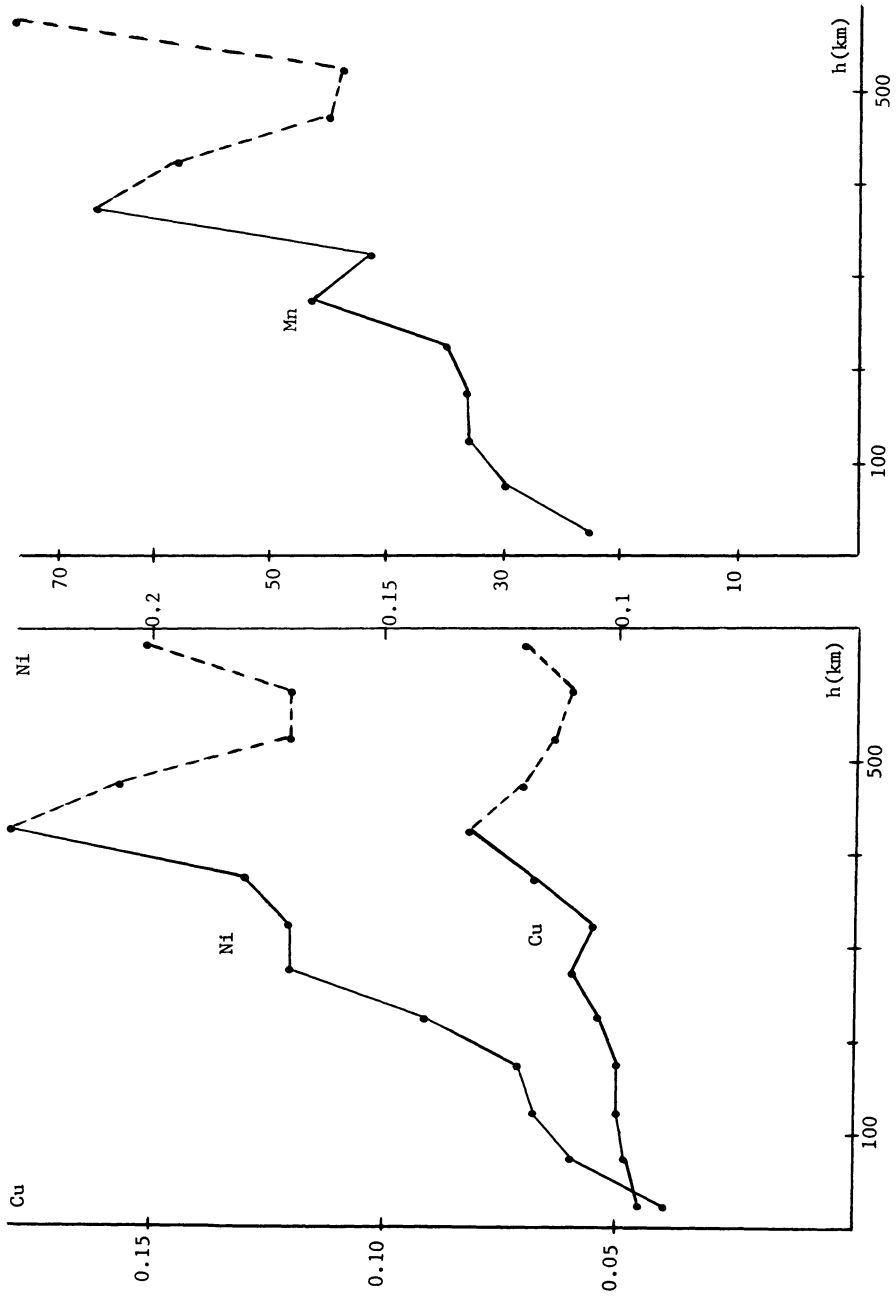


Figure 3. Variograms of data available in 1973.

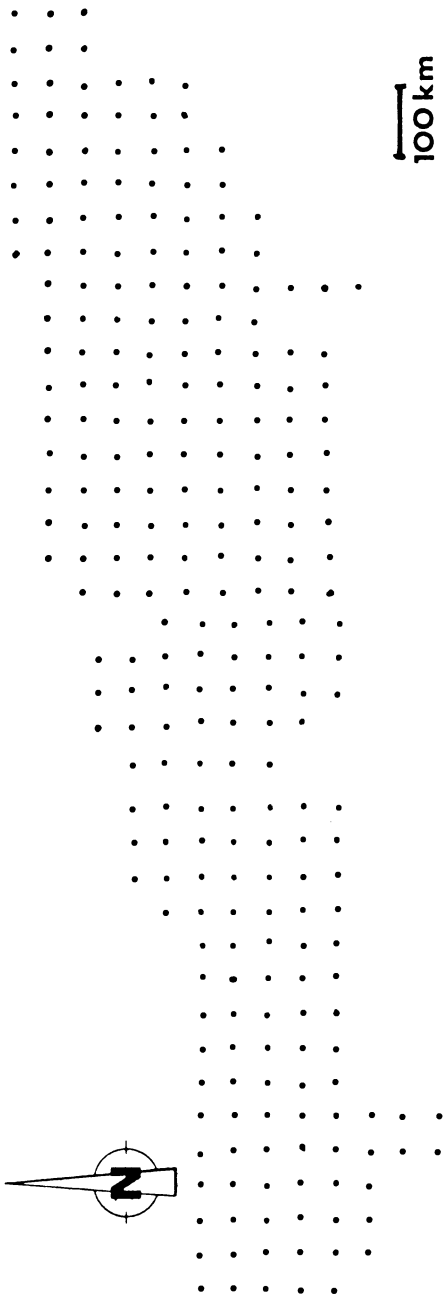


Figure 4. The grid pattern of the prospected area of 2.25 million km^2.

-- sodium, calcium, potassium, and magnesium present as they normally are in sea water;

-- cobalt, nickel, and copper considered with manganese as the main economic justification for the search for nodules as their grades could be in the range of the corresponding on-shore ores of the same mineral; and

-- titanium, barium, lead, and some others such as molybdenum could also be profitable.

Nodules look like black round balls of tomato size. Their morphology varies considerably from point to point and is somewhat linked with their chemical composition; thus, photographs give some clue to their composition without having to sample and analyze the nodules.

They lie on the sea floor, sometimes deeply imbedded in the sediment, and can be completely hidden. Their concentration or abundance is expressed in kilograms per square meter and varies from zero to 20 kg/sq.m.

They were discovered at the end of the nineteenth century during the scientific cruise of H. M. S. Challenger around the world. But until the late 1960's, they were considered as scientific curiosities only, at which time mining companies were attracted to the commercial possibilities.

As in most mineral exploration stories, the first step is a compilation of all available data. This compilation shows a world-wide distribution of nodule occurence (see Figure 1). The apparent homogeneity of this distribution is overrated and biased by the heterogeneity and the poor quality of the sampling. In fact,

(i) Most of the information is from areas close to the main oceanographic laboratories or on shipping lanes.

(ii) Samples were taken with various types of apparatus some of which could not give an accurate prediction of the nodule abundance because the coring devices had a diameter of the same size as a nodule.

The quality of the nodules differs considerably from one area to another. If expressed in nickel content--which is well correlated with copper and manganese grade--then the Pacific Ocean is much more interesting than the Atlantic Ocean. From these studies, it was found that nodules rich in nickel, copper, and

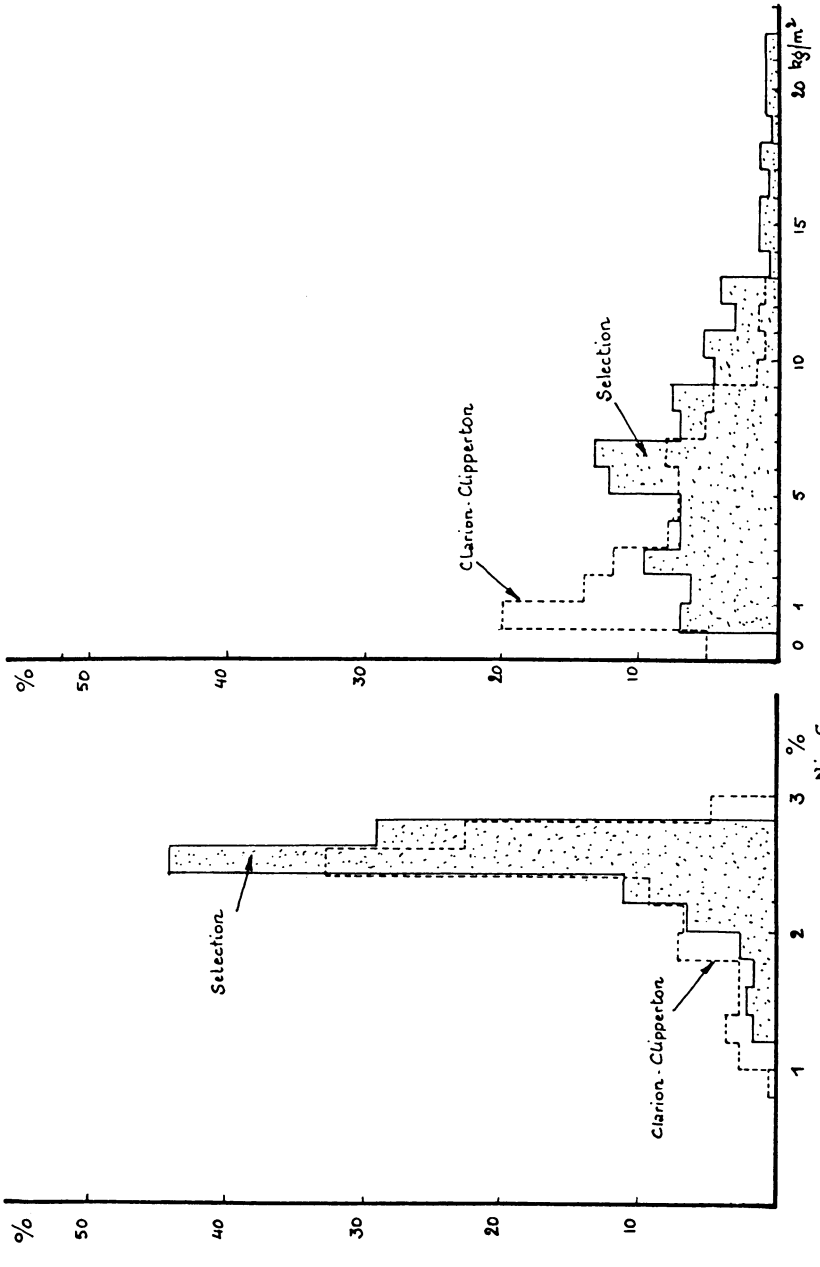


Figure 5. Compared characteristics (content and concentration) of nodules in the selected area and in the Clarion-Clipperton area.

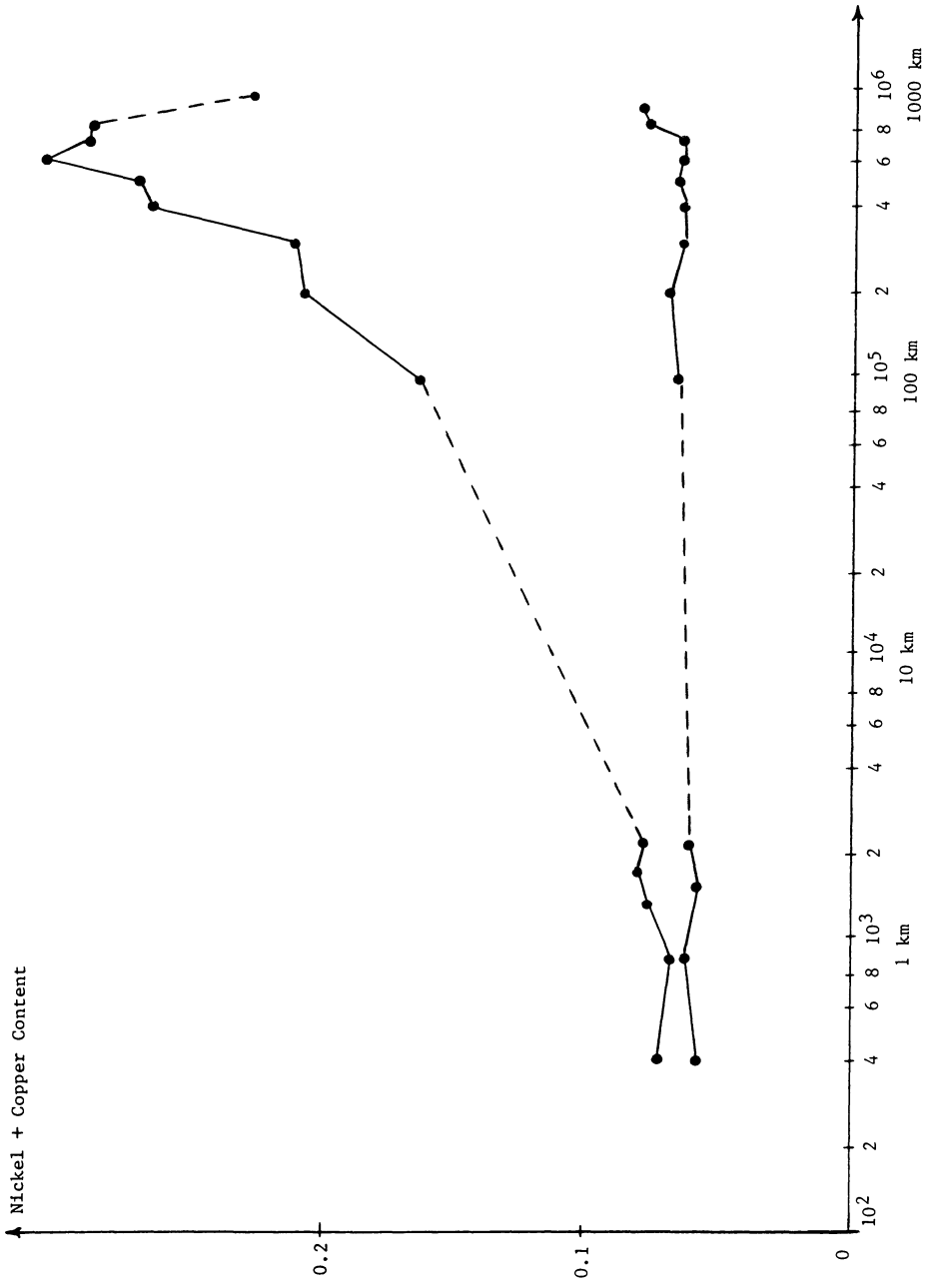


Figure 6. Nickel plus copper variograms.

The metal content variograms computed in 1973 showed a constant growth over 300 km. distance (see Figure 3). It was forecast that a 100 km. spacing between stations would be sufficient to delineate large areas where the nodules are chemically alike (see Figure 4).

This prediction proved to be correct, and a 350,000 sq.km. area was localized from a 2.5 million sq.km. prospected area where both the grade and the concentration of the nodules was higher than in the prospected area as a whole (see Figure 5). Further investigations, with the spacing between the observations reduced to 30 km. inside the selected area, confirmed the previous results.

It was also shown that inside the selected area, the metal content variograms were flat proving a random distribution between some 100 meters to several hundred kms. existed (see Figure 6). Independently, the concentration variogram shows a double-step shape with the first level corresponding to some areas which seem to be the size of nodule field units.

Although some stochastic approaches have been used, the techniques of search theory have not yet been employed. New prospecting methods, using towed fish with continuous pictures of the bottom and even side-scan sonar, are being introduced, and it is hoped that the theory of search can now be utilized.

Reference

Horn, D. R., Ewing, M., Horn, B. M. & Delach, M. N. [1972] Worldwide Distribution of Manganese Nodules, Ocean Ind. 7, pp. 26-29.

THE SEARCH FOR NODULES

G. Friedrich

Institut für Mineralogie und Lagerstättenlehre

J. P. Lenoble

Centre National pour l'Exploitation des Océans

J. G. Fergusson

Operational Research Executive, National Coal Board

Manganese nodules are found on the bottom of the ocean. They are widely but unevenly spread so that in some areas they may cover 100% of the sea floor and in others be completely absent. Theories about their origins are currently being discussed by experts.

There is a large area in the Pacific, about 100 km. north to south by more than 500 km. from east to west, which is the most promising prospect. The density of the nodules in this prospect ranges from 0.5 kg/sq.m. to 38 kg/sq.m.

Typically, nodules are about the size and internal shape of an onion and contain nickel, copper, cobalt, and iron as well as manganese as the most important constituents.

Average percentage composition of dried manganese nodules from the Pacific:

Mn	29.8
Fe	4.8
Ni	1.36
Cu	1.2
Co	0.2
Si_2	13.0
Zn	0.12
Pb	0.05
Mo	0.05
Other	49.52
	100.00

There are a number of exploration methods which can be used. These include T.V., camera sled, and grabs with cameras. All these can be used to gain information about the location and size of the nodules, but the grabs have to be used to gain knowledge about their mineralogical and chemical composition.

The weight and quality distributions of the nodules can be estimated from samples taken by means of the grab over a known area.

In order to assess the value of mining the nodules, a number of factors must be estimated, including variations in these factors to estimate the revenue.

Cost estimates require a knowledge of the topography since it is impossible to mine in certain areas because of sea mountains. Soil mechanics can also affect the efficiency of the mining process. The depth of water which is normally 5,000 m affects cost--it being cheaper to mine in shallower waters. The total investment for phase 1 of nodule mining might be about $500 m.

The problem, where it is hoped that the theory of search might help, is the planning of the exploration program so that the most beneficial information can be obtained for the purpose of assessing the economics of mining the nodules.

SEARCH THEORY IN THE EXPLORATION FOR AND MINING OF POLYMETALLIC NODULES ON THE OCEAN FLOOR

Daniel H. Wagner

Daniel H. Wagner, Associates

In this paper, we discuss search-related operations research problems which appear to be important to the exploration for, and mining of, polymetallic nodules on the ocean floor.

In the first two sections, we review fundamental features of exploration equipment and mining equipment. These provide a setting for the O.R. problems of interest here. We then outline in the next two sections some apparently relevant O.R. problems in the exploration and mining phases, respectively. Exploration and mining strategies are highly interdependent. In combination, they bear significantly on design requirements of the expensive mining vehicles.

Exploration Equipment Tools

The principal types of equipment used, and envisaged for use, in exploration are the "free fall grab" (FFG), spade corer, and a towed sled or fish on which is mounted a still or TV camera and possibly a side-looking sonar (except for sonar shown in Figure 1 taken from Metallgesellschaft [1975]). Magnetometers are used for structure information. At a given point location, five to seven FFG samples plus a spade corer sample are obtained. These yield very high quality data. Each FFG is lowered from the surface and recovered three hours later (at 5,000 m water depth) with a bottom sample 30 cm x 30 cm of loose surface material, together with a photograph 1 m x 1 m centered on the sample. It is reputed to grab a nodule in its sample if the nodule density exceeds 2% by area. Each grab device costs $3,000 to $5,000, and they are lost on about 2% of the drops. The surface vessel may move among drop

Schematic drawing of the equipment on the research vessel VALDIVIA.

Navigation

1 stars
2 satellites
3 radio navigation
4 navigation buoy (transponder/radar)

Bathymetry

5 narrow beam sounder and sediment echograph
6 various depth recorders

Reflection seismic

7 airgun
8 streamer with hydrophones, analogue and digital registration

Oceanographic survey

9 underwater measuring chain with localizable buoy
10 current meter
11 thermometer
12 water pressure gauge
13 cut-off anchor
14 bathysonde (continuous measurement of temperature, salinity, sound velocity, pressure)

Survey of ore deposits

15 deep diving probe with TV camera, still camera and lights
16 depressor platform
17 corer for sampling sediment with nodules
18 freefall sampler
19 bulk sampling of nodules for metallurgical tests

Localization of launched survey gauges

20 pinger
21 hydrophone
22 transponder

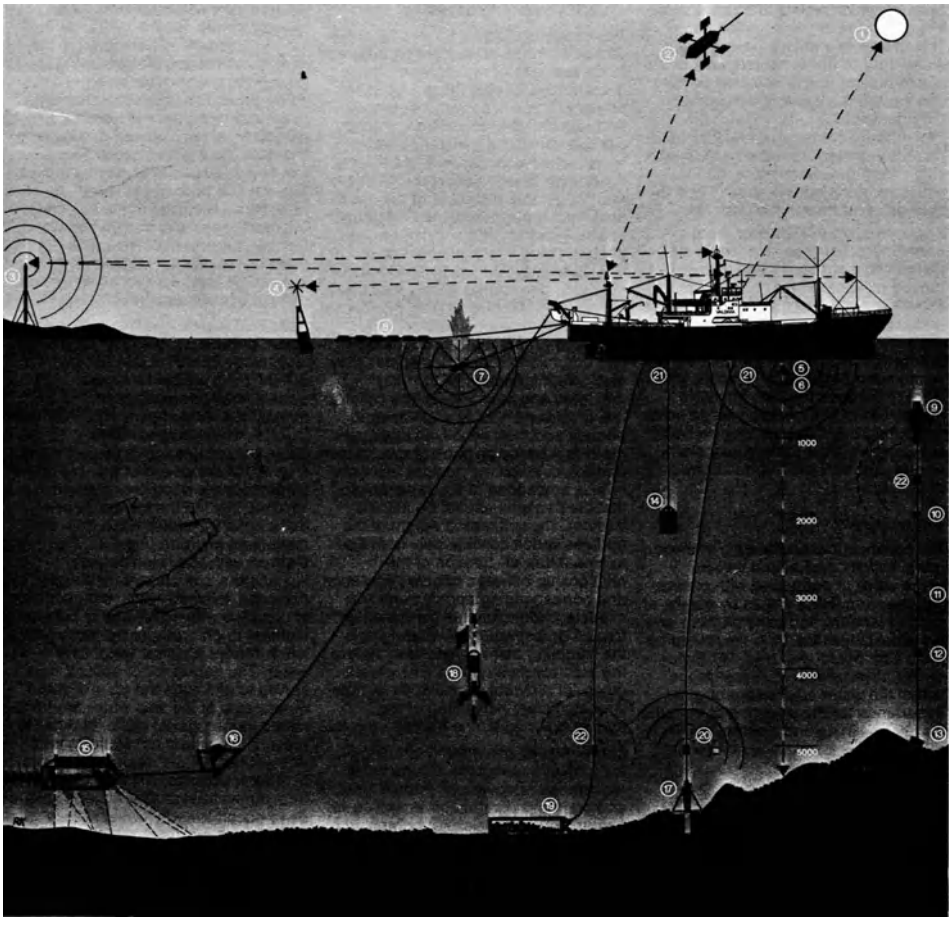


Figure 1. Exploration and mining equipment (from Metallgesellschaft [1975]).

points while waiting for recovery.

A much more mobile sensor platform is a towed sled or fish. Various stages of improvement are contemplated. For all of these, tow speed is estimated to be 1 to 2 knots (average 1.5 knots) with 4 hours to lower and 4 hours to recover, to 5,000 m depth, and 2 hours to reverse course. The stages envisaged are as follows:

(1) A current sled skids on the ocean bottom and carries a battery-operated camera, which looks ahead to photograph a 3 m x 3 m area.

(2) A current fish is towed 10 m above the bottom and from a mounted camera photographs vertically over a 10 m x 10 m field. Battery power limits endurance; two cameras each loaded with 6,ppp exposures enable 12,000 sequential exposures to be taken. Exposure rates may be varied but must be set before lowering. The photo processing and interpretation loads are heavy and might be limiting factors. Photo interpretation enables discrimination of mineral nodules with a high degree of confidence but is rather subjective as to quality. (Good quality assessment requires grab samples.)

(3) A later fish will carry a cable-powered TV, continuously monitored in real time at the surface. It detects nodules with the same high confidence as a camera and with the same problems of ascertaining their quality. Real time detection is an important advantage.

(4) Still later, a fish will be equipped with a high resolution side-looking sonar towed about 100 m above the bottom. This might look 100 m to either side, with a blind region in the middle about 50 m to 100 m wide. Obviously, the sonar will have a much higher sweep rate than the TV or camera, but it also has a much higher false target rate (unknown). It will require cable power. Probably, a TV or camera will accompany the sonar with the tow being in two tiers

Each RV or sonar wet will cost at least $250,000 (after development) so they may not be used where there are underwater hazards such as mountains or rough terrain.

