SPECIAL ISSUE: APPLICATION OF ADVANCED OPERATIONS RESEARCH METHODS TO MOBILITY MODELING

# Military Operations Research

*Winter 1996*
V1, N4

A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY OPERATIONS RESEARCH SOCIETY



MORS
MILITARY OPERATIONS RESEARCH SOCIETY

# Military Operations Research

*A publication of the Military Operations Research Society*

The Military Operations Research Society is a professional society incorporated under the laws of Virginia. The Society conducts a classified symposium and several other meetings annually. It publishes proceedings, monographs, a quarterly bulletin, *PHALANX*, and a quarterly journal, *Military Operations Research*, for professional exchange and peer criticism among students, theoreticians, practitioners and users of military operations research. The Society does not make or advocate official policy nor does it attempt to influence the formation of policy. Matters discussed or statements made in the course of MORS symposia or printed in its publications represent the opinions of the authors and not the Society.



*Military Operations Research*, A Journal of the Military Operations Research Society (ISSN 1082-5983) is published quarterly by the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304-3418. The domestic subscription price is $40 for one year and $75 for two years; international rates are $80 for one year and $150 for two years. Application to Mail at Second-Class Postage Rates is Pending at Alexandria, VA.

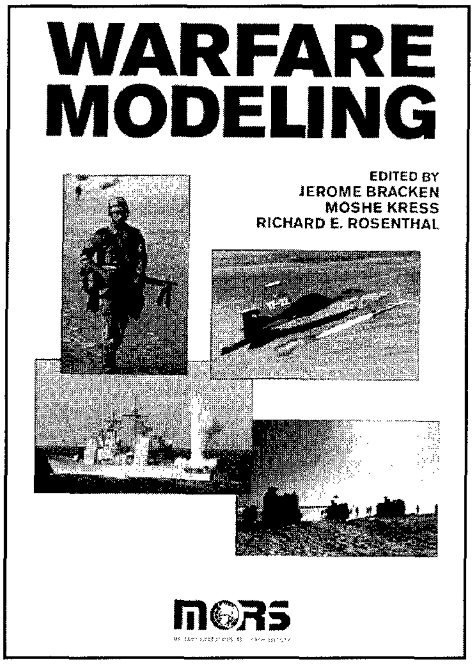
**POSTMASTER:** Send address changes to *Military Operations Research*, A Journal of the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304.

# Table of Contents

*Volume 1, Number 4*

*Winter 1996*

# MORS To Publish *Warfare Modeling*




As mentioned in the December 1995 *PHALANX* article "Let's Publish," one objective of the Military Operations Research Society Publications Committee is to ensure that publications of enduring value are made available to the MORS membership. *Warfare Modeling* falls in this category and will be available in June 1996.

*Warfare Modeling* is a single volume of three special issues of *Naval Research Logistics, An International Journal*, along with a foreword by **Wayne Hughes, FS**. This collection of military operations research articles, written by experts in their respective fields, was edited by **Jerome Bracken**, Yale University; **Moshe Kress**, Center for Military Analysis, Israel; and **Richard Rosenthal**, Naval Postgraduate School.

The book is comprised of seven sections:
(1) Theater-Level Modeling
(2) Mathematical Models of Combat
(3) Historical Analysis
(4) Weapon System Analysis
(5) Command, Control, Communications and Intelligence
(6) Cost Effectiveness Analysis
(7) Modeling in Support of Operations Desert Shield and Desert Storm

The criteria for deciding to undertake this new offering was for the publication to have enduring value. *Warfare Modeling* was evaluated as having enduring value by a number of MORS members who read the articles, by professors who use them as class reference material, and by the editors of the publication. Each of the twenty-five chapters represents the state of the art in various aspects of warfare modeling.

MORS is excited about the opportunity to update our members, and their bookshelves, with current reference material that addresses many dimensions of military operations research.

*Warfare Modeling* will be available at the 64th Symposium for $35. Please call the MORS office for more details about this publication. (703) 751-7290

---

☐ Please send me ______ copies of *Warfare Modeling*. I have enclosed $40 for each copy, which includes the freight. VA residents include $1.58 sales tax for each copy.

Name: ____________________
Address: ____________________
City: ____________________ State: __________ Zip: __________

**Payment must be made in US dollars drawn on a US bank.**

# From the Editor

**Professor Peter Purdue**
*Naval Postgraduate School*

This issue of Military Operations Research represents two changes: it is the first issue produced by a guest editor and is the last issue to appear under my editorship.

Professor Yupo Chan has done a great job in selecting for publication some of the papers that were presented at the "Third Mobility Modeling and Simulation Conference." The conference was held at the Air Force Institute of Technology, Wright Patterson AFB, 25–26 May 1995. All of the papers appearing in this issue of the journal were subjected to a rigorous reviewing process. I appreciate all of the work that Yupo, his reviewers, and the authors have put into this project so that the results of the conference could appear in a timely fashion. I also thank Brig Gen Thomas Case, USAF for providing the foreword to this issue.

On 31 December 1995 I completed my three-year term as editor of Military Operations Research. My main goal in accepting the position was to provide the military operations researcher with a high quality, professional outlet for publishing interesting and significant work in the field. I believed, and continue to believe, that MORS should publish the definitive journal in support of the military operations research professional.

In the final analysis, the quality of any journal depends upon the cooperation and support of the community it serves. The editors and reviewers can establish the criteria that papers must meet to be accepted for publication, but first and foremost they must have papers to review. The military OR community has allowed MOR the luxury of setting high standards by submitting high quality papers for consideration. As a result we have been able to produce a high quality, interesting and readable journal (Yes, I know that I am biased but all of the hard work was done by the authors and reviewers.)

My thanks to the whole community for the support I have received in what was for me a challenging but very delightful project. My only regret is that in my overly zealous desire to personally read every paper, some papers failed to receive the timely and through review they deserved. To the authors: my thanks for your patience and apologies for the delays you had to endure. Many thanks to Michael Cronin of the MORS office for the superb support he provided ever since he joined the MORS staff. Without his dedicated efforts to get me to stay somewhere close to schedule, and his excellent work with printers, we would still be waiting for some of the early issues. I would also like to thank Dick Wiles and the rest of the MORS staff for supporting the concept of a new journal from the very beginning. Finally, this whole enterprise would not have gotten off the ground without the support of the Board of Directors and a couple of MORS presidents who were willing to take a chance; we all owe them a big vote of thanks!

It has been a fun ride! I wish the new editor, Greg Parnell, all the best and look forward to an ever improving publication.

# RIST PRIZE CALL FOR PAPERS

MORS offers two prizes for best papers—the **Barchi** Prize and the **Rist** Prize. The Rist Prize will be awarded to the best paper in military operations research submitted in response to this Call for Papers. The Barchi Prize will be awarded to the best paper from the entire 64th symposium, including Working Groups, Composite Groups, and General Sessions.

**David Rist Prize:** Papers submitted in response to this call will be eligible for consideration for the **Rist Prize**. The committee will select the prize-winning paper from those submitted and award the prize at the 65th MORSS. If selected, the author(s) will be invited to present the paper at the 65th MORSS and to prepare it for publication in the MORS journal, *Military Operations Research*. The cash prize is $1000. To be considered, the paper must be mailed to the MORS office and postmarked no later than **September 30, 1996.** Please send the original, three copies and the disk.

**Richard H. Barchi Prize:** Author(s) of those papers selected as the best from their respective Working Group or Composite Group, and those of the General Sessions at the 64th MORSS will be invited to submit their paper for consideration for the **Barchi Prize**. The committee will select the prize-winning paper from among those presented, nominated and submitted. The prize will be presented at the 65th MORSS. The cash prize is $1000. To be considered, the paper must be mailed to the MORS office and postmarked no later than November 30, 1996. Please send the original, three copies and a disk.

### Prize Criteria

The criteria for selection for both prizes are valuable guidelines for presentation and/or submission of any MORS paper. To be eligible for either award, a paper must, at a minimum:

- Be original and a self-contained contribution to systems analysis or operations research;
- Demonstrate an application of analysis or methodology, either actual or prospective;
- Prove recognizable new insight into the problem or its solution; and
- Not previously been awarded either the Rist Prize or the Barchi Prize (the same paper may compete for but cannot win both prizes.)

Eligible papers are judged according to the following criteria:

### Professional Quality

- Problem definition
- Citation of related work
- Description of approach
- Statement of assumptions
- Explanation of methodology
- Analysis of data and sources
- Sensitivity of analyses (where appropriate)
- Logical development of analysis and conclusions
- Summary of presentation and results

### Contribution to Military Operations Research

- Importance of problem
- Contribution to insight or solution of the problem
- Power of generality of the result
- Originality and innovation

# Foreword

**Thomas R. Case,**
**Brig Gen, USAF**
*Director of Modeling,*
*Simulation and Analysis*
*DCS, Plans and Operations*

This issue of the *Military Operations Research* journal represents many things. First, it highlights the level of work being conducted by very talented professionals throughout the mobility community. It is important in this era of decreasing defense budgets and reductions in overseas presence that our mobility forces be sized to fit the most demanding but realistic future conflict. To do this, the Air Force requires the latest mobility modeling and problem solving techniques such as those presented in this journal. Second, this special issue of *Military Operations Research* subscribes closely to the goals of the journal, to: "establish channels of communication that link government, industry and academia; and to facilitate the interchange of ideas among practitioners, academics, and policy makers." The articles in this special issue are only a small, but representative, portion of a wide spectrum of presentations by operational, analytical and academic experts that were presented at the Third Air Force Mobility M&S Users' Group. Finally, this journal signifies the importance of such bodies as the Air Force Mobility M&S Users' Group. This group brings together operators, analysts and academia on a regular basis to pursue excellence in mobility modeling, simulation and analysis. The results have manifested themselves already in the development of a greater awareness of mobility M&S activities DoD-wide, the acceptance of a suite of mobility models and the pursuit of future mobility model architectures.

My congratulations to the individuals whose papers were selected for this special issue of *Military Operations Research*. The great intellect and capability of these individuals reflect the character of the entire mobility community. As we transition to a smaller, leaner force, these mobility modeling, simulation and analytical capabilities will ensure the Air Force and its sister services make the right decisions for the right reasons.

## 63rd MORS Best Symposium Papers—64th Barchi Prize Nominations

**WG1—** *An Improved Solution Methodology for the Arsenal Exchange Model (AEM)* by Capt **Jeffery D. Weir**, Capt **Michael G. Stoecker**, and LtCol **James T. Moore**, AFIT

**WG2—** *Impact of Theater Ballistic Missile Defense on the Joint Campaign* by **Alan Zimm**, Johns Hopkins University/APL

**WG3—** *A Methodology for Evaluating Military Systems in Counter Proliferation* by Capt **Stanley Stafira, Jr.**, AFSAA/SAG, Dr. **Gregory S. Parnell**, VCU and LtCol **James T. Moore**, AFIT

**WG4—** *The Revolution in Military Affairs: A Primer* by **Barry Watts**, Northrop-Grumman, Dr. **Andrew F. Krepinevich**, Defense Budget Project Office and **Michael Vickers**, OSD/NA

**WG5—** *MLR Supplemental Analysis, MV-22 Wargame* by **Edward A. Smyth**, Johns Hopkins University/APL

**WG6—** *The Nearland Test (NLT)* by Dr. **Jeff Lutz**, CNA representative to JADO/JEZ/ASCIET

**WG7—** *Effects of Tularemia on Human Performance* by **George Anno**, Pacific-Sierra Research Corporation and **Arthur P. Deverill**, ARES Corporation

**WG9—** *AMRAAM P3I COEA Results* by LtCol **Martin Allen**, AFSAA/SAGW and Maj **Eileen A. Bjorkman**

**WG10—***We Have Met the Enemy and...* by Maj **Cyrus Holliday**, HQ FORSCOM

**WG11—***USSOCOM COEA of the Advanced Multi-Mission Vertical Lift Aircraft (MV-X)* by **William C. Fite**, ANSER

**WG12—***FAADS C3I COEA* by **Ronald Magee** and **Frank Lawrence,** US Army TRADOC Analysis Command

**WG13—***Optimization of Shipboard Self-Protection ECM Systems Using SCE Techniques* by Prof **Phil Pace,** LT **Michael S. Moreno, B. H. Nishimura**, NPS and **W. Morris** and **R. E. Surratt**, Naval Research Lab.

**WG14—***Naval Theater Level Model (NTLM)* by LCDR **Jeffrey Cares**, CFC-KOREA, Operations Analysis Branch

**WG15—***A Bayesian Perspective of Dominant Battlefield Awareness* by COL **Raymond E. Franck, Jr.**, Defense Intelligence Agency

**WG18—***Exploring the Relationship Between Tactical Intel and Battle Results* by Prof **Don Barr** and CPT **Todd Sherril**, USMA

**WG19—***The Value of Electronic Warfare: In Search of the Magic Metric* by **Robert J. Meyer**, Naval Air Warfare Center

**WG20—***System Understanding and Statistical Uncertainty Bounds from Limited Test Data* by **James C. Spall**, JHU/APL

**WG21—***Exploring Unmanned Ground Vehicle Utility Using Technology Seminar Wargaming* by MAJ **Harvey Graf**, USAMSAA

**WG22—***Do These Costs Make Any Sense? The Use and Abuse of Costs in Defense Acquisition Analysis* by **Michael W. Smith** and **Henry L. Eskew**, CNA

**WG23—***Precision Strike Capability/Joint Direct Attack Munition (JDAM), Product Improvement Program (PIP), Accuracy Requirements Study* by **William V. Beatovich**, Veda, Inc., and Maj **Jay Kreighbaum**, HQ ACC/DRPW

**WG24—***A Computer Simulation and Analysis of the Forward Surgical Team* by MAJ **Robert Syvertson**, USA, MSC

**WG25—***The Use of Non-Parametric Statistics in Marine Corps Area Assessments* by CAPT **Gregory K. Cohen**, USMC, MCCDC

**WG26—***Force Analysis Spreadsheet Tool OOTW Requirements (FASTOR)* by LTC **Joseph J. Manzo**, US Army CAA

**WG27—***Choosing Force Structures: Modeling Interactions among Wartime Requirements, Peacetime Basing Options, and Manpower and Personnel Policies* by **Craig Moore, James Kakalik, Deena Benjamin,** and **Richard Stanton**, RAND

**WG28—***Contractor Indirect Costs* by **John Cloos**, IDA

**WG29—***Analysis of America's Readiness-Based Aviation Consolidated Allowance List* by **Anne J. Hale**, CNA

**WG30—***C-17 Paratroop Jump Separation Analysis* by **Daniel D. Dassow**, McDonnell Douglas Aerospace

**WG31—***Re-Engineering Legacy Computer Wargames* by CAPT **Al Wanski**, CADRE

**WG32—***The Effects of Decision Making Quality and Timeliness on the Response Surface of a Simple Combat Simulation* by Dr. **John B. Gilmer, Jr.**, Wilkes University

**Composite Group VI—***A New Approach for Performing Cost-Benefit Analysis* by **John James, Wolf Kohn**, Sagent Corp., **Anil Nerode**, Cornell University, **Benjamin Cummings**, Army Research Lab, and **Jagdish Chandra**, Army Research Office

# Introduction

**Yupo Chan**
*Guest Editor & Chair, Conference Facility Committee*
*Department of Operational Sciences*
*Graduate School of Engineering*
*Air Force Institute of Technology*

Under the sponsorship of the Air Force Office of Scientific Research (Dr. Neal Glassman), Air Force Studies and Analysis Agency (Col. Glenn S. Geary, chair), the Air Mobility Command (Col. Craig Northrop), the Military Applications Section (Dr. Stephen Balut) and Transportation Science Section (Dr. Hani Mahmassani) of the Institute for Operations Research and Management Science (INFORMS), the Third Mobility Modelling and Simulation Users' Group Conference was held at the Air Force Institute of Technology on the 25th and 26th of May 1995, hosted by the Department of Operational Sciences (Lt. Col. Paul Auclair), Graduate School of Engineering. One of the foci of the workshop is to uncover emerging operations-research methods in solving air-mobility problems.

## I. BACKGROUND

There has been a perceived chasm between the applications community and the research community regarding the ways and means to resolve air-mobility and transportation problems in general. While there is a need for high-quality analysis to be performed on a day-to-day basis, the state-of-the-art technologies are often not brought to bear upon the problem. At the same time, the research community is equally frustrated about the lack of sophistication in analysis performed on a real-time basis, which prevents important insights to be gained and timely decisions to be made. It has been said that many of the technological advances of operations research are at least twenty years ahead of applications. In response to this problem, the conference is geared toward narrowing the gap between research and applications by having a meaningful, structured dialogue between the two sides.

The participants strive to integrate mobility issues into existing campaign analysis. This will bring mobility to the theater level, in which real-time, stochastic events are explicitly modelled. Also of importance is multimodal transportation systems, wherein lift capacity is provided by a combination of aircraft, trucks, rail, as well as water-borne vessels. This is in response to the "new world order", wherein the strategic confrontation between the East and the West is now replaced by regional conflicts which can flare up at a moment's notice. Strategic-mobility requirements are now over shadowed by tactical transportation demands.

## II. APPLICATION PERSPECTIVE

During the two-day conference, operationally-focused presentations were split between mobility models and their enhancements, the emerging intra-theater modelling-and-analysis efforts, and war-gaming. For example, the Air Mobility Command's (AMC) Mobility Analysis Support Systems (MASS) continues to grow in fidelity and the number of users. An airlift-loading module has interfaced with the Airlift Flow Module (AFM), the detailed air-mobility simulation-module in MASS. Graphical user-interfaces are being developed by AMC to improve export ability and analytical utility. THRUPUT II, Air Force Studies and Analysis Agency's (AFSAA) quick-turnaround mobility-model, has also been improved. A time-dimension has been added to this mathematical-optimization model of global-transportation networks to increase its applicability to ongoing and future analysis. The Airlift Loading Model (ALM) is becoming a desktop windows-oriented analytic-tool with improved user interfaces and post processors. Efforts are on the way to link a mobility model to a campaign model. For example, the Aeronautical Systems Center is integrating the Generalized Air Mobility Model (GAMM) and TAC THUNDER, a two sided, theater-level combat-simulation model and war-gaming model. The Users' Group evaluated the Timed Phased Force Deployment Documents (TPFDD) to determine feasibility within the real-world constraints of global airfield-infrastructure and aircraft availability. Finally, accurate representation of airfield capacity remains one of the challenges yet to be conquered.

While there are unique requirements on *air* mobility (such as air re-fueling), Air-Force efforts need to be merged with modelling efforts of the Joint Services. For example, the following three efforts need to be coordinated:

1. Joint Flow And Analysis System For Transportation (JFAST) of the US Transportation Command (TRANSCOM),

2. Enhanced Logistics Intra Theater Support Tool (ELIST) of the Military Transportation Management Command, and
3. Global Deployment Analysis System (GDAS) of the Army Concepts and Analysis Agency.

Effort such as these which have joint utility need to be integrated. Alternative initiatives to integrate the efforts include the Analysis of Mobility Platform program of TRANSCOM and the Joint Analysis Model Improvement Program. Ultimately, these models need to be verified, validated and accredited by the defense community at large.

While initiatives may be modified and new initiatives may emerge from time to time, supporting all these efforts is a common database, which is the fundamental, unchanging element that make or break any integration plan. While efforts such as The Modeling, Analysis, Simulation and Training Data Base (MASTR DB) are underway in the AFSAA, much more needs to be done to take care of the eventual integration among the Joint Services, particularly when two or more models are coupled together. The key is to identify common data-files and structures across organizational lines, which allow the output of one model to serve as input to another. There are also the unique features of *transportation* data that need to be identified to support anything from campaign analysis to multimodal mobility requirements, recognizing that there are many aspects of theater tactics, logistics, and mobility that are interrelated. Eventually, such a database needs to be disseminated among all interested agencies across the Department of Defense.

## III. RESEARCH PERSPECTIVE

From a research stand-point, these are the areas identified for further work:

1. Vehicle routing-and-scheduling,
2. Stochastic facility-location,
3. Terminal operations,
4. Modal-share analysis,
5. Spatial gaming, and
6. Real-time-information systems.

We have included here the many thoughtful comments of the fifty-odd participants in the conference. The participants span both the operational and research communities. They represent the Air Force, Army, Navy, national research labs, consulting firms, and universities. The entire attendee-list is appended for reference, and so is the program of the two-day conference. It was pointed out that fundamental to all mobility analysis is the movement of people or cargo on board vehicles—whether they be aircraft, trucks, trains or water-borne vessels—from origin to destination. This requires the fundamental models of *routing-and-scheduling,* particularly the real-time execution of these algorithms. It can be argued that the final product of any mobility exercise is a schedule that is constantly updated.

Given today's regional conflicts, more and more elements of surprise are embedded into mobility requirements. With base closures around the globe, there is a more-than-ever requirement on "global reach." Preposition and routing decisions have to take these stochastic elements into account, giving rise to *stochastic facility-location* models. The participants pointed out that in solving vehicle-routing-and-scheduling problems a critical element is crew scheduling, which has long been glossed over in the suite of mobility models in both facility-location and routing. Crew-duty-days may be imbedded as a constraint and a crew may be traced throughout the system in existing models, but there is little crew-scheduling optimization performed as one sees in the airline industry. There is always trade-off between approximate versus exact solutions, the use of stochastic network-optimization techniques, the use of decomposition solution-techniques such as column- and row-generation and set partitioning, the issue of integrality-of-the-solution, alternatives to linear and integer programs (such as the use of heuristics,) hierarchical-decomposition in real-time control, the inclusion of inventory-control, the use of asymptotic approximation when heuristics are integrated with analytical models, multicriteria considerations in routing (such as the consideration of both travel-time and risk factors,) and the employment of generalized network-solver for computational efficiency. Much work remains before the current state-of-research can meet the needs of the operational community.

The bottleneck of any transportation system is often found in the *terminal* environment. An example in the airlift world is the well publicized "max on ground" number, which determines how much can be ultimately delivered. It so turns out that a terminal's capacity is determined by several factors that interact in complex, stochastic ways that include the entire gamut of operating procedures from "metering and spacing" of vehicles to "loading and off-loading" operations. Terminal operation remains one of the most challenging of mobility analysis. Mobility models in the past have glossed over the complexity of terminal operations by using such planning factors as max-on-ground. In practice, terminal capacity is determined by a number of factors that include both the airside capacity and ground-side capacity. A fair amount of research has already been accomplished in the civilian and aviation sector. The participants strongly recommend this research be made widely available for critique and review by the defense analytical-community.

A fundamental tenet of "Jointness" is the recognition that troops and supplies can be delivered by a combination of modes—air, ground and water-borne—to the theater of operation. The question then arises—which of the deliveries and how much of them are to be shipped by air, ground or water-borne transport—or *modal-share* analysis in short. While there is a rich repertoire of knowledge in modal-share analysis, recent developments in supply-chain management place it in a larger framework, including inventory-control considerations. Modes are selected not only on a basis of their levels-of-service, but also their potential in making just-in-time (or at least ahead of schedule) deliveries so as to minimize total stockout, storage and transportation cost.

To the extent that campaign analysis needs to be integrated with mobility analysis, we have to replace the traditional one-dimensional "piston" or "linear" combat models with models that identify the geographic position of each combat unit. This gives rise to war-games that explicitly recognize the spatial dimension. The integration of combat models into mobility models involves *spatial gaming*, and most of the existing analysis techniques are based on stochastic and deterministic simulations. Recently we have seen the employment of semantic control, multistage optimization, ordinal optimization, and new paradigms based on real-time spatial-information in general. Much of the economic theories on oligopolistic competition (such the Cournot-Nash equilibrium) can potentially be carried over to combat modelling as long as the spatial dimension can be included (as illustrated in the classic Hotelling location model.) Voronoi diagrams have been proposed as a technique to solve such spatial-gaming models.