Navigation is a considerable problem, particularly for towed vehicles, in nodule exploration. A combination of satellite and bottom transponder interrogation will be used. For broad

exploration, bottom transponders are not economical, and navigation errors of 2 km are typical using satellites. For localized mapping, bottom transponders are economical. These may be left in place for a year and then replaced to afford good repeatability of location even though geographical reference errors are large. Even though a "mini-base" on the surface vessel is used to locate the fish relative to the ship, these errors are dominated by the ship position error (in rms combination with the ship's position error). Either the ship or the fish may be located to 10 m relative to a group of three or more transponders. Bathymetric navigation is sometimes used.

Mining Equipment

Mining vessels remain to be built and equipped, although the envisaged methods have been tested experimentally. Separate trials have collected 1,000 tons in 48 hours and 500 tons in 18 hours.

The concept is as follows: A large vessel will trail a suction device at 1 knot which will bring to the surface 80% (optimistically) of the nodules in a swept path 30 m wide. A course reversal will take several days.

Large investments are needed; capital costs (here, M = 1,000,000) are $500 M for a vessel, $600 M for a treatment plant (ashore), and $200 M for transport to ports, port facilities, and a plant to dispose of processed waste. Operating costs are estimated at $200 M per year. Since revenue is currently estimated at $200 M per year, substantial metal price rises are needed as well as consideration of inflation before nodule mining becomes economical.

In the meantime, oceanographic research is collateral to commercial potential in motivating continuing exploration. Exploitability is depth-dependent, but high density nodule fields are usually found at between 4,000 m and 5,500 m depth. Distance to a shore-based processing plant is also relevant.

Analysis of Mining Stage

In this section, we outline search problems and other operations research that pertain directly to the mining stage.

To begin a mining operation (in fact, to make the mining capital commitments), we must have identified high density patches of sufficiently high quality nodules. These patches must also be readily locatable with high confidence. We presently formulate these requirements in quantifiable fashion.

To make a decision to begin mining a region and to choose the best way to proceed, one needs a description of the potential of the region which might be given as Figure 2.

Figure 2 shows a field of 12 patches of nodules. A topographical barrier, which may not be crossed during an operation of the mining equipment, divides it into two groups which might as well be considered as separate fields. Let us confine our attention to the group of five patches on the left of the barrier.

Each patch is annotated with its average density of metallic yield per unit area (in \$/sq m to combine the values for the separate metals). This quantity is the average density of nodules per sq m multiplied by the average value per nodule. We denote this quantity by D_i for the ith patch.

When good information as in Figure 1 is at hand, determination of good tracks to be followed by the mining vehicle appears to be best done by preliminary graphical trials in myopic fashion. One lays out various tracks which are feasible in the light of severe turning limitations. Each track is evaluated by its expected dollar yield per time consumed estimated by

$$\frac{\Sigma_i W\, D_i L_i V}{T_r + \Sigma_i (L_i/V + T_i)},$$

where W is the sweep width of the pickup device at the bottom (including an efficiency factor), V is speed while mining a patch, T_r is turnaround time between tracks, and for the ith patch D_i is the density, L_i is the track length covered within the patch, and T_i is transit time from the ith patch to the (i+1)st patch. The transit speed is presumed higher than the mining speed.

In Figure 2, track A is possibly attractive because of the high D values and low transit distances. Track B might also be attractive even though it entails a shallow turn between patches.

When good tracks are chosen, much or most of the patches that they intersect will best be mined by tracks parallel to the chosen ones, since W is small compared with the presumed patch dimensions.

While the mining strategy and its evaluation, given in Figure 2, and the characteristics of the mining vehicle are straightforward, it is important that this analysis be carefully done since it bears on both required accuracy in the exploration

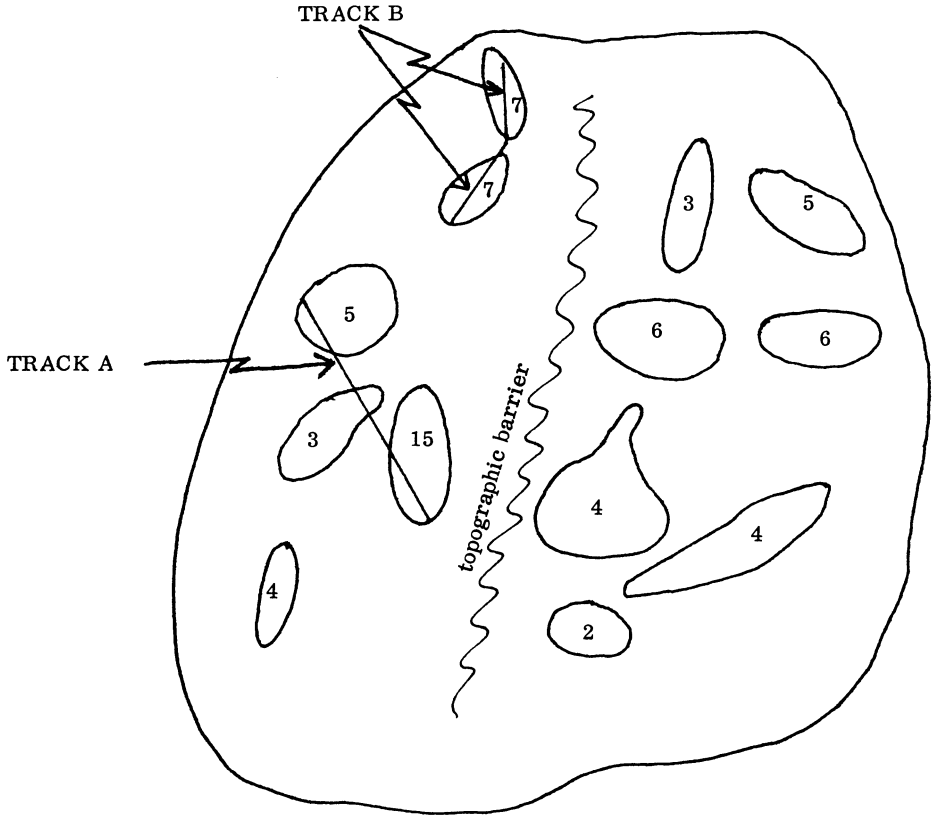


Figure 2. Illustrative description of a well-mapped nodule field.

phase and operational requirements in the vehicle design phase. For example, it can probably be shown that there is a high premium to be gained by reducing T_r and T_i with a more expensive design. The spacing of parallel tracks should be chosen by using Re/er's [1956] results to avoid undesirable overlaps and excessive gaps, i.e., to achieve a prescribed thoroughness of coverage under the estimated navigation accuracy.

The foregoind discussion of mining strategy assumes a myopic choice of tracks, i.e., at each stage one chooses the most cost-effective track available at that stage. In fact, one probably should not proceed myopically or one may be left with scattered high-yield pieces which are unexploitable because of their scatter but which could have been profitably included by a better sequence of mining tracks. That is, the choice of the next track should consider not only the cost-effectiveness of that track but also the exploitability of what remains (in analogy with optimal search for moving targets).

In full precision, this appears to be a profound problem in constrained finite optimization, possibly akin to the traveling salesman problem. In practice, brute force trials would probably do fairly well. Between these extremes, development of clever algorithms and use of machine computation will probably be worthwhile.

Analysis of Exploration Phase

In this section, we outline an analysis of the exploration phase. This phase begins with geophysical surveys and eventually leads to a description like Figure 2, which is the objective of the phase.

Before probing by sensors, identification of broad ocean areas believed to have good nodule content is obtained from marine geoscience knowledge, i.e., geophysics, geochemistry, and bathymetry The areas thus identified are several thousand km in extent. The most attractive to date appear to be in the Pacific, north and south (less in the south) of the equator, but not including high sedimentation areas near the equator. Figure 3 of Mero [1965] illustrates some of the information obtained in this preliminary fashion. We understand that better maps have since become available.

Let us suppose that by geological surveys beyond those illustrated by Mero [1965], we have identified a region of several hundred km in diameter as attractive for exploration. How do we

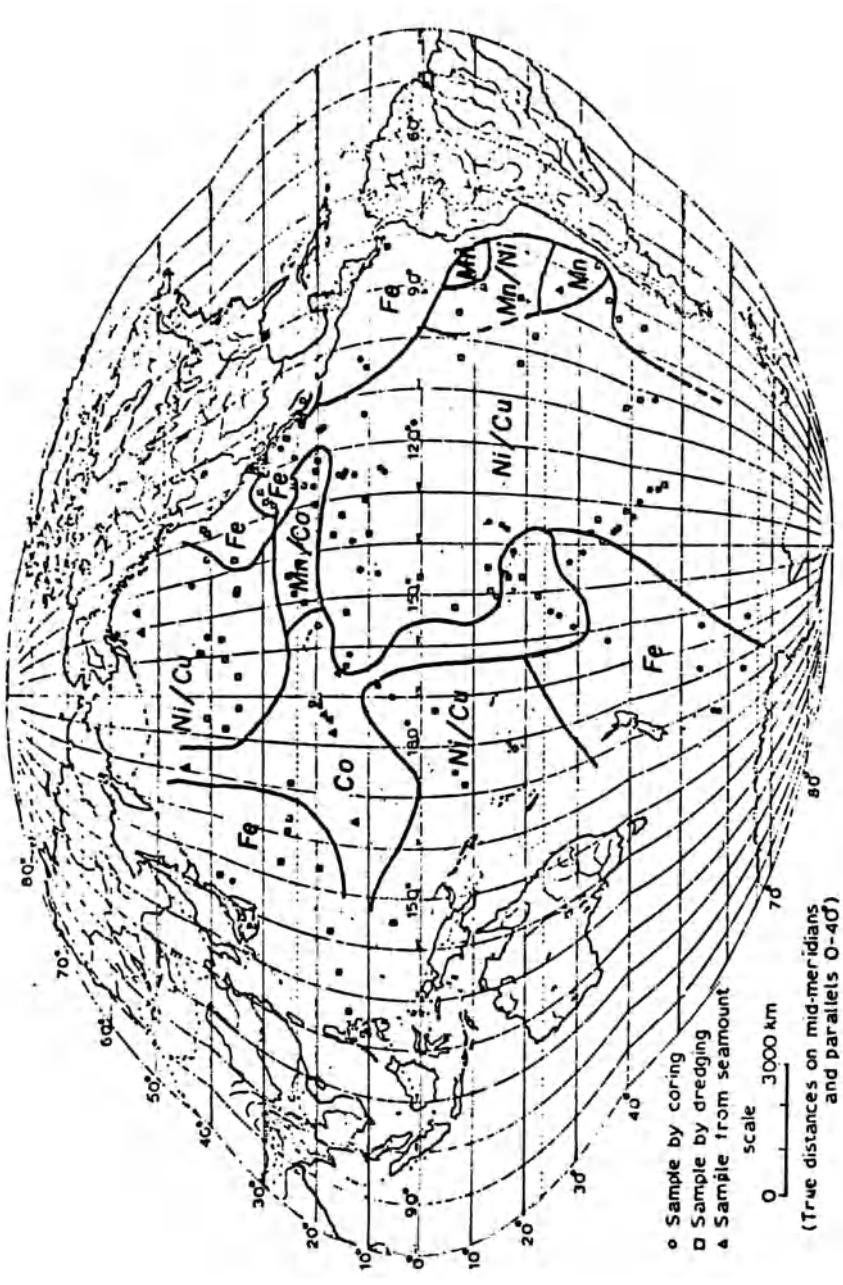


Figure 3. Compositional regions in the Pacific Ocean showing areas of significant elemental concentrations in manganese nodules (after Mero [1965]).

develop a map such as Figure 2? Successive refinements of information by FFGs and towed fish are needed.

In recent years, one has proceeded in successive stages by beginning with widely spaced probes (50 - 100 km apart) and then refining the favorable indications with finer spacing. Parallel sweeps are generally employed. Figure 4 (taken from Metallgesellschaft [1975]) refers to FFGs only.

The general form of the procedure in Figure 4 appears to be excellent. Results can be improved first of all by supplementing FFGs with a towed fish. Probably, they can be further improved by analysis of spacing and by Bayesian or other adaptive analysis which among other things adds useful inference gained from unsuccessful search.

Let us describe an adaptive approach which might be useful in choosing successive groups of FFG probe points. It is adaptive in that at any stage it uses the results of FFG probes to date to plan the ensuing probes. Any exploration planning would use such adaptive reasoning of some sort, but it is proposed that the method presented below may provide a better analytic basis for probe planning than might be otherwise available. The proposed approach is not really Bayesian but is somewhat similar in spirit, and it may well be desirable to generalize this approach to be Bayesian in particular respects.

We first form a description of the region to be explored, making full use of the causal knowledge obtained from geological and bathymetric surveys, which description should include a gross estimate of the potential for mining at each point in the region that is to be a candidate for an FFG probe. Since the estimate will generally be uncertain, it must be in probabilistic terms. If m probe points are to be chosen for FFG probes, part of the basis of choice would be high estimated potential; that alone as a basis of choice would be much too shortsighted, since one must look ahead to subsequent refinement of information about the region. The m probe points having been chosen, probing is conducted accordingly, and the samples obtained give one essentially certain information at those points. Now one must revise the probabilistic description of the exploration region in light of this new information. The revision is used as a basis for planning the next choices of probe points, and these probes lead to a new revision, etc. We now describe a model which might be used along those lines.

For each point z in the surveyed area, let $D(z)$ be the (unknown) yield density (in $\$/m^2$), corresponding to D above. Let $l(z) = \ln D(z)$, and let $\mu(z)$ be the estimate of $l(z)$ obtained from the surveys and $\sigma(z)$ the standard deviation of this estimate. It seems plausible that this estimate of $l(z)$ has a normal

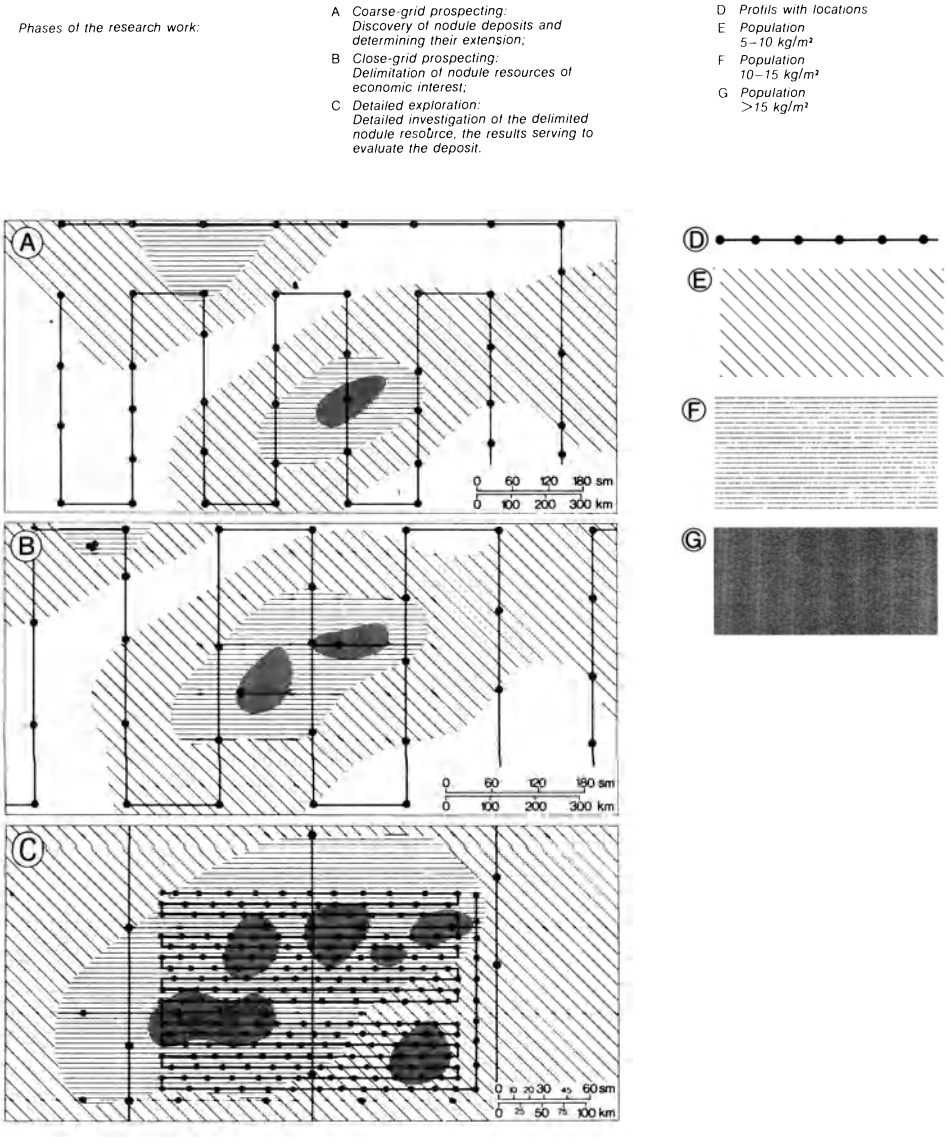


Figure 4. Exploration stages (from Metallgesellschaft [1975]).

distribution, i.e., that the estimate of $D(z)$ is distributed log-normally. It is also plausible that if x_1 and x_2 are two points, the correlation coefficient between the estimates of $l(x_1)$ and $l(x_2)$ decreases as the distance, $d(x_1,x_2)$, between x_1 and x_2 increases; let us conjecture that this coefficient is $\exp(-\alpha d(x_1,x_2))$ for a fixed $\alpha > 0$. The corresponding covariance is $\sigma(x_1)\ \sigma(x_2)\ \exp(-\alpha d(x_1,x_2))$, which we denote $\gamma(x_1,x_2)$. We have hereby embodied the presumption that if $D(x_1)$ is relatively high, and x_1 is near x_2, then one expects $D(x_2)$ to be high also.

The preceding paragraph is a somewhat heuristic description of how one might proceed in casting in probabilistic terms the uncertainties in physical knowledge that are present. We now make a more complete statement by regarding l as a sample path of a spatial-parameter stochastic process L and by postulating L in a way which is consistent with the above probabilistic assumptions.

We specify L by specifying all of its multivariate distributions. For $n = 1, 2, \ldots$, and points $x_1, \ldots, x_n$ in the exploration region, let $x = (x_1,\ldots,x_n)$, $L^n(x) = (L(x_1),\ldots,L(x_n))$, $\mu^n(x) = (\mu(x_1),\ldots,\mu(x_n))$, and $\Gamma^n(x)$ be the $n \times n$ matrix whose (i,j) entry is $\gamma(x_i,x_j)$. Here, μ and γ are defined above as estimates based on prior physical information. Then $L^n(x)$ is an n-dimensional random variable. We assume that it has a multivariate normal distribution with mean $\mu^n(x)$ and covariance matrix $\Gamma^n(x)$. Thus specified, L is a spatial-parameter Gauss-Markov process defined on the exploration region.

For computation, we have the following: For $n = 1, 2, \ldots$, if A is a (measurable) subset of n-space, then the probability that $L^n(x)$ is in A is (a vector is a row matrix)

$$(1) \qquad \int_A [(2\pi)^n \det\Gamma^n(x)]^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(a-\mu^n(x))\ \Gamma^n(x)^{-1}(a-\mu^n(x))'\}da.$$

The integrand is, of course, the probability density of $L^n(x)$ at a.

Now suppose that for $k = 1, \ldots, m$, an FFG probe is made at the point y_k, as a result of which it is determined that $l(y_k) = b_k$; denote $b = (b_1,\ldots,b_m)$ and $l^m(y) = b$. For each point z in the exploration region, we wish to form a new estimate of $l(z)$, given this new information. For this estimate, we take the mean, call it $\mu^*(z)$, of the conditional distribution of $L(z)$ given the information $l^m(y) = b$. From 3.4 of Morrison [1967], this is given by

$$(2) \qquad \mu^*(z) = \mu(z) + [\gamma(z,y_1), \ldots, \gamma(z,y_m)]\ \Gamma^m(y)^{-1}\ [b-\mu(y)]'.$$

By equation (2), $\mu^*(z)$ may be computed for various points z of interest throughout the exploration region. The μ^* values thus obtained may be used as part of the basis for choosing the probe points in the next phase of exploration, in the same way that μ was used initially. This new probing provides new information $1^m(y^*) = b^*$, where y^* and b^* are m^*-tuples of points and numbers with $m^* > m$ and $y_k = y_k^*$ and $b_k = b_k^*$ for $k = 1,\ldots,m$, i.e., the previous observation information is, of course, retained. We again apply equation (2) with y^* and b^* substituted for y and b. The procedure is repeatable further.

As an example of the possible usefulness of the foregoing, consider part A of Figure 4. The author's present knowledge of the reasoning that was employed in that exploration example is confined to this page excerpted from Metallgesellschaft [1975]. The reasoning appears to be excellent and evidently succeeded in defining high-yield patches in the center of the region. However, questions probably remain as to which additional subregions should be probed further, e.g., the far right or left upper portions of part A. It may not be obvious that they should or should not be probed further. The methods above could provide a partial basis for answering these questions and for comparative evaluation with numerous alternative probe conditions that are present.

While the above approach appears to be attractive as a model and computation procedure, several observations are in order, primarily as to needs for further information and analysis:

(1) The most important need is for information on the realism of the Gauss-Markov model of yield density of nodule fields, especially on the spatial correlation structure. One should either confirm the plausibility of this model or find a better one.

(2) Analysis of spacing of FFG probes is needed, using in part current μ values at each stage, to develop systematically the identification of high-yield density patches and/or regions of some 25 km in extent for exploration by towed fish.

(3) Adaptive analysis methods in the towed fish phases are needed-- these would differ from the above methods in that the "new information" would be swept paths of good but uncertain data compared to discretely spaced near-certain data from FFGs. Perhaps, this analysis would be closer to traditional search theory.

One should borrow from search theory by employing Reber's methods [1956] of spacing parallel sweeps to

achieve a prescribed thoroughness of coverage under a given navigation accuracy. Note that for this purpose, the effective sweep width is the average diameter of the search targets' high-yield patches of nodules, rather than the much smaller sweep width of the sensor.

(4) Computation of equation (2) for a given z is dominated by inversion of $\Gamma^m(y)$; the number of multiplications required is, in general, of the order of $2m^3$, which should be no great problem if $m \leq 100$. After a given sequence of observations, the inversion is computed once and the inverse multiplied by $b - \mu^m(y)$, independent of z, but the rest of the formula must be computed for each point z under consideration. When y, the sequence of probe points, is enlarged, the previous evaluation of $\Gamma^m(y)^{-1}$ is utilized to advantage in inverting the new enlarged matrix.

(5) Probably, Bayesian estimation of $\mu(z)$, for each candidate point z, is in order. This requires better knowledge of the physical processes, postulation of a prior distribution of $\mu(z)$, and probably a more complicated updating analysis than the above. A similar comment applies to $\sigma(z)$ and α.

(6) In making revisions in response to observations, one probably also wants the revised variance of L(z), as well as $\mu^*(z)$, which will then give the complete distribution of L(z), conditioned on the results of probing. This is included in the following full description of the revised characterization of L. From 3.4 of Morrison [1967] for x an n-tuple of points, the conditional distribution of $L^n(x)$ given $1^m(y) = b$ is again multivariate normal with mean

$$\mu^{n*}(x) = \mu^n(x) + \Gamma^n(x|y)\ \Gamma^m(y)^{-1}\ [b-\mu^m(y)]$$

(generalizing equation (2)) and covariance matrix

$$\Gamma^{n*}(x) = \Gamma^n(x) - \Gamma(x|y)\ \Gamma^m(y)^{-1}\ \Gamma(x|y)',$$

where

$$\Gamma(x|y) = \begin{pmatrix} \gamma(x_1,y_1), & \ldots, & \gamma(x_1,y_m) \\ \gamma(x_n,y_1), & \ldots, & \gamma(x_n,y_m) \end{pmatrix}.$$

Acknowledgement

Background information on nodule exploration and mining was very informatively supplied by Professor Günther Friedrich and Dr. W. L. Plüger, Institut für Mineralogie und Lagerstättenlehre, Aachen University, and M. Jean-Pierre Lenoble, Chef du Département "Ressources Minérales," Centre Nationale pour l'Exploitation des Océans, Paris. The author's colleagues, Dr. Bernard J. McCabe and Dr. Thomas L. Corwin, were very helpful in formulating the use of a Gauss-Markov model in the final section.

References

Archer, A. A. [1976] The Prospects for the Exploitation of Manganese Nodules: The Main Technical, Economic and Legal Problems, Inst. Geol. Sci. IGS 1589 200.