Finally, the fundamental basic-building-block of all the models is the database, and we emphasize that such a database has to support not only strategic-mobility decisions, but also tactical ones on a real-time basis. There are some very unique features of spatial data that are different than just any database. These features need to be identified and integrated with modelling requirements. The ultimate application of mobility models hinges upon the availability of timely information. This is the question of "what do you know" and "when you know it." With today's Global Positioning System, remote-sensing technology and geographic-information system, real-time information is not just a fantasy but a reality. The challenge, however, is to manage the voluminous data that can be collected to support the modelling efforts and ultimately mobility decisions. In vehicle routing it has been shown that real-time diversion-strategies based on spatial data is an example of the astute use of such an information base. Other examples of the data-based analysis include robust-optimization solvers based on nonanticipativity constraints. Through special data-partitioning, parallel processing and object-oriented-programming techniques can also be used for real-time mobility analysis. While the AMC Deployment Analysis System (ADANS) is a step toward the right directions, much more can be done along this line.

## IV. PRELIMINARY RESULTS

With the assistance of the following referees, we are happy to report some research results in response to the challenges posed above.

1. Jeffrey Camm, University of Cincinnati
2. Mark Daskin, Northwestern University
3. Richard Deckro, Air Force Institute of Technology.
4. Neal Glassman, Air Force Office of Scientific Research
5. Y C Ho, Harvard University
6. Pitu Mirchandani, University of Arizona
7. Warren Powell, Princeton University
8. Morton O'Kelly, Ohio State University
9. David Simchi-Levi, Northwestern University

Included in this special issue are five quality papers. The paper by Busch and Mulvey is entitled "A Quick Course-Of Action Evaluation Toolkit". The paper provides a powerful method for optimizing real-time operations. While the methodology may be straightforward, the philosophy behind such rapid modelling tools is worth preaching.

Chan, through his paper entitled "Real-Time Information and Transportation Decisions," introduces the readers to the potentials of future mobility models that are specifically organized around a structured, spatial database. The thesis is that careful linkage between models and data can provide quick and timely response to tactical—in addition to strategic—mobility problems on various levels of detail.

Morton et al., in the paper entitled "Optimization Modelling of Airlift Mobility", provides a linear-programming model for airlift deployment, which is used to address strategic issues related to fleet size and types, airfield capabilities, and the identification of bottlenecks. Aggregation was used for computational efficiency, including discretization of the time axis.

Powell, in his paper entitled "Real Time Control of Logistics," provides a taxonomy and notional scheme that can be used to represent a large number of real-time logistical-control problems. It highlights the importance, capabilities and weaknesses of sophisticated decision-support-systems for solving large-scale resource-management problems.

Yang et al., in their paper "Modeling and Optimization of Mobility Analysis", present a promising model for developing resource requirements to move cargo to the specific destination, satisfying a particular desired closure-schedule. Specifically, it solves a pickup/delivery vehicle-routing-and-scheduling problem with time-window constraints.

## V. FUTURE PLANS

Given the richness of these research agendas, it is clear that this special issue is only a modest beginning for further research in this area. Through the publication of this special journal issue, it is hoped that we further stimulate the many excellent efforts that have already started in many quarters of the defense community. The ultimate viability of any research result is its implementation in the field. The Users' Group discussed the need for improvements in mobility and logistical play in war-games. In war-gaming exercises, the primary contribution has to be made in the crisis-action phases, evaluating the TPFDDs to determine feasibility, within the real-world constraints, of global airfield infrastructure and aircraft availability. The next challenge is to incorporate the lessons learned in enhancing mobility play into future war-games. In a related effort, improvements have to be made in the mobility and logistic modules in war-game specific models, or to have war-game-specific models access the requisite information from cooperative, stand-alone mobility and logistics models. To facilitate future dialogue, a list server has been hosted at the Air Force Institute of Technology. All participants and invitees of previous Mobility Users Group Conferences have been put on this server. Anyone else who wishes to be included can e-mail his/her request to Yupo Chan at YCHAN@AFIT.AF.MI

## ACKNOWLEDGEMENT

We wish to thank the many individuals who made this special issue of the journal and the Third Mobility Modeling and Simulation Conference a resounding success. First and foremost, the research funding from the Air Force Office of Scientific Research (AFOSR) under the grant "Application of advanced operations research methods to mobility modeling" (AFOSR-PO-95-0006) allowed the principal investigator, Yupo Chan, to invite the best minds in the country to participate in this most worthwhile effort. The result is an outstanding collection of preliminary results as documented in this journal issue.

In alphabetical order, we wish to acknowledge particularly these tireless individuals who are an integral part of working team:

Daniel Briand—Our contact at the Air Force Studies and Analysis Agency, which instituted this series of mobility conferences,

Peter Purdue—Editor of the *Military Operations Research* journal who granted us the privilege to publish this special issue of the journal, and

Anthony Waisanen—our contact at the Air Mobility Command which is a close partner of the conference effort.

It goes without saying that the staff of the Military Operations Research Society, in their time-honored tradition, rendered the greatest support in the publication of this document.

For local arrangements, we are in debt to the support of Richard Deckro and Price Smith of the Operational Sciences Department, Air Force Institute of Technology, who provided the greatest assistance as members of the Conference Facilities Committee. We also owe our gratitude to the AFIT Civil Engineering and Service School, which made available their excellent auditorium for our use. Naturally, the support of everyone in the Graduate School of Engineering, particularly Dean Calico is also gratefully acknowledged.

It goes without saying that the views expressed in this document are those of the individuals and in no way reflect the position of the sponsors, the U.S. Air Force or the U.S. Government.

## 1995 RIST PAPERS

*The **Rist** and **Barchi** Prizes will be announced during the opening ceremonies of the 64th MORS Symposium. This special session provides the opportunity for the prize winners to present their papers. The Prize Committee Chairs will discuss the selection process and pertinent points from selected non-winning papers. The following papers were submitted in response to the 1995 call for papers:*

1. *Combined Forces Command—Decision Support Modeling, Vol I, II, III,* LTC **Patrick Guinnane,** USA CAA, MAJ **Paul Buhl**, USA CAA, Dr. **Elizabeth Abbe**, USA CAA, **Renee Carlucci**, USA CAA, **Louise McLean,** USA CAA, **Richard Poulos,** USA CAA, **John De Palma** USA CAA, and MAJ **Douglas Herr,** USA CAA.

2. *AMRAAM Upgrade Analysis: The Benefits of Aggressive Model Validation*, Maj **Eileen Bjorkman,** 846th Test Squadron, Holloman AFB, and LtCol **Martin Allen,** AFSAA.

3. *Evaluating the Effectiveness of Shoot-Look-Shoot Tactics in the Presence of Incomplete Damage Information*, **Yossi Aviv** Ben Gurion University of the Negrev, and **Moshe Kress,** Ben Gurion University of the Negrev.

4. *Anti-Armor Advanced Technology Demonstration Experiment*, **Mark A. Burrough,** USAMSAA, **Kent Butler,** USAMSAA, **Dwayne W. Nuzman,** USAMSAA, **Floyd C. Wofford,** USAMSAA.

5. *De Physica Belli: An Introduction to Lanchestrian Attrition Mechanics*, **Bruce W. Fowler,** USA Missile Command.

6. *Foundations of the Theory of Volley Fire*, **Robert L. Helmbold,** USA CAA.

7. *Toward a Paradigm for Validating Man-in-the-Loop Simulations*, Maj **William C. Hopkinson** and **Jose Sepulveda**.

8. *Optimal Distribution of Army Officers*, Maj **Doug L. McAllaster,** CGSC.

9. *A Quick Response Approach to Assessing the Operational Performance of the XM93E1 NBCRS Through the Use of Modeling and Validation Testing*, **Richard W. McMahon,** USA Research Lab.

10. *A New Weapon in the Information War*, MAJ **David H. Olwell,** USMA.

11. *Military Force Structure and Realignment "Sharpening the Edge" through Dynamic Simulation,* **Stephen R. Parker**, USA CAA.

12. *Exploring a Relationship Between Tactical Intelligence and Battle Results*, **Todd E. Sherrill,** USMA, and **Donald R. Barr,** USMA.

# QCOA: A Quick Course-of-Action Evaluation Toolkit[1]

**Ingrid K. Busch**
*Center for Transportation Analysis*
*Oak Ridge National Laboratory*

**Steve Mulvey**
*Tanker Airlift Control Center*
*Air Mobility Command*
*Scott AFB, IL*

## INTRODUCTION

At the start of a contingency, airlift flow planners have very minimal information on what is to be transported or what resources are available for utilization. Yet even at that stage, the planners need the capability to calculate accurate estimates of the efficacy of air mobility transportation. Until the advent of the Quick Course-of-Action (QCOA) Toolkit, much of those determinations were made without the assistance of a decision support system. QCOA has been developed to fill that void as part of the AMC Deployment Analysis System (ADANS) which provides the Air Mobility Command (AMC) with scheduling and analysis tools to assist them in accomplishing their taskings.

## CONTINGENCY PLANNING

Contingency operations executed by AMC are unique airlift missions characterized by rapid evolution, no-notice taskings, varied requirements, shifting priorities, and distant and diverse operating environments. Routine AMC channel missions have pre-determined requirements that are supported by established route and support structures. Airlift support for JCS exercises is identified and planned months in advance. The time sensitive nature of a contingency combined with heightened national and global interest make contingency planning and execution a challenging and demanding task.

Contingency operations support situations ranging from humanitarian relief (Somalia, Rwanda, South Florida/Hurricane Andrew), to peacekeeping (Haiti, Bosnia), to the urgent deployment of combat forces to counter a regional threat (VIGILANT WARRIOR the operation in response to the aggression against Kuwait by Iraq in the Fall of 1994.). With missions this varied, the airlift support structure required to project global reach becomes as unique as each contingency.

Contingencies begin in response to events taking place anywhere in the world. The National Command Authorities (NCA) discuss available courses of action with the supported CINC. USTRANSCOM evaluates potential air, land, and sea movement requirements. TRANSCOM directs all airlift issues to AMC. AMC forwards all available information to the Contingency Operations division of the Tanker Airlift Control Center (TACC). In the first hours of a contingency, when significant political and military decisions are being made, accurate information must be provided in a timely fashion to the NCA.

Contingency directors are experienced mobility planners who blend years of operational flying experience with knowledge of worldwide command, control, and airlift mission planning. The challenge of contingency planning is to optimize the mix of elements making up a global reach laydown package (Technical Airlift Control Element (TALCE) support, air refueling requirements, airspace issues, diplomatic clearance restrictions, and liaisons with supported commands and international relief agencies) in a *fly by the seat of your pants* environment.

Technology can assist the contingency director in determining broad theater-to-theater capabilities and the mobility assets required. This is vital information for the senior leadership that is assessing the courses of action. Present AMC airlift capability models (ADANS automated scheduler, JFAST, MASS) are excellent tools, but require extensive data input, and their processing time can be several hours. Clearly, a system that provides quick estimates of capability can improve the contingency planning process.

The Quick Course-of-Action (QCOA) toolkit provides such a capability. It is a flexible, easy-to-use system that permits the user to state requirements and solve problems in many ways. Questions and direction from the senior leadership vary with each contingency. Two recent examples highlight this fact. In July 1994 AMC deployed support to ease the refugee crisis in and around Rwanda. Contingency planners were directed to establish a mobility infrastructure that would deliver five C-5 and nine C-141 missions into Central Africa while supporting any non-governmental agencies operating in the region. Three months later, while contingency planners were supporting US forces in Haiti, Iraqi provocations required a deployment of combat forces to the Persian Gulf. The USCENTCOM requirement was to deliver 2000 short tons of cargo and 2000 passengers to the Gulf each day. Contingency planners were tasked to establish a global reach laydown package for that requirement while ensuring no shortfall of airlift support for forces in Haiti. QCOA has the ability to quickly provide the broad-based answers no matter how the requirement has been defined.

The input of data to QCOA is straightforward. The planner needs to enter a movement requirement, an airlift network and the types of aircraft to be used. If the requirement includes a date by which it must be in the theater (the closure date), QCOA will determine the airlift mix required to meet closure. If a closure date is not provided, QCOA calculates one based on the airlift network and aircraft apportionment. Additionally, it calculates how many air refueling tankers and airlift crews will be required to sustain the air refueling based airlift. The contingency director provides the network, which is often determined by considering such factors as flight times, diplomatic clearances, air refueling tracks, and en route support bases. With only minimal information, QCOA can quickly analyze various networks comparing closure, required airlift aircraft, tankers, and crews.

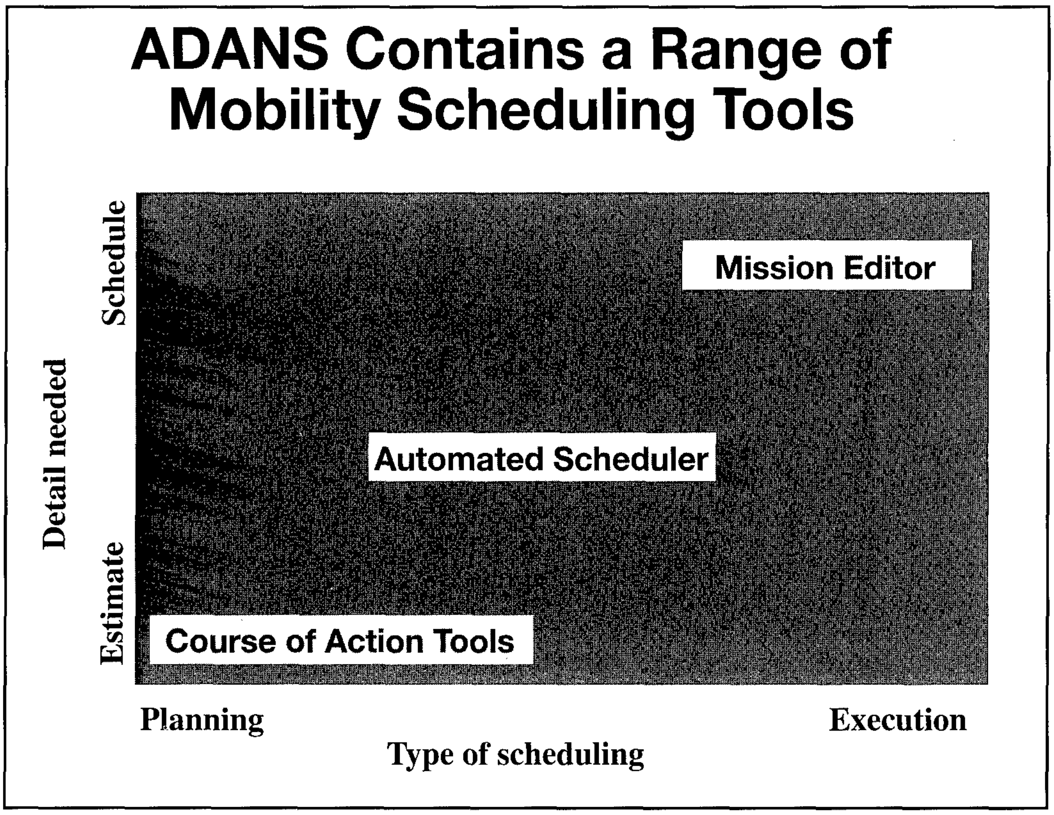
## ADANS

ADANS is a scheduling and analysis system that is being developed for AMC by the Oak Ridge National Laboratory. The goal of ADANS is to integrate planning and scheduling for both peacetime and wartime operations. As such, ADANS is utilized by AMC organizations which have diverse tasks.

The component of ADANS that deals with the subject of this paper is the planning component, which provides scheduling and analysis capability for the deliberate and execution planning communities of AMC. Deliberate planners are tasked with the evaluation of large-scale operations plans to determine whether they are transportation-feasible. Execution planners direct airlift exercises as well as plan the airlift for real-world contingencies.

The planning component of ADANS provides planners with a number of scheduling tools, each appropriate for a particular task in terms of the detail that is needed and the type of scheduling that is being done. (See Figure 1.) An automated scheduler is used by the deliberate planners to produce a set of missions for a plan of operations. The information that is needed to run the automated scheduler is quite extensive. The cargo and passenger requirements that are to be moved are input electronically; a large plan may contain several thousand individual requirements. The planner needs to indicate what types of planes can be used to transport these requirements, and in what configurations these planes may fly. He also needs to build a network which comprises the routes over which the planes can fly and to indicate (through the setting of permissions) where planes are allowed to onload, offload, en route, and refuel, and where crews are allowed to rest or be exchanged. After the plan has been set up, a run of the automated scheduler produces a set of routings and schedules for the aircraft. ADANS provides textual and graphical tools and printed reports to assist the planner in evaluating the schedule that was produced. He may then modify the setup and rerun the scheduler until the plan has been scheduled satisfactorily. Depending on the size of the plan, a run of the scheduler may finish in a few minutes or it may take an hour or longer, making the process of iteratively rerunning the automated scheduler time-consuming.

When an exercise or contingency is being executed, planners require the capability to specify exactly how the mission is to fly. The requirement to be moved and the routing and schedule to be followed are known *a priori*. ADANS contains a mission editor that allows the planners to type in this information, and store it in the database. After coordination of the mission has been completed, the planner will transmit the detailed mission information to AMC's command and control system, the Global Decision Support System (GDSS).




However, when the possibility of a contingency first arises, many of the details of the scenario may not be known. For example, the specific onload and offload stations may not be known, though the countries of origin and destination likely are. Which planes will be available for the contingency may not be known, though a planner may be told that he would have access to, say, 20 C-141's to fly in the contingency. The en route stations may not be known either, though in most cases, direct flight from onload to offload will not be possible, so the use of en routes and possibly air refueling locations would then be necessary. In short, the fuzziness of the scenario when it first is proposed reduces the effectiveness of the ADANS automated scheduler in evaluating the airlift capability available. Forcing such a tentative scenario into the automated scheduler would require the planner to make decisions to dictate details of the scenario that were not known, in most cases placing on the problem constraints that did not actually exist.

By providing tools that allow planners to model scenarios at the level of detail of the data that is available to them at the outset of a contingency, QCOA completes the set of ADANS planning and scheduling tools. Planners now have, within one seamless system, scheduling tools that take them from deliberate planning to contingency planning to contingency execution.

## QCOA INTERFACE

QCOA offers the planner access to the model via a graphical user interface, which relies on the mouse for input and graphical output as much as possible.

The requirements screen offers the planner the ability to work with the data at two levels of detail. In the limited mode, requirement data is limited to the onload and offload station and the amount of cargo and number of passengers to be moved. In case the onload and offload stations are not known, notional stations may be used in their stead. When requirement data is input in the limited mode, the assumption made by the model is that the cargo is of the oversize cargo class, and that the requirement will need to be delivered to the offload on the first day of the scenario. If these assumptions do not adequately model the scenario, the planner has the ability to switch to the more detailed level of data display and modify the information.

The aircraft screen allows the planner to specify the types of aircraft that will be available for utilization in the scenario. The allocation of aircraft is indicated by the manipulation of a graph which indicates by day how many individual planes of a specific aircraft type are available for use. A configuration screen is also available which allows the planner to specify in what configurations an aircraft type may fly, and how many passengers and how much and what classes of cargo a configuration can carry. All configuration information, as well as the default configuration for an aircraft type, is pulled in from the database.

The path screen allows the planner to indicate what paths should be used to fly from an onload to an offload, and which aircraft are allowed on individual paths. The model will calculate the flying times along the path; the planner is allowed to edit these values, as well as the ground times at the stations. Additionally on this screen, the planner can indicate where a crew change or crew rest will take place.

## THE QCOA MODELS

QCOA makes a number of assumptions in order to make the solution of the models tractable. The first is that the en route capability of a station is modeled as a single constraint. This constraint aggregates a myriad of resources that are required at an en route station, such as ramp space and fueling equipment. The combination of all of these resources is termed "parking." As resources are generally consumed at a rate that is proportional to the size of the aircraft, such an aggregation provides a good approximation to the en route capability of a station without requiring the planner to provide detailed information. For each station, the planner specifies the instantaneous parking capability of the station, as well as the amount of parking that would be required by a narrow- and by a wide-body aircraft. These data are used to construct the en route constraints in the model. For example, if a station were open for 12 hours per day, and the parking at the station were rated as 10 parking spaces, the daily parking capability of that station would be 120 parking hours. If a narrow-body plane were to take 1 parking space and require a 2 hour stop, that would be a utilization of 2 parking hours. On the other hand, if a wide-body plane were to take 2 parking spaces and be on the ground for 3 hours, that would be a utilization of 6 parking hours.

Another assumption is that an estimate need not have as fine a scheduling granularity as a scheduling algorithm provides. The time periods in QCOA may be adjusted to a range of settings from 1 hour to 24 hours. Using a 24 hour time period decreases the run time of the estimate, but a 1 hour time period improves its accuracy.

To streamline the exposition of the models, statement of the formulation is deferred to the appendix. However, presentation of the variables utilized in QCOA will clarify the concepts involved in the discussions of the models that follow.

QCOA makes use of four classes of variables. The variable $x_{icpk}^{t}$ will indicate the number of aircraft of configuration $k$ that fly along path $p$ carrying cargo class $c$ of onload/offload pair $I$ arriving at the offload at time $t$, where $c$ is a preferred cargo class for configuration $k$ (i.e., is a cargo class for a cargo configuration or a passenger class for a passenger configuration).

Nearly similar to $x_{icpk}^{t}$, the variable $y_{icpk}^{t}$ s the number of aircraft of configuration $k$ that fly along path $p$ carrying cargo class $c$ of onload/offload pair $I$ arriving at the offload at time $t$, where $c$ is not a preferred cargo class for configuration $k$ (i.e., is a passenger class for a cargo configuration or a cargo class for a passenger configuration).

Note that the difference between the $x$ and $y$ variables is that the $x$ variables correspond to the preferred class for an aircraft configuration. Oftentimes, a cargo configured aircraft can also carry passengers, and utilizing this capability is essential to an efficient airlift. However, the accompanying passengers will not be flown without sufficient cargo to justify the flight. Hence, QCOA treats accompanying passengers on cargo configured aircraft (and similarly, accompanying cargo on passenger configured aircraft) with a separate set of variables.

The variable $z_{ic}^{t}$ will indicate the total amount of shortfall that has occurred for cargo class $c$ of onload/offload pair $I$ by time $t$.

Finally, $w_m$ will indicate the number of aircraft of type $m$ needed in the scenario. (This variable is only used in the models that are to determine the number of aircraft needed. In all other models, the number of available aircraft is specified by the planner.)

## ESTIMATE OF A SCHEDULE

QCOA allows the planner to estimate the schedule that would be produced by the ADANS automated scheduler. The objective for this linear program weights the shortfall variables heavily. The movement variables $x_{icpk}^{t}$ are assigned weights according to the $t$ parameter, so that early movement of cargo and passengers is encouraged.

There are constraints in this estimate to ensure that requirements are not moved before they are available, and that each requirement reaches its destination on or before its latest allowed arrival date. To decrease the size of the linear program that is to be solved, these requirement constraints are based on the onload/offload pair and the cargo class (*i.e.*, outsize, oversize, bulk, passenger), not on the individual requirement. These constraints guarantee that all requirements are accounted for, either as deliveries or as shortfall.

Other constraints in the model require that the number of planes that are utilized on any particular day does not exceed the number that have been allocated by the planner. The en route capability of a station is modeled with a single constraint as described above, and a thruput constraint limits the amount of cargo handling (i.e., onloading and offloading) that can be done at a station on a day.

Finally, there is a set of logical constraints that serve to ensure that there is a sufficient amount of cargo to justify the accompanying passengers (or, in the case of a passenger configured aircraft, sufficient numbers of passengers to support the accompanying cargo). A formal presentation of this model is provided in the appendix.

While each movement requirement in an operation plan has a date by which it must be delivered to its destination, it is often not possible to meet all of these demands on time. QCOA provides the planner with the ability to determine the effect of allowing this latest delivery date to be violated. This is accomplished in two ways. In the first, the planner can iteratively solve the model, where in each iteration the latest delivery date of each requirement is increased by another day. In the second, the planner can ask QCOA to determine the minimum date of last delivery (known as closure of the plan). It may be that a plan will never close. This would happen, for example, if there were a requirement in the scenario for which no suitable plane were available. In that case, QCOA would provide the last delivery date of the requirements that could be delivered.

## ESTIMATE OF AIRCRAFT NEEDED

The aircraft estimate minimizes the number of planes that are needed to deliver the requirements, taking into account the fact that a plane can fly several missions over the course of the scenario. For this estimate, the planner specifies the movement requirements and the types of aircraft that could be utilized, as well as their possible configurations.

Performing this estimate involves solving two linear programs. In the first, the model determines how much of the requirements can be delivered, with no limitations on the number of planes used. The objective of this linear program is to minimize the shortfall. There are requirement constraints as in the schedule estimation problem, and the en route and thruput constraints are added at the discretion of the planner. He may elect to ignore these constraints if he believes that the necessary support at the stations in the network will be provided, or he may simply not know what the support will be and does not want to impose artificial constraints on the system. Finally, the logical constraints are also imposed here, as accompanying passengers and accompanying cargo often play an important role in determining the best mix of planes in the fleet. In the second linear program, the shortfall is fixed from the solution of the first linear program. The objective is now a weighted function of the number of planes needed in the scenario, with the weights characterized by the type of the aircraft. The constraints are as in the first linear program, with additional constraints that determine the number of each type of aircraft that will be needed.

QCOA also provides another estimate, that of determining the minimum number of planes needed to support a constant delivery rate. Solving this estimate is similar to that above, though the time element is taken away: the requirement amounts for each onload/offload pair are entered as a daily delivery amount. The set of requirement constraints is replaced with a single constraint for each onload/offload pair and cargo class.

## ESTIMATE OF NETWORK CAPABILITY

The final set of estimates available in QCOA allows the planner to determine the capability of the network. Oftentimes it is the limitations imposed by the en route stations, the air refueling tracks, and the offload stations that will constrain the airlift flow. Obtaining quick determinations of these limitations provides the planners with the information needed to move equipment and manpower into the key points of the network to circumvent these bottlenecks.

The estimate of network capability determines the maximum average daily amount that can be delivered to the theater. As such, no requirements are entered by the planner, but he must indicate which onload/offload pairs are to be considered. He also indicates the types of aircraft that are to be considered for flying, and the configurations in which they are allowed to fly. Finally, and perhaps most importantly for this estimate, he specifies the paths over which the planes are to fly, the en route and thruput capabilities of the stations, and the separation time that is required on the air refueling tracks. There is also a mechanism by which the planner can direct the model to give preferential treatment to passengers or to cargo.

The objective of this model is to maximize the amount that is delivered. As there are no actual movement requirements involved in this estimate, there are no requirement constraints. The en route and thruput constraints are included in the linear program, as are constraints to ensure that a minimum separation time is enforced on the air refueling tracks. The planner has the option of adding plane availability constraints if he so chooses. The logical constraints that control the relationship between the cargo and accompanying passengers (or the passengers and accompanying cargo) are also included; although the planner's main question in this case is likely that of how much of the cargo can be delivered on an average day, knowing the number of "free seats" that are available is valuable information as well.

## SOLUTION OF THE MODELS

The models described above are solved through the use of the CPLEX linear programming package. The planner runs QCOA locally on a Sun workstation. When he asks for an estimate, QCOA writes the MPS file to the disk and calls the CPLEX solver on another Sun processor. When that processor has arrived at the solution, the local workstation displays the solution.

## SUMMARY

QCOA includes a number of models, each one looking at a different view of contingency airlift. This is necessary, since every contingency is slightly different in terms of the information that is known, and the objective of the use of airlift in the operation plan. This flexibility, as well as its speed and easy-to-read graphical display of networks, cargo delivery, and airlift missions required is what makes QCOA valuable to the planners. Many times in recent months the TACC commander has tasked the contingency cell to provide options and capabilities for real-world *what if* situations. Within minutes this accurate information is in the hands of the AMC vice commander and CINCTRANS.

## APPENDIX

Presented below is the formulation that is used by the QCOA model in estimating a schedule. This first requires the specification of data.

There are data corresponding to the requirements. The possible classes of a requirement are *bulk, oversize, outsize, passenger,* and *litter*; a requirement may include more than one class.

$C_1$ = {*outsize, oversize, bulk*}.
$C_2$ = {*pax, litter*}.
$s_c$ = the amount of class $c$ in requirement $s$.
$S_{ic}$ = $\{s \mid$ requirement $s$ is associated with onload/offload pair $i$ and $s_c > 0\}$.
$ead(s)$ = the earliest delivery date for requirement $s$.
$lad(s)$ = the latest delivery date for requirement $s$.
$EAD_{ic}$ = $\{ead(s) \mid s \in S_{ic}\}$.
$LAD_{ic}$ = $\{lad(s) \mid s \in S_{ic}\}$.

$$T_{ic} = \max_{t \in LAD_{ic}} t.$$

$$q_{ic}^{t} = \sum_{t' < t,\ s \in S_{ic},\ ead(s) = t'} s_c$$

$$r_{ic}^{t} = \sum_{t' \le t,\ s \in S_{ic},\ lad(s) = t'} s_c.$$

That is, $q_{ic}^{t}$ is the amount of cargo class $c$ associated with onload/offload pair $i$ that is available to be delivered before time $t$, and $r_{ic}^{t}$ is the amount of cargo class $c$ associated with onload/offload pair $i$ that must be delivered by time $t$.

There are data corresponding to the aircraft:

$M_m$ = $\{k \mid k$ is an allowed configuration for aircraft type $m\}$.
$K_1$ = $\{k \mid k$ is a cargo configuration$\}$.
$K_2$ = $\{k \mid k$ is a passenger configuration$\}$.
$p_m^{t}$ = the number of aircraft of type $m$ available on day $t$.
$\gamma_{pk}$ = the amount of cargo that a plane of configuration $k$ can carry on path $p$.
$\phi_{pk}$ = the number of passengers that a plane of configuration $k$ can carry on path $p$.

There are data corresponding to the stations:

$e_n$ = the parking capacity of station $n$.
$b_n$ = the cargo handling capability of station $n$.
$b_n'$ = the passenger handling capability of station $n$.
$\mu_{nk}$ = the number of parking hours required by a plane of configuration $k$ at station $n$.

Finally, there are data corresponding to the paths:

$N_p$ = $\{n \mid \text{station } n \text{ is on path } p\}$.
$\alpha_{pk}$ = the amount of time for a plane of configuration $k$ to reach the onload along path $p$.
$\beta_{pk}$ = the amount of time for a plane of configuration $k$ to reach the offload along path $p$.
$\tau_{pk}$ = the amount of time for a plane of configuration $k$ to traverse path $p$.
$\sigma_{npk}$ = the amount of time for a plane of configuration $k$ to reach station $n$ along path $p$.
$\rho_{ic}^{t}$ = the penalty for shortfall of class $c$ of onload/offload pair $i$ with LAD of $t$.

$$\text{Minimize} \quad \sum_{i,c,p,k,t} \lambda_{ic}^{t}\, x_{icpk}^{t} + \sum_{i,c,t} \rho_{ic}^{t}\, z_{ic}^{t}$$

subject to

$$\sum_{k \in K_1, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_2, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} \le q_{ic}^{t} \quad \forall t \in EAD_{ic}\ \forall i, c \in C_1 \tag{1.1}$$

$$\sum_{k \in K_2, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_1, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} \le q_{ic}^{t} \quad \forall t \in EAD_{ic}\ \forall i, c \in C_2 \tag{1.2}$$

$$\sum_{k \in K_1, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_2, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} \ge r_{ic}^{t} \quad \forall t \in LAD_{ic}\ \forall i, c \in C_1 \tag{1.3}$$

$$\sum_{k \in K_2, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_1, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} \ge r_{ic}^{t} \quad \forall t \in LAD_{ic}\ \forall i, c \in C_2 \tag{1.4}$$

$$\sum_{k \in K_1, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_2, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} = r_{ic}^{t} \quad t = T_{ic}\ \forall i, c \in C_1 \tag{1.5}$$

$$\sum_{k \in K_2, p, t' \le t} \gamma_{pk} x_{icpk}^{t'} + \sum_{k \in K_1, p, t' \le t} \gamma_{pk} y_{icpk}^{t'} + \sum_{t' \le t} z_{ic}^{t'} = r_{ic}^{t} \quad t = T_{ic}\ \forall i, c \in C_2 \tag{1.6}$$

$$\sum_{i,c,p,k \in M_m,\ t' \ge t + \beta_{pk} - \tau_{pk},\ t' \le t + \beta_{pk}} x_{icpk}^{t'} \le p_m^{t} \quad \forall m, t \tag{1.7}$$

$$\sum_{i,c,p,k} \mu_{nk}\, x_{icpk}^{t+\beta_{pk}-\sigma_{npk}} \le e_n^{t} \quad \forall n, t \tag{1.8}$$

$$\sum_{i,c\in C_1,p,k\in K_1} \gamma_{pk}\, x_{icpk}^{t} + \sum_{i,c\in C_2,p,k\in K_2} \gamma_{pk}\, y_{icpk}^{t} + \sum_{i,c\in C_1,p,k\in K_1} \gamma_{pk}\, x_{icpk}^{t+\beta_{pk}-\alpha_{pk}} + \sum_{i,c\in C_2,p,k\in K_2} \gamma_{pk}\, y_{icpk}^{t+\beta_{pk}-\alpha_{pk}} \le b_n \quad \forall n, t \tag{1.9}$$

$$\sum_{i,c\in C_2,p,k\in K_2} \gamma_{pk}\, x_{icpk}^{t} + \sum_{i,c\in C_1,p,k\in K_1} \gamma_{pk}\, y_{icpk}^{t} + \sum_{i,c\in C_2,p,k\in K_2} \gamma_{pk}\, x_{icpk}^{t+\beta_{pk}-\alpha_{pk}} + \sum_{i,c\in C_1,p,k\in K_1} \gamma_{pk}\, y_{icpk}^{t+\beta_{pk}-\alpha_{pk}} \le b_n' \quad \forall n, t \tag{1.10}$$

$$\sum_{c} x_{icpk}^{t} \ge \sum_{c} y_{icpk}^{t} \quad \forall i, p, k, t \tag{1.11}$$

Constraints (1.1) and (1.2) ensure that the delivery of the requirements associated with an onload/offload pair does not exceed the amount eligible for delivery. Constraints (1.3) and (1.4) require that the delivery of the requirements associated with an onload/offload pair does not fall behind the latest arrival date curve. This is accomplished by assigning the shortfall to the $z$ variables when necessary. Constraints (1.5) and (1.6) guarantee that all delivery requirements have been either delivered or assigned to the shortfall variables.

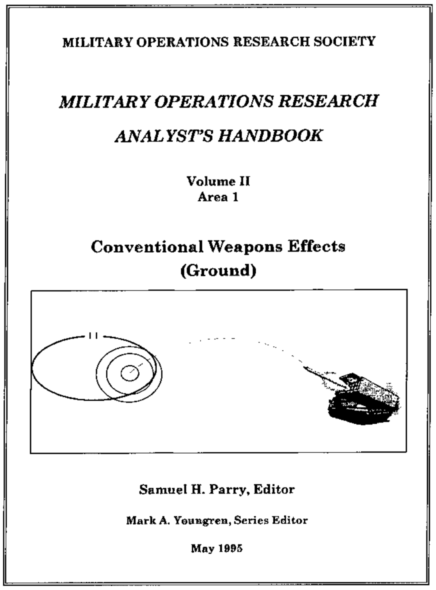
Constraints (1.7) limit the number of planes that are flying at a particular time $t$ to not exceed the number of aircraft that have been allocated for that time. Constraints (1.8) control the en route capability of the stations. Constraints (1.9) and (1.10) limit the thruput of cargo and passengers, respectively at the stations. Constraints (1.11) enforce the logical relationships between the $x$ and $y$ variables.

This is the linear program that is solved when the planner requests an estimate of the schedule. The other models are solved via linear programs that are based on this same formulation. Those models, naturally, have alternate objective functions and a variation of the stated constraint set. However, the basic structure remains the same, which allows QCOA to make use of a previously solved problem to develop the starting basis for a subsequent solution of the current linear program.

## ENDNOTES

1 Submitted June 1995; In final form July 1995.

# ANALYST'S HANDBOOK SUPPLEMENT NOW AVAILABLE




Area I of Volume II, *Conventional Weapons Effects*, is now available through the MORS office. It is intended to be part of the initial notebook in the *Handbook* series, along with Volume I (*Terrain, Unit Movement, and Environment*) and Volume II, Area II (*Search, Detection and Tracking*) (To be published.)

The *Conventional Weapons Effects* (*Ground*) area was edited by Sam Parry, who has been educating and advising Army OR analysts for many years. It is designed to provide a quick reference for models that represent the effects of a conventional attack against ground targets. It emphasizes the ground battle; publication of algorithms for naval and air engagements is planned for the future. It cannot, of course, include every algorithm used by military OR practitioners, but it is an attempt to characterize those that are most commonly used.

The cost of Volume II Area I is $8.00. As always, Volume I, which includes a three-ring binder for future supplements, is available for $25.00. **Combine both and save: $30.00.** All prices include freight.

**Please call the MORS Office for international shipping rates.**
**VA Residents add 4.5% sales tax.**

✂
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Volume II Area I: ________ Volume I: ________
Volume I and Volume II, Area I: __________

Total Enclosed: __________ **Make checks payable to MORS**

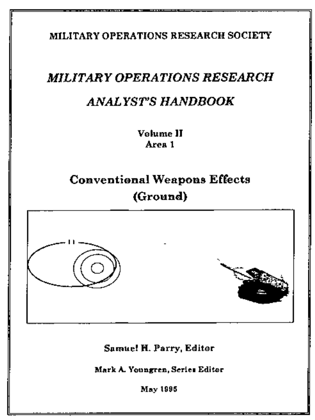
Name: ___________________________________

Shipping Address: ____________________________

_________________________________________

_________________________________________

City: ____________ State: ____ Zip Code: ________




Military Operations Research Society • 101 S. Whiting Street • Suite 202 • Alexandria VA 22304 • (703) 751-7290

# Real-Time Information and Transportation Decisions: An Analysis of Spatial Data[1]

**Yupo Chan**
*Professor, Department of Operational Sciences, Air Force Institute of Technology*

## I. ABSTRACT

Critical to making timely transportation decisions is current information. This paper differs from other information-system discussions in several ways. First, we include remote-sensing technology such as satellite-image processing. It is felt that oftentimes such a data-gathering procedure has not been adequately examined. For example, current technology allows very inexpensive roof-top antennas to collect real-time satellite-sensing data. A second focus of this paper is on the application of Geographic Information System (GIS) toward transportation analysis. This bridges the gap between data availability and modelling—a gap that has long existed but little has been done about it. A third emphasis, judged to be the most important one, is on applications. Oftentimes remote sensing and GIS are viewed as gadgets rather than as means toward analysis. For this reason, remote sensing and GIS have not been fully exploited as a tool for more innovative real-time decisions. Illustrative applications, particularly built upon the marriage between data and modelling, will be highlighted. These applications range from transport scheduling-routing, resource management to contingency-facility location. They show that when data are properly organized and algorithms are tailored toward the database, sophisticated transportation analysis can be performed in a tactical—in addition to a strategic—environment. The purpose of this paper is to review the relationship between remote sensing GIS and transportation modeling efforts using such an information base.

## II. INTRODUCTION

It is clear that one of the most demanding parts of any transportation analysis is the collection of data. In urban-transportation applications, for example, the resource requirement for data assembly is any where from 30 to 50 percent of the study. Furthermore, data usually come from a diversity of sources. Population and employment data—the driving force behind urban travel—are compiled usually on a census tract, while educational information are compiled by school districts. These geographic subunits are different, making the merging of data particularly difficult. Geographic information systems (GIS) allow a much more consistent format for the assembly of such data. In addition, a GIS typically relies on land-use information consisting of survey data supplemented by aerial photographs, which greatly enrich the database. In American urban applications, there is a Standard Land Use Coding Manual promulgated by the U.S. Department of Housing and Urban Development that provides guidelines for the preparation of such database. Recent standardization includes Topologically Integrated Geographic Encoding and Referencing (TIGER) files of the U.S. Census Bureau and spatial-data-transfer standard promulgated by the U.S. Department of Transportation. Increasingly, GIS also capitalizes on the recent introduction of satellite raster or pixelized imagery, which rivals and in some cases exceeds the fidelity and format of aerial photos. The reliance upon a systematic database becomes increasingly critical in logistical applications. Industries rely more and more heavily upon satellites to track their moving assets such as trains and trucks. In mitigating the ill effects of natural and manmade disasters, evacuation of population from disaster sites is of paramount importance. Remote sensing offers a means to predict such disasters as hurricanes and earthquakes.

In the military, the post cold-war era introduces a new challenge in resolving regional conflicts which can arise within a moment's notice. Instead of a "linear" or "piston" model today's combat models call for tactical decisions to be made on a real-time basis. Similarly, strategic east-west nuclear exchange gives way to surgical actions in redressing regional confrontations. These decisions often involve spatial attributes such as the whereabouts of a tank, an aircraft and troop movements. Most importantly, the question arises as to how to re-supply these combat units on a real-time basis. Today's information technology provides the platform for such timely decision to be made. It is the intent of this paper to outline some of the ways to make this happen even more rapidly and more effectively than it has been, and to remove some of the road blocks that have proven to be insurmountable thus far. Database to support timely transportation-analysis has indeed come of age.

# III. OBJECTIVES

Technology transfer is a often an overworked term, suggesting the tremendous benefits of sharing know-how among various segments of society. We have heard of the transfer of our cutting-edge knowledge in aerospace and defense toward civilian use. Our contention is that cross fertilization works both ways. In transportation science, the advances in industry and the civilian sector in general have been truly amazing over the last few decades. With the tremendous investment in this field by large corporations and federal agencies such as the U.S. Department of Transportation, the state-of-the-art is certainly way ahead of real-world applications. Fortunately, the transportation problems we try to solve are often common among both the defense and civilian communities, which paves the ground for two-way transfers. For example, both seek a closer tie between strategic and tactical decisions. In the civilian sector, the planning studies of the 1960's have given way to "Intelligent Transportation Systems" which seek to monitor and control anywhere from urban-travel congestion to commercial-vehicle operations. Similarly, the defense community, in the post-Cold-War era, is facing regional conflicts that require not only advanced multimodal lift strategies, but also theater-level tactics. Both can benefit significantly from the availability of real-time information and analysis procedures that can be executed as events unfold.

To reiterate, the purpose of this paper is to review the relationship between remote-sensing, GIS, and transportation-modelling efforts using such an information base. We illustrate the analysis process starting from data collection via satellite. This includes geometric correction (from different viewing angles of sensors) and radiometric enhancement due to differences in sensor instrumentation. If we are looking at land cover, for example, the boundaries of lakes, forests, man-made objects, and ocean need to be delineated once correction and enhancement are accomplished. Then these land-cover data are to be integrated with other ancillary data such as those gathered from the census or other sources of intelligence. The ultimate use of data is obviously problem solving. We will round out our discussion here in this paper by pointing out how the GIS database can support modelling efforts such as transportation analysis, analysis that rely heavily on *spatial* data. This is, in the author's opinion, the most interesting part of the paper (Marble and Peuquet 1988). The focus is really on the relationship between data *collection, organization* and *problem solving.* Within the limited space available, we wish to illustrate the integration that is possible between these three areas.

Recent effort by Kwan and Golledge (1995), for example, show how a GIS can be used to calibrate a computational-process model for activity scheduling. Psaraftis (forthcoming) points out that computer-based technologies, including GIS, have significantly enhanced dynamic routing and opened interesting direction for research. The impetus is further accelerated by the introduction of Global Position Systems (GPS) as afforded by geosynchronized satellite constellations, not to say electronic data-interchanges as afforded by the Internet. All things considered, the integration effort is really at its infancy. While a fair amount is known about data collection and organization, little is available regarding the interface between database and model-solution algorithms. Most modellers tend to assume data are available to support the analysis. Likewise, most data-collectors also take the attitude of comprehensiveness, or the collection of *any* pieces of information available irrespective of their relevancy. The result is that models become increasingly more sophisticated, often with a zealous appetite for data—data that may not be available. At the other end, reams and reams of data may be downloaded from satellites and stored away, with no real prospects that any but a minute fraction of them would ever be used (if at all). Sophisticated models are not only data-hungry, they are equally demanding on execution time. Thus vehicle-routing models may take days to execute on the latest computer, while logistical decisions are eagerly waiting to be made in the field. This does not even include the time required to assemble the data required for meaningful model application in the first place!

It is our contention that relevant data need to be collected in a format directly amenable to model application. Notice this is not a trivial task. To the extent that the repertoire of models have been developed with the mindset that data are available to support its execution, a new paradigm needs need to be constructed that truly configure models around the available databases. A simple example may make this clear. Recently, operations research (OR) models have been coded on a spreadsheet. For the first time, a database tool such as the spreadsheet has been integrated with model construction. What is more gratifying is that oftentimes, the execution speeds of these spreadsheet-based algorithms rivals (and in some cases exceed) the conventional algebraic ones. This is the tip of an iceberg that will blossom into its full potential if conscientious effort is made to remove some of the major obstacles. To the extent that GIS is the "spreadsheet" for transportation analysis, we argue that there should be more integration between the two sides—much as the synergy that has taken place between spreadsheet and OR models.

## IV. GEOGRAPHIC INFORMATION SYSTEM

The advantages of GIS are well known. They include the ability to integrate layers of spatially oriented data through a variety of analytical approaches. An advantage of GIS is that the data can be easily retrieved, and it allows for the interaction between facility location, land-use, and transportation analysis. For example, the implication of siting a hazardous facility in terms of environmental impacts among surrounding land, including the transportation of hazardous materials to and from the site, can be easily assessed using GIS. In emergency management, the evacuation of civilian population in the event of manmade or natural hazards can also be facilitated by the timely availability of data. Another advantage of GIS is that a large amount of data can be processed rather quickly. The process of automation allows scale and projection changes to be readily performed. Image distortions can be easily removed and coordinate rotation and translation is at the push of a key. In summary, GIS (including remote sensing) allows for the ready application of empirical and quantitative models.