Holster, A. F. [1976] Manganese Nodule Resources and Mine Site Availability, Prof. Staff Study Ocean Mining Administration.

Mero, J. L. [1965] "The Mineral Resources of the Sea," Elsevier, New York.

Metallgesellshaft [1975] Manganese Nodules--Metals from the Sea, Review of the Activities, Edition 18.

Moore, R. [1973] Identification of Technical Gaps in Exploration of Marine Ferromanganese Deposits, Sea Grant Advisory Report WIS-SG-T3-404, Sea Grant Commission Office, Madison, Wisconsin, USA.

Morrison, D. F. [1967] "Multivariate Statistical Methods," McGraw-Hill, New York.

Reber, R. K. [1956] A Theoretical Evaluation of Various Search/Salvage Procedures for Use With Narrow Path Locators, Part I: Area and Channel Searching, Bureau of Ships, Minesweeping Branch Technical Report No. 117.

AN APPLICATION OF SEARCH THEORY IN MEDICINE:

THE DETECTION OF GLAUCOMA

Peter Kolesar

Columbia University

There is substantial medical interest in improving screening and monitoring of large populations for the presence of diseases such as cancer, hypertension, and glaucoma for which prognosis is much better with early detection. Operations research models can play a role in the design and evaluation of screening programs (Klein and Kirch [1974], Nichols and Weinstein [1978], Rabin and Kolesar [1978], and Schwartz [1978]). The nature of these problems suggests that the concepts or methods of search and screening theory might be useful, and we illustrate this with an application of search theory to the detection of glaucoma.

Glaucoma is a disease of the eye in which the sufferer gradually and usually painlessly loses vision. The afflicted person is often unaware of the disease until a very substantial and irremedial amount of vision has been lost. It is generally believed that glaucoma is caused by elevated pressure of the clear fluid that fills the eye. This pressure, it is conjectured, interferes with the local supply of blood to the optic nerve. Portions of the nerve cease to function and the sufferer loses vision at specific areas called scotoma. A simple test will detect the high pressure, but since vision loss occurs unpredictably and since all known treatments have possibly serious side effects, it is not standard practice to treat for high pressure until actual vision loss occurs. At best, the treatments arrest the progress of the disease. Hence, there is keen medical interest in the important problem of early detection of vision loss itself.

Vision loss may be detected via a visual field examination in which the subject fixes vision on a particular point on a distant screen. Other light spots are sequentially presented to the

subject in order to map out where the subject has vision. A normal subject will have only one small blind spot located on the temporal side of the eye about 15^{o} out from the center of fixation. Most of us are completely unaware of this blind spot which occurs where the optic nerve connects to the eye. In the glaucoma sufferer, other blind spots appear and are also unknown to the sufferer until vision is almost totally lost. They will be detected if the subject fails to respond to a light stimulus that falls on the afflicted area. By moving the stimulus probe around in the field of vision, one searches for blind spots.

Such a test can be very time-consuming and, therefore, trying on the patient and uneconomical for a medical center to employ on the many subjects who should be tested and retested frequently. In this paper, we discuss an optimization approach to specifying efficient testing sequences.

Problem Formulation

In the spirit of much search theory, let us discretize the problem by specifying a grid of candidate test points $C = \{j:j = 1,2,\ldots,N\}$ from which we wish to select an optimal subset of size $T << N$ which will maximize the probability of detecting at least one defect (scotoma). This is an appropriate criterion since although there may be more than one defect or several test points may be contained within the same scotoma, if we make one hit our detection phase will be a success. We would then pass on to a thorough diagnostic and exploratory examination. A patient for whom there was no hit would be passed as normal--at least until the next examination.

Because the scotoma are often quite large in area compared to the size of both the probe (test spot) and to the intergrid spacing, the events, vision, or not at neighboring points are highly dependent. Data are not available to estimate the correlation structure of this dependency, so instead we take a convenient but somewhat fraudulently deterministic approach to the problem. Our approach could be viewed as a simulation somewhat similar in spirit to the use that CASP makes of its scenarios (see the paper by Richardson and Corwin and the SAR working group report in this volume).

In order to formulate the problem of selection of an optimal subset of the N candidate test points, imagine that we have a sample of M eyes of early glaucoma sufferers and we test them at each of the N candidate test points. We can define an early glaucoma sufferer as a person who has high intraocular fluid pressure, has had a visual field exam that showed no vision loss, and then on a follow-up exam, say within six months thereafter,

showed vision loss at one or more points. Represent the data by

$$d_{ij} = \begin{cases} 1 & \text{if eye } i \text{ has a vision defect at point } j, \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1,2,\ldots,M,\ j = 1,2,\ldots,N.$$

We represent the decision of where to test in the future by the variables

$$x_j = \begin{cases} 1 & \text{if we test at point } j, \\ 0 & \text{if not} \end{cases} \quad \text{for } j = 1,2,\ldots,N.$$

Notice that

$$1 - x_j d_{ij} = \begin{cases} 1 & \text{if we miss a defect or if there is no defect at point } j \text{ for eye } i, \\ 0 & \text{otherwise,} \end{cases}$$

and that

$$\prod_{j=1}^{N} (1-x_j d_{ij}) = \begin{cases} 1 & \text{if we miss all defects for eye } i \text{ or if eye } i \text{ has none,} \\ 0 & \text{otherwise.} \end{cases}$$

Our problem can be stated as finding a set of T test points that will miss as few defective eyes as possible. Namely,

find $x_1, x_2, \ldots, x_N$ to

minimize

$$\sum_{i=1}^{M} \prod_{j=1}^{N} (1-x_j d_{ij})$$

subject to

$$\sum_{j=1}^{N} x_j = T,$$

$$x_j = 0 \text{ or } 1 \text{ for } j = 1,2,\ldots,N.$$

This nonlinear integer optimization problem is as difficult to solve as it is nonseparable. Incrementally optimal search plans are not necessarily totally optimal. For contrast with classical

problems, see Chapter IV of Stone [1975]. An efficient heuristic has been developed which, by finding alternative optima, has fortunately enabled us to generate totally optimal plans in all actual applications. The problem falls into a class of optimization problems that have been called submodular set function problems. Little is known about efficient methods of solving these problems. Algorithms that are optimal from a worst case point of view have been found by Nemhauser and his colleagues (Nemhauser and Wolsey [1978]). Since the heuristic cannot demonstrate optimality, the following linear formulation is of interest: Let

$$y_i = \begin{cases} 1 & \text{if eye } i \text{ is detected,} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 1,2,\ldots,M.$$

Find x_j, $j = 1,2,\ldots,N$, and y_i, $i = 1,2,\ldots,M$, to

maximize

$$\sum_{i=1}^{M} y_i$$

subject to

$$\sum_{j=1}^{N} y_i \leq \sum_{j=1}^{N} d_{ij}x_j, \quad i = 1,2,\ldots,M,$$

$$\sum x_j = T,$$

$$x_j \text{ and } y_i = 0 \text{ or } 1.$$

Results

This problem formulation was applied to a retrospective simulation of such data. My collaborator, S. Rabin, M.D., obtained and interpreted plots of visual field examinations of appropriate glaucoma patients, and by simulating the testing procedure he generated a $\{d_{ij}\}$ matrix. The results of the optimization produced very striking results. Where it has been common to use 72 test points (Rock, et.al. [1971]), and some researchers have commented that thousands of points might be necessary (Greve [1973]), our data indicate that only ten points, when optimally selected and sequenced, may yield a 95% detection probability. We are currently validating these findings with other independent retrospective data and with data from clinical tests on actual patients. The results of these experiments are not all in yet, but to date they appear very encouraging.

Certainly, the cost to either patient or doctor of the visual field exam is not a linear function of the number of test points. Yet, many medical experts feel that if the test can be really shortened without substantial loss in reliability, it will be more frequently and more broadly applied. This could well be achieved by a combination of computer-assisted testing used together with optimal sequences of the test points (Heijl [1977]). If done properly, this could result in preserving substantial amounts of vision that would otherwise be lost. Work on the timing of these mass screenings is now underway.

Conclusion

The glaucoma detection problem, although similar in spirit to classical search problems, is different in ways that bear significantly on the methods of solution. Instead of a single small lost object, we search for one of possibly many large objects. Moreover, the search is with a probe of very small sweep width--small as compared to the size of the objects. These differences lead to a mathematical formulation that does not have a separable objective function and, hence, is intrinsically harder to solve than the classical case. Moreover, it is not necessarily true that uniformly optimal search plans exist for the glaucoma problem.

The discussions in the ARI working groups on exploration suggested that petroleum exploration shares some characteristics of the glaucoma problem (for background, see Cozzolino [1972]). This is true--at least superficially--and I am at present working on a problem formulation of a drill site location problem along these lines. The fact that interesting medical problems and resource exploration problems share these characteristics will, I hope, motivate search theorists to work on these problems.

Acknowledgements

My medical collaborator, S. Rabin, M.D., has contributed greatly to this work. Our joint efforts are reported in some detail in Rabin and Kolesar [1978]. This research was funded in part by a National Library of Medicine Training Grant 1T15LM07008BLR and by a National Eye Institute Grant IR01EY02836-01 to Mt. Sinai Medical Center, New York City. This manuscript was prepared while the author was a Resident Visitor at the Operations Research Center, Bell Laboratories.

Nearly 20 years ago, while a graduate student at Columbia, I was fortunate to take a course called "Mathematical Methods of Operations Research." The course was offered by Bernard Koopman and he might well have called it "Mathematical Methods of Search Theory" for that was its main content. Although it was almost 20

years before I got to use any search theory, I have always relied on Bernard Koopman's model of how a scientist should think. It is a great pleasure to acknowledge here his deep influence on my approach to problems--be they search or anything else.

References

Cozzolino, J. M. [1972] Sequential Search for an Unknown Number of Objects of Non-Uniform Size, Oper. Res. 20, pp. 293-308.

Greve, E. L. [1973] "Single and Multiple Stimulus Static Perimetry in Glaucoma: The Two Phases of Perimetry," B. V. Junk Publishers, the Hague.

Heijl, A. [1977] Studies on Computerized Perimetry, Acta Ophthalmologica. Supplementum 132, pp. 1-42.

Klein, M. & Kirch, R. L. A. [1974] Surveillance Schedules for Medical Examinations, Management Sci. 20, pp. 1403-1407.

Nemhauser, G. L. & Wolsey, L. A. [1978] Best Algorithms for Approximating the Maximum of a Submodular Set Function, Math. Oper. Res. 3, pp. 177-188.

Nichols, A. & Weinstein, M. [1978] Optimal Resource Allocations in Community Hypertension Programs, Management Sci. 24, pp. 1526-1537.

Rabin, S. & Kolesar, P. [1978] Mathematical Optimization of Glaucoma Visual Field Screening Protocols, Documenta Opthalmologica. 45-2, pp. 361-380.

Rock, W. J., Drance, S. M. & Morgan, R. W. [1971] Modifications of the Armaly Visual Field Screening Technique for Glaucoma, Can. J. Ophthalmology. pp. 283-292.

Schwartz, M. [1978] Mathematical Model for Generalized Breast Cancer Screening, Oper. Res. 26, pp. 937-955.

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

INDUSTRIAL APPLICATIONS OF SEARCH THEORY

J. B. Kadane

Carnegie-Mellon University

"Search theory" means different things to different groups of authors. In fact, the 1977 Science Citation Index reveals about 2,900 articles having "search" in their titles, of which only a few appear to be the sort of search theory to be discussed in this conference. Consequently, the kinds of search theory whose applications will be considered in this paper are now discussed.

Only applications of two discrete models of search theory are given. In the first model, there is a single object hidden in one of n boxes. The kth box contains the object with probability p_k, ($\Sigma\ p_k = 1$) has overlook probability α_k, and cost of search c_k. A strategy to find it is a sequence indicating which box is to be searched first, which to be searched second if the first was unsuccessful, etc. A typical result in this theory is that the expected cost of the search is minimized if the jth search of box k is in place i, if among the numbers $p_k\alpha_k^{j-1}\ (1-\alpha_k)/c_k$, the (j,k)th is the ith largest (Black [1965], Blackwell, see Matula [1964]).

For the second model, there can be any number of objects from zero to n; each box k has an object with probability p_k independently of whether the others do. This is the difference between the two models: in the first model, the events E_k that box k has an object were disjoint and exhaustive (one and only one obtains), where in the second model they are independent. In the second model, again, the kth box can be taken to have overlook probability α_k and cost of search c_k. A search strategy takes the same form as before. One minimizes the expected cost of finding an object by putting the jth search of box k in place i if among the numbers $[\{p_k\alpha_k^{j-1}(1-\alpha_k)\}^{-1}-1]^{-1}/c_k$, the (j,k)th is the ith largest (Kadane [1969]).

The purpose of this paper is to help readers recognize applications of these two models and to describe some applications that have already been made.

Recognizing a Search Problem

These problems will appear in many guises in the next sections. The following guide is intended to help the reader recognize whether an applied problem is one of these two search problems.

Suppose in the first model that it is decided to search first box 1, then box 2, if it is not found in box 1, then box 3, if it has not yet been found, ... The expected cost of this search strategy is

$$(1) \quad X(1,2,3,\ldots) = p_1 c_1 + p_2(c_1+c_2) + p_3(c_1+c_2+c_3) + \ldots,$$

where here p_i is the probability that the search of box i is successful (for simplicity, think of the overlook probability $\alpha_i = 0$ for all boxes i, so that looking in the right box will certainly find the object). Also in (1) payment is made whether or not the object is found.

Now suppose the same strategy is used in the second problem. Then the expected cost is

$$(2) \quad V(1,2,3,\ldots) = q_1 c_1 + q_1 q_2 c_2 + q_1 q_2 q_3 c_3 + \ldots,$$

where again for simplicity take the overlook probabilities to be zero. Here, payment for a search is only made if it was unsuccessful and $q_i = 1-p_i$. Equations (1) and (2) are fundamental to the comparison suggested. A problem obeying equation (1) is a search problem of the first type; a problem obeying equation (2) is a search problem of the second type. This is true even if the problem comes dressed in very different "clothes" which makes it not seem like a search problem at all.

Parenthetically, from equation (1) and (2) it is easy to see how the optimal strategies are derived. Suppose the strategy (1,2,...,n) is compared with the one where the ith and (i+1)st boxes are interchanged. Then the difference in expected cost of search by these two strategies is

$$X(1,2,\ldots,i-1,i,i+1,i+2,\ldots,n) - X(1,2,\ldots,i-1,i+1,i,i+2,\ldots,n)$$

$$= p_i(c_1+\ldots+c_{i-1}+c_i) + p_{i+1}(c_1+c_2+\ldots+c_{i-1}+c_i+c_{i+1})$$

$$- p_{i+1}(c_1+\ldots+c_{i-1}+c_{i+1}) - p_i(c_1+c_2+\ldots+c_{i-1}+c_{i+1}+c_i)$$

$$= p_ic_i + p_{i+1}(c_i+c_{i+1}) - p_{i+1}c_{i+1}-p_i(c_{i+1}+c_i)$$

$$= p_{i+1}c_i-p_ic_{i+1} = c_ic_{i+1}(p_{i+1}/c_{i+1}-p_i/c_i).$$

Hence, the strategy $(1,2,\ldots,i-1,i,i+2,\ldots,n)$ has smaller expected cost if and only if $p_i/c_i > p_{i+1}/c_{i+1}$. Hence, ordering the boxes by the function (p/c) yields the optimal strategy. Introduction of overlook probabilities does not change the algebra; the only change necessary is to think of each search of each box as contributing a summand to the expected cost.

Similarly, for the second model of search theory,

$$V(1,2,\ldots,i-1,i,i+1,i+2,\ldots,n) - V(1,2,\ldots,i-1,i+1,i,i+2,\ldots,n)$$

$$= (q_1,\ldots,q_{i-1}) [\, q_ic_i + q_iq_{i+1}c_{i+1} - q_{i+1}c_{i+1} - q_{i+1}q_ic_i \,]$$

$$= (q_1,\ldots,q_{i-1}) [\, p_{i+1}q_ic_i - p_iq_{i+1}c_{i+1} \,]$$

$$= (q_1,\ldots,q_{i-1}) \times [\, p_{i+1}/\{q_{i+1}c_{i+1}\} - p_i/\{q_ic_i\}] \times (q_iq_{i+1}c_ic_{i+1}).$$

Hence, optimally, one searches that box for which p/qc is largest or, conversely, for which qc/p is smallest. The similarity in the solutions (3) and (4) has been of interest to the author and is explored more fully in Kadane [1978]. The full modern theory of these problems allows for constraints on the order in which boxes can be searched. So far, the optimal strategies have been explored in the context of constraints in which certain searches of certain boxes must be completed before others can begin, Kadane and Simon [1977]. However, it is not yet known, to the best of my knowledge, how to handle constraints of the type "to search box i, you must immediately before search box k or box l or box m." This type of constraint, very natural in a special context of contiguous boxes, imposes very difficult mathematical problems.

Equations (1) and (2) are important in determining the type of problems.

Industrial Quality-Control Testing

Suppose an item is subject to n different tests, all of which it must pass to be acceptable. Suppose these tests are characterized by two constants each: C_j, the cost per item of

conducting the jth test, and R_j, the probability of rejecting the item in the jth test. The tests may be run in any order, and the constants do not depend on the order. Specifically, this means that success on the different tests are probabilistically independent. An item is subjected to each test in the sequence as long as the tests accept the item; if the item is rejected by a test, no further tests in the sequence are performed. The expected cost of testing an item, if the tests are performed in the sequence 1,2,3,..., is

$$C_1 + (1-R_1)C_2 + \cdots,$$

which is seen to be in the form (2) with

$$C_i = q_i c_i \quad \text{and} \quad R_i = p_i.$$

Then the optimal strategy is to choose tests according to the function

$$\frac{qc}{p} = \frac{(1-R_j)\ (C_j/(1-R_j))}{R_j} = C_j/R_j,$$

smallest first. This application is explored by Mitten [1960].

One might not have thought of this problem as having any relationship to search theory at all. The key to understanding the connection is that if the item is good, the full price $C_1+C_2+\cdots+C_n$ will be paid for testing the item. If it is defective, however, one saves cost by detecting it as cheaply and quickly as possible. Hence, one is searching for the defect and chooses those tests to do first which are least costly (smallest C_j) or most likely to reveal a defect (largest R_j).

Discovering a Defective Component

Suppose that a machine breaks because one or more of n components are defective. If each component j has a Poisson distribution of failures with parameter λ_j, the probability that it will fail in a time interval Δt is $\lambda_j \Delta t$. Assume that the failure times of components are independent of each other. The probability of two components failing in a given time period goes to zero as Δt goes to zero. Hence, given that failure has occurred in some time period, the probability that, in fact, it is component j that has failed is

$$\gamma_j = \lambda_j \ /\{\lambda_1 + \lambda_2 + \cdots + \lambda_n\}.$$

Suppose each component j costs β_j to test. Then the expected cost of finding the failed components if they are tested in the order 1,2,3,..., is

$$\gamma_1\beta_1 + \gamma_2\{\beta_1 + \beta_2\} + \ldots,$$

which has the form of (1). Hence, the optimal strategy is to test components in the order (γ/β), largest first. This problem was studied by Denby [1967].

Parenthetically, Denby suggests evaluating all n! permutations because he did not know the mathematics we now do!

Managing a Research and Development Project

Often in structuring a research and development project, several distinct parts of the technology have to "go right" in order for the project to be successful. Suppose that these are n in number, and that the jth subproject costs k_j to investigate and has probability s_j of success. Suppose that these subprojects can be done in any order, and that success on any subproject is independent of success on any other. For the project to succeed, all the subprojects must succeed. In this case, the expected cost of success is

$$k_1 + s_1k_2 + s_1s_2k_3 + \ldots,$$

which has the form of equation (2) with $k_i = q_ic_i$ and $s_i = q_i$. Then the optimal strategy is first to try those subprojects for which $k_i/(1-s_i)$ is smallest. This problem was studied by Joyce [1971].

Note that this strategy involves trying the subprojects first which are least likely to succeed (if costs are equal). Joyce points out that this is not what ordinary managers of R and D projects tend to do, since if they did they could well find out sooner that their project is destined to fail with possible negative influence on their careers. Nevertheless, for the organization as a whole, the strategy above would be optimal.

As in the case of the quality control example, in this problem one essentially searches for the first failure and hopes that it comes quickly. If the R and D project is successful, then the full cost $k_1 + k_2 + \ldots + k_n$ will be paid, no matter what strategy is taken. Hence, the optimal strategy kills off early projects sure to be unsuccessful in the end.

Scheduling Oil Well Workovers

In this problem, there are n oil wells which could benefit from "workover" by an oil rig. Each well has two parameters associated with it: t, the number of days required to perform the well workover and p, the expected increase in productivity in barrels per day resulting from the workover. The loss of a particular schedule is the number of barrels of oil not recovered because the workover has not yet occurred in that well. Then the loss from the assignment 1,2,..., is

$$p_1t_1 + p_2(t_1 + t_2) + \ldots \quad .$$

Again, this is in the form of (1), and the optimal strategy is to workover wells in the order p/t, largest first. This problem is studied by Barnes, Brennan, and Knapp [1977], who are especially concerned about the problem when m rigs are available, for which only approximate solutions have been found to date.

References

Barnes, J. W., Brennan, J. J. & Knapp, R. M. [1977] Scheduling a Backlog of Oil Well Workovers, J. Petroleum Tech. pp. 161-165.

Black, W. [1965] Discrete Sequential Search, Inform. and Control. 8, pp. 156-162.

Denby, D. C. [1967] Minimum Downtime as a Function of Reliability and Priority in a Component Repair, J. Indust. Eng. 18, pp. 436-439.

Joyce, W. B. [1971] Organization of Unsuccessful R & D Programs, IEEE Trans. Engr. Management. 18, pp. 57-65.

Kadane, J. B. [1969] Quiz Show Problem, J. Math. Anal. Appl. 27, pp. 609-623.

Kadane, J. B. [1978] A Characterization of the Rau Class of Sequential Problems, Math. Oper. Res. 3, pp. 42-56.

Kadane, J. B. & Simon, H. A. [1977] Optimal Strategies for a Class of Constrained Problems, Ann. Statist. 5, pp. 237-255.

Matula, D. [1964] A Periodic Optimal Search, Amer. Math. Monthly. 71, pp. 15-21.

Mitten, L. G. [1960] An Analytical Solution to the Least Cost Testing Sequence Problem, J. Indust. Eng. 11, p. 17.

Smith, W. E. [1956] Various Optimizers for Single-Stage Production, Naval Res. Logist. Quart. 3, pp. 59-66.

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

THE SEARCH FOR COSMOS-954

R.L. Grasty

Geological Survey of Canada

On January 24th, 1978, COSMOS-954, a nuclear powered Russian satellite, disintegrated on reentering the earth's atmosphere and scattered radioactive debris over a large area of Canada's Northwest Territories. The U.S. Government offered the Canadian Government the assistance of their Department of Energy (DOE) Nuclear Emergency Search Team (NEST) in the search and recovery of any radioactive debris, and that same evening approximately 120 NEST scientists and technicians arrived in Edmonton, the base of the search operations. The Canadian Department of National Defense (DND) was assigned the lead role in the search operation and the Atomic Energy Control Board (AECB), the responsibility for the recovery of any radioactive debris. The following day the Geological Survey of Canada (GSC) gamma ray spectrometer (Bristow, 1978) was flown to Edmonton from Ottawa, and, subsequently, approximately 30 scientists and technicians from the AECB, Department of Energy, Mines, and Resources (EM&R), worked with the Canadian Forces and the U.S. NEST team in the search and recovery operations.

Based on computer reentry prediction, the probable impact zone was defined by trajectory and ballistic experts from the U.S. team and covered an area 800 km long by 50 km wide stretching from the east end of Great Slave Lake, northeast to Baker Lake near Hudson Bay. For logistic reasons, this initial search area was divided into 8 sectors (Figure 1). Later on in the search, the area was extended west and south to include sector 9 to 14. To cover this large area, four C-130 Hercules aircraft, with gamma ray detection equipment, carried out a reconnaissance type of survey by flying grid lines 1 nautical mile apart at an altitude of 500 m (1,500 feet) above ground level. The first of three U.S. gamma ray

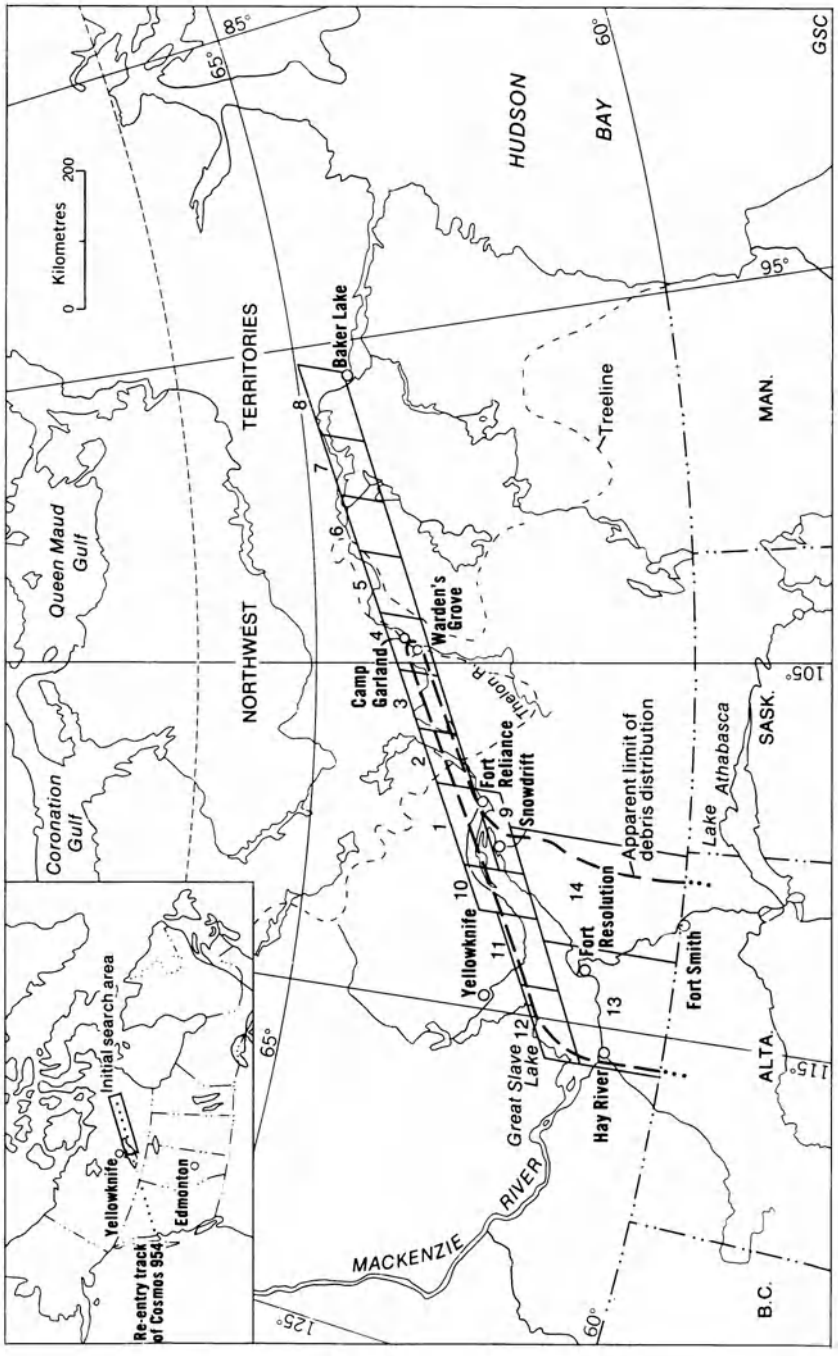


Figure 1. Cosmos-954 search area.

spectrometers was airborne early on the morning of January 25th. The Canadian system was in the air 24 hours later.

On the night of the 26/27th January, the Canadian spectrometer, while searching sector 1, detected a radioactive source on the ice at the east end of Great Slave Lake, 27 km north of Fort Reliance. Subsequent computer analysis of the data collected on magnetic tape confirmed the presence of gamma rays characteristic of the products of nuclear fission. The procedure used to detect the radioactive debris is described in the appendix.

On January 28, a radio message was received from six men camping for the winter at Warden Grove in sector 4 (Figure 1) that two of their group had seen a metal object, which later proved to be part of the satellite imbedded in a refrozen crater on the ice of the Thelon River. Subsequent analysis of the refrozen ice and snow showed the presence of lithium, an element commonly used as a neutron shield. The lithium had apparently reacted violently with the ice and snow, producing sufficient heat to form the crater. Overflying Hercules aircraft soon discovered several more radioactive anomalies in the same area, which later proved to be of geological origin. For logistics reasons, a winter camp was set up nearby at Camp Garland (Figure 1) so that these sources of radioactivity could be investigated by the Canadian Forces Nuclear Accident Support Team (NAST) operating under the supervision of the AECB.

By January 31st, the Hercules search aircraft had completed the general search of the area, which had been extended to include sectors 9 and 10, and all radioactive fragments from the satellite had fallen within a narrow ten kilometer wide strip which extended part way along the center of the predicted reentry path. Navigation, however, had been a major problem the initial search being carried out using a combination of visual navigation, doppler, and omega. Much of the area was north of the tree line, with few reference points (Figure 1), and considerable difficulty was found in pinpointing the location of radioactive debris identified by the overflying Hercules aircraft. In order to provide accurate and systematic coverage of the narrow band where the radioactive fragments had been found, two microwave ranging systems (MRS) were employed. The MRS uses two portable land beacons positioned 20 to 50 km apart which communicate automatically with units in the aircraft to enable the pilot to fly specific paths. Detailed coverage of this redefined narrow search area was flown by one U.S. system and the Canadian system at a line spacing of 500 m (1,500 feet) at a nominal terrain clearance of 250 m (750 feet).

On February 10th, a helicopter with detection equipment on board discovered many radioactive sources on the ice near Snowdrift in sector 9. These sources were too weak to be detected at the altitude flown by the Hercules. Later analysis of these minute pieces showed them to be part of the missing reactor core. In the next few days, other low flying helicopter flights detected many more radioactive particles along the south shore of Great Slave Lake. The reactor core had apparently disintegrated on entering the earth's atmosphere and minute pieces of the core, carried in the northerly winds of the morning of January 24th, had drifted southward to dust a large area south of the satellite track. The area extended from Hay River in the west to a line from Fort Smith to Fort Reliance in the east (Figure 1). During the early part of March, communities on the south shore of Great Slave Lake were surveyed by members of the NAST team and scientists from the AECB and numerous particles collected. At the end of March, a systematic gamma ray helicopter survey was carried out over the ice of Great Slave Lake by commercially built Canadian equipment which replaced the American spectrometers. Figure 2, taken from Grasty [1978], shows the distribution of radioactive debris on the ice and that the level of radioactivity decreases significantly at the south shore of the lake. Analysis of the data provide further evidence that the reactor core had completely disintegrated on entering the earth's atmosphere.

During the course of the summer of 1978, the AECB conducted a further monitoring and clean-up operation primarily to ensure that all hazardous material had been removed. Approximately 3,500 minute particles were collected from communities, including some as far as 300 miles south of the satellite track.

APPENDIX

The Detection of Radioactive Satellite Debris

It is a characteristic of fission products of a reactor that they emit gamma radiation with a higher proportion of low energy gamma rays compared to those emitted by naturally occurring radioactive isotopes. Gamma rays from the neutron activation of elements present in steel are also predominantly low in energy. Table 1 shows the major gamma rays emitted by fission and neutron activation products which are most likely to be observed soon after reactor shutdown. The half-lives of the various isotopes are also indicated. Figure 3 shows a typical natural gamma ray spectrum and a fission product spectrum for comparison. These spectra were obtained with the GSC gamma ray spectrometer (Bristow [1978]).

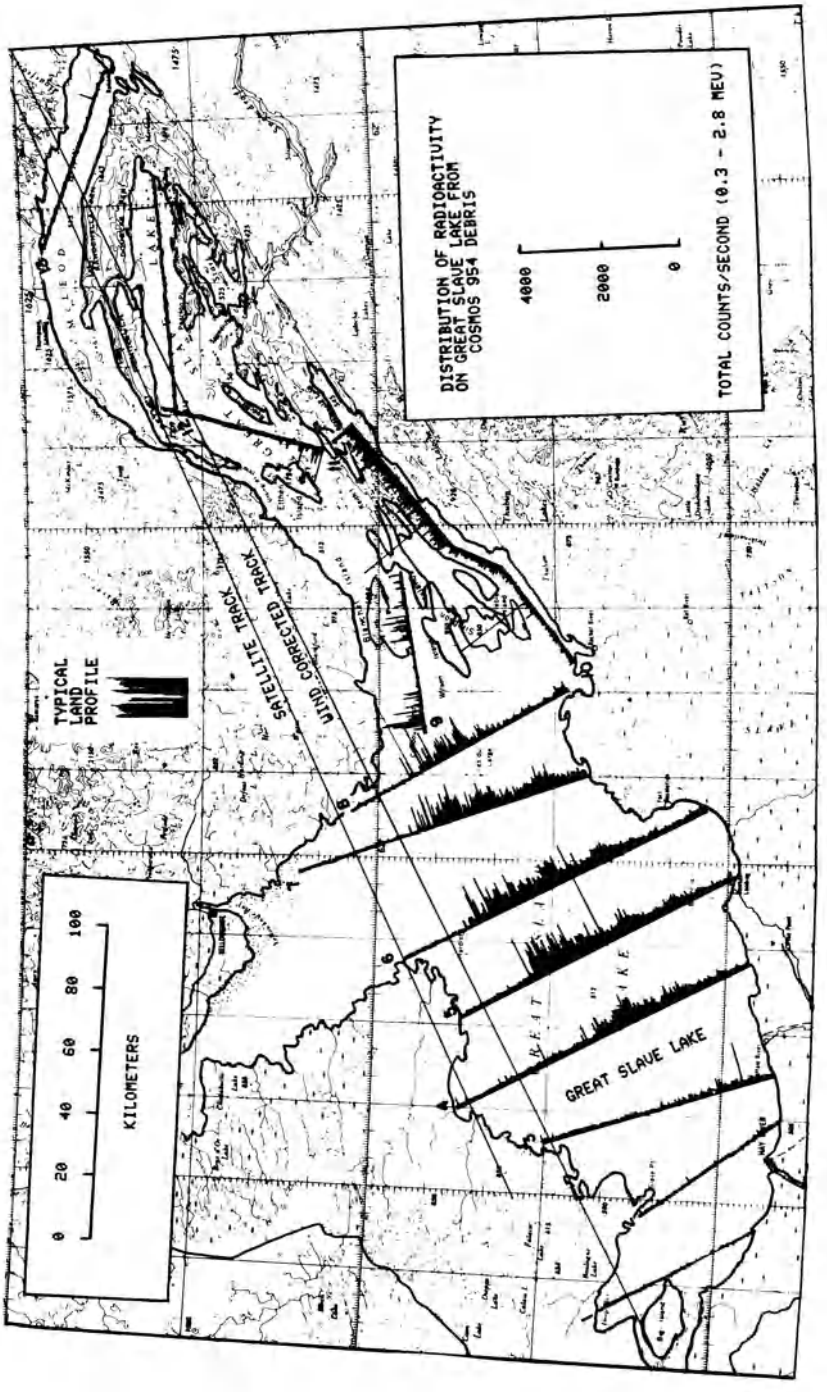


Figure 2, Distribution of radioactivity on Great Slave Lake from Cosmos-954 debris.

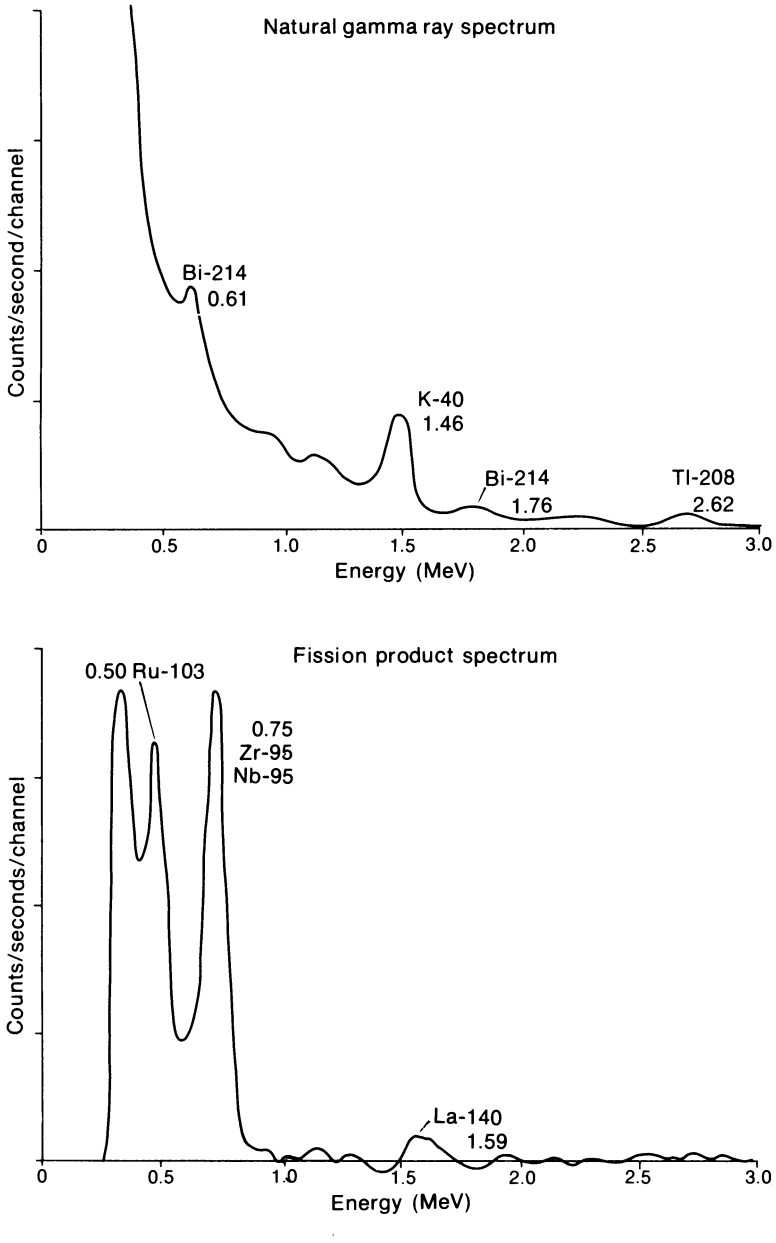


Figure 3. Typical natural and fission product gamma ray spectra.

Table 1. Major gamma ray emitting isotopes from fission and neutron activation.

Isotope	Gamma Ray Energy (keV)	Half-Life (Days)	Origin
Zr-95	724,756	65	Fission Product
Nb-95	765	35	Fission Product (daughter of Zr-95)
Ru-103	497	40	Fission Product
Co-58	810	71	Neutron Activation of Ni-58
Mn-54	835	300	Neutron Activation of Fe-54

Because of the high levels of radioactivity of rocks in the search area, mainly due to potassium-40 and decay products in the thorium series, a ratio technique was used to distinguish between natural and artificial sources of radiation. This ratio technique makes use of the fact that the radioactive satellite debris is dominated by low energy gamma rays. The ratio of all gamma rays detected with energies between 300 and 1,400 keV to those from 1,400 to 2,800 keV was found to remain relatively constant for all natural sources of radioactivity even while passing from water to land (Figure 4). The radioactivity over water arises from cosmic radiation, the radioactivity of the air, and the aircraft and its equipment (radium dials, etc.). Radioactive satellite debris is indicated by a significant increase in the low/high energy ratio (Figure 4). Additional confirmation is provided by the increase in total radioactivity. Figure 5 shows the total count profile of the first radioactive piece found at the end of Great Slave Lake. The gamma ray spectrum of the source, obtained by subtracting the spectrum from neighboring points, shows the predominance of low energy gamma radiation.

Bristow [1978] describes in more detail the procedures that were used to detect the radioactive satellite debris. He describes a ratio technique using a low energy region from 300 to 900 keV and high energy of 900 to 1,500 keV. However, for the GSC airborne system, this value did not remain constant when passing from water to land. This disadvantage was also observed for other systems flown in the operation. Further work is recommended to optimize the procedures used to detect man-made sources of radioactivity.

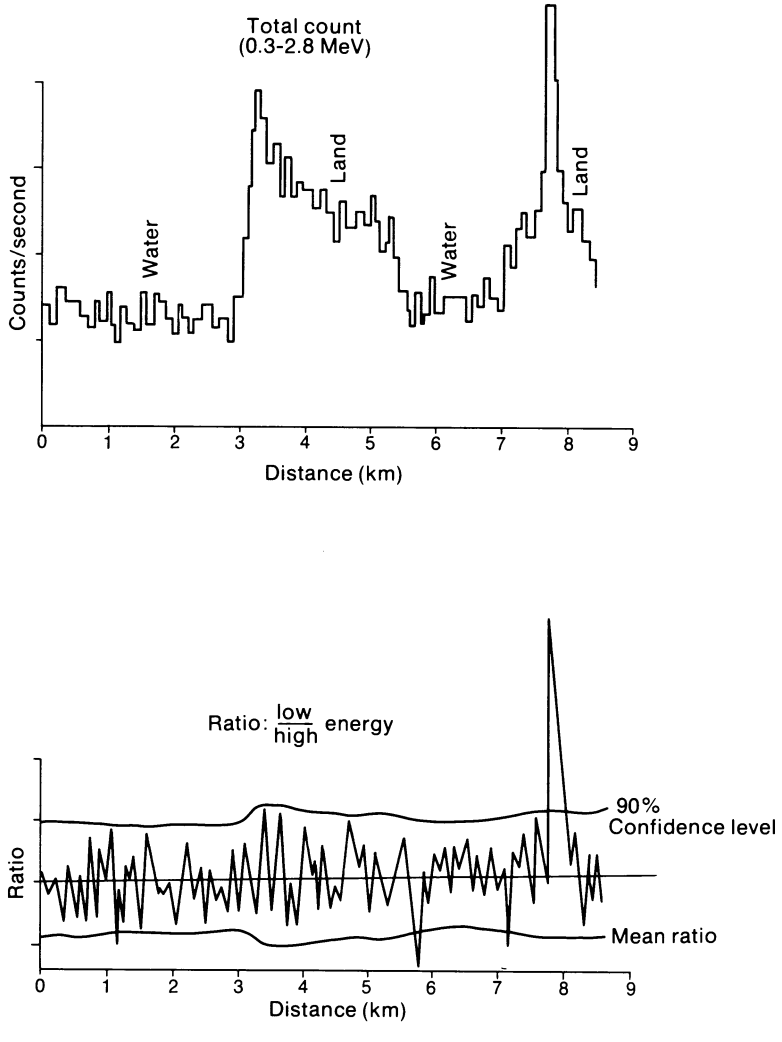


Figure 4. Example of procedure to detect radioactive satellite debris.

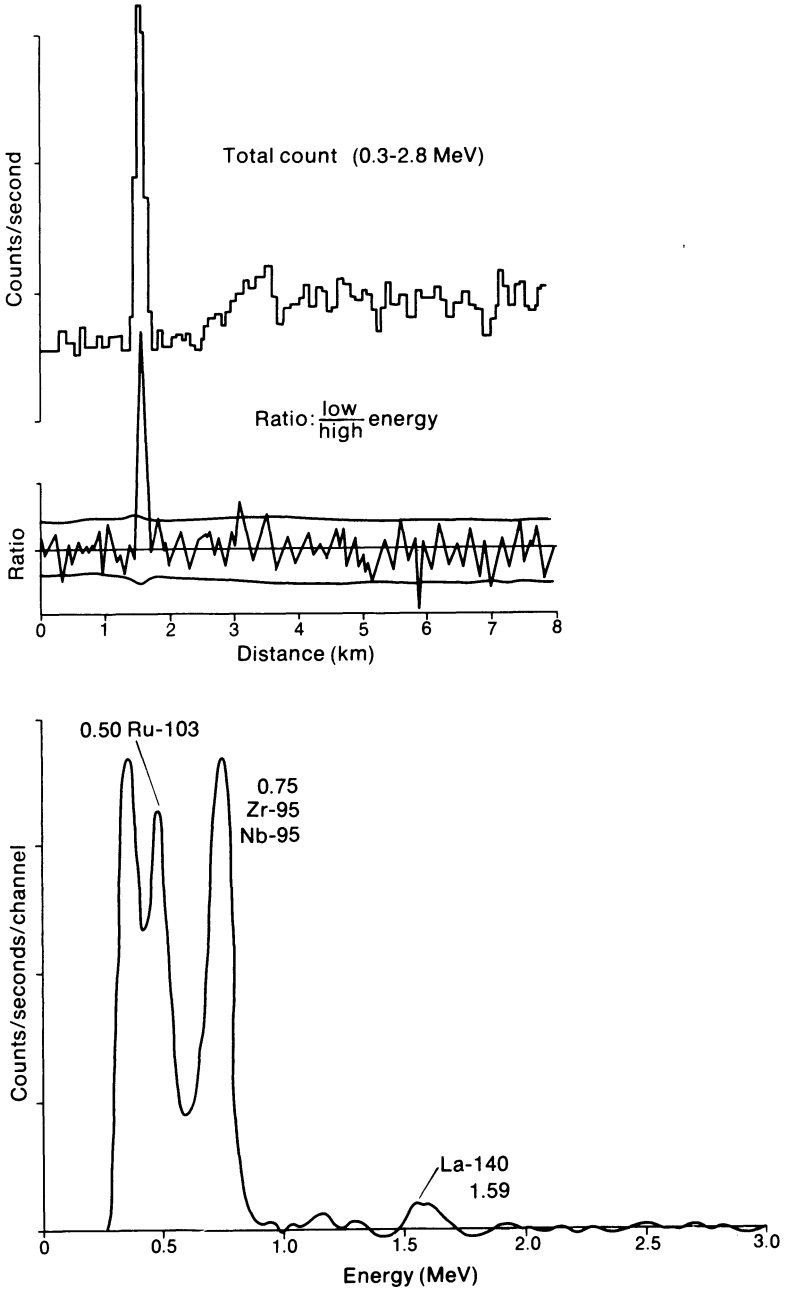


Figure 5. Total count profile and gamma ray spectrum of first confirmed debris from Cosmos-954.

Acknowledgements

I am grateful to Jalle Tammenmaa who has worked with me on the problems of identifying artificial sources of radioactivity. He has also carried out the computing work involved in this paper.

References

Bristow, Q. [1978] The Application of Airborne Gamma Ray Spectrometry in the Search for Radioactive Debris from the Russian Satellite COSMOS-954 (Operation Morninglight), Geol. Surv. Can. paper 78-1B.

Grasty, R. L. [1978] Estimating the Fallout on Great Slave Lake from COSMOS-954, Trans. Am. Nucl. Soc. Fall Meeting, Washington, Nov. 12-16.

OPTIMAL SEARCH FOR A MOVING TARGET IN DISCRETE TIME AND SPACE WITH AN EXPONENTIAL DETECTION FUNCTION

Scott Shorey Brown

Daniel H. Wagner, Associates

Consider search for a target which occupies one of a finite number of cells during each of a finite number of time intervals. We allocate a limited supply of search effort at each time interval to maximize the overall probability of detection. Call this an optimal search plan. Under the hypotheses stated below, a plan is optimal if and only if it cannot be improved by reallocating the search effort at any interval. The problem of optimally reallocating effort at interval i is equivalent to optimal search for a stationary target which occupies cell c with probability equal to the joint probability that the original target occupies cell c at time i and is not detected by the original search at another time interval.

An algorithm to compute optimal search plans by iteratively reallocating search at each time interval is stated. We show that the process converges. An example is given which shows rapid convergence, and other examples confirm that this is generally the case. We show how to implement the algorithm so that it can handle practical problems with moderate computation time.

Brown [1977] and [1980] describe these results and use them to investigate optimal search plans in a variety of examples. Also, they prove an existence and uniqueness theorem. Stone [1979] generalizes this characterization to all combinations of continuous and discrete space and time and to more general detection functions. Washburn [1979] generalizes the necessary conditions to an arbitrary detection function with independent glimpses at each time interval, states an algorithm to solve the necessary conditions, and gives an example where the conditions are not sufficient for optimality. Stewart [1979] used the

algorithm given below in a branch and bound algorithm for problems where the searcher's motion is restricted.

The Search Problem

Consider a target which occupies one of a finite set of cells, C, during each of n time intervals. The target follows path $b = (b_1, b_2, \ldots, b_n) \in C^n$ with probability p(b). We allow the possibility $\Sigma_b p(b) < 1$, but we assume that the target's path is unaffected by the search.