An early GIS based on remote-sensing data is the Image Based Information System (IBIS) of the Jet Propulsion Laboratory. Most data in this system are in *raster* or image-based format, complimented by ancillary and tabular information. GIS needs accurate update of the various spatial-data elements, and remote-sensing systems provide precise ground data on a most timely basis for updating purposes. An illustration of IBIS is shown in Figure 1, where the typical "pancake stack" configuration contains a planimetric base, a geo-reference plane, land-cover data, and other image-based data. It should be noted that the planimetric base usually comes from a LANDSAT imagery and the geo-reference plane comes from (say) census-tract maps. Both of these serve as reference mats for other layers of data. To show how other data can be referenced against these mats, we illustrate with an interface file in Figure 1, consisting of a district name, the number of image elements (pixels) within that district, and the grey values of these pixels. Such an interface file can also provide an accurate accounting of (say) census information such as population and employment density. Most of us have seen examples of an IBIS-like GIS. Weather channels on television often show an image from weather satellites, with an overlay of the state and national boundaries. In this example, the pixel based clouds and land cover illustrates the planimetric base, while the state and national boundaries illustrate the geo-reference plane. Obviously, there are more sophisticated applications beyond the "pancake stack" illustrations. Recent advances in relational databases and object-oriented programming have greatly expanded our ability to "fuse" diverse number of formats together for various applications.

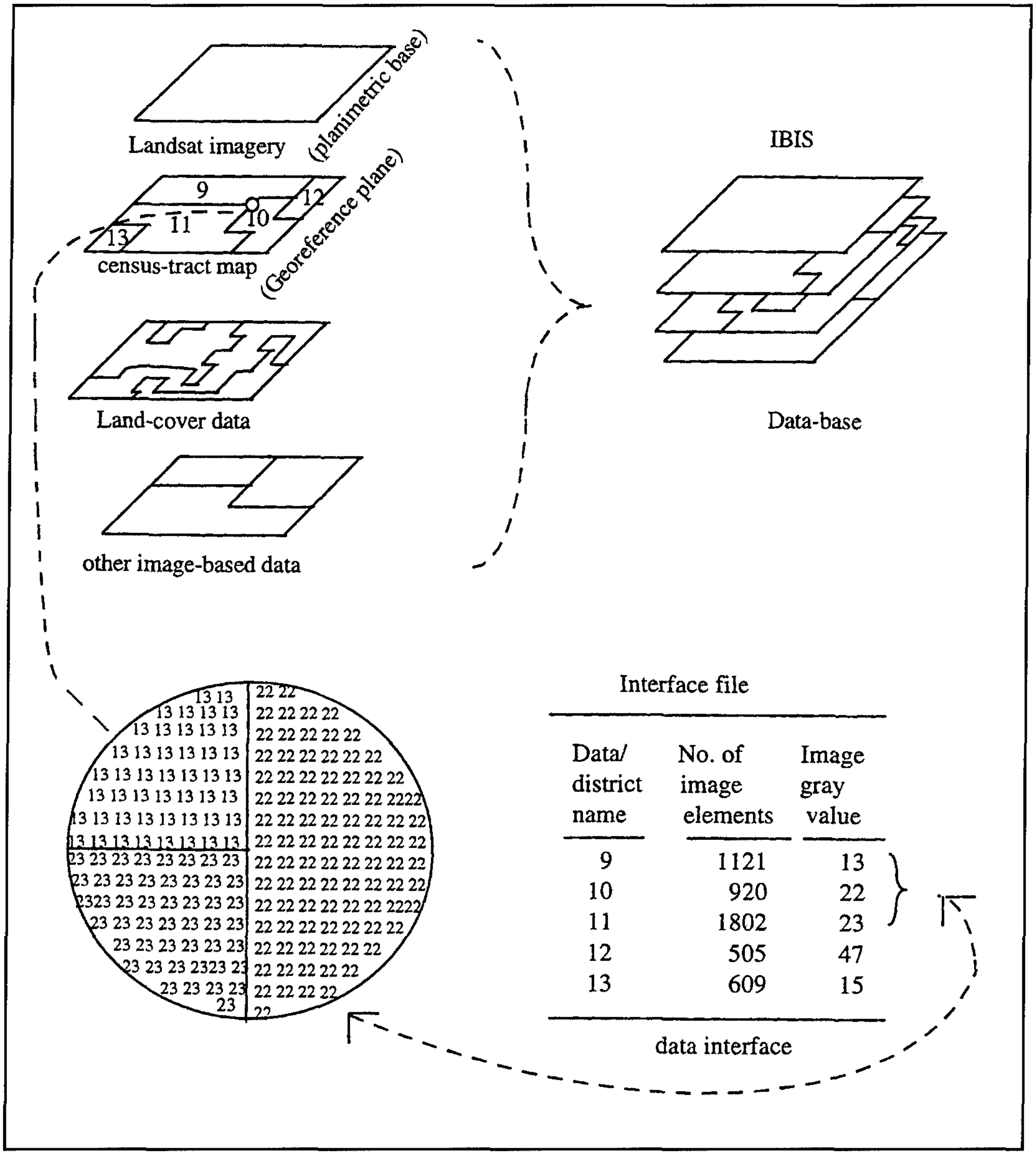
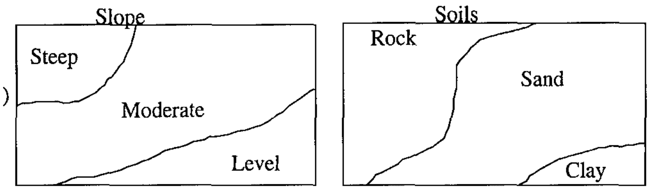



**Figure 1—Formation of an IBIS data base (Adapted from Marble and Peuquet [1988])**

As mentioned, the ultimate use of a GIS is for problem solving. Depicted in Figure 2 is a good example of how data can be merged to show useful information. Consider a raster-based GIS in which two of the files indicate the slope and the soil conditions of an area. Through a look-up table, trafficability information can be shown based on the combination of slope and soil conditions. For example, when the ground is level and the soil condition is rocky, it provides easy passage. On the other hand, steep slope combined with clay soil makes it hard to traverse. An output raster file can thus be produced from the two input raster files showing the trafficability of the land. One would agree that this output information—often accomplished through a mathematical model—satisfies those who wish to travel among this area, who can make informed travel decisions based on Figure 2(c).

(a)




(b)

| Slope \ Soil Type | Rock | Sand | Clay |
|---|---|---|---|
| Level | Easy | Easy | Easy |
| Moderate | Easy | Hard | Fair |
| Steep | Fair | Hard | Hard |

(c)

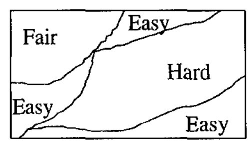


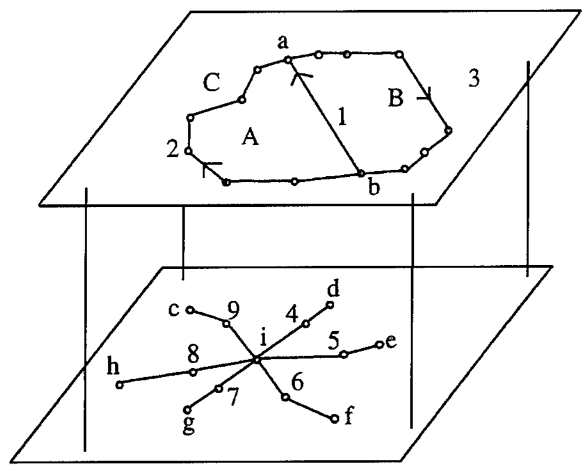

**Figure 2—"Manipulation and analysis (Star and Estes 1990)"**

The raster format is not the only way of storing geographic information. *Vector* and relational databases are two other very common formats. Shown in Figure 3 are two layers of information, consisting of geographic districts in the first layer and highway network on the second. One can see how geographic districts, represented as polygons, can be encoded by a *chain list*. For example, area A is bounded by a ring consisting of arc 1 and arc 2. At the same time, the relationship between arc 1 and other districts can be shown in an accompanying table where area B is indicated to be on the right-hand side in addition to area A being on its left-hand side. Furthermore, the highway network can be encoded in reference to the district map by virtue of the x-y coordinates. Thus any information base can and are often represented by a combination of vector, relational, and raster formats. A vector format is often more efficient in terms of storage space, while the raster format is more amenable to data merging among diverse databases.

To show the tremendous potential of GIS in problem solving, we cite a facility-location example due to Densham and Rushton (1992), who reported that the processing costs for algorithms can be drastically reduced by exploiting spatial structure. They follow a strategy in which the interpoint-distance data were preprocessed as both site-candidate and transportation-demand strings. These strings are then used to update an allocation table between sites and demands. The result is the ability to solve large problems of up to 3,000 nodes on a personal computer. Most importantly, solution time is a linear function of a problem size [O(n)], meaning that the execution speed is in the order of the number of nodes $n$. This is particularly exciting considering the non-deterministic computational complexity (or the exponential time of algorithmic execution) of most location problems. The speed with which the Densham-and-Rushton algorithm executes has direct bearing upon tactical application such as prepositioning supply depots in an emergency-relief operation or a wartime conflict.

Geographic districts

Highways




Geographic districts - Polygons

| Area ID | Ring (chain list)* |
|---|---|
| A | -1, 2 |
| B | 3, 1 |

Geographic district - boundary chains

| Boundary chain ID | From node | To node | Left area ID | Right area ID | x-y coordinate |
|---|---|---|---|---|---|
| 1 | b | a | A | B | $x_1\, y_1 \ldots x_n\, y_n$ |
| 2 | b | a | C | A | $x_1\, y_1 \ldots x_n\, y_n$ |
| 3 | a | b | C | B | $x_1\, y_1 \ldots x_n\, y_n$ |

Highway Network chains

| Network chain ID | From node | To node | x-y coordinates |
|---|---|---|---|
| 4 | i | d | $x_1\, y_1 \ldots x_n\, y_n$ |
| 5 | i | e | $x_1\, y_1 \ldots x_n\, y_n$ |
| 6 | i | f | $x_1\, y_1 \ldots x_n\, y_n$ |
| 7 | i | g | $x_1\, y_1 \ldots x_n\, y_n$ |
| 8 | i | h | $x_1\, y_1 \ldots x_n\, y_n$ |
| 9 | i | c | $x_1\, y_1 \ldots x_n\, y_n$ |

* Negative entries indicate reverse order

**Figure 3—Chain and Polygon data records for GIS (Nyerges and Dueker 1988)**

*A. Location example.* Consider the problem of locating *p* facilities or finding the *p-medians*, where p = 1, 2, .... In this case they are to be placed in the continuous plane (rather than at discrete locations in a network). We assume that travel occurs according to the *rectilinear metric* in the presence of impenetrable barrier to travel (Larson and Sadiq 1983), meaning that travel can only take place in the east-west and north-south directions and in avoidance of such intraversable regions as a lake. The number of demand points is finite and that the demands are proportional to the share of total population. Figure 4 shows an example of a case with three equally weighted demand-points and one diamond-shape barrier. The objective is that one wishes to locate *p* facilities in a manner that minimizes the average travel-distance to all random demands, assuming that each demand is served by its closest facility and that facilities are always available to service demands. Such a problem arises in a number of contexts, including the location of obnoxious facilities in an urban area and the associated routing of hazardous materials around the population, which defines the "barrier" region to travel.

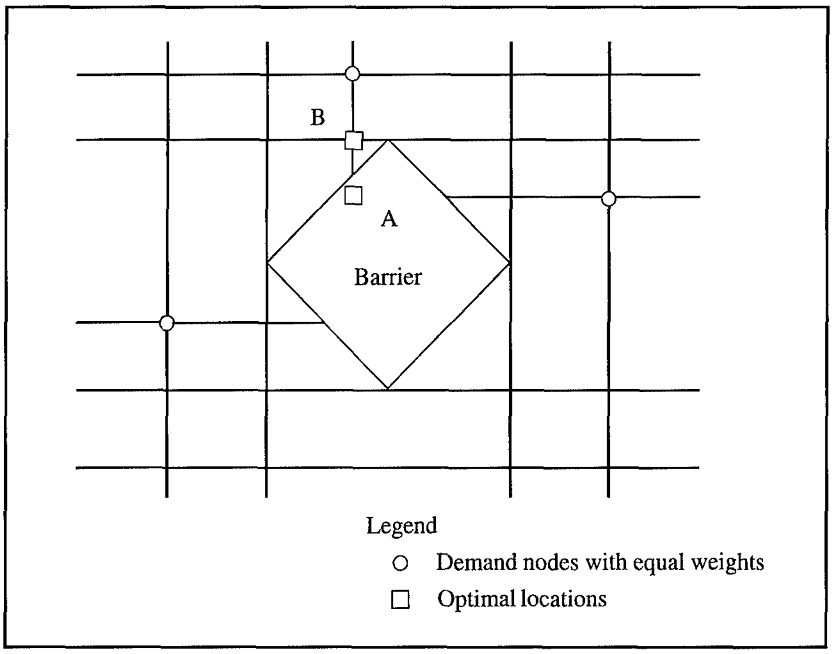



**Figure 4—Illustrating Larson and Sadiq's model (Ding et al. 1994)**

An optimal selection of facility locations can be drawn from a finite set of easily identifiable candidate points. This set of points is determined by drawing grid lines—lines parallel to the two travel directions through each demand point and lines tangent to the barrier boundaries in the two rectilinear travel-directions, with all lines terminated when intersecting a barrier. Intersection points of grid lines and demand nodes themselves provide the candidate locations. We note that the actual number of candidate locations depends upon the number and shapes of the barriers and the location of demand points relative to the barriers. However, the number is finite and typically not too much larger than the square of the number of demand points. Returning to the example, the 1-median objective—or locating the most proximal facility—if evaluated at each candidate point gives point B as the optimal solution. We note that this is a different location than the median solution if the barrier is assumed to be absent in which case point A will be selected. Because of the obviously geometric nature of this algorithm, it can be verified that this model can be easily implemented in a GIS (Ding et al. 1994).

*B. Location-routing example.* Another example of GIS application is drawn from location-routing applications, in which a facility is located in explicit consideration of the access and egress routing. Many of us know that location-routing analysis can be performed readily by Space Filling Curves (Bartholdi and Platzman 1988). Shown in Figure 5 is a three-dimensional Space Filling Curve where the x-y dimensions correspond to latitude and longitude of a map and the z's represents the "service" rendered. By constructing a *Sierpinski's curve* through the fractal cubicals (shown as the bolded dashed line in Figure 5), one can discern the logical clusters of locations where services can be delivered by a single vehicle-tour. In the medical-evacuation problem shown in Figure 5, the Space Filling Curve identifies the hospitals to which the wounded U.S. soldiers can be delivered from the Charlotte hub where the wounded from overseas are initially dropped off. In this example, a planimetric base with latitude and longitude information can be combined with ancillary data such as the hospital beds available at each location to complete the entire analysis, as shown in the following table (Carter 1990):

| *Location i* | *Hospital* | *Latitude $x_i$* | *Longitude $y_i$* | *Patients $z_i$* | *Sierpinski Function* |
|---|---|---|---|---|---|
| 1 | Charlotte | 35.21 | 80.44 | 0 | 0.03125 |
| 2 | Ft Gordon | 33.37 | 81.97 | 39 | 0.81250 |
| 3 | Ft Bragg | 35.17 | 79.02 | 234 | 0.83940 |
| 4 | Ft Jackson | 33.94 | 81.12 | 44 | 0.90630 |
| 5 | Charleston, SC | 32.90 | 80.04 | 29 | 0.95310 |

When the three-dimensional space-filling curve (or Sierpinski's curve) is constructed as in Figure 5, the cluster of points on the "spaghetti-strand like" curve—Fort Gordon, Fort Bragg, Fort Jackson, and Charleston—constitute a tour. Remember that the spatial separation between the depot at Charlotte and the four hospitals listed are represented as distances on the wrap-around curve via the Sierpinski function. Granted that the use of Space Filling Curve (SFC) is not as precise as an analytical model based on mathematical programming, one can readily agree that this approach exploits the existing data structure in GIS, and the solution has been shown to be asymptotically within 25 percent of the optimal as the number of hospitals grows. Most importantly, the SFC algorithm executes with a computational complexity of O(n log n).

The generic SFC heuristic works as follows:

*Step 1.* Depending on whether we are solving a two-or three-dimensional problem, transform the problem in the unit square or cube, via an SFC, to a problem on the normalized unit interval. For example, given the $n$ three-dimensional coordinates$(d_{1i}, d_{2i}, d_{3i})$ of the nodes i (demand points), compute the Sierpinski number $\psi(d_{1i}, d_{2i}, d_{3i})$ for each node.

*Step 2.* Solve the easier location or/and routing problem on the unit interval $0 \leq \psi \leq 1$. ■

SFC is particularly adept in measuring spatial separation. For example, it is common to apply the SFC to routing problems, such as the *travelling-salesman problem* (TSP), defined as the tour that visits specified demand points from a depot and returns in the shortest length. Such a tour can be approximated by the Sierpinski curve by sorting the numbers $\psi$ in ascending order and visit the cluster of nodes in the same order producing a tour. In a logistical application, any point in the tour can be the depot, as long as this 'warehouse' is the most logical considering the locations of these warehouses vis-a-vis the central depot from where the warehouses receive their supplies (and in turn deliver to the final demand points.) A distinct advantage of the SFC is that it executes in seconds—a far cry from the formal solution of a

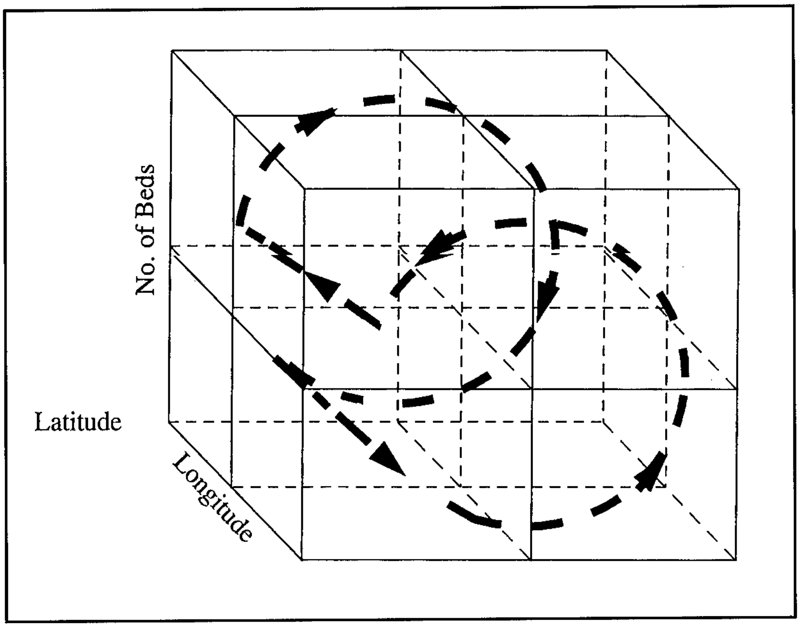



**Figure 5—Three-dimensional space filling curve**

TSP! The speed allows a large number of replications to be computed very quickly corresponding to a variety of stochastic demand scenarios.

*C. Real-time diversion example.* As yet another example, Regan et al. (1995) investigated the real-time diversion strategies for commercial-vehicle operations. Under idealized geometries directly discernable from GIS, the authors outlined the scenarios where diversions may be beneficial. The following fundamental question is asked, "While a driver is en route to a load, information about another load to be moved becomes available. What is the probability, given various diversion-decision rules, that the driver will be diverted to serve the new load first? What is the probability that following such diversion-decision rules will result in a reduction of overall distance traveled? And, what are the associated expected restrictions in travel?" Simulation experiments show that the *greedy* or myopic look-ahead strategy consistently lead to a reduction in overall travel, even with very limited real-time information. In other words, the primary contribution is derived from the database organization, which facilitates the use of such real-time information (including that from the Global Positioning System which provides the precise location of a vehicle for navigational purposes).

Consider the example in Figure 6, a vehicle that begins at the center C of a circle and moves toward a load origin $x_1$ that is uniformly and randomly generated over the area of the circle. Given a diversion-point P some fraction of the distance from the center C of the circle and origin $x_1$, what is the probability that the distance between the diversion point P, to a new origin $x_3$ will be less than the distance from P to origin $x_1$? Let $\alpha$, $0 \leq \alpha \leq 1$, denote the fraction of the distance from C to $x_1$ travelled to reach P. It can be shown that the probability that the distance from P to the new origin is less than to the old origin is given by $(1-\alpha)^2/2$. Following a myopic strategy of diverting to the new demand origin $x_3$ if it is closer to the diversion point than origin $x_1$, then $(1-\alpha)^2/2$ represents the fraction of loads for which we actually divert.

A more plausible diversion-strategy would also consider the relative distances between the destination point of the first movement $x_3$ and $x_1$ and the origin point of the next load $x_2$ and $x_4$. In Figure 6, these are given by $C(x_2,x_3)$ and $C(x_4,x_1)$ respectively. In this case, diversion is chosen if $C(P,x_3)+C(x_4,x_1)<C(P,x_1)+C(x_2,x_3)$. Analytic derivation of the corresponding diversion probability under this strategy is no longer straightforward because the respective distances are not independent. Correspondingly, the diversion likelihood and associated expected benefit are evaluated using a simulation program. The analysis can be further extended to consider several demands in a particular decision to divert and to look at demands that are uniformly generated in space but arrive according to a Poisson stream as well as from a uniform distribution. In addition, Regan et al. have explored operational constraints in which every demand must be served, and those in which there is freedom to accept or reject demands according to the cost of serving them. The exploration of idealized scenarios suggests that a reduction of overall travel distance of between five to ten percent would not be unreasonable. In lieu of re-routing, optimal re-sequencing of the first few queued demands results in an attractive shipping strategy. Notice this is analogons to metering and spacing in the Air-Traffic-Control System in which the takeoff time from origin airport A is delayed to avoid the congestion at destination airport B.

When a TSP routing is considered, the problem is termed the dynamic travelling-repairman problem (DTRP). Asymptotic analytical-results of the DTRP are reported in Psaraftis (forthcoming) for both the single- and multiple-vehicle cases. Healy and Moll (1995) showed that extension can be made to the local-improvement algorithm. Inasmuch as the local-search is often characterized by uneven neighborhood structure (instead of a circle as shown in Figure 6), a universal secondary metric, neighborhood size, can be used as a proxy metric for the search algorithm. The strategy is to move away from a primary-metric local-optimum to a neighboring solution if the new solution's primary-metric evaluation is not much worse, and the new solution has a larger feasible local-neighborhood. Federgmen and Tzur (1996) proposed an exact solution to this routing problem. One thing is clear, all algorithms rely on the availability of geographic information, whether it be a neighborhood defined by a circle or one with irregular shape.

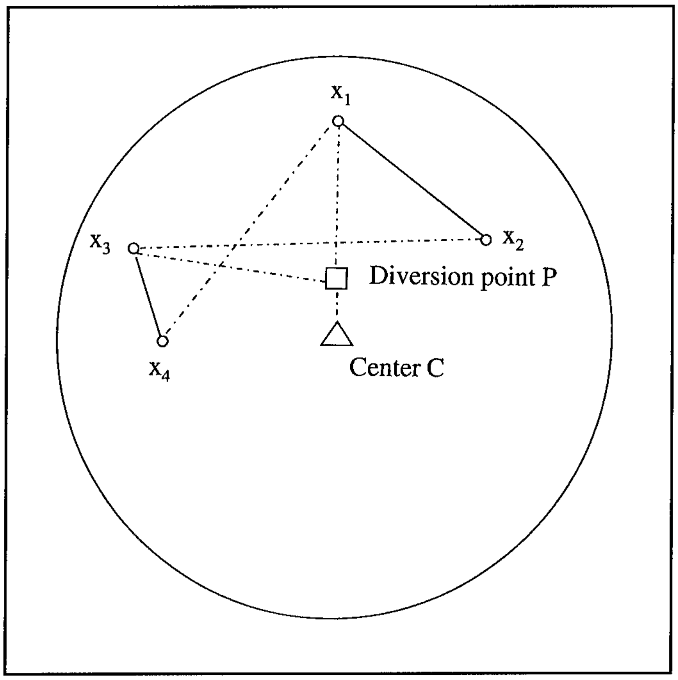



**Figure 6—Diversion example (Regan et al. 1995)**

# V. REMOTE SENSING

Remote sensing can serve as an excellent source of timely (and even real-time) information for GIS. The weather satellite, for example, has resolution of one kilometer, and a single pass of such satellites can cover areas the size of a continent. Very inexpensive roof-top antennas can be used to collect such data on a real-time basis. LANDSAT has a finer resolution of 80 by 80 meters, and recent versions have pixels as detailed as 30 by 30 meters. Pixels smaller than one meter are available in military reconnaissance (Zimmerman 1988). Currently in the commercial market, SPOT is able to offer 10-meter resolution in black-and-white and 20 meters in color. In July 1987, the then Soviets offered a 6-meter resolution satellite. Sweden is now considering a satellite offering as fine as one-meter resolution. LANDSAT and certain Russian satellites can detect long-wave-length radiation produced by heat sources. Both LANDSAT and SPOT can "see" short-wave-length infrared radiation. Being able to sense heat sources and infrared radiation is a tremendous asset of remote sensing not affordable by regular sources of information such as census and traditional aerial photography. There is a physical limitation about satellites however. They cannot orbit the earth faster than once every 90 minutes or aerodynamic drag would become insurmountable. They can only photograph a limited swath of the earth during each revolution, and oftentimes many swaths have to be pieced together as in aerial photography to make up a complete picture of a study area. Considering its orbits around the earth, an average lag-time of half-of-a-day is required to acquire a specific picture. Geosynchronized satellites such as GOES overcome this problem, but at a tremendous cost. Generally speaking, the best satellite would require a full day to photograph the entire earth given the speed with which it can orbit.

*A. Sensors.* To turn satellite images into useful information, it is necessary to classify pixels into land-cover "objects" such as forests, lakes, oceans, and man-made objects such as airfields and buildings. Such image classification is often referred to as *pattern recognition. Spectral* pattern recognition refers to a set of decision rules that are based solely on the spectral radiances observed in the data. Thus, grass emits a different radiance in comparison with rock, and we say that grass can be spectrally recognized and distinguished from rock on this basis. *Spatial* pattern recognition, on the other hand, is a set of decision rules based on the geometrical shapes, sizes, and patterns present in the data. A contiguous block of pixels of certain grey values, for example, can represent a lake. If we know the shape of an object in addition, such as a *rectangular* runway, we can start to look for a patch of concrete with such an elongated shape as well. The remote-sensing characteristics of the Advanced Very High Resolution Radiometer (AVHRR) of the National Oceanographic and Atmospheric Administration satellites are facilitated by five spectral bands (Sabins 1987). The first band is best for the detection of clouds and vegetation. Band two is most suitable for detecting shorelines and vegetation. Band three is used to detect hot targets such as fires and volcanos and so on. With five or more spectral bands, most modern satellites can gather a wealth of information unimaginable just a couple of decades ago. This information allows much better-informed decisions ranging from manmade-or-natural-hazard management to environmental monitoring.

To illustrate the sophistication of remote-sensing in image classification, we have included in Figure 7, shown as an elongated rectangle lined up with 17th Street, a portion of the Washington, D.C. mall. SPOT images of this area are shown in Figure 8, including channels one, two, and three. It can be seen that channel one is particularly adept in detecting bodies of water such as that found in the Constitution Garden, the Reflecting Pool and the Tidal Basin. Channels two and three, on the other hand, are very poor sensors of water while they could very well be excellent sensors of other land-cover types (such as pavement). By a combination of more than one sensor, one can improve identification and recognize land-cover features with better confidence (Amrine 1992). Depending on the problem we are solving, the pertinent sensor can be exploited correspondingly. By a combination of remote-sensing images, including satellite and aerial photography, one can, in turn, construct a variety of image scales. For example LANDSAT multi-spectral scanner images provides scales of one in 250,000 or smaller. At the other end of the spectrum, low altitude aerial photography provide images larger than one in 20,000 scale (Anderson 1976). One thing is clear, the information afforded by remote sensing is extremely rich, providing multi-level classification, such as Figure 8, of images for various applications. Combined with other data in a traditional GIS, the modelling procedure can be further enhanced.

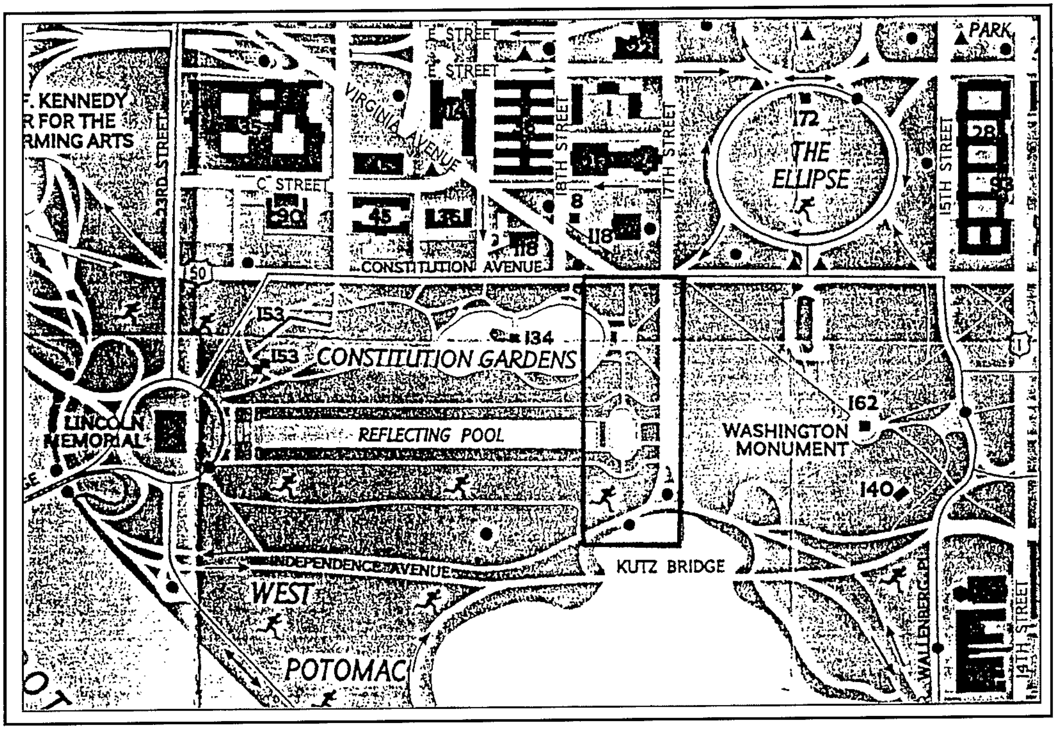



**Figure 7—Portion of the Washington D.C. Mall (National Geographic 1994)**

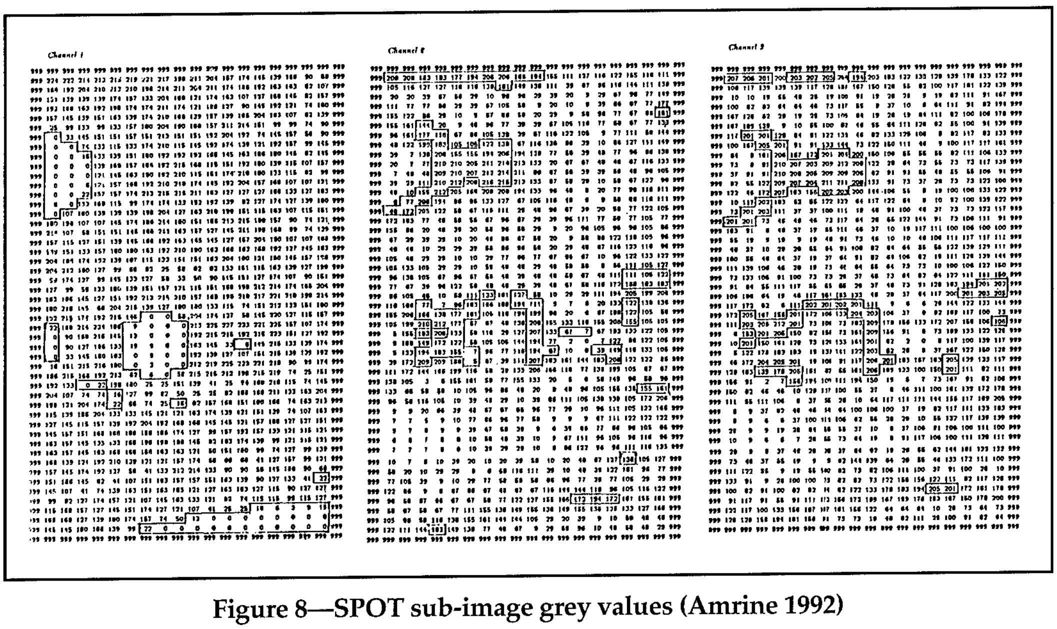


**Figure 8—SPOT sub-image grey values (Amrine 1992)**

*B. Advanced information.* Aside from target-recognition and intelligence applications, remote sensing offers advanced information that allows for pre-planning to take place. Given a two-dimensional data set, such as a satellite picture, if we have elevation information in addition, which can be derived from a series of satellite pictures at different viewing angles, a geometric model can be constructed in the computer. The output image will be a three-dimensional representation of what started as a two-dimensional scene. This type of image manipulation has many possible uses. American B-1 bomber pilots could rehearse in simulators low-level bombing missions, becoming familiar with enemy terrain without ever getting near it. The same can be performed for other air-mobility missions in the theater. In this case, *advanced,* instead of real-time, information is available for vehicle-routing purposes. Such advanced information allows planning in anticipation of future events. In Intelligent Transportation Systems, remote-sensing can monitor highway-traffic flow, offering advanced warnings to motorists about potential congestion and advising them about diversion strategies using the appropriate exit and entrance ramps.

Let us discuss this traffic-monitoring example in further detail. McCord (1995) experimented with the use of remote-sensing devices to monitor traffic flow. Specifically, he explored the use of geosynchronized satellites to measure traffic-flow via the fundamental equations governing flow, density, and velocity (Morlok 1978). He found that the latest (commercially available) technology does not have the one-meter resolution detailed enough for such applications. Currently, no geosynchronized satellites offer one-meter resolution, only orbiting satellites offer this capability. This poses the additional challenge as to how many highways one can "see" per unit time. He did, however, design a methodology using *nonlinear programming* to outline how this can be performed once the appropriate resolution is available. He found out that one can cover about one percent of the highways in the Continental United States per day. Sensitivity analysis shows that data transmission and resolution are the limiting factors to increase coverage. To the extent that the fundamental flow-equations govern all modes of transportation, including air traffic, such a methodology is general enough to monitor other traffic-flow patterns. It represents a worthy research agenda that will eventually provide real-time strategies for traffic control. This is particularly cogent for the military, which has the technology to gather less than one-meter resolution images and the communication capacity for electronic data-transfer.

## VI. SPATIAL-INFORMATION PROCESSING AND MODELLING

As mentioned, the fundamental idea behind image classification is that different land-cover types are associated with different *combinations* of digital numbers (DNs) based on the inherent spectral reflectance and emittance properties on each sensor-scanner band. We often refer to such combinations as *signatures* of different land-cover types, which allow for their accurate identification. Once spectrally identified, spatial pattern-recognition involves the categorization of image pixels on the basis of their spatial relationship with pixels surrounding them. Spatial classifiers might consider such aspects as image texture, pixel proximity, feature size, shape, directionality, repetition, and context. *Noncontextual* classification is based only on spectral pattern-recognition while *contextual* classification is based on both spectral and spatial pattern-recognition. Both of these two classification-results can be compared with the "ground truth." More sophisticated classifiers employ *Bayesian statistics* exploiting updated information and a *logistic-discriminant* model which considers the context of a pixel. It can be verified that contextual classification often yields a much more satisfactory result than its noncontextual counterpart (McLachlan 1992).

While these classification techniques are well established, a more interesting, tempored extension can be shown in a deforestation study at the eastern part of Texas. We trace a single pixel of a weather-satellite image of the Texas-gulf-coast forest over a 144-week period. This time-series information includes approximately three years of data depicting decline of the vegetation in the fall and winter and greening in the spring. Typical *time-series analysis* can be performed on this set of data by removing the seasonality and analyzing the resulting "stationary" data. We show an example series for pixel 191 in Figure 9. Sophisticated analysis can be performed on this time series to discern whether the vegetation is healthy or deforestation is taking place. The idea is to trace deforestation spatially as well as temporally discerning the possible "spread" of any disease.

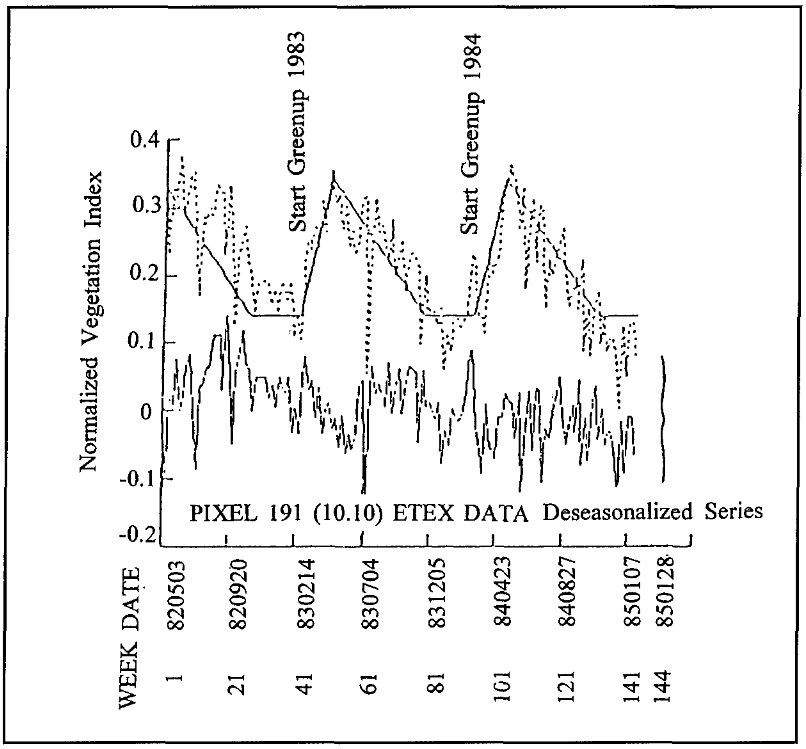



**Figure 9—Time series of a pixel in Eastern Texas (Robinson 1987)**

*A. Transportation Planning Models.* An extension of such an analysis can be applied to transportation applications, in which the migration of population or employment can be tracked in the same way. To model this type of economic-activity and land-use information, we show the fundamental ideas in Figure 10. Here a subject pixel is shown at the center of the image, which could very well represent the source of deforestation or downtown employment. Similar to the spread of deforestation, residential-location preferences can be modelled as the first-, second-, third-, and fourth-order neighbors—sometimes known as the rook and bishop contiguities and so on. It is not at all difficult to construct a "weight matrix" out of such contiguity information, as shown by the zero-one entries in Figure 11. Thus if one examines the row "e" of the weight matrix, the first-order neighbors b, d, f and h are marked with a "1". Once such weight matrix has been constructed, we have the center piece of most spatial analysis. The most common spatial-interaction functions, manifested typically in *gravity models*, are simply modifications of such a weight matrix (Chan-forthcoming).

Figure 12 shows a spatial function where spatial cost is calibrated by a power-function of beta ($\beta$). When beta is equal to one we have a Euclidian-distance measure between a target pixel and its surrounding neighbors as commonly illustrated and approximated in the first, second,

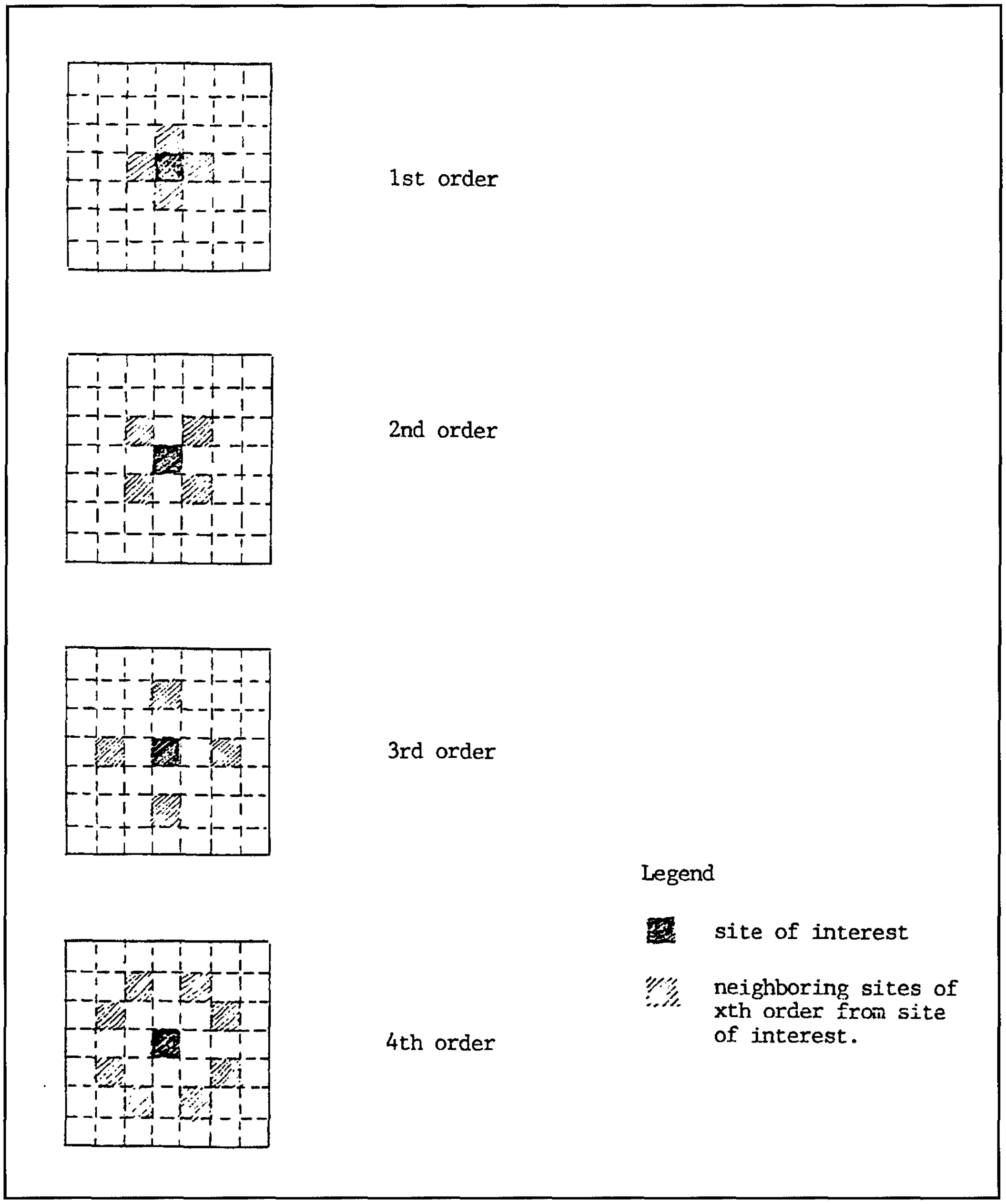



**Figure 10—Spatial order in a grid**

third, fourth-order (or higher-order) contiguity relationship. When beta is equal to two, a Newtonian form of the gravity model results. Instead of a zero/one weight matrix, entries of the weight matrix are now generalized to fractional values, broadening the contiguity concept to cover an amazingly rich number of spatial models, including many forms of transportation analysis. Thus row "e" of the weight matrix $\mathbf{W}=[w_{ij}]$ may assume the form

$$w_{ej}=\frac{C_{ej}^{\beta}}{C_{eb}^{\beta}+C_{ed}^{\beta}+C_{ef}^{\beta}+C_{eh}^{\beta}} \qquad (j=b,d,f,h) \tag{1}$$

for first-order neighbors and

$$w_{ej}=\frac{C_{ej}^{\beta}}{C_{ea}^{\beta}+C_{ec}^{\beta}+C_{eg}^{\beta}+C_{ei}^{\beta}} \qquad (j=a,c,g,i) \tag{2}$$

for second-order neighbors. For normalization purposes, there may be a requirement to have the row sum of spatial weights equal 1:

$$w_{ea}+w_{eb}+w_{ec}+w_{ed}+w_{ee}+w_{ef}+w_{eg}=1. \tag{3}$$

It turns out that this simple idea is behind all spatial interactions, including spatial competition in modelling market shares of competing retail outlets. A common use of such weight matrix is in forecasting travel demand, such as destination choice, transport-mode choice and route-choice (Oppenheim 1994). Although less well-publicized, it is behind the analysis of spatial conflicts such as *Cournot-Nash games* as well, of which predicting the outcome of a regional conflict is a natural application (Harker 1986).

| | | |
|---|---|---|
| a | b | c |
| d | e | f |
| g | h | i |

| W | a | b | c | d | e | f | g | h | i |
|---|---|---|---|---|---|---|---|---|---|
| a | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| b | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| c | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| d | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| e | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| f | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| g | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| h | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| i | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

**Figure 11—Illustrating the rook weight matrix**

Shaw (1993) examines the GIS requirements for integrating urban-travel-demand models consisting of trip generation (normally performed via linear regression), destination choice, mode choice and route-choice. He suggested that an ideal GIS for spatial analysis is a true integration of full GIS and modelling capabilities within a single system. The main difficulty associated with this approach is the differences of data requirements and analysis procedures between the GIS functions and various modelling procedures. He conjectured that a full integration of GIS and modelling capabilities is unlikely to take place, at least in the foreseeable future, because of two major reasons. First, a GIS-data model that could handle complex spatial relationships of spatial entities is yet to be developed. Second, the range of analysis procedures for various modelling applications is so wide, if not unbounded, that it

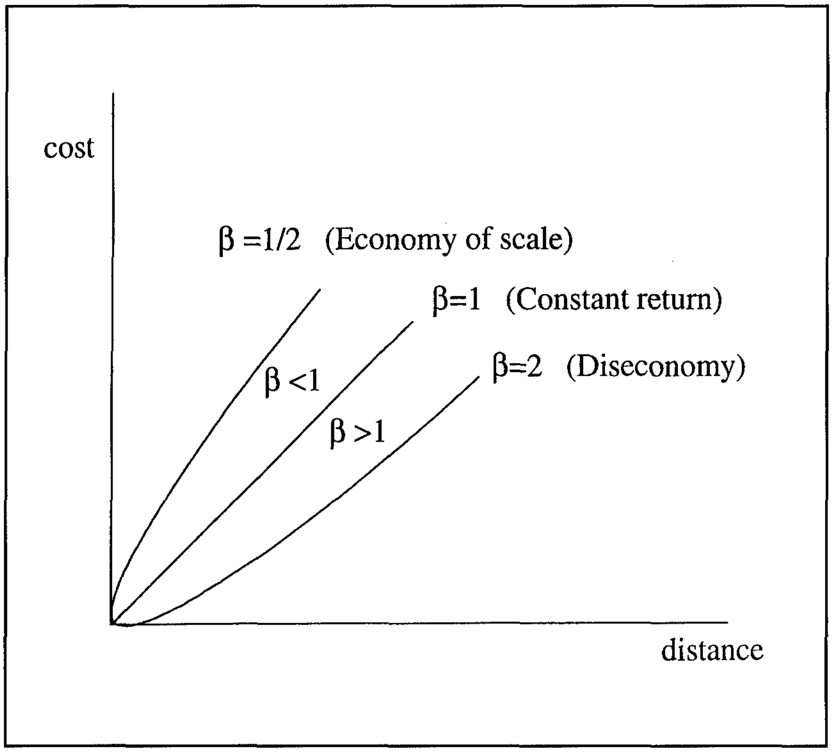



**Figure 12—Spatial function**

is not feasible to rely on a single data model to support all of them. I am a great deal more optimistic about the prospects, however, given the commonalities among spatial analysis as pointed out throughout this paper, and there is no reason why a GIS-data model cannot be constructed to capture this. We will come back to this point at the conclusion of this paper.