The searcher has L(i) units of search effort during the ith time interval which he may divide between the cells in arbitrary proportions. Units of search effort may be hours of search, area searched, or another measure of coverage. A search plan is a function $X: C \times [1,\ldots,n] \to [0,\infty)$ satisfying $\Sigma_c X(c,i) \leq L(i)$ for i = 1 to n with X(c,i) interpreted as the number of units of search put in cell c at time i.

We assume that the probability that the searcher will detect the target at time i given cell c contains the target at i equals $1-\exp(-w(c,i)X(c,i))$ where $w(c,i) > 0$ is a constant reflecting search conditions in cell c at time i. We postulate independent detection opportunities at each time interval so the probability that plan X fails to detect the target is

$$f(X) = \Sigma_b p(b) \prod_{i=1}^{n} \exp\{-w(b_i,i)X(b_i,i)\}.$$

Call the problem of finding a search plan X which minimizes f(X) the moving target problem.

Search for a Stationary Target

Call the special case of the moving target problem where n = 1 the stationary target problem. The stationary target problem has a unique solution which may be obtained as follows:

Algorithm 1

(1) Set X(c,1) = 0, for all c in C.

(2) For each c in C, compute the marginal payoff from taking plan X and adding effort to cell c, i.e., compute $w(c,1)\ p(c)\ \exp\{-w(c,1)\ X(c,1)\}$.

(3) Let $d_{max} \in C$ be the highest marginal payoff computed in (2) and let C_{max} be the set of cells with marginal payoff d_{max}. Let d_{sec} be the next highest marginal payoff computed in (2). Set $d_{sec} = 0$ if $C_{max} = C$.

(4) Add search effort to the cells in C_{max} in amounts proportional to $1/w(c,1)$ until either the marginal payoff from adding effort in the cells of C_{max} is reduced to d_{sec} or all available effort is expended, i.e., let

$$x = \min\left(\log(d_{max}/d_{sec}), \frac{\{L(i) - \Sigma_{c\varepsilon C}X(x,1)\}}{\{\Sigma_{c\varepsilon C}1/w(c,1)\}}\right);$$

then increase $X(c)$ by $x/w(c,1)$ for $C \varepsilon C_{max}$.

(5) If $\Sigma_{c\varepsilon C}X(c) < L(i)$, then go back to step (2).

(6) Stop. The answer is X.

Incremental Search Plans

If Algorithm 1 is rerun with L(1) increased, then each cell will clearly receive at least as much effort as it originally received. Thus, if L(1) units of search have been optimally allocated and the searcher receives additonal units of search, then the new search can be optimally utilized without reallocating the original search effort. In particular, assume $n > 1$, but $w(c,i) = w(c,1)$ and $p(b) > 0$ implies $b_i = b_1$ for all c and i. This is just a stationary target problem divided into time intervals. Call a plan which allocates search at each time interval to get the best payoff at the end of that time interval an incremental plan. We have shown that an incremental plan is optimal in the long term for this problem.

Incremental plans are useful because they can be computed incrementally. The formula in step (2) of Algorithm 1 shows that after a search according to plan X, further search is allocated as if one were just starting a search for a target with defective distribution having probability $p(c)\exp(-w(c,1)X(c,1))$ in cell c. This is just the probability that the target is in cell c and was not detected by the past search. Thus, only this distribution need be retained between time intervals.

When search conditions change with time, then an optimal plan must account for future as well as past search and an incremental plan need not be optimal. For example, assume a stationary target is equally likely to be in either cell a or cell d, the search is divided into two hours, search units are hours of search, and $w(a,1) = w(d,1) = w(a,2) = \log(2)$ but the weather in cell d during the second hour makes search there completely ineffective so $w(d,2) = 0$. The incremental plan puts one half hour of search in each cell during hour 1 and one hour of search

in cell a during hour 2. Its probability of detection is $1-3/(4\sqrt{2}) = .47$. If, however, the search in cell a during hour 2 is anticipated, the probabilities that the target is in each cell and is not detected during hour 2 are 1/4 for cell a and 1/2 for cell d. Using these probabilities, the full hour of search is reallocated to cell d during hour 1 and the probability of detection rises to .5. This plan is, in fact, optimal.

Target motion can cause the same effect. Assume a search is again divided into two hours with $w(c,i) = \log(2)$ for all c and i, and the target paths among the cells $C = \{a,d,e\}$ have probabilities $p(a,a) = 1/2$, $p(d,d) = 1/4$, and $p(d,e) = 1/4$. The incremental plan and its improvement to the optimal plan coincide with those in the previous example. In each case, the optimal plan puts more effort in cell d during hour 1 than the incremental plan does because a target which starts in cell d becomes more difficult to find during hour 2 than a target which starts in cell a.

Search for a Moving Target

In order to solve the moving target problem, consider the following related problem called the time i reallocation problem for X: Given a search plan X and a time interval i, find a search plan Y which minimizes f(Y) subject to the constraint $Y(c,j) = X(c,j)$ for all $c \in C$ and $j = 1$ to n except $j = i$. The theory of conditional probabilities shows

$$f(Y) = \Sigma_{c \in C} P(c,i,X) \exp\{-w(c,i)\ Y(c,i)\},$$

where

$P(c,i,X)$ = the probability that the target is in cell c at time i and is not detected by search according to plan X at any time interval other than i.

Thus, the time i reallocation problem is a stationary search problem. Let $R_i(X)$ be its unique solution.

Proposition 1: A search plan X solves the moving target problem if and only if for all $i = 1$ to n, $X = R_i(X)$.

Proof: The only if statement is immediate from the definitions. The if statement is an application of the Kuhn-Tucker theorem. Brown [1979] gives the details. QED.

Proposition 1 leads to the following algorithm:

Algorithm 2

(1) Let X_0 be any search plan.

(2) Set j = 0.

(3) Perform steps (4)-(5) for i = 1 to n.

(4) Let $X_{nj+1} = R_i(X_{nj+i-1})$

(5) Go on to the next i.

(6) Increase j by 1 and go back to step (3).

We generally take $X_0 = 0$ in step (1). In this case, X_n is just the incremental search plan.

Algorithm 2 generates a sequence of search plans. The definition of R shows that

$$f(X_0) \geq f(X_1) \geq f(X_2) \geq \ldots \geq 0.$$

Thus, the sequence $\{f(X_j)\}$ converges. Indeed, examples similar to practical examples show that convergence is generally very rapid with $f(X_{3n})$ close to the limit.

Let m be the number of cells and identify each search plan X with the point of $\mathbb{R}^{mn}$ having components X(c,i). The sequence $\{X_j\}$ need not converge unless the moving target problem has a unique solution. This is generally the case (see Brown [1980]). In any event, we can show:

Proposition 2:
(1) $\lim_{j\to\infty} f(X_j) = \min\{f(X): X \text{ is a search plan}\}$
(2) $\{X_j\}$ has a convergent subsequence and the limit of any convergent subsequence solves the moving target problem.

Proof. The subset of $\mathbb{R}^{mn}$ consisting of search plans is compact so there is an infinite set of positive integers J' such that $\{X_{j'}: j' \in J'\}$ converges to a search plan X*. We show below that X* solves the moving target problem. Since f is continuous, $\lim f(X_j) = f(X^*)$ and (1) is proved. The sequence $\{X_{j'}: j' \in J'\}$ is a convergent subsequence of $\{X_j\}$. If Y is the limit of any convergent subsequence of $\{X_j\}$, then compactness shows Y is a search plan and $f(Y) = \lim f(X_j) = f(X^*)$, so Y solves the moving target problem. This proves (2).

In order to show that X* solves the moving target problem, let R be the composition $R = R_n R_{n-1} \ldots R_1$, so $X_{nj+1} = R(X_{nj})$.

Since R_i is continuous, so is R and

$$f(X^*) = \lim\{f(X_j'):j'\varepsilon J'\} = \lim\{f(R(X_j')):j'\varepsilon J'\} = f(R(X^*)).$$

In particular, $f(X^*) = f(R_1(X^*))$, so $X^* = R_1(X^*)$.

Likewise, $X^* = R_i(X^*)$ for all i and Proposition 1 completes the proof. QED.

Implementation of the Algorithm

In order to implement Algorithm 2, we need a means of specifying the parameters of the moving target problem and a method of calculating $R_i(X_{nj+i-1})$ using the specification. The only difficult function to specify is p. Several schemes are presented below. In each case, we just show how to calculate $P(*,i,X_{nj+i-1})$ since Algorithm 1 then yields $R_i(X_{nj+i-1})$.

When $\{b:p(b)>0\}$ is small enough to be enumerated practically, we may calculate directly. We call this case conditionally deterministic motion.

Let $\tilde{p}(b,X)=p(b)\exp\{-\Sigma_{j=1}^{n}w(b_j,j)X(b_j,j)\}$. One maintains a table of $\{(b,\tilde{p}(b,X_{nj+i-1})):p(b)>0\}$ during the computation so that

$$P(c,i,X_{nj+i-1}) = \Sigma_i\{p(b,X_{nj+i-1}) : b_i=c\}/\exp(-w(c,i)\ X(c,i)).$$

Monte Carlo Simulation provides an approximate but powerful method for generating a tabular specification of p from almost any description of target motion.

Markov Motion allows the user to enter the probabilities r(c) that the target starts in cell c and t(c,d) that the target moves from cell c to cell d between successive time intervals, i.e., we set

$$p(b) = r(b_1)\ t(b_1,b_2)\ t(b_2,t_3),\ \ldots,\ t(b_{n-1},b_n);$$

also let

reach(c,i,X) = the probability that the target is in cell c at time i and is not detected by search according to plan X before time interval i

and

survive(c,i,X) = the probability that the target is not detected by search according to plan X after time interval i given the target is in cell c at time i,

so

$$P(c,i,X) = \text{reach}\ (c,i,X)\ \text{survive}\ (c,i,X).$$

For i = 2 to n, we have

$$\text{reach}\ (c,1,X_{nj}) = r(c)$$

$$\text{reach}(c,i,X_{nj+i-1})$$

$$= \Sigma_{d\varepsilon C}\text{reach}(d,i-1,X_{nj+i-1})\exp\{-w(d,i-1)X_{nj+i-1}(d,i-1)\}t(d,c)$$

$$\text{survive}(c,n,X_{nj}) = 1$$

$$\text{survive}(c,i-1,X_{nj})$$

$$= \Sigma_{d\varepsilon C}t(c,d)\exp\{-w(d,i)X_{nj}(d,i)\}\text{survive}(d,i,X_{nj}).$$

These recursive formulas for reach and survive permit an efficient calculation of $P(c,i,X_{nj+i-1})$. Calculate $\text{survive}(*,*,X_{nj})$ at the beginning of step (3) in Algorithm 2 and note that $\text{survive}\ (c,i,X_{nj+i-1}) = \text{survive}(c,i,X_{nj})$ for $i < n$. Calculate $\text{reach}(*,i,X_{nj+i-1})$ in step (4) to complete the recursion.

Markov motion models are appropriate when target probability diffuses. When the target's general destination is known or can be guessed, then it is appropriate to set

$$p(b) = r(b_1)\ t(b_1,b_2)\ \ldots\ t(b_{n-1},b_n)\ s(b_n),$$

where s(c) is an independent estimate of the target's final position. We call this constrained Markov motion. One computes P(c,i,X) as for Markov motion but with $\text{survive}(c,n,X_{nj}) = s(c)$ replacing the previous definition. Further constraints on the Markov motion may be made at intermediate time periods. If, for example, the target must pass through a choke point at time interval k, we might set

$$p(t) = r_1(b_1)\ t_1(b_1,b_2)\ \ldots\ t_1(b_{k-1},b_k)$$

$$r_2(t_2)\ t_2(b_k,b_{k+1})\ \ldots\ t_2(b_{n-1},b_n)\ r_3(b_n).$$

Again, P(c,i,X) can be computed recursively.

Finally, if we specify motion models p_1 through p_I as above, we may give positive weights W_1 through W_I to each possibility and let $p = \Sigma_{i=1}^{I}W_ip_i$. If P_i is the function P for the ith model, then $P = \Sigma_iW_iP_i$. This is called a multiscenario motion model. Use of several multiply constrained Markovian scenarios allows a high degree of flexibility in target motion specification and still permits rapid computation for practical problems.

Example

We give an example using a variation of multiscenario Markov motion. The cells are a grid of squares, three miles on a side, oriented north-south and east-west. We take n = 8 hours and let w(c,i) = 1 for all c and L(i) = 10 for all i. The target's initial distribution is a circular normal distribution truncated to a 15 by 15 mile square. The target chooses a base heading uniformly from $\{157\frac{1}{2}^0, 172\frac{1}{2}^0, 187\frac{1}{2}^0, 202\frac{1}{2}^0\}$. At the beginning of each hour, it retains its old base heading with probability 90 percent and chooses a new base heading uniformly from the other three headings with probability 10 percent. It chooses its actual heading for the hour uniformly from whichever of the following intervals contains the base heading: $[150^0, 165^0]$, $[165^0, 180^0]$, $[180^0, 195^0]$, or $[195^0, 210^0]$. It chooses its speed between 6 and 12 mph with 9 mph 1.5 times as likely as 6 or 12 mph and with the density at intermediate speeds varying linearly. The chosen course and speed are retained for the hour. Thus, the target executes Markovian motion along one of four base courses with each base course retained a geometrically distributed length of time.

By the end of the search, The target density spreads to more than 200 cells. The detection probabilities at each iteration through step (3) of Algorithm 2 are 74.646%, 75.757%, 75.780%, 75.787%, 75.790%, 75.791%, and 75.792%, which repeats. Convergence is rapid and the first probability, the probability for the incremental plan, is nearly optimal. This occurs because the target density is spreading out at nearly the same rate everywhere. Brown [1977] gives other examples demonstrating rapid covergence even when the incremental plan is significantly suboptimal. Unlike an incremental plan, the optimal plan surrounds the target early in the search. The centrally located target density can be searched more easily later.

Acknowledgement

This work was supported by ONR contract N00014-76-C-0696.

References

Brown, S. S. [1977] Optimal and Near Optimal Search for a Target With Multiple Scenario Markovian, Constrained Markovian, or Geometric Memory Motion in Discrete Time and Space, Daniel H. Wagner, Associates Memorandum Report to the Office of Naval Research (case 611).

Brown, S. S. [1980] Optimal Search for a Moving Target in Discrete Time and Space, Oper. Res. (to appear).

Stewart, T. J. [1979] Search for a Moving Target When Searcher Motion is Restricted, Computers and Oper. Res. 6, pp. 129-140.

Stone, L. D. [1979] Necessary and Sufficient Conditions for Optimal Search Plans for Moving Targets, Math. Oper. Res. 4, pp. 431-440.

Washburn, A. R. [1979] On Search for a Moving Target (submitted for publication).

Zangwill, W. I. [1969] "Non-Linear Programming: A Unified Approach," Prentice-Hall, Englewood Cliffs, New Jersey.

SEARCH FOR A MOVING TARGET:

UPPER BOUND ON DETECTION PROBABILITY

A. R. Washburn

Naval Postgraduate School

Introduction

The object is to detect a randomly moving target at one of the discrete times $0,1,\ldots,\tau$. The searcher determines a non-negative effort distribution, $\psi(x,t)$, such that the total effort applied at time t does not exceed $m(t)$. Our purpose here is to establish an upper bound on the detection probability for every effort distribution. This is intended to supplement existing iterative procedures that develop sequences of effort distributions that improve monotonically.

The General Case

Let $w(x,t)$ be an "effectiveness coefficient" for search effort applied at position x at time t. If X is the position of the target at time t, then the probability of detection depends, we assume, only on the total effective search effort

$$Z = \Sigma_{t=0}^{\tau} w(X_t,t)\ \psi(X_t,t).$$

Specifically, the probability of detection is $P(\psi) = E(b(Z))$, where the expectation operator is needed because X is a stochastic process. We assume that $b(z') - b(z) \leq s(z)\ (z'-z)$ for some function $s(z)$ and for all z'; if the detection function b is concave and differentiable, then $s(z)$ is just the derivative of b. In most applications, $b(z) = 1 - \exp(-z)$ and $s(z) = \exp(-z)$.

Let ψ and ψ' be two effort distributions; then

(1) $P(\psi') - P(\psi) = E\{b(Z') - b(Z)\} \leq E\{s(Z)\ (Z'-Z)\}$.

We also have

(2) $E\{s(Z)(Z'-Z)\} = \Sigma_{t=0}^{\tau} E\{s(Z)\ w(X_t,t)\ [\psi'(X_t,t) - \psi(X_t,t)]\}$.

We now consider two cases, depending on whether X is discrete or continuous. If X_t is discrete, we let $p_t(x)$ be the probability mass function of X_t and require

$$\Sigma_x\ \psi'(x,t) = \Sigma_x\ \psi(x,t) = m(t).$$

We define

(3) $D_\tau(\psi,x,t) = w(x,t)\ p_t(x)\ E\{s(Z)\,|\,X_t=x\}$.

Then (2) can be written

(4) $E\{s(Z)\ (Z'-Z)\} = \Sigma_{t=0}^{\tau}\ \Sigma_x\ D_t(\psi,x,t)\ [\psi'(x,t) - \psi(x,t)]$.

If X_t is continuous, let $p_t(x)$ be the probability density function of X_t and require

$$\int\psi'(x,t)dx = \int\psi(x,t)dx = m(t).$$

Then (2) can be written

(4') $E\{s(Z)(Z'-Z)\} = \Sigma_{t=0}^{\tau} \int D_\tau(\psi,x,t)[\psi'(x,t) - \psi(x,t)]\ dx$.

In either case, suppose that $D_\tau(\psi,x,t) < \bar{\lambda}(t)$ for all x, and $D_\tau(\psi,x,t) > \underline{\lambda}(t)$ when $\psi(x,t) > 0$. In the discrete case, from (4),

(5) $E\{s(Z)(Z'-Z)\} < \Sigma_{t=0}^{\tau}[\Sigma_x\ \bar{\lambda}(t)\ \psi'(x,t) - \Sigma_x\ \underline{\lambda}(t)\ \psi(x,t)]$.

Combining (5) and (1),

(6) $P(\psi') - P(\psi) \leq \Sigma_{t=0}^{\tau}\ \{\bar{\lambda}(t) - \underline{\lambda}(t)\}\ m(t) \equiv \Delta(\psi)$.

A formula similar to (5) shows that (6) must also hold in the continuous case.

Now ψ' is not needed to compute any of the quantities on the right-hand side of (6), so every effort density ψ' has associated with it an upper bound on the detection probability $P(\psi) + \Delta(\psi)$. In the event $\bar{\lambda}(t) = \underline{\lambda}(t)$ for all t, ψ must actually be optimal--this has been observed by Stone [1979], and, in

fact, our whole development is a modification of his sufficiency proof. The main issue is now computational: Is determination of $\Delta(\psi)$ worth the effort?

Markov Motion and Exponential Detection Function

Let

$$Z = Z_t^- + w(X_t,t)\ \psi(X_t,t) + Z_t^+ \quad \text{for } 0 \le t \le \tau,$$

where

$$Z_t^- = \Sigma_{u=0}^{t-1}\ w(X_u,u)\ \psi(X_u,u)$$

and

$$Z_t^+ = \Sigma_{u=t+1}^{\tau}\ w(X_u,u)\ \psi(X_u,u); \quad Z_0^- = Z_\tau^+ = 0.$$

Then

$$D_\tau(\psi,x,t) = w(x,t)p_t(x)E\{s(Z_t^- + w(x,t)\ \psi(x,t) + Z_t^+)\,|\,X_t = x\}.$$

If the motion is Markov, then Z_t^+ is independent of Z_t^- when $X_t = x$ is given. If $b(z) = 1 - \exp(-z)$, then $s(z) = \exp(-z)$. If both conditions hold, then

$$D(\psi,x,t) = w(x,t)\ P(\psi,x,t)\ \exp(-w(x,t)\ \psi(x,t))\ Q(\psi,x,t),$$

where

$$P(\psi,x,t) \equiv p_t(x)\ E\{s(Z_t^-)\,|\,X_t = x\} \text{ and}$$

$$Q(\psi,x,t) \equiv E\{s(Z_t^+)\,|\,X_t = x\},$$

and $P(\psi,x,t)$ is the joint probability that $X_t = x$ and that the target is not detected by any of the searches at $0,1,\ldots,t-1$; note that $P(\psi,x,t)$ does not depend on $\psi(y,u)$ for $u > t$. $Q(\psi,x,t)$ is the conditional probability that the target is not detected by any of the searches at $t+1,\ldots,\tau$ given that $X_t = x$; note that $Q(\psi,x,t)$ does not depend on $\psi(y,u)$ for $u \le t$. Given ψ, $P(\psi,x,t)$ and $Q(\psi,x,t)$ can be easily computed recursively. $P(\psi,x,0)$ is a given initial distribution, and $P(\psi,x,t+1)$ can be obtained from $P(\psi,x,t)$ using the Markov transition rule and $\psi(\cdot,t)$. Similarly, $Q(\psi,x,\tau) \equiv 1$ and $Q(\psi,x,t-1)$ can be obtained from $Q(\psi,x,t)$ using the Markov transition rule and $\psi(\cdot,t)$. After obtaining $P(\psi,x,t)$ and $Q(\psi,x,t)$, it is a simple matter to compute $D_\tau(\psi,x,t)$ and then $\Delta(\psi)$.

Algorithms for finding the optimal effort distribution ψ^* typically operate by generating a sequence $\psi_1,\psi_2,\ldots,$ that approaches ψ^*. The method of computation is such that $\Delta(\psi_i)$ can be computed with only slightly more effort. Consider the discrete

case. For any t, the probability of detection is

$$(7) \quad P(\psi) = 1 - \Sigma_x P(\psi,x,t) \exp\{-w(x,t)\ \psi(x,t)\}\ Q(\psi,x,t).$$

In order to find ψ^*, we first make an initial guess $Q^0(x,t)$ and then calculate ψ^n, the function that minimizes

$$\Sigma_x P(\psi,x,t) \exp\{-w(x,t)\ \psi(x,t)\}\ Q^{n-1}(x,t) \quad \text{for } n \geq 1$$

with $Q^n(x,t) = Q(\psi^n,x,t)$ for $n \geq 1$. Each of the minimization problems is relatively simple, and it can be shown that $P(\psi^n)$ increases with n. See Stone, et.al. [1978], Brown [1980], and Washburn [1979].

For each of the minimization problems, there must exist a function $\lambda_n(t)$ such that $D_n(x,t) = \lambda_n(t)$ when $\psi^n(x,t) > 0$ and $D_n(x,t) \leq \lambda_n(t)$ when $\psi^n(x,t) = 0$, where

$$D_n(x,t) = w(x,t)\ P(\psi^n,x,t) \exp\{-w(x,t)\ \psi^n(x,t)\}\ Q^{n-1}(x,t).$$

These are simply the Kuhn-Tucker conditions; in practice, some sort of a search is made until $\lambda_n(t)$ is found such that $\Sigma_x \psi(x,t) = m(t)$. But

$$(8) \quad D_\tau(\psi^n,x,t) \begin{cases} = \lambda_n(t)\ Q^n(x,t)/Q^{n-1}(x,t) \text{ when } \psi^n(x,t) > 0 \\ \leq \lambda_n(t)\ Q^n(x,t)/Q^{n-1}(x,t) \text{ when } \psi^n(x,t) = 0, \end{cases}$$

so computation of $D_\tau(\psi^n,x,t)$ is a small burden given that $\lambda_n(t)$, $Q^n(x,t)$, and $Q^{n-1}(x,t)$ have to be computed in any case.

In the continuous case, Σ_x is replaced by $\int dx$; otherwise, the development remains the same.

An Example

We set $\tau = 79$, so there are 80 looks. The target does a random walk over the cells 1,...,67, starting at cell 34. At each opportunity, it moves left with probability .3, right with probability .3, or else does not move. The boundary is reflecting. We set $w(x,t) = .001625$ and $m(t) = 100$; if all 100 units of search effort are used in a single cell, the probability of detection would be $1 - \exp(-.1625) = .15$ at each of the 80 opportunities. We also set $Q^0(x,t) \equiv 1$.