*B. Intelligence Application.* Combining spatial with temporal techniques, one can model a spatial time-series as the intelligence application illustrated in Figure 13, which tracks almost 80 months of data for 20 regions in the study area and shows the spatial distribution of an activity ranging from economic activities to natural-and-manmade disasters, including regional conflicts. In a study of a world-wide sensor system, Greene (1992) investigated the use of *spatial-temporal-autoregressive-moving-average* technique to forecast incidents that may happen around the globe. Building upon the spatial-order relationship as shown in Figure 10, the relationship between geo-political events among the 20 regions is modelled. When the spatial relationship is projected forward in time, it allows the U.S. Department of Defense to *anticipate* incidents and to initiate the necessary tasks to handle the situation. The following model is used:

$$z(t)=\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}z(t-k)-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}a(t-k)+a(t) \qquad (4)$$

Here $\mathbf{z}(t)$ is the 20-entry activity vector corresponding to the 20 regions being monitored, $\phi$ and $\theta$ are calibration constants, $\mathbf{W}^{(l)}$ is the 20x20 weight matrix for spatial order $l$ (with $\mathbf{W}^{(0)}=\mathbf{I}$). Matrix $\mathbf{W}^{(l)}$ has nonzero elements only for those pairs of sites that have been defined to be $l$th order neighbors. First-order neighbors are understood to be closer than second-order neighbors, which are closer than third-order neighbors, etc. The modeler specifies exogenously the order and magnitude of a nonzero entry for a particular pair of sites.

Instead of a "natural" projection similar to the East-Texas Vegetation Study, precipitous events may take place. A typical precipitous event is a "doctrinal change," defined as an abrupt switch over of operations. After the model included a doctrinal change of the adversary, it achieved significant success in forecasting incidents at region 11. Thus the univariate version of Equation(4) for the *11*th entry of the activity vector took on the form of a moving-average model (or the second group of terms in Equation (4) above) with the first group of autoregressive terms dropped out:

$$\begin{aligned} z_{11}(t) &= \sum_{k=1}^{2}\sum_{l=0}^{1} \theta_{kl} W_{(l)} a(t-k) - \theta_{12,0} W_{(0)} a(t-12) + a(t) \\ &= \theta_{1,0} W_{(0)} a(t-1) - \theta_{1,1} W_{(1)} a(t-1) - \theta_{2,0} W_{(0)} a(t-2) - \theta_{2,1} W_{(1)} a(t-2) - \theta_{12,0} W_{(0)} a(t-12) + a(t) \\ &= -0.424 a(t-1) + 0.160 W_{(1)} a(t-1) - 0.396 a(t-2) + 0.131 W_{(1)} a(t-2) + 0.5865 a(t-12) + a(t) \end{aligned} \qquad (5)$$

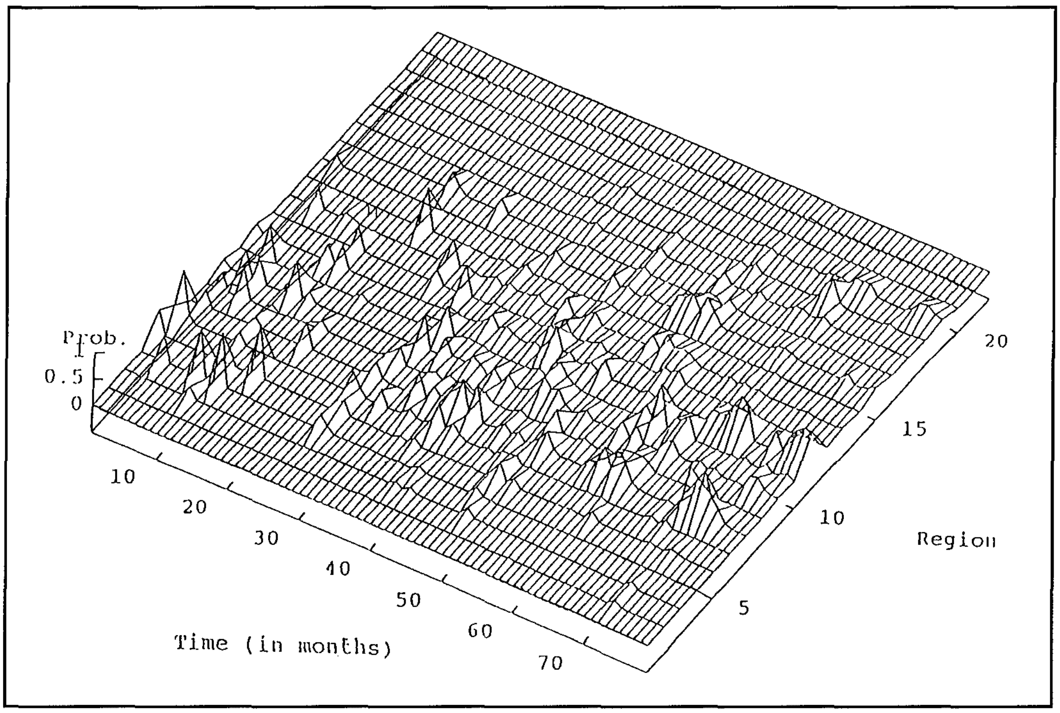



**Figure 13—"Univariate STARMA for region 11 [Greene 1992]"**

where $\mathbf{W}_{(l)}$ is the *11*th-row vector of $\mathbf{W}^{(l)}$. Here a 12-month seasonal-pattern was detected, and correspondingly modelled by the extra term introduced beyond the model specification in Equation (4). This extra term essentially removes the seasonal pattern similar to that shown graphically in Figure 9. Example weights between regions 11 and its first-order neighbors look like

| | | | | | | |
|---|---|---|---|---|---|---|
| Region | 7 | 9 | 13 | 15 | 17 | 19 |
| Weight | 0.68 | 0.17 | 0.12 | 0.01 | 0.01 | 0.01 |

where the sum of the weights equals unity as suggested by Equation (3). It can be seen that one only needs some very common-place statistical-functions such as *autocorrelation* computations—the temporal or spatial analogue of the regular Pearson-correlation coefficient used in ordinary-least-squares regression—to effect such a model. Such computations can be added to a GIS package in a straightforward manner. This three-dimensional plot illustrates the tremendous potential of GIS in model construction even when the time dimension is included.

*B. Routing revisited.* It turns out that the spatial-cost function as shown in Figure 12 has profound implications on spatial organization, since it is really the "price system" that allocate resources spatially. Aside from locating pixels, these resources include anything from transportation terminals, depots, to population and employment. For example, under Euclidean distance ($\beta=1$), nodes on a network are often optimal locations to place a depot. This is a well known result in discrete facility-location, often attributed to Hakami (1964). When $\beta$ is no longer unity, such optimal locations could very well be on an arc (or link) of the network and nodal optimality condition is destroyed. When a discrete network is replaced by a continuous plane, it can be shown that an optimal location may not even exist—as shown in the classic *Steiner-Weber problem* (Weslowsky 1993). As alluded to already, there are totally different locational implications when the spatial cost-function, f(distance)=$f(C)=C^{\beta}$, has the form of constant return-to-scale, $f(C)=C^1$, increasing return-to-scale, $f(C)=C^{\beta}$ for $\beta<1$, and decreasing return-to-scale, $\beta>1$.

Distance is often measured in the Minkowski metric, $l_p(C)$—as is common for approximating highway distances between two points on a map (Love et al. 1988). The spatial cost-function now becomes f(distance)=$[l_p(C)]^{\beta}=[(\sum_i C_i^p)^{1/p}]^{\beta}$, where $C_i$ is the rectilinear metric as measured in the east-west direction or north-south direction (or for that matter, any third or fourth dimension and so on). In this case, spatial cost $f$ no longer satisfies the triangular inequality $f_{ik} \leq f_{ij}+f_{jk}$ when $\beta>1$ (Love et al. 1988). This in turn invalidates a number of important procedures for solving travelling-salesman and vehicle-routing problems, including column-generation schemes commonly used in solution algorithms. Most importantly, performance guarantees no longer exists. And this applies to other heuristics algorithms such as the classic Clarke-Wright and spanning-tree procedure for solving travelling-salesman problem (Lawler et al. 1985). The space-filling-curve heuristic mentioned above, however, maybe robust enough to be still applicable (Bartholdi and Platzman 1988).

## VII. SPATIAL ANALYSIS

Carrying the analysis concept of integrating spatial-data with analysis, more sophisticated models can be built on the database and data structure. In fact the general family of models illustrated in Figure 9 and Figure 13 is referred to as the *auto-regressive-moving-average model* (ARMA). When we take the spatial dimension into account, such as in the previously mentioned deforestation study in eastern Texas, the model becomes the *space-time-auto-regressive-moving-average model* (STARMA). In the case of tracking a single pixel we had no spatial dimension and henceforth no spatial weights are employed. With the spatial dimension included (and hence a spatial weight matrix,) the model becomes a lot more versatile and robust. It allows us to analyze, for example, whether a point source of pollution has spread among vegetation spatially. Thus we are relating deforestation not only to time, but also how it spreads geographically. We can readily extend this idea to many facility-location, land-use, and transportation models. Placed in a broader methodological framework, STARMA is a version of the general vector-time-series models that has accumulated a wealth of knowledge base over the years. A specialization of STARMA is the econometric models that forecast economic activities (Chan-forthcoming). To the extent that future economic and geo-political activities drive transportation and mobility requirements, we show how the advocated spatial-structure will help such forecasts.

*A. An econometric model.* In the 1970s and 80s, a very widely disseminated economic forecasting-model EMPIRIC is in fact a *mixed-regressive-spatial-autoregressive model.* An examination of the following equations would verify this. It can also be shown to be a one-lag specification of the STARMA model—specifically the mixed-regressive-spatial-auto-regressive model: $\mathbf{z}=\boldsymbol{\Phi}\mathbf{W}\mathbf{z}+\mathbf{Z}\beta+\epsilon$ (Anselin 1988), where Z is a matrix of exogenous variables measuring $n$ x $K$ or 3 x 3 in the following example. $\Phi=\begin{bmatrix}\phi_1 & 0 & 0\\ 0 & \phi_2 & 0\\ 0 & 0 & \phi_3\end{bmatrix}$ is an $n$ x $n$ or 3x3 matrix consisting of autoregressive coefficients at its diagonal, and $\beta$ is a vector of calibration coefficients measuring $K$ x *1* or 3 x 1. The reader can verify that this is a specialization of Equation (4) when the moving-average terms are replaced by the regressive terms $\mathbf{Z}\beta$. Population, white-collar (w.c.) and blue-collar (b.c.) employment and hence travel demand have been successfully forecast in seven cities of North America using this sample EMPIRIC equation set.

$$(\Delta pop)_{t+1}=0.32(\Delta w.c.emp)_{t+1}-0.006(pop)_t+1.93(\Delta access\ to\ emp)_{t+1}$$

$$(\Delta w.c.emp)_{t+1}=0.42(\Delta pop)_{t+1}-0.006(w.c.emp)_t+0.96(\Delta access\ to\ pop)_{t+1}$$

$$(\Delta b.c.emp)_{t+1}=0.16(\Delta pop)_{t+1}-0.013(b.c.emp)_t+1.00(\Delta access\ to\ pop)_{t+1}$$

Here $\mathbf{z}=(\Delta pop,\ \Delta w.c.emp,\ \Delta b.c.emp)^T$, $\mathbf{Z}\beta=\mathbf{Z}_1\beta_1+\mathbf{Z}_2\beta_2$, $Z_1=\begin{bmatrix}pop_t & 0 & 0\\ 0 & w.c.emp_t & 0\\ 0 & 0 & b.c.emp_t\end{bmatrix}$, where

$Z_2=\begin{bmatrix}\Delta emp\ access_{t+1} & 0 & 0\\ 0 & \Delta pop\ access_{t+1} & 0\\ 0 & 0 & \Delta pop\ access_{t+1}\end{bmatrix}$, $\beta_1=(-0.006\ -0.006\ -0.013)^T$, $\beta_2=(1.93\ 0.96\ 1.00)^T$, $\Phi=\begin{bmatrix}0.32 & 0 & 0\\ 0 & 0.42 & 0\\ 0 & 0 & 0.16\end{bmatrix}$, $\epsilon$ is the error term in forecasting and the weight matrix is simply $W=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 1 & 0 & 0\end{bmatrix}$. The weight matrix in this case goes well beyond geographic contiguity. It represents "causal analysis" among economic sectors, showing the interdependency between population, white-collar employment and blue-collar employment.

The point is now clear, the desirable development of future GIS should keep in mind the synthesis between data collection and model development. We have shown by very simple examples that by judicious organization of data structure, facility-location, land-use, and transportation models can be readily built on top of a GIS. Standard spatial-statistics tools is a natural processor to be built upon the data pool, affording such computations as spatial auto-correlation, spatial interaction (the gravity model), entropy (an alternative to the gravity model in spatial allocation and organization), spatial dominance, and uncertainty (Pooler 1992). The last two—spatial dominance and uncertainty—are particularly relevant toward games and competition in which the physical positions of participants become important in their strategic advantage or lack of advantage over other participants. Available literature points to the fact that these specific techniques are nothing but special cases of the weight matrix. Specialization of the weight matrix will therefore allow these functions to be implemented very readily. Combined with the spatial *homogeneity* assumption—the assumption that each geographic unit behaves similarly—a whole host of transportation models can possibly be operationalized in real time, including *spatial gaming* and competition.

*B. Voronoi diagrams.* Another very exciting spatial-analysis development is in *Voronoi polygons* and its dual graph, *Delaunay triangles,* where both terms refer to tile-like patterns of partitioning x-y space. These are examples of spatial tessellation which has been identified as one of the most promising in a number of analysis, including spatial and transportation modelling (Okabe et al. 1992). Built upon the *primal-dual* concepts, Figure 14 shows the fundamental ideas behind this paradigm. Overlaid on top of the tile-like Voronoi polygons is the dual graph: Delaunay triangles. It can be seen that the triangles form the shortest paths linking up the "centroids" of the polygons. They form the natural channels for traffic flow between these subregions in the study area. Instead of the conventional square grids, such a pattern can be overlaid over a map to facilitate vehicle routing, including combat-aircraft routing (Information Warfare Center 1995). Figure 15 shows the very convincing power of these techniques in replicating the county boundaries in Ireland. Thus human settlement can possibly be "explained" in terms of such a spatial-analysis tool. The readers can verify easily that spatial-tessellation techniques—a systematic way of partitioning space—can be implemented quite straightforwardly given the database organized around a GIS, using a vector data-structure such as the one shown in Figure 3.

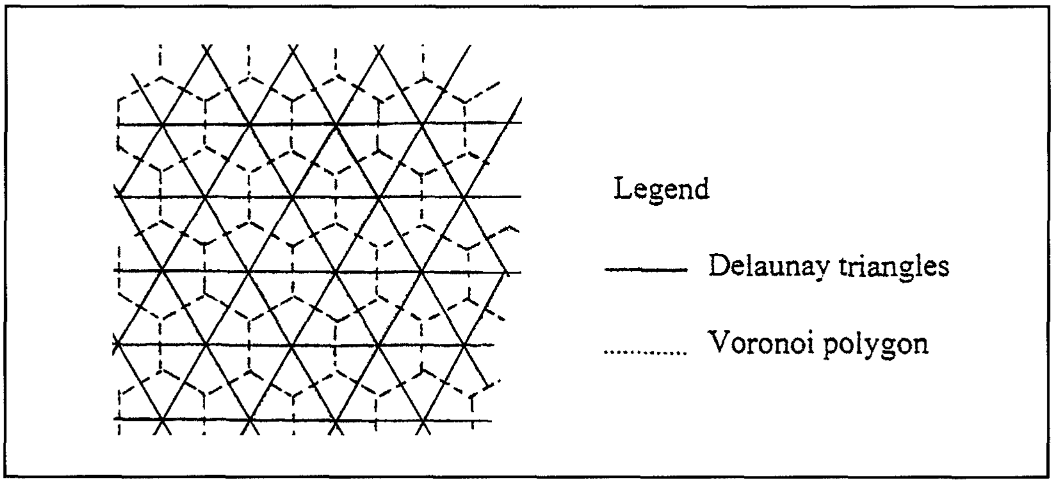



**Figure 14—Voronoi polygons and dual Delaunay triangles**

In spatial economics, a considerable number of studies have been carried out to examine market-area stability. Firms are located to compete against each other in order to maximize their profits. Suppose there are $n$ firms located at $x_1,...,x_n$ in a region $R_T$, and these firms are selling the same products with the same mill price—price not including a delivery cost. We assume that the delivery cost from a firm at $x_i$ to a consumer at $\mathbf{x}$ is proportional to the Euclidean distance $\| x_i - \mathbf{x} \|$, and that consumers buy the products from the firm that quotes the lowest delivered price (the mill price plus the delivery cost). Under these assumptions the configuration of $n$ market areas is represented by the ordinary Voronoi diagram $R(J)=\{R(x_1),..., R(x_n)\}$, and the market area of firm $i$ is represented by the Voronoi polygon $R(x_i)$. We further assume that the demand for the products is uniformly distributed over region $R_T$; the marginal cost of the products is the same for all firms; and relocation cost is negligibly small. Then the profit of firm $i$ is proportional to the area $R(x_i)$. The firms compete in terms of their locations to maximize their profits. As a result we observe a spatial-competition process of $n$ firms over time. The Voronoi diagram (or the market areas) may

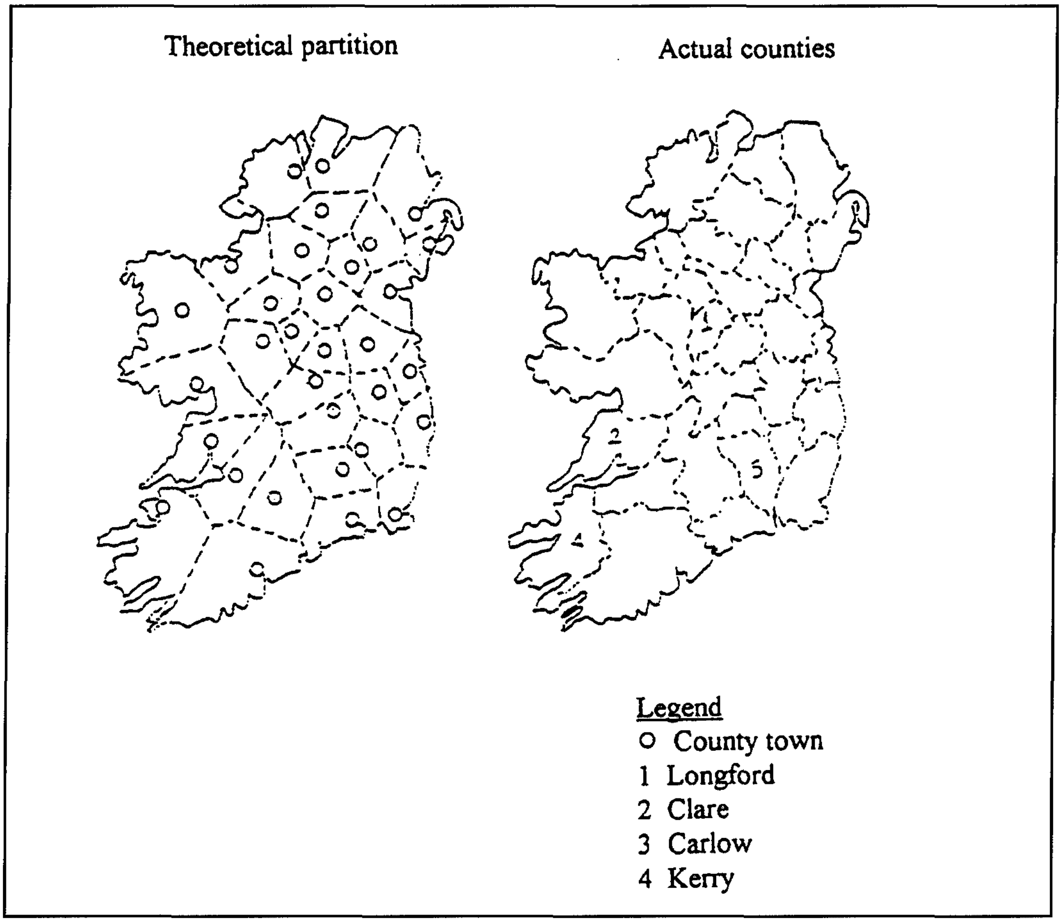



**Figure 15—"Defining Irish Counties using Voronoi diagrams [Okabe et al 1992]"**

change over time, or it may eventually reach an equilibrium state in which all firms have no incentive to relocate (a Nash equilibrium). To be precise, the configuration of the $n$ firms is in *global equilibrium* if and only if no firm can find a more profitable location than the present location, that is, the Voronoi diagram $R^*=\{R(x_1^*),\ldots,R(x_n^*)\}$ is in global equilibrium if and only if

$$|R(x_i^*)| \geq \max_{x_i \in R_T} \left\{ |R(x_i)| \,\big|\, [X_{-i}=X^*_{-i}, x_i \neq x_j^*, j \in (I-i)] \right\} \tag{10}$$

for any $i \in I$, where $X^*_i=\{x_1^*,\ldots,x_{i-1}^*,x_{i+1}^*,\ldots,x_n^*\}$ and I is the set of all discrete candidate-locations. It can be shown that the same concepts can be carried over to other spatial competition and conflicts, including military operations in the theater level (Electronics and Space Corporation 1994).

## VIII. CONCLUSION

In this paper, we have introduced remote-sensing as a viable source of real-time information. To the extent that aerial photography is one kind of remote sensing, the concept is not new. The large number of channels available in today's satellites, however, make available information that the bare eye cannot see, affording infrared and heat-sensing signals that are essential in transportation, reconnaissance, and target-location applications. Geosynchronized satellite constellations also provide real-time raster-images to a fine level of resolution, not to say the Global Positioning System, which is used extensively for navigation (US Department of Transportation 1995). Geographic information systems (GIS) allow the merging of data from diverse sources—from remote-sensing to survey and interview data. Modern data-processing capabilities such as relational database and object-oriented programming do not only facilitate data fusion, but also greatly streamline modelling applications, including spatial analysis. To the extent that spatial relationship is the basic building block for transportation and locational modelling, GIS becomes an integral part of today's analysis toolkit. It has been shown in this paper that a very desirable focus of GIS is problem solving. With the convenience of electronic data-transfer, GIS is also a *global* information system, affording truly distributed decision-making to take place. It is the contention of the author that through remote-sensing and through a very careful planning of the data structure, transportation, facility-location, and land-use analyses can be readily performed using models based on a spatial-oriented set of data-processing procedures including spatial statistics. Prescriptive tools can also be easily incorporated into a GIS. For example, optimization procedures based on the "generalized algebraic modelling language" builds heavily upon arrays that are organized in certain ways, representing vectors and matrices in a mathematical-programming model. These vectors and matrices (such as the "node-arc incidence" matrix) can possibly be extracted directly from a GIS through relational-data organization. The same arguments hold for recent emphasis on spreadsheet-based management-tools ranging from optimization to simulation. The developers of GIS should keep this in mind in their future endeavors.

We have demonstrated in this paper that there are some very basic principles involved in spatial-temporal analysis—a term that encompasses transportation, facility location and land use. Instead of calling upon a variety of analysis tools, data-oriented computation tools can be easily embedded into the database. The first example is the efficient space-filling-curve location-routing heuristic, which builds directly on the latitude-longitude coding of a location and an additional ancillary database (Chan-forthcoming). By preprocessing the interpoint distance data as both candidate and demand strings, an O(n) algorithm is found to locate facilities for a study area that can have up to 3,000 nodes. A third example is a simple "look-ahead" capability in a spatial database that will allow real-time diversion of vehicles in case of unexpected demands. In reconnaissance, spatial pattern-recognition builds directly upon the concept of contiguity, which is easily implemented on top of a spatial database in terms of a weight matrix. It turns out that such a weight matrix and the associated spatial-cost function is also the common vehicle to effect gaming and competition, not to say the placement of facilities, population and other economic activities such as employment on a plane or network, which in turn generate demand for transportation. Among other procedures, we have shown how the spatial-temporal-autoregressive-moving-average (STARMA) model or its specialized econometric form can be used to allocate economic activities over time. The allocation of these spatial resources is further assisted by the time-honored analysis tool called Voronoi diagrams—a technique based heavily on a spatial database. These examples provide convincing arguments for a brand new, simple and robust way of performing transportation and spatial analysis based directly on a database. Most importantly, these analysis tools can be truly executed on a real-time basis once the basic database (to be contrasted with the derived database) is in place.

## ACKNOWLEDGEMENT

Part of this research is supported by a grant from the Air Force Office of Scientific Research. The views expressed here are those of the author and do not reflect the official policy or position of the U.S. Air Force, Dept of Defense or the U.S. Government.

## REFERENCES

Amrine, J. M. (1992) "Spectral and Spatial Pattern Recognition in Digital Imagery," Master's thesis, AFIT/GSO/ENS/92D-01, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Anderson, J. R.; Hardy, E. T.; Roach, J. T.; Witmer, R E (1976) "A land use and land cover classification system for use with remote sensing data, " U.S. Geological Survey Professional Paper 964.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models,* Kluwer Academic Press

Bartholdi, J.; Platzman, L. (1988) "Heuristics based on spacefilling curves for combinatorial problems in Euclidean space," *Management Science,* Vol. 34, No. 3, pp. 291-305.

Carter, W. B. (1990) "Allocation and routing of CRAF MD80 aircraft," Master's thesis, AFIT/GST/ENS/90M-3, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Chan, Y (forthcoming) *Facility Location and Land Use: Multicriteria Analysis of Spatial-Temoral Information,* ITP/Southwestern.

Densham, P. J.; Rushton, G. (1992) "Strategies for solving large location-allocation problems by heuristic methods," *Environment and Planning A,* Vol. 24, pp. 289-304.

Ding, Y.; Baveja, A.; Batta, R. (1994) "Implementing Larson and Sadiq's location model in a geographic information system" *Computers and Operations Research, Vol. 21,* pp. 447-454.

Electronics and Space Corporation (1994) Presentation at the Second Air Force Mobility Modelling and Simulation User's Group meeting at the Department of Systems Science, Washington University, St. Louis, Missouri. (Hosted by E. Rodin on 16 November.)

Federgmen, A.; Tzur, M. (1996) "Detection of minimal forecast horizons in dynamic programs with multiple indicators of the future," *Naval Research Logistics,* Vol. 43, pp. 169-189.

Greene, K. A. (1992) "Causal Univariate Spatial-Temporal Autoregressive Moving Average Modelling of Information to Generate Tasking of a World-wide Sensor System," Master's thesis, AFIT/GOR/ENS/92M-12, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Hakimi, S. (1964) "Optimal location of switching centers and absolute center and the medians of a graph," *Operations Research,* Vol. 12, pp. 450-459.

Harker, P. T. (1986) "Alternative models of spatial competition" *Operations Research, Vol. 34,* pp. 410-425.

Healy, P.; Moll, R. (1995) "A new extension of local search applied to the Dial-a-Ride problem" *European Journal of Operational Research, Vol. 83,* pp. 83-104.

Information Warfare Center (1995) *Tactical Sensor Planner,* U.S. Air Force, San Antonio, Texas.

Kwan, M. P.; Golledge, R. G. (1995) "Integration of GIS with activity-based models in ATIS," Preprint 950419, presented at the 74th Transportation Research Board Meeting, Transportation Research Board, Washington, D.C.

Larson, R. C.; Sadiq, S. (1983) "Facility locations with the Manhattan metric in the presence of barriers to travel" *Operations Research, Vol. 31,* pp. 652-669.

Lawler, E. L.; Lenstra, J. K.; Kan, A. H. G.; Shmoys, D. B. (1985) *The Traveling Salesman Problem,* Wiley-Iuterscience.

Lillesand, M. T.; Kiefer R. W. (1987) *Remote Sensing and Image Interpretation,* pp. 610-705, Wiley.

Love, R. F.; Morris, J. G.; Weslowsky, G. O. (1988) *Facility Location Models and Methods,* North-Holland.

Marble, D. F.; Peuquet D. J. (1988) "Geographic information systems and remote sensing," in *Manual of Remote Sensing* (D. S. Simonett, Editor) Chapter 22, American Society of Photogrammetry, pp. 923-958.

McCord, M. (1995) Personal communication, Department of Civil Engineering, Ohio State University, Columbus, Ohio.

McLachlan, G. J. (1992) *Discriminant Analysis and Statistical Pattern Recognition,* Wiley-Interscience.

Morlok, E. (1978) *Introduction to Transportation Engineering and Planning,* McGraw-Hill.

*National Geographic* (1994) "Tourist Washington" Map insert.

Nyerges T. L.; Dueker K. J. (1988) "Geographic Information Systems in Transportation," Report to the U.S. Department of Transportation, Federal Highway Administration, Washington, D.C.

Okabe, A.; Boots, B.; Sugihara, K. (1992) *Spatial Tessellations - Concepts and Applications of Voronoi Diagrams,* Wiley.

Oppenheim, N. (1993) *Urban Travel: Integrated Demand Modeling and Supply Analysis,* course notes, Institute for Transportation Systems, City University of New York, New York, New York.

Pooler, J. (1992) "Spatial uncertainty and spatial dominance in interaction Modelling: a theoretical perspective on spatial competition" *Environment and Planning A,* Vol. 24, pp. 995-1008.

Psaraftis, H. N. (forthcoming) "Dynamic vehicle routing: status and prospects," *Annals of Operations Research.*

Regan, A. C.; Mahmassani, H. S.; Jaillet, P. (1995) "Improving the efficiency of commercial vehicle operations using real time information: Potential uses and assignment strategies," Preprint 950942, Presented at the 74th Transportation Research Board meeting. Transportation Research Board, Washington, D.C.

Robinson, J. N. (1987) "Identification of Spatial-Temporal Models of Remote-sensing Data," Doctoral dissertation, Department of Mechanical Engineering, University of Texas, Austin, Texas.

Sabins, F. F. (1987) *Remote Sensing: Principles and Interpretations,* Second Edition, Freeman.

Shaw, S-L (1993) "GIS for urban travel demand analysis: requirements and alternatives," *Computers, Environment and Urban Systems, Vol. 17,* pp. 15-29.

Star J.; Estes J. (1990) *Geographic Information Systems: An Introduction,* Prentice Hall.

U.S. Department of Transportation (1995) *News,* August 1, Office of the Assistant Secretary for Public Affairs, Washington, D.C.

Weslowsky, G. O. (1993) "The Weber Problem: History and Perspectives" *Location Science,* Vol. 1, No. 1, pp. 5-23.

Zimmerman, P. (1988) "Photos from space—why restrictions won't work" *Technology Review,* May-June, MIT, pp. 45-53.

## ENDNOTE

1 Submitted June 1995; In final form August, 1995

# Optimization Modeling for Airlift Mobility[1]

**David P. Morton**
*Graduate Program in Operations Research University of Texas*

**Richard E. Rosenthal**
*Operations Research Department Naval Postgraduate School*

**Captain Lim Teo Weng**
*Republic of Singapore Air Force*

## ABSTRACT

We describe a multi-period optimization model, implemented in GAMS, to help the U.S. Air Force improve logistical efficiency. It determines the maximum on-time throughput of cargo and passengers that can be transported with a given aircraft fleet over a given network, subject to appropriate physical and policy constraints. The model can be used to help answer questions about selecting airlift assets and about investing or divesting in airfield infrastructure.

## 1. INTRODUCTION

In an Operation Desert Storm type scenario, massive amounts of equipment and large numbers of personnel must be transported over long distances in a short time. The magnitude of such a deployment imposes great strains on air, land and sea mobility systems.

The U.S. military services are well aware of this problem and various optimization and simulation models have been developed to help improve the effectiveness of limited lift assets and alleviate the problem. Congress commissioned the Mobility Requirement Study (MRS) in 1991, when post-operation analysis of Desert Storm revealed a shortfall in lift capability.

Two linear programming (LP) optimization models that were developed as part of MRS and subsequent studies form the primary background of this research. They are: (1) the Mobility Optimization Model (MOM) developed for MRS by the Joint Staff's Force Structure Resource, and Assessment Directorate (J8) [Wing *et al.*, 1991] and (2) the THRUPUT Model developed by the USAF Studies and Analyses Agency (USAF/SAA) [Yost, 1994]. MOM considers both air and sea mobility, whereas THRUPUT and the model developed here cover only the air aspects of the problem. The model of this paper was first described in a Naval Postgraduate School master's thesis [Lim, 1994], which was sponsored by USAF/SAA.

In this research, the strategic airlift assets optimization problem is formulated as a multi-period, multi-commodity network-based linear programming model, with a large number of side constraints. It is implemented in the General Algebraic Modelling System (GAMS) [Brooke et al., 1992], and its purpose is to minimize late deliveries subject to physical and policy constraints, such as aircraft utilization limits and airfield handling capacities. For a given fleet and a given network, the model can help provide insight for answering many mobility questions, such as: 1) Are the aircraft and airfield assets adequate for the deployment scenario? 2) What are the impacts of shortfalls in airlift capability? 3) Where are the system bottlenecks and when will they become noticeable? This type of analysis can be used to help answer questions about selecting airlift assets and about investing or divesting in airfield infrastructure.

## 2. OVERVIEW OF MODEL

The analyses described above are accomplished through repeated runs of the model. Each run assumes a particular scenario as defined by a given set of time-phased movement requirements and a given set of available aircraft and airfield assets. It is then solved for the optimal number of missions flown and the optimal amounts of cargo and passengers carried, for each unit, by each aircraft type, via each route, in each time period.

### 2.1 Model Features

The model has been designed to handle many of the airlift system's particular features and modes of operation. For example, the payload an aircraft can carry depends on range (the shorter the range, the heavier the load), and aircraft with heavy loads may be required to make one or more enroute stops. Also, there is a need to ensure cargo-to-carrier compatibility since some military hardware is too bulky to fit into certain aircraft. These features have been incorporated in the model to make it as realistic as possible. Others, such as the use of tanker aircraft for aerial refueling of airlift aircraft are recommended as follow-on work. (See CONOP by RAND [Killingsworth and Melody, 1994] for extensive treatment of aerial refueling in another GAMS-based optimization model.) The major features of the airlift system currently captured by the model include:

- Multiple origins and destinations: In contrast to MOM, the current model allows the airlift to use multiple origin, enroute and destination airfields.
- Flexible routing structure: The air route structure supported by the model includes delivery and recovery routes with a variable number of enroute stops (usually between zero and three). This provision allows for a mixture of short-range and long-range aircraft. The model can thus analyze trade-offs between higher-payload, shorter-range flights and lower-payload, longer-range flights. For further routing flexibility, the model also allows the same aircraft to fly different delivery and recovery routes on opposite ends of the same mission.
- Aircraft-to-route restrictions: The user may impose aircraft-to-route restrictions; e.g., only military aircraft may use military airfields for enroute stops. This particular provision arises because the USAF Air Mobility Command (AMC) may call upon civilian commercial airliners to augment USAF aircraft in a deployment, under the Civil Reserve Airfleet (CRAF) program. The model distinguishes between USAF and CRAF aircraft.
- Aircraft assets can be added over time. This adds realism to the model, because CRAF and other aircraft may take time to mobilize and are typically unavailable at the start of a deployment.
- Delivery time windows: In a deployment, a unit is ready to move on its *available-to-load date (ALD)* and has to arrive at the theater by its *required-delivery-date (RDD)*. This aspect of the problem has been incorporated in the model through user-specified time windows for each unit. The model treats the time windows as "elastic" in that cargo may be delivered late, subject to a penalty.

## 2.2 Conceptual Model Formulation

This section gives a verbal description of the key components of the airlift optimization model. The mathematical formulation is covered in detail in Section 3.

The primary decision variables are the number of missions flown, and the amounts of cargo and passengers carried, for each unit, by each aircraft type, via each available route, in each time period. Additional variables are defined for the recovery flights, for aircraft inventoried at airfields, and for the possibility (at high penalty cost) of not delivering required cargos or passengers.

## 2.2.1 Objective Function

The purpose of the optimization model is to maximize the effectiveness of the given airlift assets, subject to appropriate physical and policy constraints. The measure of effectiveness is the minimization of total weighted penalties incurred for late deliveries and non-deliveries. The penalties are weighted according to two factors: the priority of the unit whose movement requirement is not delivered on time, and the degree of lateness. The penalty increases with the amount of time late, and non-delivery has the most austere penalty.

The anticipated use of the model is for situations when the given airlift resources are insufficient for making all the required deliveries on time. On the other hand, if there are enough resources for complete on-time delivery, then the model's secondary objective function is to choose a feasible solution that maximizes unused aircraft. The motivation of the secondary objective is that if the available aircraft are used as frugally as possible, while still meeting the known demands and observing the known constraints, then the mobility system will be as well prepared as it can be for unplanned breakdowns and unforeseen requirements, such as an additional nearly simultaneous regional contingency.

### 2.2.2 Constraints

The model's constraints can be grouped into the five categories: demand satisfaction, aircraft balance, aircraft capacity, aircraft utilization, and airfield handling capacity.

- Demand Satisfaction Constraints: The cargo demand constraints attempt to ensure for each unit that the correct amounts of cargo move to the required destination within the specified time window. The passenger demand constraints do the same for each unit's personnel. The demand constraints have elastic variables for late delivery and non-delivery. The optimization will seek to avoid lateness and non-deliveries if it is possible with the available assets, or to minimize them if not.
- Aircraft Balance Constraints: These constraints keep physical count of aircraft by type (e.g., C17, C5, C141, etc.) in each time period. They ensure that the aircraft assets are used only when they are available.
- Aircraft Capacity Constraints: There are three different kinds of constraints on the physical limitations of aircraft—troop carriage capacity, maximum payload, and cabin floor space—which must be observed at all times.
- Aircraft Utilization Constraints: These constraints ensure that the average flying hours consumed per aircraft per day are within AMC's established utilization rates for each aircraft type.
- Aircraft Handling Capacity at Airfields: These constraints ensure that the number of aircraft routed through each airfield each day is within the airfield's handling capacity.

## 2.3 Assumptions

Some major assumptions of the model are listed below. These are known to be sacrifices of realism, but such assumptions are needed in modeling most real-world problems due to the limitations of data availability or the need to avoid computational intractability.

- Airfield capacity is represented by Air Force planners by a measure called *Maximum-on-Ground (MOG).* The literal translation of MOG as the maximum number of planes that can be simultaneously on the ground at an airfield is somewhat misleading, because the term MOG means more than just the number of parking spaces at an airfield. In actuality, airfield capacity depends on many dimensions in addition to parking, including material handling equipment, ground services capacity and fuel availability. Some Air Force planners use the terms *parking MOG* and *working MOG* to distinguish between parking space limits and servicing capability. Working MOG is always smaller than parking MOG, and is the only MOG for which we have data. Working MOG is an approximate measure because it attempts to aggregate the capacities of several kinds of services into a single, unidimensional figure. Disaggregation of airfield capacity into separate capacities for parking spaces and for each of the specific services avail-

able would yield a more accurate model. Unfortunately, data are not currently available to support this modeling enhancement.
- Inventoried aircraft at origin and destination airfields are considered not to affect the aircraft handling capacity of the airfield. This assumption is not strictly valid since an inventoried aircraft takes up parking space, but, as noted, working MOG dominates parking MOG.
- Deterministic ground time: Aircraft turnaround times for onloading and offloading cargo and enroute refueling are assumed to be known constants, although they are naturally stochastic. This ignores the fact that deviations from the given service time can cause congestion on the ground. To offset the optimism of this assumption, an efficiency factor is used in the formulation of aircraft handling capacity constraints to cushion the impact of randomness. Better handling of stochastic ground times is a subject of ongoing research.

Other approximations of reality employed in the model for computational tractability are aggregation of airfields, discretiztion of time, and continuous decision variables. A limitation on the scope of the model is that it considers only *inter*-theater, not *intra*-theater deliveries.

## 3. OPTIMIZATION MODEL

This section gives a mathematical formulation of the conceptual optimization model discussed previously in Section 2.2.

The airlift optimization problem is formulated as a multi-period, multi-commodity network-based linear program with a large number of side constraints. Two key concepts are employed in the model. The first is the use of a time index to track the locations of aircraft for each time period. The modeling advantages of knowing when an aircraft will arrive at a particular airfield are that it enables us to model aircraft handling capacity at airfields and to determine unit closures (i.e., the time when all of a unit's deliveries are completed). This approach is in contrast to the THRUPUT model of Yost [1994], which takes a static-equilibrium or steady-state approach.

The second key concept is model reduction through data aggregation and the removal of unnecessary decision variables and constraints prior to optimization. This is necessary as the airlift problem is potentially very large. Without this model reduction step, the number of decision variables would run into the millions even for a nominal deployment. The unnecessary decision variables and constraints are removed by extensive checking of logical conditions, performed by GAMS during model generation. (See Lim [1994] for details.)

### 3.1. Indices

| | |
|---|---|
| u | indexes units, e.g., 82nd Airborne |
| a | indexes aircraft types, e.g., C5, C141 |
| t,t′ | index time periods |
| b | indexes all airfields (origins, enroutes and destinations) |
| i | indexes origin airfields |
| k | indexes destination airfields |
| r | indexes routes |

## 3.2 Index Sets

*Airfield Index Sets*

| | |
|---|---|
| $B$ | set of available airfields |
| $I \subseteq B$ | origin airfields |
| $K \subseteq B$ | destination airfields |

*Aircraft Index Sets*

| | |
|---|---|
| $A$ | set of available aircraft types |
| $A_{bulk} \subseteq A$ | aircraft capable of hauling bulk-sized cargo |
| $A_{over} \subseteq A_{bulk}$ | aircraft capable of hauling over-sized cargo |
| $A_{out} \subseteq A_{over}$ | aircraft capable of hauling out-sized cargo |

Bulk cargo is palletized on 88 x 108 inch platforms and can fit on any milatary aircraft (as well as the cargo-configured 747). Over-sized cargo is non-palletized rolling stock: it is larger than bulk cargo and can fit on a C141, C5 or C17. Out-sized cargo is very large non-palletized cargo that can fit into a C5 or C17 but not a C141.

*Route Index Sets*

| | |
|---|---|
| $R$ | set of available routes |
| $R_a \subseteq R$ | permissible routes for aircraft type a |
| $R_{ab} \subseteq R_a$ | permissible routes for aircraft type a that use airfield b |
| $R_{aik} \subseteq R_a$ | permissible routes for aircraft type a that have origin i and destination k |
| $DR_i \subseteq R$ | delivery routes that originate from origin i |
| $RR_k \subseteq R$ | recovery routes that originate from destination k |

A delivery route is a route flown from a specific unit's origin to its destination for the purpose of delivering cargo and/or passengers. A recovery route is a route flown from a unit's destination to that unit's or some other unit's origin, for the purpose of making another delivery. Since recovery flights carry much less weight than deliveries, the recovery routes from k to i may have fewer enroute stops than the delivery routes from i to k.

*Time Index Sets*

| | |
|---|---|
| $T$ | set of time periods |
| $T_{uar} \subseteq T$ | possible launch times of missions for unit u using aircraft type a and route r |

The set $T_{uar}$ covers the allowed time window for unit u, which starts on the unit's available-to-load date and ends on the unit's required delivery date, plus some extra time up to the maximum allowed lateness for the unit.