Cell	30	31	32	33	34	35	36	37	38
Time					.85				
					.85				
					.85				
				.99	.87	.99			
				.97	.90	.97			
				.96	.92	.96			
				.96	.93	.96			
				.95	.94	.95			
				.95	.94	.95			
10				.95	.94	.95			
			1.00	.95	.95	.95	1.00		
			.99	.96	.96	.96	.99		
			.98	.96	.96	.96	.98		
			.98	.96	.96	.96	.98		
			.97	.97	.96	.97	.97		
			.97	.97	.97	.97	.97		
			.97	.97	.96	.97	.97		
		1.00	.97	.97	.97	.97	.97	1.00	
		.99	.97	.97	.97	.97	.97	.99	
20		.99	.97	.98	.97	.98	.97	.99	
		.98	.97	.98	.98	.98	.97	.98	
		.98	.97	.98	.98	.98	.97	.98	
		.98	.98	.98	.98	.98	.98	.98	
		.98	.98	.98	.98	.98	.98	.98	
		.98	.98	.98	.98	.98	.98	.98	
		.97	.98	.98	.98	.98	.98	.97	
	1.00	.97	.98	.98	.98	.98	.98	.97	1.00
	1.00	.98	.98	.98	.98	.98	.98	.98	1.00
	.99	.98	.98	.98	.98	.98	.98	.98	.99
30	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.99	.98	.98	.98	.98	.98	.98	.98	.99
	.98	.98	.98	.98	.98	.98	.98	.98	.98
	.98	.98	.98	.98	.98	.98	.98	.98	.98
40	.98	.98	.98	.98	.98	.98	.98	.98	.98

Figure 1. Survival probabilities for near optimal search.

Cell	28	29	30	31	32	33	34	35	36	37	38	39	40
Time			.98	.98	.98	.98	.98	.98	.98	.98	.98		
			.98	.98	.98	.98	.98	.98	.98	.98	.98		
		1.00	.98	.98	.98	.98	.98	.98	.98	.98	.98	1.00	
		1.00	.98	.98	.98	.98	.98	.98	.98	.98	.98	1.00	
		1.00	.99	.98	.98	.98	.98	.98	.98	.98	.99	1.00	
		1.00	.99	.98	.98	.98	.98	.98	.98	.98	.99	1.00	
		.99	.99	.98	.98	.98	.98	.98	.98	.98	.99	.99	
		.99	.99	.98	.98	.98	.98	.98	.98	.98	.99	.99	
		.99	.99	.98	.98	.98	.98	.98	.98	.98	.99	.99	
50		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.99	.98	.98	.98	.99	.99	.99	.99	
60	1.00	.99	.99	.99	.99	.98	.98	.98	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.98	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
70	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
	1.00	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99	1.00
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
		.99	.99	.99	.98	.98	.98	.98	.98	.99	.99	.99	
80		1.00	.99	.98	.98	.98	.98	.98	.98	.98	.99	1.00	

Figure 1. Continued.

The first five detection probabilities (and values for $\Delta(\psi)$) are: .7054 (.0393), .7075 (.0246), .7086 (.0120), .7090 (.0067), and .7092 (.0036). The amounts of CPU time required on the NPS IBM 360/67 for computation of each of these five pairs were 2.4, 2.8, 2.4, 2.0, and 2.0 seconds, respectively. The second pass always takes a relatively long time because $Q^1(x,t)$ is radically different from $Q^0(x,t)$. The function $\exp(-w(x,t)\,\psi^5(x,t))$ is shown to two decimal places in Figure 1, with a blank indicating no search. Time reads down the page (there are 80 rows). Only the central cells are shown; the rest are not searched.

It is interesting to compare these results with those that result when all effort must be placed in a single cell at each time. The best probability of detection in that case is apparently .6994. Sufficient conditions for optimality are not known in that case, but the author has done sufficient experimentation to conjecture that .6994 is the answer. The increase from .6994 to .7092 can, of course, be attributed to the relaxation of a constraint.

References

Brown, S. S. [1980] Optimal Search for a Moving Target in Discrete Time and Space, Oper. Res. (to appear).

Stone, L. D., Brown, S. S., Buemi, R. P. & Hopkins, C. R. [1978] Numerical Optimization of Search for a Moving Target, Daniel H. Wagner, Associates Report to Office of Naval Research.

Washburn, A.R. [1979] "On Search for a Moving Target," (submitted for publication).

OPTIMIZATION ALGORITHM FOR GENERAL TARGET MOTION

L. D. Stone

Daniel H. Wagner, Associates

Basic Search Problem

In this section, we give a description of the basic search problem considered in this paper.

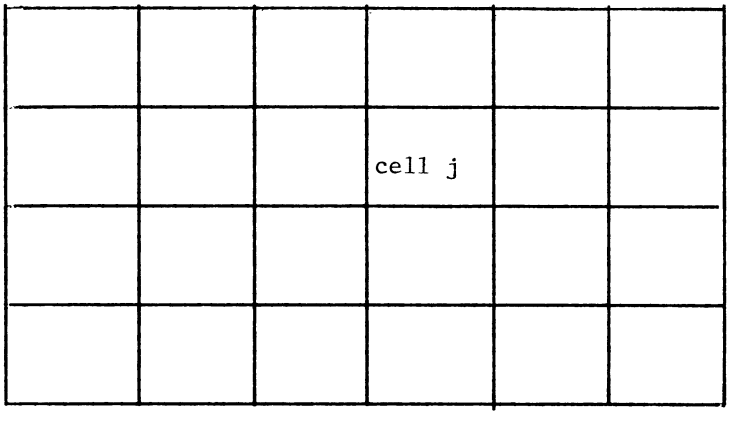
Target motion. The target's motion is represented by a stochastic process $X = \{X_t, t = 1,\ldots,T\}$, where X_t is a random variable which gives the target's position* at time t.

In a typical search problem, one is given the target's initial probability distribution and a stochastic description of the target's motion. These two are combined to produce the stochastic process which represents all possible target paths over time $t = 1,\ldots,T$.

Search grid. We assume that a grid, like the one in Figure 1, has been established on the search space which is two-dimensional and that we must allocate our effort uniformly within a given cell of the grid, although effort may vary from cell to cell. Let J denote the set of cells in the grid, and let $p_t(j) = \Pr\{X_t \text{ in cell } j\} = \Pr\{\text{target is in cell } j \text{ at time } t\}$, $A(j)$ = area of jth cell, and $W(j)$ = sensor sweep width in cell j. Although we impose this grid structure, it does not affect the target motion which may take place in either discrete or continuous space.

---

* We understand X_t to be either a point in the plane or a cell index. It will be clear from the context which is the case.



$p_t(j)$ = Pr{target is in cell j at time t}

$A(j)$ = area of j^{th} cell

$W(j)$ = sensor sweep width in j^{th} cell

J = collection of cells

Note: The search grid need not be uniform in space or time.

Figure 1. Search grid.

Class of search plans considered. A search plan is described by a nonnegative function ψ of time and space such that

$$\psi(j,t) = \text{effort placed in cell } j \text{ at time } t \text{ for } j \in J, t=1,\ldots,T,$$

and we assume that the amount of effort available is constrained so that

$$(1) \qquad \sum_{j \in J} \psi(j,t) = m(t) = \text{total effort available at time } t \text{ for } t = 1,\ldots,T.$$

The class of plans which satisfy (1) is called $\Psi(m)$.

Detection function. Let ω indicate a sample target path, and let $X_t(\omega)$ be the target's position (i.e., cell) at time t on this path. The probability of detecting the target with plan ψ given it follows path ω is

$$(2) \qquad 1 - \exp\left[-\sum_{t=1}^{T} W(X_t(\omega))\ \psi(X_t(\omega),t)/A(X_t(\omega))\right].$$

Let E denote the expectation over the sample paths of the target motion process. Then the total probability of detection by time T is

$$(3) \qquad P_T[\psi] \equiv E\left[1 - \exp\left(-\sum_{t=1}^{T} W(X_t)\ \psi(X_t,t)/A(X_t)\right)\right]$$

for any search plan ψ.

Problem statement. We seek a plan ψ^* which satisfies the effort constraint (1) and maximizes the probability of detection by time T within the class of plans satisfying (1). Such a plan is called T-optimal within $\Psi(m)$.

Basic Necessary and Sufficient Conditions

In this section, we present the basic necessary and sufficient conditions which we shall use in constructing our algorithm. The form of the conditions given here applies to discrete-time target motion processes and exponential detection functions. A generalization of these conditions to include continuous time motion processes and nonexponential detection functions is stated and proved in Stone [1979].

In order to state these conditions, we let ψ^* denote a plan which is T-optimal within $\Psi(m)$ and let

$$(4) \qquad g_{\psi}(j,t) = \Pr\left(\begin{array}{l|l}\text{target is} & \text{failure to detect at all times} \\ \text{in cell } j & \text{other than } t \text{ using plan } \psi^*\end{array}\right)$$

$$\text{for } j \in J,\ t = 1,\ldots,T.$$

Observe that the conditioning on failure in (4) includes search which takes place after time t as well as before. Let E_{jt} denote expectation conditioned on the target being in cell j at time t. Then there is a constant K(t) for t = 1,...,T such that

$$g_{\psi^*}(j,t) = K(t)\, p_t(j)\, E_{jt}[\exp(-\sum_{s \neq t} W(X_s)\, \psi^*(X_s,s)/A(X_s))]$$

for $j \in J$, $t = 1,\ldots,T$.

Basic condition. In order that $\psi^* \in \Psi(m)$ be T-optimal, it is necessary and sufficient that the allocation $\psi^*(\cdot,t)$ maximize the probability of detecting a stationary target with distribution $g_{\psi^*}(\cdot,t)$ and effort m(t) for t = 1,...,T.

This condition is a generalization to include continuous space of the conditions given by Brown [1980] for discrete space.

Description of Basic Algorithm

Using the basic conditions, we now describe an algorithm that produces a sequence of search plans which converge to the optimal plan.

First iteration. For time t = 1, find the optimal allocation of search effort m(1) for the initial target location distribution at time 1. Compute the posterior target distribution at time 2 given failure to detect the target at time 1, and find the optimal allocation of m(2) effort for this distribution. Continue this way, computing the posterior target location distribution at time t given failure to detect the target at times s = 1,...,t-1 and allocating the effort m(t) for time t in a manner which is optimal for that distribution. When time T is reached, we have found the incrementally optimal or myopic plan. That is, at each time period we have allocated effort to maximize the increase in detection probability for that time period. For most moving target problems, the myopic plan is not T-optimal for T > 1.

nth iteration (n>1). Compute the posterior probability distribution for the target's location at time t = 1, given failure to detect the target at times t = 2,...,T, using the allocation at times t = 2,...,T obtained on the previous iteration. Usually, this allocation differs from the previous allocation at time 1. For t = 2,...,T, compute the posterior target location distribution at time t given failure to detect using the allocations for times t = 1,2,...,t-1 obtained on this iteration as well as the allocations after time t obtained on the previous iteration, and reallocate the effort at time t to be optimal for this distribution.

The above algorithm converges to the optimal plan. That is, one can find a plan whose probability of detection is as close to optimal as he desires by performing a large enough number of iterations. Notice that each iteration produces a plan with a higher probability of detection than the previous one.

The above algorithm amounts to solving a sequence of stationary target problems to solve an optimal moving target problem. There are very efficient algorithms for solving stationary target problems when the detection function is exponential (see Example 2.2.8 of Stone [1975]). Thus, we can produce efficient algorithms which follow the procedure outlined above provided we can efficiently compute the posterior target distributions described above. Computation of these distributions will depend on the target motion model, i.e., the nature of the stochastic process $\{X_t, t=1,\ldots,T\}$. For example, the algorithm described by Brown [1980] computes these posterior distributions very efficiently when the target motion is Markovian in discrete space and time. The algorithm we are about to describe can be used with any Monte Carlo target motion. It is more flexible than the one for Markovian motions, but it is also slower and requires an additional approximation, i.e., that involved in replacing the stochastic process $\{X_t, t=1,\ldots,T\}$ by some large but finite number of its sample paths.

Description of the Algorithm for Monte Carlo Target Motion

We suppose that we are given N sample paths drawn from the process $\{X_t, t=1,\ldots,T\}$. The nth sample path is assumed to have the form $(x_1^n, x_2^n, \ldots, x_T^n)$, w_n, where x_s^n is the position of the target at time s and w_n is the probability that the nth sample path represents the target's actual path (usually $w_n = 1/N$).

Conceptually, the first act of the optimizer is to convert the sample paths into sequences of cell numbers. Thus, the nth sample path becomes

$$(j_1^n, j_2^n, \ldots, j_T^n), w_n,$$

where $j_s^n = c(x_s^n)$ is the index of the cell into which x_s^n falls.

Let Ξ_t be the function which acts on a search plan ψ by replacing the allocation at time t, $\psi(\cdot,t)$ with f*, an optimal allocation of m(t) effort for $g_\psi(\cdot,t)$, the posterior target distribution at time t given failure to detect at all times other than t using ψ.

That is,

$$\Xi_t\psi(j,s) = \begin{cases} \psi(j,s) & \text{for } s \neq t, \\ f^*(j) & \text{for } s = t. \end{cases}$$

First pass. Set $\psi_0(j,t) = 0$ for $j \in J$, $t = 1,\ldots,T$. We begin with $t = 1$. For each cell j, we accumulate all the probabilities w_n of paths such that $j_1^n = j$ to compute $p_1(j)$, the probability that the target is in cell j at time 1 for $j = 1,\ldots,J$.

Using a standard algorithm (e.g., the algorithm in Example 2.2.8 of Stone [1975]), we calculate the allocation of effort f^* which maximizes the probability of detection,

$$\sum_{j=1}^{J} p_1(j)\,[1 - \exp(-W(j)\, f^*(j)/A(j))],$$

for a stationary target with probability $p_1(j)$ of being in cell j for $j = 1,\ldots,J$, subject to the effort constraint $m(t)$. Let $\psi_1 = \Xi_1\psi_0$ and compute

$$\tilde{w}_n = w_n \exp(-W(j_1^n)\ \psi_1(j_1^n,1)/A(j)) \qquad \text{for } n = 1,\ldots,N.$$

The value $\tilde{w}_n$ is the probability that the target is following the nth sample path and is not detected by the search effort at time 1.

For $t = 2$, we compute $\tilde{g}_{\psi_1}(j,2)$, the probability that the target is in cell j at time 2 and not detected by the search applied at time 1 by summing the probabilities $\tilde{w}_n$ of all paths such that $j_2^n = j$. Let $g_{\psi_1}(\cdot,2)$ be the probability distribution that one obtains by normalizing $\tilde{g}_{\psi_1}(\cdot,2)$ so that it sums to 1. As above, we find an optimal allocation of $m(2)$ effort for the stationary target problem with distribution $g_{\psi_1}(\cdot,2)$, and let $\psi_2 = \Xi_2\psi_1$. The weights $\tilde{w}_n$ are then multiplied by $\exp(-W(j_2^n)\ \psi_2(j_2^n,2)/A(j_2^n))$ to obtain the revised values of $\tilde{w}_n$ for $n = 1,\ldots,N$.

This process is continued for $t = 3,\ldots,T$. At the end of the first pass, we have calculated the myopic plan ψ_T, i.e., the plan which at each time period allocates its effort to maximize the probability of detection during that time period given failure to detect the target in previous time periods.

Second pass. Let w_n^1 be the value of $\tilde{w}_n$ obtained by the end of the first pass described above, i.e., w_n^1 is the probability that the target is following the nth sample path and is not detected by the effort allocated in the first pass. For $t = 1$, we calculate

$$\hat{w}_n = w_n^1 \exp(W(j_1^n)\ \psi_T(j_1^n,1)/A(j_1^n)) \qquad \text{for } n = 1,\ldots,N,$$

and accumulate $\hat{w}_n$ into cell j_1^n to calculate $g\psi_T(j,1)$, the probability that the target is in cell j at time 1 given failure to detect at all times in the future. Generally, $p_1 \neq g\psi_T(\cdot,1)$.

We then find an optimal allocation of m(1) effort for the distribution $g\psi_T(\cdot,1)$, set $\psi_{T+1} = \Xi\psi_T$, and compute new values for $\tilde{w}_n$ by $\tilde{w}_n = \hat{w}_n \exp(-W(j_1^n)\ \psi_{T+1}(j_1^n,1)/A(j_1^n))$. For t = 2, we calculate $\hat{w}_n = \tilde{w}_n \exp(W(j_2^n)\ \psi_{T+1}(j_2^n,2)/A(j_2^n))$, and repeat the above process for time 2. This continues for t = 3,...,T, and the second pass is completed. Additional passes proceed in a similar fashion.

Observe that at the end of the kth pass,

$$1 - P_T[\psi_{kT}] = \sum_{n=1}^{N} \tilde{w}_n,$$

and that $1 - P_T[\psi_{kT}]$ is monotonically decreasing in k. It is shown in Stone, et.al. [1978] that as $k \to \infty$, there is a convergent subsequence of plans which converge to a limiting plan ψ^* which is T-optimal. Furthermore, $\lim_{k\to\infty} P_T[\psi_{kT}] = P_T[\psi^*]$.

Thus, one can come as close to the optimal plan as he wishes by making enough passes. The paper by Washburn in these proceedings describes a method for deciding how close to optimal a plan is.

Acknowledgement

This paper was prepared with the support of ONR Contract No. N00014-76-C-0696.

References

Brown, S. S. [1980] Optimal Search for a Moving Target in Discrete Time and Space, Oper. Res. (to appear).

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

Stone, L. D. [1979] Necessary and Sufficient Conditions for Optimal Search Plans for Moving Targets, Math. Oper. Res. 4, pp. 431-440.

Stone, L. D., Brown, S. S., Buemi, R. P. & Hopkins, C. R. [1978] Numerical Optimization of Search for a Moving Target, Daniel H. Wagner, Associates Report to Office of Naval Research.

EXPERIENCE WITH A BRANCH-AND-BOUND ALGORITHM FOR CONSTRAINED SEARCHER MOTION

T.J. Stweart

National Research Institute for Mathematical Science

Introduction

In a recent paper (Stewart [1979]), we have introduced the problem of search for a moving target when there are practical constraints on the rate of movement of the searcher and have suggested approaches to a solution. In the present paper, we report, in particular, on computational experience with the branch-and-bound algorithm proposed there for determining a track for a single searcher while searching for a moving target in discrete space and time. Although the method was developed for a single searcher, a heuristic adaptation was also suggested for use when there are multiple searchers.

For completeness, we shall repeat her the proposed algorithm before proceeding to a discussion of computational experience. As a vehicle for computational tests, we have again selected, and for the same reasons, a search for a one-dimensional random walker. Although better algorithms may probably be proposed for this particular problem in view of its simple structure, it appears to be a useful test problem for our algorithm, which, it is hoped, will ultimately be applied to more general problems.

Outline of the Problem and Algorithm

We assume the following:

(a) the discrete search space is represented by a set of cells denoted J;

(b) for the full duration of each time period $t(t=1,\ldots,T)$, both searcher and target are restricted to a single cell;

(c) the target path, denoted by $\omega = [\omega(1), \omega(2),\ldots, \omega(T)]$ where $\omega(t) \in J$, is unknown, but for each possible path an a priori probability p_ω is specified;

(d) the searcher path, denoted by $\sigma = [\sigma(1), \sigma(2), \ldots, \sigma(T)]$ where $\sigma(t) \in J$, has to be determined subject to a constraint on motion of the form $\sigma(t+1) \in I[\sigma(t)]$, where for each $j \in J$, $I[j] \subset J$; an initial location of the searcher is for convenience denoted as $\sigma(0)$;

(e) conditioned on no detection before t and on $\omega(t) = \sigma(t) = j$, say, the probability of detection during period t is $1-\exp[-\beta_j]$, independent of $\sigma(s)$, $\omega(s)$ for $s \neq t$;

(f) the terminal period T is specified and the objective is to maximize the probability of detection within T periods.

With this set of assumptions, the problem is a (0,1) integer programming problem with nonlinear objective functions and linear constraints. There is sufficient special structure to encourage us to seek special-purpose algorithms rather than to rely on generalized approaches.

In principle, a branch-and-bound implicit enumeration scheme can easily be formulated as follows.

Step (1): set $t=0, P_0 = \hat{P} = 1$; define the set $K[0, \sigma(0)]$ as $I[\sigma(0)]$.

Step (2): obtain, by some means, a lower bound P_B on the probability of no detection during periods $t+1,\ldots,T$, with any continuation of the partial plan $\sigma(1),\ldots, \sigma(t)$ (a null plan for $t=0$) such that $\sigma(t+1) \in K[t, \sigma(t)]$, conditional on no detection up to period t.

Step (3): if $P_t P_B < \hat{P}$, go to step (6); otherwise, it is proved that all continuations of the partial plan are nonoptimal; thus, go to step (4).

Step (4): if t=0, stop; the current best solution is optimal; otherwise, go to step (5).

Step (5): delete $\sigma(t)$ from $K(t-1, \sigma(t-1))$ and from the partial plan, and set t=t-1; if $K[t, \sigma(t)]$ is now empty, return to step (4); otherwise, return to step (2).

Step (6): select in some way (possibly on the basis of the computation in step (2)) a provisional extension of the plan to t+1 (i.e., specify a value for $\sigma(t+1)$); set t=t+1, and denote by P_t the probability of no detection up to period t with this partial plan; if t<T, define the set $K[t, \sigma(t)]$ as $I[\sigma(t)]$ and return to step (2); if t=T, go to step (7).

Step (7): if $P_T > \hat{P}$, return to step (5); otherwise, install the current solution as current best solution, set $\hat{P}=P_T$, and return to step (5).

Clearly, the critical step is (2). The most promising approach at this stage seems to be to obtain the bound by relaxing the constraints

$$\sigma(\tau) \in I[\sigma(\tau-1)] \text{ for } \tau > t \quad \text{to } \sigma(\tau) \in J_\tau,$$

where

$$J_{t+1} = K[t,\sigma(t)]$$

and for $\tau > t+1$

$$J_\tau = \cup[I(j) \mid j \in J_{\tau-1}].$$

Since this is a moving target problem without constraints on searcher motion, it is basically similar to the problem solved by Brown [1980]. Unfortunately, the integer constraint implies that the space of search plans is nonconvex and Brown's algorithm may converge to a suboptimal answer. Further relaxation of the problem to allow the search effort to be divided over all members of J_τ eliminates this problem, but the bounds are very weak. The result is that at step (3), $P_t P_B < \hat{P}$ very frequently, and a very large proportion of all possible search plans has to be explicitly enumerated.

For this reason, we propose the use of Brown's algorithm, accepting the risk of suboptimality. This risk could be reduced by repeating Brown's algorithm with a number of different starting points, but as shown in the next section, there seems little practical advantage in doing so. The effect of suboptimality in computing the bound is that a branch containing an optimal

continuation may be incorrectly eliminated at step (3). Nevertheless, even should this occur, we may at the very least expect a good solution and may even arrive at an optimal one. This is discussed in the next section.

Computational Experience

As indicated, our computational experience is based on the one-dimensional random walker search (McCabe [1974]) as a test problem. In our earlier paper (Stewart [1979]), we show how this problem can be formulated as in the last section. Note that I[j] always consists of two elements only.

The first set of tests was aimed at testing the approach to optimality with the algorithm. For this reason, a large number of different cases were selected, all with T=10, so that the true optimal solution could be obtained by total enumeration. The cases were defined by various combinations of p_ω and β_j. In the algorithm, the starting point was taken to be the myopic solution.

Out of a total of 105 test problems, the true optimal solution was obtained in all except four cases. Of more interest is the severity of nonoptimality in these four cases. The mean and maximum ratios of probability of detection to the corresponding minimum probability for the four nonoptimal solutions is shown in Table 1 (under the heading "basic branch-and-bound algorithm").

Although the approach to optimality is clearly satisfactory, computational efficiency is not consistently so. The mean number of branches which had to be evaluated per test problem was 165.7 out of a total of 1,022. The standard deviation of this number was 104.0, and the maximum number of evaluations found in any one problem was 497. At this latter level, it is better to explicitly evaluate all solutions! Thus, although, as we shall see, the proportion of branches required to be evaluated falls rapidly with increasing problem size, there is a need to seek a modification to the algorithm which will at least prevent such excessively large numbers of evaluations being required. Two modifications have been formulated and tested. The first (denoted in the tables as "modified branch-and-bound") is obtained by following the same procedure as before, but in the unconstrained problem taking periods in pairs and considering all feasible pairs of allocations for each pair of periods. Thus, in consecutive pairs of periods $(\tau, \tau + 1)$, $(\tau + 2, \tau + 3)$, we would have $\sigma(\tau) \in J_\tau$, $\sigma(\tau + 1) \in I[\sigma(\tau)]$, $\sigma(\tau + 2) \in J_{\tau+2}$, and $\sigma(\tau + 3) \in I[\sigma(\tau + 2)]$. The solution would then be closer to the constrained solution than before, and a better bound may be expected. By considering each feasible pair of allocations as a single "hyperallocation," Brown's algorithm is easily adapted

Table 1. Approach to optimality with the three algorithms.

	Percentage of solutions which are non-optimal	Mean (amongst non-optimal solutions) of: prob. of no detection / minimum prob.	Maximum value of: prob. of no detection / minimum prob.
Basic branch-and bound	3.8	1.0029	1.0041
Modified branch-and-bound	9.3	1.0152	1.0474
Myopic branch-and bound	10.3	1.0147	1.0474

but, of course, with a larger number of options in each period pair, raising the possibility of more local optima.

As shown in Table 1, the approach to optimality is not substantially worse than before. The mean number of branches evaluated was 131.3 with standard deviation 73.5. This is a marked improvement on the basic algorithm, although the maximum number of evaluations was still as high as 390.

The second modification is initially aimed more at reducing computation per branch than at reducing the number of evaluations. It has this effect as well, however. The approach is to use the myopic solution of the unconstrained problem directly as an approximate bound without attempting to improve it further by Brown's algorithm. In view of the previous results, fairly good results may be expected, but the approach to optimality shown in Table 1 ("myopic branch-and-bound") is quite surprising and hardly worse than before. The mean number of branch evaluations per problem was 98.9 with standard deviation 47.4 while the maximum number of evaluations was 215. This is a substantial improvement on the other approaches and suggests that this "myopic branch-and-bound" approach may be the best practical method. It may be thought that in view of this success, direct myopic choice of searcher path may also be near-optimal. This is not, in fact, true, and it has been found that the myopic path yields up to twice the probability of no detection as the optimal value (see, for example, Table 2).

The test problems above all had T=10, in which case total enumeration of all paths is a convenient approach. As T increases, however, total enumeration quickly becomes unfeasible (estimated computing times on the CDC Cyber 174 being 1 hour at T=20 and 10^3

Table 2. Computational behavior for

Problem 1: β_j=0.2 throughout; probability of either direction =0.5

Problem 2: β_j=0.25 throughout; probability of one direction =0.75

	Basic branch-and-bound			Modified branch-and-bound		
	No. of branches evaluated	Computer time (secs)	Minimal probability	No. of branches evaluated	Computer time (secs)	Minimal probability
Problem 1						
T = 10	125	3.2	0.2074	107	3.1	0.2074
15	253	14.3	0.1279	216	13.2	0.1279
20	452	42.6	0.0851	186	21.3	0.0851
30	1611	390.4	0.0431	299	81.3	0.0431
40	8000*	1282.1*	0.0300*	6000*	1282.4*	0.0300*
Problem 2						
T = 10	313	6.2	0.0249	151	3.4	0.0249
15	1398	58.4	0.0172	1091	47.5	0.0172
20	5330	453.9	0.0131	4326	397.7	0.0131
30	4000*	1282.2*	0.0078*	3000*	1282.0*	0.0078*

* Best solution after elapsed time : algorithm had not terminated

hours at T=30). We wish, therefore, to investigate computational requirements for the branch-an-bound approaches for larger T. Examples of such requirements are shown in Table 2 for two cases, one a symmetrical case in which the random walker is equally likely to choose either direction and the other an asymmetrical case in which one direction is chosen with probability 0.75. All computing times refer to the Control Data Cyber 174 system.

It is clear that in the symmetrical case the branch-and-bound approach is feasible up to T=40. The asymmetrical case presents greater difficulties but still provides solutions for situations in which total enumeration is unfeasible and in which the myopic path can be quite substantially nonoptimal.

Concluding Comments

We have demonstrated the effectiveness and computational efficiency of the branch-and-bound approach for the one-dimensional random walker. The intention is now to try to use these algorithms to investigate searcher paths in the plane for various scenarios. The aim of this work will be primarily to derive recurrent patterns in search plans generated, which patterns can be incorportated into on-site manuals, as at this stage it seems unlikely that direct use of the algorithms on-site will be practicable.

References

Brown, S. S. [1980] Optimal Search for a Moving Target in Discrete Time and Space, Oper. Res. (to appear).

McCabe, B. J. [1974] Searching for a One-Dimensional Random Walker, J. Appl. Probab. 11, pp. 86-93.

Stewart, T. J. [1979] Search for a Moving Target When Searcher Motion is Restricted, Computers and Oper. Res. 6, pp. 129-140.

OPTIMAL STOPPING OF THE INVESTIGATION SEARCH

Takasi Kisi

National Defense Academy, Japan

A two-stage search problem is dealt with in this paper. The first stage is search of preliminary nature, where false alarms are inevitable. When a contact telling a possible object is obtained, the second stage of detailed search is started, which investigates whether the possible object is the actual one being sought. The investigating search is expensive or time-consuming, and so if the contact under investigation turns out to be suspicious, the investigation should stop. In this paper, the optimal stopping rule of the second stage of search is studied.

Model

It is assumed that a searcher is trying to detect as many objects as possible, the location of which is equally probable in a given search area. Random search is assumed. The time X between the start of the search and a contact has an exponential distribution with mean time denoted as $1/\lambda$, i.e.,

$$(1) \qquad P(X \leq t) = 1 - \exp(-\lambda t), \quad t>0.$$

Let the parameter λ have a constant value, i.e., the arrival of contacts is assumed to be a Poisson process with rate λ which does not vary in time and is not influenced by the past detection of objects.

A contact does not mean the detection of an object. Sometimes it is a false contact. Let the probability of false alarm be q. The probability that a contact is true, $1 - q$, is denoted as p.

When a contact comes, the second stage of search is started immediately. The second-stage search or the investigating search is often carried out by other search units equipped with detectors suitable for investigation. In our model, no specific assumptions are made as to the searcher and the detector in the second stage but only assume that if the contact is true, the time for detecting the target object has a density function h(t). Later, an exponential distribution will be assumed for h(t), i.e.,

$$h(t) = \mu \exp(-\mu t), \quad t>0, \tag{2}$$

where $1/\mu$ is the mean detection time.

If the target object is not detected in a certain time of investigating a contact, the investigation should be stopped because the contact is now suspicious; if the contact was true, the target object would have been detected earlier.

The purpose of this paper is to determine the optimal duration of investigation of a contact. First, the normal case where the remaining time for search and investigation is known to the seacher is considered, and then the case where the remaining time is uncertain to the searcher is treated for the sake of comparison.

It is noted here that three assumptions are crucial in the model: First, if the contact was false, assume that no positive information indicating that it was a false target will be obtained by the investigation. In general, the false alarm may originate from some physical objects which are not the target object being sought. In this case, the investigating search will eventually yield some positive evidence showing that the contact was false. In this model, however, the false alarm of this kind is disregarded, and only false contacts caused by noise are assumed to be present.

The second assumption is that the preliminary search produces no information when the second-stage search is in progress. It does not matter whether the preliminary search is stopped when a contact is made or if it continues during the second stage of search because the information is wasted as long as the investigation is continued. When the investigating search is stopped, the first stage of search is resumed again.

Finally, it is assumed that the possibility of obtaining a detection from a contact is completely lost when one stops investigating the contact. That is to say, resumption of investigation after a stop never yields success even if the contact was true.

Restricted Time for Search and Investigation

Let $g(t)$ denote the maximum expected number of detections obtainable when a time t remains for search and investigation. Similarly, the conditional maximum expected number $f(t,\tau)$ is also defined; $f(t,\tau)$ means that a contact has arrived at the instant t and the searcher employs an arbitrary stopping time τ for the investigation of this first contact. The expected number $f(t,\tau)$ is made up of two terms corresponding to the case where the first contact yields a detection and the case of stopping without any detections:

$$(3) \quad f(t,\tau) = p \int_0^\tau h(u)\ (1 + g(t-u))du + (1 - p \int_0^\tau h(u)du)\ g(t-\tau).$$

The maximum expected number $f(t)$ given that a contact is made when time t remains before the termination of the whole operation is, therefore,

$$(4) \quad f(t) = \max_\tau f(t,\tau).$$

Of course, there is another relationship between $g(t)$ and $f(t)$:

$$(5) \quad g(t) = \int_0^t \lambda\exp(-\lambda u)\ f(t-u)du.$$

Equations (3), (4), and (5) determine the functions $g(t)$, $f(t)$, and $\tau = \tau(t)$, the optimal stopping rule.

Differentiating (3) with respect to τ gives the necessary condition for the local maximum of $f(t,\tau)$, i.e.,

$$(6) \quad g'(t-\tau) = ph(\tau)/\{1 - p \int_0^\tau h(u)du\}.$$

In the following, the exponential distribution (2) is assumed for the detection time. First, an example solution to a discrete version model is shown in Figure 1, where the time increment is assumed to be unity and the parameters are $\lambda = 0.1$, $\mu = 0.05$, and $p = 0.2$. It should be pointed out that an abrupt change in $\tau = \tau(t)$ followed by "ripples" is observed, whereas there are no such discontinuities in both curves $f = f(t)$ and $g = g(t)$.

When t is sufficiently large, one can show that the optimal $\tau(t)$ approaches a limiting value τ, which is given by the next equation:

$$(7) \quad \exp(\mu\tau) - \mu\tau - 1 = \mu/\lambda q.$$

Let the unit of time be $1/\mu$, the mean detection time in the investigation search. Then $\mu\tau$ is the optimal stopping time, and $\mu/\lambda q$ in the right-hand side is the expected interarrival time

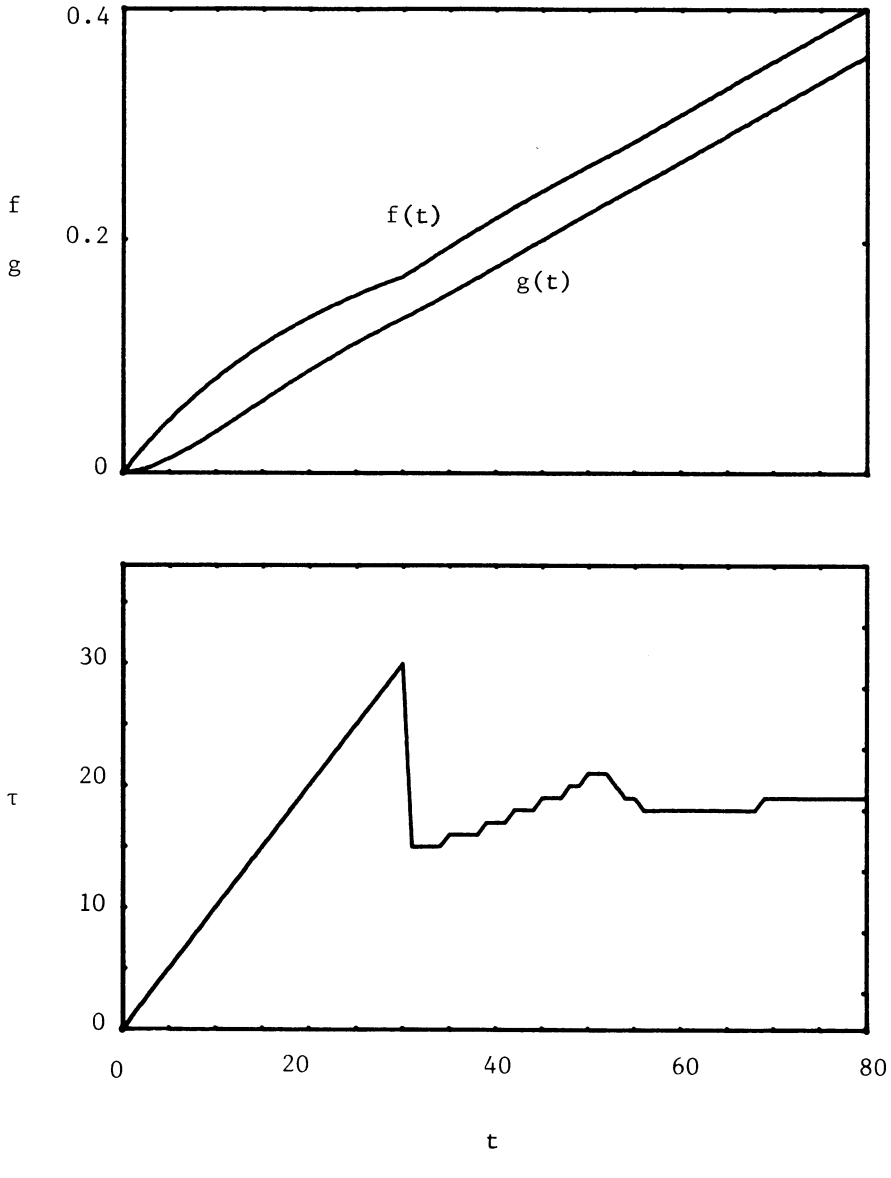


Figure 1. Sample solution to discrete time problem.

of false alarms in this unit. The curve in Figure 2 shows the relation (7). For small values of $1/\lambda q$, the optimal stopping time is very small, but it grows rapidly when $1/\lambda q$ is increased. The rate of rise of the curve, then, decreases gradually as is seen in the figure.

Uncertain Time Remaining for Search and Investigation

It is now assumed that the searcher has no exact knowledge of the remaining time for search and investigation but is informed of only its probability distribution. The exponential distribution with mean T is assumed. Therefore, the whole operation can come to an end in the next interval Δt of time with probability $\Delta t/T$ irrespective of the past history.

The maximum expected number of detections at an arbitarary instant g and the maximum conditional expected number f (assuming a contact has arrived) are defined in a similar way as in the last section. The third function $f(\tau)$ is also defined, where τ is the stopping time of the investigating search, the optimal value of which is determined in the following. As before,

$$f = \max_{\tau} f(\tau). \tag{8}$$

Now, at any instant, we can expect that a contact will arrive before the termination of the operation with a probability $\lambda/(\lambda + 1/T)$, which is readily obtained by the use of the assumption of exponential distribution of both arrivals. Therefore, instead of (5), we have

$$g = f\lambda T/\{\lambda T + 1\}. \tag{9}$$

The next relationship is the counterpart of (3):

$$f(\tau) = p(1+g)\int_0^{\tau} h(u)\exp(-u/T)du + g\exp(-\tau/T)\{1-p\int_0^{\tau} h(u)du\}. \tag{10}$$

The first term in the right-hand side corresponds to the case where the contact was true and then the target object is detected before both the termination of the whole operation and the stopping time. The second term means that if the investigation is stopped without any detections, and if the operation has not come to the end in that time, the expected number that can be obtained in the search and investigation that follow is given by g.

The necessary condition for the optimal τ is, from (10),

$$g/T = ph(\tau)/\{1-p\int_0^{\tau} h(u)du\}, \tag{11}$$

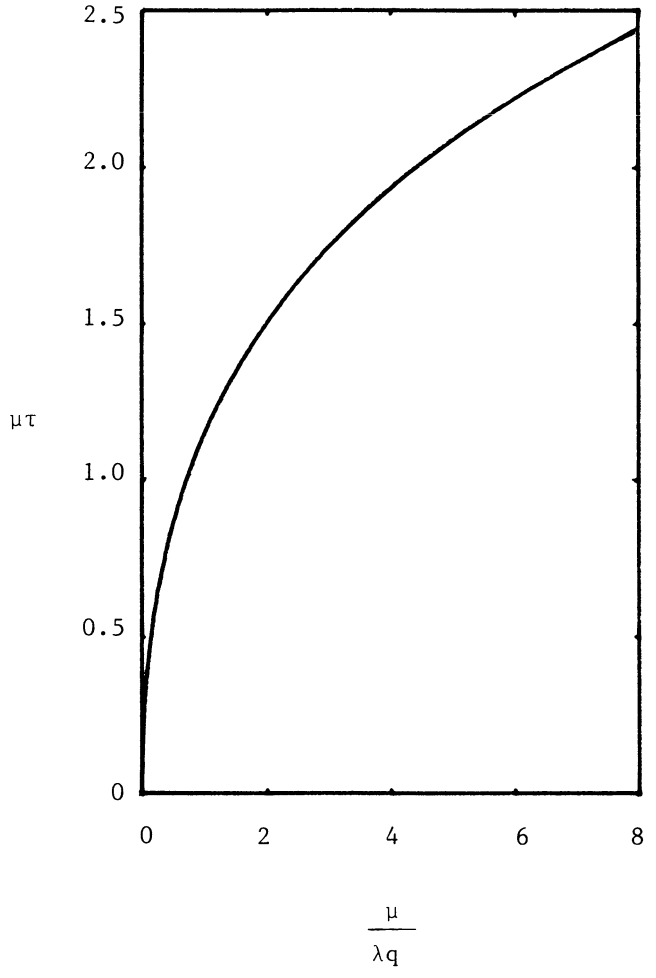


Figure 2. Optimal stopping time.

where its similarity to (6) should be pointed out. When h(t) is the exponential distribution given by (2), the following equation in τ is derived after some calculations:

$$(12) \quad \exp(\mu\tau) - \mu T\{1 - \exp(-\tau/T)\} - 1 = \{1 + 1/\mu T\}\{\mu/\lambda q\}.$$

It is easily seen that this equation reduces to (7) if the parameter T is increased to infinity. For a finite value of T, the optimal stopping time τ is found to be always larger than the corresponding value given by (7).

A SOLUTION FOR THE OPTIMAL MULTIPLE RECTANGLE SEARCH

J. H. Discenza

United States Coast Guard

The U. S. Coast Guard's lifesaving mission often requires large-scale ocean searches using multiple aircraft. A recent addition to the methods of planning such searches provides a two-dimensional probability "map" of the target's location. This simulation system (Computer-Assisted Search Planning System or CASP) provides a rational evaluation of complex scenario information but not a search plan. When search effort can be subdivided without limit, an optimal search plan can be calculated directly from a map. This analytical solution provides some valuable information but is not a practical plan, since real-life searches are composed of large rectangles of uniform effort densities. The multiple-rectangle search problem can be formulated and solved as an integer linear program, but the computational requirements prohibit its routine application. There is a subset of the problem which can be solved as a longest path problem in an acyclic network. The longest path version is the foundation of a complete operational program suitable for routine use by search planners to solve actual problems.

The process of planning a multiaircraft search is a complex one. First, the search planner must decide on the most probable position of the search target and decide on the most effective area of search. He must then construct a series of rectangles, one for each aircraft.

Each aircraft will have some fixed time endurance on scene which, at standard search speed, will fix the number of miles (L)

which can be expended in searching. If the total area in the rectangle is A, then the track spacing (S) which uses all the aircraft's time is

(1) $$S = A/L.$$

With fixed endurance, choosing a larger search area requires a larger track spacing, and vice versa. The probability of detecting the target in the search area is a function of the track spacing; as track spacing decreases (tracks closer together), the probability of detecting a target inside the area increases.

For the safety of the search aircraft, the planner does not have complete freedom in assigning search areas. Most importantly, he cannot assign different aircraft overlapping search areas.

The initial problem of determining the most probable location of the target was traditionally solved by heuristic procedures, without explicit definitions of probability.

In 1974, the Coast Guard introduced the Computer-Assisted Search Planning System (CASP), a search target simulation system, which was developed by Richardson and Stone for D. H. Wagner, Associates under Coast Guard contract. (This system is described by Richardson and Discenza [1980].) This system produces a completely general probability "map." It shows the search planner an irregular probability distribution resulting from a target simulation based on scenario data he supplies to the computer. Figure 1 shows a sample of such a map.

Using the CASP system, the search planner has explicit data concerning target location probability. He can see where the high probability regions are and can thereby estimate where he should search. But he still must manually make the tradeoff between search size and effort density.

The detection capability for a searcher is expressed as sweep width, W, as developed by Koopman [1946]. The conditional detection probability at any point of application is a function of the effort density z at that point. There are several operational functions, but the simplest and most powerful is called the Random Search Law, as developed by Koopman, i.e.,

(2) $$b(z) = 1 - \exp(-z).$$

Both W and L are given for each aircraft as part of the problem, and their product is called available effort, Z=WL. Effort density can then be alternatively expressed:

(3) $$z = W/S = WL/A = Z/A.$$

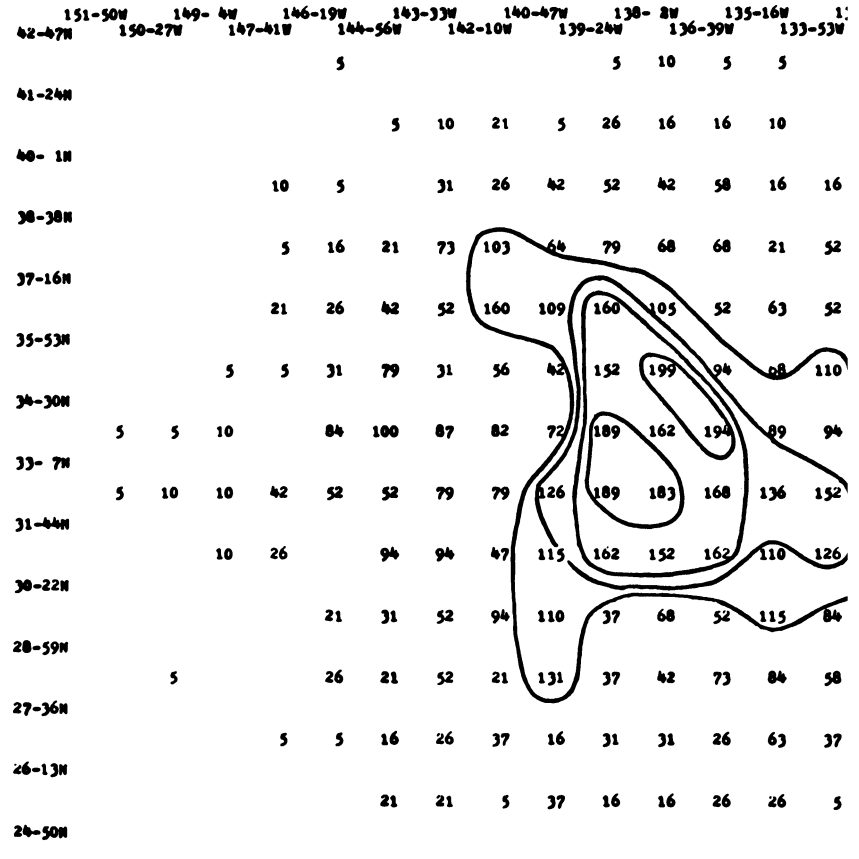


Figure 1. Sample CASP probability map.

Notes: 1. The number in each cel
probability of the tar
2. Contours were drawn by

The search planner s problem is to find the location and size ot a rectangle for each aircratt in the search. He must balance a desire to cover a large area against the need to obtain a high conditional detection probability.

If one could apply a different etfort density z(x,y) at every point in the search area, a theoretical derivation trom Koopman [1946] would provide the optimal etfort distribution. Koopman s derivation is based on random search detection probability from (2),

$$b\{z(x,y)\} = 1 - \exp\{-z(x,y)\}. \tag{4}$$

The optimal plan under these conditions has the etfort distribution:

$$z^*(x,y) = \begin{cases} \ln\{p(x,y)/\lambda\} & \text{if } p(x,y) \geq \lambda \\ 0 & \text{if } p(x,y) < \lambda \end{cases} \tag{5}$$

with the interpretation that there is some critical probability λ, a tunction ot the target probability distribution and the total etfort z. No searching is done where the target's probability density is less than λ.

If the target distribution is irregular but can be described by a grid ot n cells, where p(i) is the probability that the target is in cell (i), then there is a straightforward technique described by Charnes and Cooper [1958] which can be used to tind λ in (5). The method is described in algorithmic form by Stone [1975], p.57.

Equation (5) could provide a practical plan for CASP only when it is possible to assign a different etfort to each grid cell in a CASP map. Although such a search plan is generally not practical, the computation ot this plan will provide an upper bound on the total probability ot success for a given amount of effort. Really practical plans using the same amount of etfort will have smaller probabilities of success, but a plan might be considered satistactory if its success probability approaches this upper bound.

If there is only one aircratt with etfort z available for search, the integer interpretation allows the following method of solution: Generate all possible rectangular combinations ot integer numbers ot grid cells. Let each rectangular combination be designated search area h, h=1,2,...,H. Each search area will be composed of exactly N_h cells. We assume that each cell contains

equal area a, so that the area of each search rectangle is aN_h. Thus, each search area will have uniform effort density z_h,

$$z_h = Z/(aN_h).$$

Then each search rectangle will have an associated probability of success, C_h,

$$C_h = b\{Z/(aN_h)\} \sum_{i \epsilon N_h} p(i).$$

We then solve the problem by choosing a rectangle h which provides

$$C_{h^*} \geq C_h \qquad \text{for } h = 1,2,\ldots,H.$$

But suppose there are K aircraft available, each with a different amount of effort z_k. This variation in effort can arise by having different types of aircraft or by using aircraft based in different locations so that they have unequal search time left after deducting transit time.

We still generate the same H different rectangles, but now for each of the rectangles there are K different success probabilities. This is because each rectangle may be assigned to any one of the K aircraft. There are then m = KH possible aircraft-rectangle assignments.

Let C_j be the success probability associated with each aircraft-rectangle assignment. Let the variable X_j be equal to 1 if assignment j is in the search plan, and let $X_j = 0$ if it is not. The objective function then becomes

$$\text{(6)} \quad \text{maximize} \quad \sum_{j=1}^{m} C_j X_j.$$

Overlapping search areas are not permitted. In other words, no cell may appear more than once in the final search plan. Form the cell-search incidence matrix A_1 as follows: if cell i is in search j, let $a_{ij} = 1$; otherwise, $a_{ij} = 0$. The nonoverlapping constraints are then

$$\text{(7)} \quad \sum_{j=1}^{m} a_{ij} X_j \leq 1 \qquad \text{for } i = 1,2,\ldots,n.$$

Further, no aircraft may be assigned more than once in the search plan. Form the aircraft-search incidence matrix A_2 as follows: if search j is associated with aircraft k, let $a_{n+k,j} = 1$; otherwise, $a_{n+k,j} = 0$. The aircraft assignment

constraints are then

$$(8) \qquad \sum_{j=1}^{m} a_{n+k,j} X_j \leq 1 \qquad \text{for } k = 1,\ldots,K.$$

If e represents an appropriately dimensioned vector of 1's, the matrix representation of the integer program formulated thus far becomes

$$\begin{aligned} &\text{maximize} && CX \\ &\text{subject to} && \\ &(9) && A_1X \leq e \\ &(9) && A_2X \leq e \\ &(9) && X_j = 0 \text{ or } 1. \end{aligned}$$

The maximum possible number of rectangles H is a function of the size of the grid chosen. If a rectangular grid has I cells in one direction and J cells in the other, the maximum number of different rectangles is

$$H = \tfrac{1}{2}J(1+J)\ \tfrac{1}{2}I(1+I).$$

A 20x30-cell grid will then generate 97,650 rectangles. With 5 aircraft, the integer program would have a maximum of 488,250 variables, with 20x30+5 = 605 constraints.

While it may be possible to obtain a solution to an integer programming problem of this magnitude, the prospect of having to solve it several times a day is not very attractive. If, however, this problem or an equivalent formulation can be proved always to provide integer answers to the associated linear program, we might still have a computationally feasible formulation.

If the constraint matrix of a linear programming problem has certain qualities, all the extreme points on the constraint polyhedron are integer-valued.

Hoffman and Kruskal [1958] first showed that given a constraint set $[x \mid A\,x \leq b, x \geq 0]$ with arbitrary integer b, total unimodularity of A is necessary and sufficient for all extreme points to be integer.

Total unimodularity is the strongest criterion for integrality because it holds for any arbitrary integer vector b. For certain problems which restrict A and b, less demanding criteria will also guarantee integrality.

Berge [1972] defined a balanced hypergraph. A (0,1)-matrix is said to be balanced if it is the incidence matrix of a balanced hypergraph. A totally unimodular matrix is balanced, but the converse is not true.

Padberg [1974] characterizes perfect matrices in terms of forbidden submatrices.

Problems of this class where the right hand sides are all 1's are called set-packing problems, and the matrix A is called perfect if every vertex of the constraint polyhedron is integer.

A balanced matrix is perfect, but again the reverse statement is not true. Since the formulation in (9) corresponds to the set-packing problem, it may fail to be totally unimodular or balanced and still have all integer solutions if it is perfect. To show that (9) does not have the integer property, it must be shown that the (n+K) by m matrix of coefficients (a_{ij}) has at least one of Padberg's forbidden submatrices.

Figure 2 shows a diagram of search rectangles which represent a submatrix having Padberg's forbidden property. Here, the characteristics of the forbidden submatrices can be related to the original physical problem. In this way, we can see how to restrict the orginal problem to eliminate these forbidden submatrices, and thus guarantee the integer property.

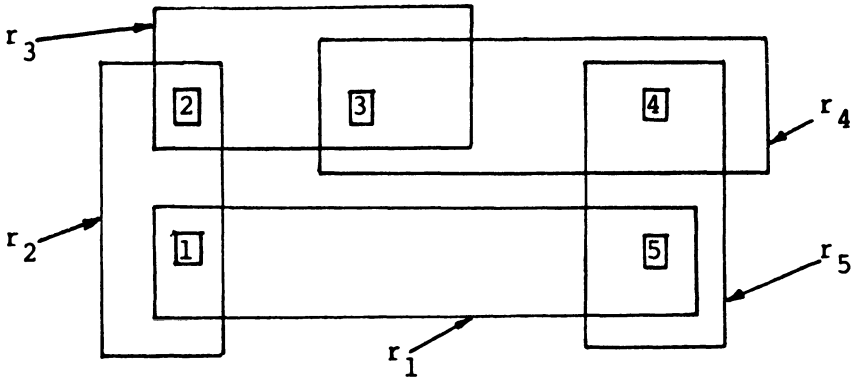


Figure 2. Rectangle arrangement which is a subset of the linear program solution set of (9), which has the fractional property. This does not violate any constraints if the variables corresponding to searches r_1, ..., r_5 have the value ½.

Rule: For each different aircraft, choose a region of the grid to contain all its assigned search areas. Select one pair of adjacent rows in that region for that aircraft (no two aircraft with the same pair) and permit for that aircraft only search areas occupying at least one of those rows. Let no other aircraft's search area occupy both those rows.

The resulting constraint matrix is balanced. Observe the corresponding hypergraph where the vertices representing rows from A are placed between the two critical rows of the Rule, as in the example of Figure 3. This hypergraph is balanced because it is a family of intervals of points on a straight line (Berge [1972]).

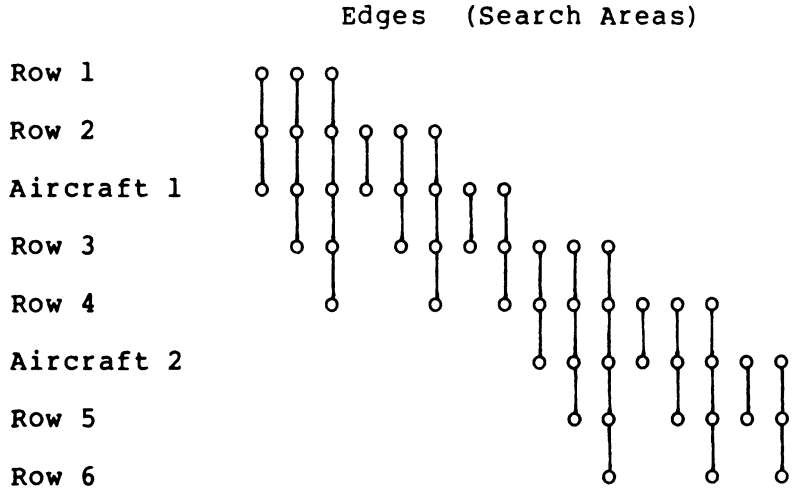


Figure 3. Edges of the hypergraph of constraint matrix when the Rule is observed. In this problem there are 6 probability grid rows and 2 aircraft. The vertices representing the aircraft constraints have been placed between the vertices representing each aircraft's two critical rows, (2,3) and (4,5), in this example. The hypergraph is balanced.

The constraint set A of a family of points along a straight line, when associated with a problem

$$
\begin{array}{lll}
\text{Max} & CX & (\text{or MIN} \quad CX \\
\text{s.t.} & AX \leq e & \text{s.t.} \quad AX > e \\
X = 0 \text{ or } 1 & & X = 0 \text{ or } 1)
\end{array} \tag{10}
$$

converts directly to a longest (or shortest) path problem by manipulation similar to the employment scheduling problem in Wagner [1975], page 23. Where Wagner derives his clue to this manipulation from the "triangular appearance of the coefficients," the real criterion is that (perhaps after some rearrangement of rows) the 1's in each column form an uninterrupted string.

The transformation to a longest (or shortest) path proceeds as follows: Add two dummy rows (0) and (1'), consisting of all zeroes, to the top and bottom of the constraint matrix. Form a new matrix using every row (1) to (1') minus the row above it as shown in Figure 4. The resulting longest path problem can be related to the original optimization problem directly and forms the basis for an efficient algorithm to solve the simplified problem.

Figure 5 shows a set of search plan solutions based on the CASP probability distribution of Figure 1 for four, six, and nine aircraft using the longest path algorithm.

The technique of examining a constraint matrix for the integral property is common practice in integer programming studies. Unfortunately, there are no easily tested necessary and sufficient conditions for the totally unimodular, balanced, or perfect qualities of a general matrix.

In the search problem, we have examined the characteristics of forbidden submatrices (according to Padberg) as they apply to the original problem. That portion of a solution represented by a forbidden submatrix has a fractional property. Elimination of certain portions of the original problem to deny specific classes of fractional solutions has eventually resulted in a subproblem with the integral property.

All three qualities (totally unimodular, balanced, and perfect) have forbidden submatrices which constitute necessary and sufficient conditions. The strategy demonstrated here, of relating the forbidden submatrices to the original problem, may be very useful for simplifying a wide variety of difficult combinatorial problems.

The opinions or assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Commandant or the Coast Guard at large.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	S1	S2	S7	S3	S4	S8	S5	S6	
(1)	1	1	1														1								= 1
(2)	1	1	1	1	1	1												1							= 1
(7)	1	1	1	1	1	1	1	1											1						= 1
(3)		1	1		1	1	1	1	1	1	1									1					= 1
(4)			1			1		1	1	1	1	1	1	1							1				= 1
(8)									1	1	1	1	1	1	1	1						1			= 1
(5)										1	1		1	1	1	1							1		= 1
(6)											1			1		1								1	= 1
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	0	0	0	0	0	0	0	0	Max

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	S1	S2	S7	S3	S4	S8	S5	S6	
(1)	1	1	1														1								= 1
(2)				1	1	1											-1	1							= 0
(7)							1	1										-1	1						= 0
(3)	-1			-1					1	1	1								-1	1					= 0
(4)		-1			-1		-1					1	1	1						-1	1				= 0
(8)			-1			-1		-1							1	1					-1	1			= 0
(5)									-1			-1										-1	1		= 0
(6)										-1			-1		-1								-1	1	= 0
(1')											-1			-1		-1								-1	=-1
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	0	0	0	0	0	0	0	0	Max

Figure 4. Constraint matrix from Figure 3 transformed into a longest path problem in an acyclic network.

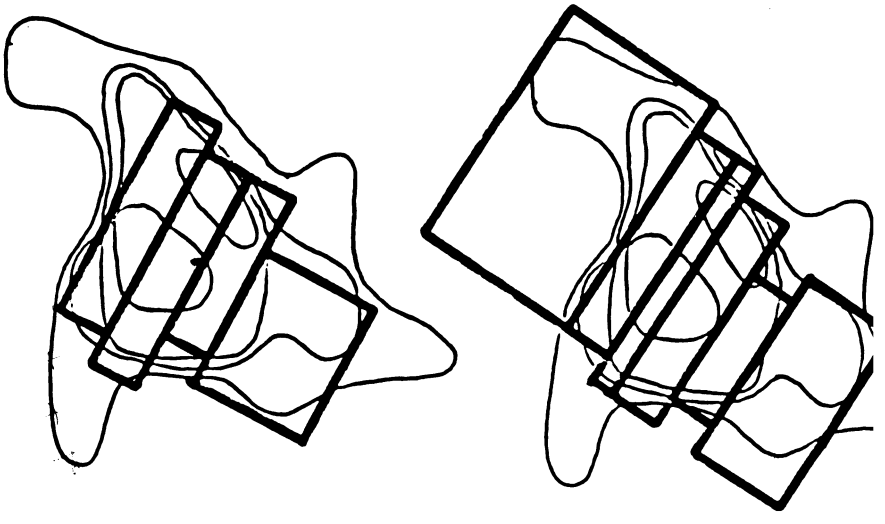


Figure 5. Four, six, and nine aircraft search plan

References

Berge, C. [1972] Balanced Matrices, Math. Programming. 2, pp. 19-31.

Charnes, A. & Cooper, W. [1958] The Theory of Search: Optimum Distribution of Search Effort, Management Sci. 5, pp. 44-50.

Hoffman, A. & Kruskal, J. [1956] Integral Boundary Points of Convex Polyhedra in "Linear Inequalities and Related Systems," H. Kuhn and A. Tucker, eds., Annals of Maths. Study No. 38, Princeton Univ. Press.

Koopman, B. O. [1946] Search and Screening, OEG Report No. 56, Washington, D. C. (a modernized edition has been published by Pergamon Press, Inc., New York, 1980).

Koopman, B. O. [1956] The Theory of Search, Parts I, II, and III, Oper. Res. 4, pp. 324-348 and 503-531; 5 [1957] pp. 613-623.

Lawler, E. [1976] "Combinatorial Optimization: Networks and Matroids," Holt, Rinehard and Winston, New York.

Padberg, M. [1974] Perfect Zero-One Matrices, Math. Programming. 6, pp. 180-196.

Richardson, H. R. & Discenza, J. H. [1980] The United States Coast Guard Computer-Assisted Search Planning System (CASP), Naval Res. Logist. Quart. (to appear).

Stone, L. D. [1975] "Theory of Optimal Search," Academic Press, New York.

Wagner, H. M. [1975] "Principles of Operations Research," Prentice-Hall, Englewood Cliffs, New Jersey.

PARTICIPANTS

BAYRAKTAR, Dr. Bulent, Scientific Affairs Division, NATO, Brussels, Belgium.
BERTHON, M., Ingénieur d'études, Société AERO, 3 Avenue de l'Opéra, 75001 Paris, France.
BRESSON, M., AERO, 3 Avenue de l'Opéra, 75001 Paris, France.
BROWN, Dr. Scott S., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.
BUTERA, Col. James L., U. S. Air Force, Director Inland Search and Rescue, Headquarters ARRS, Scott AFB, Illinois 62225 USA.
CORWIN, Dr. Thomas L., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.
COZZOLINO, Prof. John M., University of Pennsylvania, Philadelphia, Pennsylvania 19104 USA.
DISCENZA, Cdr. Joseph H., U. S. Coast Guard, 7506 Mendota Place, Springfield, Virginia 22150 USA.
FERGUSSON, J. G., O. R. Executive, National Coal Board, Coal House, Lyon Road, Harrow, Middlesex, HA1 2EX England.
FIELD, E. A., Esso Petroleum Co., Ltd., Esso House, Victoria Street, London SW1 England.
FRIEDRICH, Prof. Dr. G., Institut für Mineralogie und Lagerstättenlehre, RWH Aachen, 51 Aachen, Germany.
GRASTY, R. L., Geological Survey of Canada, Department of Energy, Mines and Resources, 601 Booth Street, Ottawa, Ontario, Canada KIA OE8.
HALEY, Prof. K. Brian, Department of Engineering Production, the University of Birmingham, Edgbaston, Birmingham B15 2TT England.
HANNA, LCdr. John, U. S. Coast Guard, Governor's Island, New York 10004 USA.
HARTLEY, Dr. D. A., Royal Military College of Science, Shrivenham, England.
HAYWOOD, Keith H., The Whitefish Authority, St. Andrew's Dock, Hull Yorkshire HU3 4QE England.
HILLBORN, Ray W., Institute of Animal Resource Ecology, 2075 Wesbrook Mall, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1W5.

HYPHER, R. P., Cmd. Operations Research Adv., Air Transport Group, Operations Research Branch, Canadian Forces Base, Trenton, Astre, Ontario, Canada KOK 1BO.

JEAN, Prof. Michel, Collège Militaire Royal de Saint-Jean, Saint-Jean, Quebec, Canada JOJ1RO.

KADANE, Prof. Jospeh B., Carnegie-Mellon Univ., Pittsbg. Pa.15213USA

KISI, Prof. Takasi, Department of Applied Physics, National Defense Academy, Yokosuka, Japan.

KOLESAR, Prof. Peter, Graduate School of Business, Columbia University, New York, New York 10027 USA.

KOOPMAN, Prof. Bernard O., Randolph Hill Rd., Randolph, N.H.03570USA

KUWAHARA, Dr. Ronald H., Defense Research Establishment Pacific FMO Victoria, B. C. VOS IBO Canada.

LENOBLE, J. P., Centre National pour L'Exploitation des Océans, 66 Avenue d'Iena, 75763 Paris Cedex 16, France.

LEONARDO, Dr. Martin, Naval Air Devlopmt Ctr.,Warminister, Pa.18974USA

LEVINE, Docteur, Centre d'Automatique et d'Informatique, Ecole des Mines 357 rue Saint Honore, 77300 Fountainbleau, France.

MANGEL, Dr. Marc, Center for Naval Analyses, 2000 N. Beauregard Street, P. O. Box 11280, Alexandria, Virginia 22311 USA.

MATTSON, Lt. Col. Robert, U. S. Air Force, Chief SAR Activities HQ CAP-USAF/DOSS, Maxwell AFB, Alabama 36112 USA.

PAYNE, C. D., Home Office, Horseferry House, Dean Ryle Street, London SW1P 2AW England.

POLLOCK, Prof. Stephen M., Department of Industrial Engineering, University of Michigan, Ann Arbor, Michigan 48104 USA.

REISS, Manfred, c/o IABG mbH, Abt. Systemstudien SOM, Einsteinstr. 20, D-8012 Ottobrunn, Germany.

RICHARDSON, Dr. Henry R., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.

SOLOMON, Dr. Herbert, Chief Scientist, Office of Naval Research, Branch Office, London 222/231 Old Marylebone Road, London NW1 5TH England.

STEWART, Dr. T. J., Head Operations Research and Statistics Division, National Research Institute for Mathematical Sciences, P. O. Box 395, Pretoria, 0001 South Africa.

STONE, Dr. Lawrence D., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.

STRUMPFER, Dr. Johan, Institute of Maritime Technology, P. O. Box 181, Simonstown 7995 South Africa.

WAGNER, Dr. Daniel H., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.

WASHBURN, Prof. Alan, Naval Postgraduate School, Monterey, California 93940 USA.

WEISINGER, Dr. James R., Daniel H. Wagner, Associates, Station Square One, Paoli, Pennsylvania 19301 USA.

ZIMMERMAN, Prof. Dr. H.-J., Institut für Wirtschaftenwissenschaften, Lehrstuhl fur Unternehmensforschung, Rheinish-Westfälische Technische Hochshule, Aachen, Germany.

INDEX

Aircraft, 9, 10, 50, 73, 71ff, 87, 93, 111, 127, 261, 255
Algorithm, 221, 233, 239, 242, 247-253
Area Scenario, 30

Bathymetric, 193
Bayes, 24, 103, 207
Border Patrols, 107
Borel, 39
Boustrophedon, 14
Branch and Bound, 247
Brownian Bridge, 40

Calculus of Variation, 19
Canada, 73-77, 211
Coal, 152ff
Coast Guard (US), 6, 50, 61ff, 261
Colliery, 166
Computers, 21, 47, 88, 127, 133, 234, 250, 264
CAS (Computer Assisted Search), 6, 25-40, 59
CASP, 6, 30, 61ff, 84, 201, 261
Contraint (Effort), 3, 11, 20, 257
Constraint (Motion), 241
Convoy, 17
Cost, 15, 80, 162, 176, 185
Cosmos-954, 211

Data, 66
Decision Rules, 56, 57
Defectives, 208
Density in Search, 17
Detection Function, 1, 11, 19, 20, 119, 221, 233, 241
Detection Equipment, 13, 183, 185
Detection Probabilities, 18, 83, 119, 147, 203
Discrete Time, 123
Dynamic Programming, 124

Early Warning, 104
Effort, 11, 109, 257
FLT (Emergency Locator Transmitter), 51, 79, 88
ENIGMA, 16
Enforcement, 109
Equipment, 59
Euler, 21
Exploration, 7, 155, 159-185
Experimental, 60
Exploitation Plan, 169, 176, 185
Eye, 199
Exponential Detection, 221, 233

False Contact, 36, 63, 118, 255
Free Fall Grab, 185
Fisheries, 107, 109

Gaming, 167
Gauss-Markov, 170
Geological, 156, 167, 171
Glaucoma, 199

Heuristic, 203
Hidden Object, 205
Hierarchical Surveillance, 102
Human Factors, 46
Hypergraph, 268

Infra-red, 141
Information Processing, 25 etseq
  Negative, 24
  Positive, 36
Industrial Application, 205
Interactive System, 47

Lagrange Multiplier, 21
Land Search, 6, 87, 94
Linear Programming, 266

Magnetometer, 185
MAD (Magnetic Anomoly Detectors), 13, 127
Manganese, 183, 185
Man Machine, 47
Marine, 65, 77, 141, 183-185
Markov, 39, 226, 233
Measures of Performance, 58
Medicine, 199
Monte-Carlo, 25, 199, 243
Motion (Relative), 13
Moving Target, 221, 231, 249-250, 253
Minerals, 165, 173, 183, 185
Mining, 166, 185
Military, 12
Mission Activity, 46
Multiple Search, 12
Multi-Scenario, 22ff
Multiple Criteria, 56
Multiple Sensor, 64

Navigation, 187
Navy (US), 6
Nuclear Emergency, 211
Negative Result, 24
Nodules, 173-185-185
Non-reactive Search, 12
Non-linear Programming, 202

Operational Gaming, 167
Oil Exploration, 7, 161
Oil Wells, 210
Objectives, 56, 163
Optimal Search Plan, 3-5, 20, 141, 221-245, 247-253, 257
Organisation, 93
OTV (Offset and Track Variables), 80

Padberg, 267
Parallel Sweeps, 14, 55
Passive Search, 12, 17
Patterns of Search, 55, 111, 142
Performance Measures, 58
Persons, 87, 93, 106
Planning Search, 5, 54, 162
Priorities, 93
Prior Distribution, 156, 195
Positive Information, 36
Police, 106
Pollution, 106
Polymetallic Nodules, 173-183
Probability Distribuiton (Target), 1, 10, 18, 30, 156
Probability Map Display, 28
Probability of Success, 57, 83, 124, 147, 231
Problem Definition, 160
Progress of Theory, 3

Quality Control, 207
Quantum Theory, 13

R and D, 209
Radioactive, 214
Radiolaria, 176
Radar, 13, 141
Random Search Law, 15, 255, 262
Range Law, 22, 128
Region of Search, 22
Rescue (Land), 6
Real Time Systems, 47
Random Walk, 237
Resource Allocation, 52, 54, 84

SAR Search and Rescue, 45-66a, 73-77
Satellites, 79, 211
Scorpion, 28
Sea, 17, 185
Scotoma, 199
Search
  Cost, 15, 80
  Active, 12
  Effort, 11
  Types, 12
  Non-reactive, 12

Search (continued)
  Passive, 12, 17
  Single Try, 13
  Repetative, 14
  Subdivisable, 14, 17
  Random Law, 15
  Additative, 14, 17
  Density, 18
  Region, 22
  Management, 46
  Length, 50
  Plan, 54, 223, 241
  Patterns, 55, 111, 142, 239, 264
  Grid, 239, 264
Search Master Co-ordinator, 48, 80, 94
SEP (Search Effectiveness Probabilities), 35
Sensors, 17, 51, 61, 79, 88, 127, 141, 185
Ships, 141, 188
Sonar Escort Screen, 17, 185
Submarines, 6, 9, 10, 17, 28, 127
Stochastic Process Models, 38, 239
Stopping, 51, 91, 257
Starting, 51
Sweep Width, 20, 60, 142, 189, 262
System Variables, 17
Simulation, 55, 127, 142, 226
Stationary Target, 222

Target
  Stationary, 222
  Location, 1, 33, 221, 231, 239
  Distribution, 10, 147, 264
  Multiple, 53
  multistate, 53
  Identification, 101
  Destination, 162
Tracking Scenario, 30
Track Length, 83, 189
Training, 48, 80, 139
TV, 184, 185

U-boat (see submarines)
Under Sea Exploration, 173, 183, 185

Variograms, 176
Vision Loss, 199

Wald, 125
Wiener Process, 38