## 3.3 Given Data

*Movement Requirements Data*

| | |
|---|---|
| $MovePAX_{uik}$ | Troop movement requirement for unit u from origin i to destination k |
| $MoveUE_{uik}$ | Equipment movement requirement in short tons (stons) for unit u from origin i to destination k |
| $ProBulk_u$ | Proportion of unit u cargo that is bulk-sized |
| $ProOver_u$ | Proportion of unit u cargo that is over-sized |
| $ProOut_u$ | Proportion of unit u cargo that is out-sized |

*Penalty Data*

| | |
|---|---|
| $LatePenUE_u$ | Lateness penalty (per ston per day) for unit u equipment |
| $LatePenPAX_u$ | Lateness penalty (per soldier per day) for unit u troops |
| $NoGoPenUE_u$ | Non-delivery penalty (per ston) for unit u equipment |
| $NoGoPenPAX_u$ | Non-delivery penalty (per soldier) for unit u troops |
| $MaxLate_u$ | Maximum allowed lateness (in days) for delivery |
| $Preserve_{at}$ | Penalty (small artificial cost) for keeping aircraft type a in mobility system at time t |

*Cargo Data*

| | |
|---|---|
| $UESqFt_u$ | Average cargo floor space (in sq. ft.) per ston of unit u equipment |
| $PAXWt_u$ | Average weight of a unit u soldier inclusive of personal equipment |

*Aircraft Data*

| | |
|---|---|
| $Supply_{at}$ | Number of aircraft of type a that become available at time t |
| $MaxPAX_a$ | Maximum troop carriage capacity of aircraft type a |
| $PAXSqFt_{ua}$ | Average cargo space (in sq. ft.) consumed by a unit u soldier for aircraft type a |
| $ACSqFt_a$ | Cargo floor space (in sq. ft.) of aircraft type a |
| $LoadEff_a$ | Cargo space loading efficiency (<1) for aircraft type a. This accounts for the fact that it is not possible in practice to fully utilize the cargo space. |
| $URate_a$ | Established utilization rate (flying hours per aircraft per day) for aircraft type a |

*Airfield Data*

| | |
|---|---|
| $MOGCapb_t$ | Aircraft capacity (in narrow-body equivalents) at airfield b in time t |
| $MOGReq_{ab}$ | Conversion factor to narrow-body equivalents for one aircraft of type a µat airfield b |
| $MOGEff_{bt}$ | MOG efficiency factor (<1), to account for the fact that it is impossible to fully utilize available MOG capacity due to randomness of ground times |

*Aircraft Route Performance Data*

| | |
|---|---|
| $MaxLoad_{ar}$ | Maximum payload (in stons) for aircraft type a flying route r. |
| $GTime_{abr}$ | Aircraft ground time (due to onload or offload of cargo, refueling, maintenance, etc.) needed for aircraft type a at airfield b on route r |
| $DTime_{abr}$ | Cumulative time (flight time plus ground time) taken by aircraft type a to reach airfield b along route r |
| $FltTime_{ar}$ | Total flying hours consumed by aircraft type a on route r |
| $CTime_{ar}$ | Cumulative time (flight time plus ground time) taken by aircraft type a on route r |
| $DaysLateu_{art}$ | Number of days late unit u's requirement would be if delivered by aircraft type a via route r with mission start time t |

## 3.4 Decision Variables

*Mission Variables*

| | |
|---|---|
| $X^u{}_{art}$ | Number of aircraft of type a that airlift unit u via route r with mission start time during period t |
| $Y_{art}$ | Number of aircraft of type a that recover from a destination airfield via route r with start time during period t |

*Aircraft Allocation and De-allocation Variables*

| | |
|---|---|
| $Allot_{ait}$ | Number of aircraft of type a that are allocated to origin i at time t |
| $Release_{ait}$ | Number of aircraft of type a that were allocated to origin i prior to time t but are not scheduled for any missions from time t on |

*Aircraft Inventory Variables*

| | |
|---|---|
| $H_{ait}$ | Number of aircraft of type a inventoried at origin i at time t |
| $HP_{akt}$ | Number of aircraft of type a inventoried at destination k at time t |
| $NPlanes_{at}$ | Number of aircraft of type *a* in the air mobility system at time *t* |

*Airlift Quantity Variables*

| | |
|---|---|
| $TonsUE_{uart}$ | Total stons of unit *u* equipment airlifted by aircraft of type *a* via route *r* with mission start time during period *t* |
| $TPAX_{uart}$ | Total number of unit *u* troops airlifted by aircraft of type *a* via route *r* with mission start time during period *t* |

*Elastic (Nondelivery) Variables*

| | |
|---|---|
| $UENoGo_{uik}$ | Total stons of unit *u* equipment with origin *i* and destination *k* that is not delivered in the prescribed time frame |
| $PAXNoGo_{uik}$ | Number of unit *u* troops with origin *i* and destination *k* who are not delivered in the prescribed time frame |

## 3.5 Formulation of the Objective Function

*Minimize*

$$\sum_{u} \sum_{a} \sum_{r \in R_a} \sum_{t \in T_{uar}} LatePenUE_u * DaysLate_{uart} * TonsUE_{uart}$$

$$+ \sum_{u} \sum_{a} \sum_{r \in R_a} \sum_{t \in T_{uar}} LatePenPAX_u * DaysLate_{uart} * TPAX_{uart}$$

$$+ \sum_{u} \sum_{i} \sum_{k} (NoGoPenUE_u * UENoGo_{uik} + NoGoPenPAX_u * PAXNoGo_{uik})$$

$$+ \sum_{a} \sum_{t} Preserve_{at} * NPlanes_{at}$$

The $DaysLate_{uart}$ penalty parameter has value zero if $t+CTime_{ar}$ is within the prescribed time window for unit $u$. Thus, the first two terms of the objective function take effect only when a delivery is late. The third term in the objective function corresponds to cargo and passengers that cannot be delivered even within the permitted lateness. Late delivery and non-delivery occur only when airlift assets are insufficient for on-time delivery.

The reason for including elastic variables that allow late delivery and non-delivery is to ensure that the model produces useful information even when the given assets are inadequate for the given movement requirements. The alternative of using an inelastic model (i.e., a model with hard constraints that insist upon complete on-time delivery) is inferior because it would report infeasibility without giving any insight about what *can* be done with the assets available.

A useful modeling excursion that is made possible by the elastic variables is to vary the number of time periods. As the horizon is shortened, it is interesting to observe the increase in lateness and non-delivery.

As noted, the model's anticipated use is in cases when the airlift assets are insufficient for full on-time delivery. In the opposite case, the model will be governed by the fourth term of the objective function, which rewards asset preservation for the reasons given in Section 2.2.1.

Some care must be taken in selecting the lateness and non-delivery penalties and the aircraft preservation rewards to ensure consistency. Late delivery should be preferred to non-delivery. The weights will be consistent with this preference provided the late penalty (per ston per day) is less than the corresponding non-delivery penalty (per ston) divided by the maximum allowed lateness (in days).

## 3.6 Formulation of the Constraints

As noted in the conceptual model, there are five categories of constraints. Their mathematical formulations are as follows.

### 3.6.1 Demand Satisfaction Constraints

There are four different kinds of demand constraints, corresponding to troops and the three classes of cargo (bulk, over-sized and out-sized). Separate constraints are required for the different cargo types to ensure cargo-carrier compatibility. For example, a carrier of over-sized cargo cannot be used to carry the larger out-sized cargo. On the other hand, it is possible to use a carrier of out-sized cargo to carry over-sized cargo. The model accounts for this asymmetry.

The demand constraints also account for the desired delivery time-windows by use of the index sets $T_{uar}$ and the lateness parameters $DaysLate_{uart}$.

*Demand Satisfaction Constraints for All Classes of Cargo:*

$$\sum_{a \in A_{bulk}} \sum_{r \in R_{aik}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_{uik} = MoveUE_{uik} \quad \forall\ u,i,k\text{: } MoveUE_{uik} > 0$$

*Demand Satisfaction Constraints for Out-Sized Cargo:*

$$\sum_{a \in A_{out}} \sum_{r \in R_{aik}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_{uik} \geq ProOut_u * MoveUE_{uik}$$

$$\forall\ u,i,k\text{: } MoveUE_{uik} > 0$$

*Demand Satisfaction Constraints for Over-Sized Cargo:*

$$\sum_{a \in A_{ovr}} \sum_{r \in R_{aik}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_{uik} \geq (ProOver_u + ProOut_u) * MoveUE_{uik}$$

$$\forall\ u,i,k\text{: } MoveUE_{uik} > 0$$

*Demand Satisfaction Constraints for Troops:*

$$\sum_{a} \sum_{r \in R_{aik}} \sum_{t \in T_{uar}} TPAX_{uart} + PAXNoGo_{uik} = MovePAX_{uik} \quad \forall\ u,i,k\text{: } MovePAX_{uik} > 0$$

### 3.6.2 Aircraft Balance Constraints

There are five kinds of aircraft balance constraints enforced for each aircraft type in each time period. At origin airfields, they ensure that the number of aircraft assigned for delivery missions plus those inventoried for later use plus those put in the released status equal the number inventoried from the previous period plus recoveries from earlier missions and the new supply of aircraft that is allocated to the origin.

The meaning of *releasing*, or *de-allocating*, an airplane in period $t$ is that it is not flown on any missions from period $t$ through the end of the horizon. In practice, the analyst can interpret a release in the model's solution in a variety of ways. It can mean, as in the case of the civilian CRAF aircraft, that the plane is literally sent back to its owner, but not necessarily. The aircraft can also be kept in the mobility system, available as a replacement in case of breakdowns or for unforeseen demands.

The second kind of aircraft balance constraints concerns destinations. They are similar to the first kind except releases are not allowed and the roles of delivery and recovery missions are reversed. The third kind of aircraft balance constraint ensures that if any new planes become available in period $t$, they are allotted appropriately among the origins. There is a potential gain in efficiency to allow the optimizer to make these allocation decisions, rather than relying on the user to preassign them to origin airfields. The fourth type of aircraft balance constraints is a set of accounting equations for defining the $NPlanes_{at}$ variables based on cumulative allocations and releases.

*Aircraft Balance Constraints at Origin Airfields:*

$$\sum_{u} \sum_{r \in DR_i} X_{uart} + H_{ait} + Release_{ait} = H_{ai,t-1} + Allot_{ait} + \sum_{r \in R_{ai}} \sum_{t' + [CTime_{ar}] = t} Y_{art'} \quad \forall\ a,i,t$$

where $[CTime_{ar}]$ is $CTime_{ar}$ rounded to the nearest integer.

*Aircraft Balance Constraints at Destination Airfields:*

$$\sum_{r \in RR_k} Y_{art} + HP_{akt} = HP_{ak,t-1} + \sum_{u} \sum_{r \in R_{ak}} \sum_{\substack{t' \in T_{uar} \\ t' + [CTime_{ar}] = t}} X_{uart'} \quad \forall\ a,k,t$$

*Aircraft Balance Constraints for Allocations to Origins:*

$$\sum_{t'=1}^{t} \sum_{i} Allot_{ait} \le \sum_{t'=1}^{t} Supply_{at} \quad \forall\ a,t$$

This constraint is in the cumulative form, rather than in the simpler form $\sum_i Allot_{ait} \le Supply_{at}$, to allow aircraft that become available in period $t$ to be put into service at a later period.

*Aircraft Balance Constraints Accounting for Allocations and Releases:*

$$NPlanes_{at} = \sum_{t'=1}^{t} \sum_{i} Allot_{ait'} - \sum_{t'=1}^{t} \sum_{i} Release_{ait'} \quad \forall\ a,t$$

The fifth and final set of aircraft balance constraints helps to correct the discretization error that can result from rounding $CTime_{ar}$ to $[CTime_{ar}]$, the nearest integer, in the other balance constraints. For example, suppose $CTime_{ar}$ is less than half a day for some aircraft $a$ and route $r$. When this time is rounded to zero in the balance constraints of the route's origin and destination, these constraints unrealistically permit an unlimited number of missions per day on that route. Solving the model with this deficiency would yield overly optimistic results.

One way to fix this problem would be to insist that $CTime_{ar}$ be rounded up to a higher integer. Then the model would be overly pessimistic, because it would rule out the possibility of an aircraft flying two or more missions in a day even when this is possible. This sort of problem is common in mathematical modeling whenever time is discretized. The approach taken here is to enforce the following additional constraints, based on the cumulative plane-days available.

*Cumulative Aircraft Balance Constraints:*

$$\sum_{r\in R_a}\sum_{t'=1}^{t}\sum_{u} K_{artt'}\, X_{uart'} + \sum_{r\in R_a}\sum_{t'=1}^{t} K_{artt'}\, Y_{art'} + \sum_{i}\sum_{t'=1}^{t} H_{ait'}$$

$$+ \sum_{k}\sum_{t'=1}^{t} HP_{akt'} \le \sum_{t'=1}^{t} NPlanes_{at'} \quad \forall\ a,t$$

$$K_{artt'} = \begin{cases} t-t'+1 & if\ \ t' \le t < t' + CTime_{ar} - 1 \\ CTime_{ar} & if\ \ t \ge t' + CTime_{ar} - 1 \end{cases}$$

The right-hand-side indicates the cumulative number of plane-days available for type $a$ aircraft up to day $t$. The left-hand-side accounts for all possible plane activities up to day $t$, whether flying or inventoried. The inventory terms are straightforward. The delivery and recovery terms work as follows: if a delivery initiated on day $t'$ is completed by the end of day $t$, then the entire time $CTime_{ar}$ (which may be integer or fractional) is included in the left-hand-side of the cumulative balance constraint for day $t$. On the other hand, if a delivery initiated on day $t'$ is not completed by the end of day $t$, then only the time expended so far, $t-t'+1$, is counted in the day $t$ constraint.

An experiment attesting to the value of the cumulative aircraft balance constraints is described in Section 5.4. If the $CTime_{ar}$'s were all integer, these constraints would be redundant and could be omitted.

### 3.6.3 Aircraft Capacity Constraints

*Troop Carriage Capacity Constraints:*

$$TPAX_{uart} \le MaxPAX_a * X_{uart} \quad \forall\ u,a,r,t\!: t \in T_{uar}$$

*Maximum Payload Constraints:*

$$TonsUE_{uart} + PAXWt * TPAX_{uart} \le MaxLoad_{ar} * X_{uart} \quad \forall\ u,a,r,t\!: t \in T_{uar}$$

*Cargo Floor Space Constraints:*

$$PAXSqFt_a * TPAX_{uart} + UESqFt_u * TonsUE_{uart} \le ACSqFt_a * LoadEff_a * X_{uart}$$

$$\forall\ u,a,r,t\!: t \in T_{uar}$$

### 3.6.4 Aircraft Utilization Constraints

The aircraft utilization constraints ensure that the total flying hours consumed by the fleets of each aircraft type over the planning horizon are within AMC's established utilization rates [Wilson, 1985; Gearing *et al.*, 1988]. These rates are meant to capture spares availability, aircraft reliability, crew availability, and other factors. The utilization constraints are formulated by comparing the flying hours consumed by an aircraft fleet in delivery and recovery flights to the maximum achievable flying hours for the fleet according to the utilization rate.

$$\sum_{u} \sum_{r \in R_a} \sum_{t \in T_{uar}} FltTime_{ar} * X_{uart} + \sum_{r \in R_a} \sum_{t} FltTime_{ar} * Y_{art}$$

$$\leq \sum_{t} URate_a * NPlanes_{at} \quad \forall a$$

As an illustration of the above equation, consider a fleet of 5 aircraft of the same type made available from day 11. If the utilization rate for this aircraft type is 10 flying hours per aircraft per day and the horizon is 30 days, then the maximum achievable flying is 1000 hours *(10 hours/plane-day x 20 days x 5 planes)*. This total may not be exceeded for the whole fleet over the entire planning horizon, however, it is not unusual for a subset of aircraft to exceed utilization rates over a subset of the horizon, particularly during the early (surge) stage of a deployment.

### 3.6.5 Aircraft Handling Capacity of Airfields (MOG Constraint)

The aircraft handling constraints at airfields, commonly called MOG constraints, are perhaps the most difficult to model. This is because of two complicating factors that necessitate approximations. First, there is no airfield capacity data available that provides separate accounting of parking spaces and all the various services (refueling, maintenance, etc.). The MOG data provided by the Air Force is an approximation, attempting to aggregate all these services. Thus, the units of $MOGCap_{bt}$ are an idealized notion of airfield parking spaces (normalized to narrow-body sized aircraft), not a precisely defined physical quantity.

The second complicating factor in modeling airfield capacity is the congestion caused by the uncertainty of arrival times and ground times. A deterministic, time-discretized optimization model cannot accurately treat events occurring within a time period. For example, suppose the time period of the model is one day and an airfield has 20 landings per day. How much congestion occurs depends on when the landings occur during the day, a phenomenon not captured in the daily model. It is possible to attack these concerns with stochastic modeling techniques, however, the existing simulation and optimization models for air mobility have made very limited progress to date in this area [Morton and Rosenthal, 1994]. The MOG efficiency factor *MOGEff* is introduced to cushion the effect of not explicitly modeling uncertainty. The MOG constraints are formulated for each airfield and time period as follows:

$$\sum_{u} \sum_{a} \sum_{r \in R_a} \sum_{\substack{t' \in T_{uar} \\ t' + [DTime_{abr}] = t}} (MOGReq_{ab} * GTime_{abr} / 24) * X_{uart'}$$

$$+ \sum_{a} \sum_{r \in R_a} \sum_{t' + [DTime_{abr}] = t} (MOGReq_{ab} * GTime_{abr} / 24) * Y_{art'}$$

$$\leq MOGEff_{bt} * MOGCap_{bt} \quad \forall\ b,t$$

Dimensional analysis is useful for understanding these constraints. The right-hand-side is in the units of narrow-body parking spaces, because $MOGCap_{bt}$ is in those units and $MOGEff_{bt}$ is dimensionless. The first term on the left-hand-side accounts for airfield capacity consumed by all delivery missions that pass through airfield $b$ during period $t$. The second term on the left does the same thing for recovery missions. The dimension of $MOGReq_{ab}$ is narrow-body parking spaces per plane, the dimension of $GTIME_{abr}/24$ is days, and the dimensions of $X_{uart'}$ and $Y_{uart'}$ are planes per day; thus, the MOG constraints are dimensionally balanced.

Aircraft inventoried at origin or destination airfields do not consume any MOG capacity in the above formulation. This is not a mathematical limitation, but rather a modeling choice taken because inventoried planes do not consume ground services. It can be easily modified if data for "parking space MOG" and various "ground service MOG's" become available.

## 4. PERFORMANCE

The performance of the optimization model is relatively fast. On an IBM RS6000 model 590 workstation with GAMS/OSL, it takes about 100 seconds to generate and an additional 100 seconds to solve a sample problem with 20 units, 7 aircraft types, 17 airfields and 30 time periods. A 486/66 laptop computer running the same software on the same problem takes about 28 minutes. After extensive variable and constraint reduction, the sample problem has 11,516 decision variables, 6,970 constraints and 189,351 nonzero coefficients. The data entry time for the sample problem is about one and a half hours. Excursions from a base model run take considerably less time to prepare. In short, turnaround time for the optimization model is significantly faster than simulation models commonly used in the Air Force [Morton and Rosenthal, 1994].[2]

## 5. ANALYTIC INSIGHTS

We now describe some examples of modeling excursions and the resulting analytic insights. The base case scenario, developed by the U.S. Air Force Studies and Analyses Agency, notionally resembles a Desert Storm scenario. This is the same problem instance whose dimensions (after model reductions) are given in the Performance section.

## 5.1 Diversion of Ramstein-Riyadh Demand to Dhahran

In the base case scenario, there are twenty origin-to-destination demand pairs, but they are dominated by the demand for airlifting two Army mechanized units from Ramstein, Germany to Riyadh, Saudi Arabia. These two units combined account for 66,400 short tons (stons), or 48%, of all unit equipment to be moved. When the base case is optimized, the given fleet delivers only 67% of the total unit equipment. The shortfall is due entirely to 45,000 undelivered stons of Ramstein-Riyadh demand, and a critical constraint appears to be MOG limitations at Riyadh's airfield.

In one modeling excursion, we examine the effects on the airlift system of changing the destination for one of the Ramstein-based mechanized units to Dhahran, Saudi Arabia, which is 250 miles northeast of Riyadh and closer to Kuwait and Iraq. Re-optimizing with this one change, the same fleet delivers 85% of all unit equipment, a dramatic improvement from 67%. However, the shortfall of 20,000 stons of unit equipment from Ramstein may still be a serious impediment to the Army's effectiveness, necessitating a re-evaluation of the scenario's war plans or augmentation of the mobility system.

The graphs in Figure 1 show a summary of this modeling excursion over time. The unit equipment demand profile has jumps at the required delivery dates (RDD's). Cumulative delivery profiles are shown for the base case and the excursion. When the demand curve is higher than the delivery profile, shortfalls occur. All passenger demands, though not shown in the figure, are delivered on time in both cases.

## 5.2 Required-Delivery-Date Sensitivity

As a second excursion, after shifting some of the Ramstein demand to Dhahran, we investigated the effect of changes in the required delivery date for the unit whose equipment could not be delivered. With the given RDD, the total unit equipment delivered is 85%, as noted. If extra days are allowed, delivery increases as follows:

| *Extra Days Allowed* | *Percent Unit Equipment Delivered* | *Objective Function Value* |
|---|---|---|
| 0 | 85% | 12.45 |
| 2 | 88% | 11.35 |
| 4 | 93% | 10.13 |
| 6 | 99% | 8.56 |

The maximum allowed lateness is four days in all these runs. However, around 99% of all the deliveries made are on time.

## 5.3 Identifying Critical Resources

The overall performance of the air mobility system in our optimization runs can be characterized as having three phases. During the first third of the thirty days modeled, the system is airframe constrained. During the middle third (plus or minus a few days depending on location), the system is airfield-capacity constrained. During the final third, the system is in a sustainment phase with diminished demands. Neither airframes nor airfield capacities are critical resources, and it is too late to deliver cargos that were undelivered earlier.

After looking at Figure 1, one might disagree with the assertion that the mobility system is airframe-bound in the first phase, because there are no significant shortfalls until Day 16. This would be a mistake, however. In fact, all available aircraft are used to the maximum from the earliest available-to-load date (Day 1) through Day 11 (when a large portion of the military aircraft first become available), and the dual multipliers indicate that additional airframe assets in the first phase would have high marginal value. This is because if more aircraft were available earlier, then the optimization model would have made more early deliveries to prevent the shortfalls that it foresees but cannot avoid later in the middle phase.

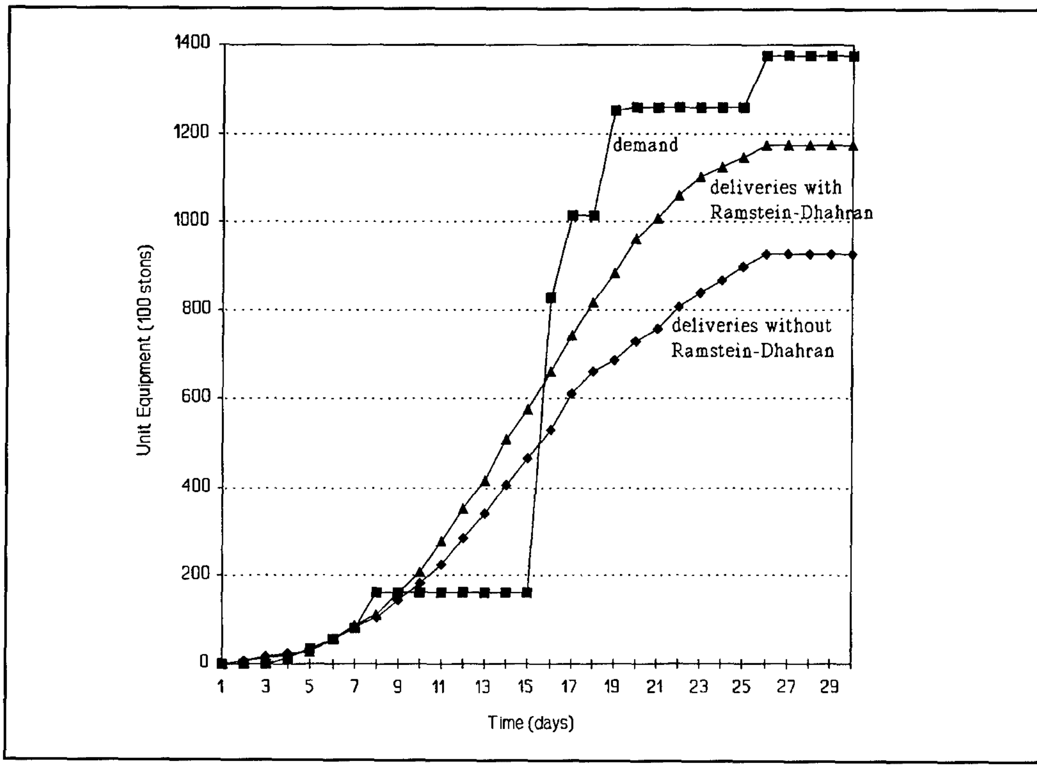



**Figure 1. A modeling excursion: after changing a Ramstein-based unit's destination from Riyadh to Dhahran, the amount of undelivered cargo decreases from 45,000 stons to 20,000 stons.**

The middle phase of the airlift has more overall flights than the first phase, because there are more aircraft in the system and demand is sufficient to keep them flying. The middle phase also has a higher percentage of the shorter Germany-to-Saudi flights, as compared to the longer CONUS-to-Saudi flights which predominate in the first phase. With more flights and with shorter flights (which consume MOG at a faster rate per plane), the mobility system becomes airfield-capacity constrained.

One might be tempted to conclude that adding more planes to the system during the middle phase would be unproductive. This would also be a mistake: the dual multipliers on aircraft consumption indicate that additional C17's and C5's would have high marginal value in the middle phase. Why does the optimization say that adding more planes would help the mobility system when airfield capacities are already hitting their limits?

The answer is that the optimization advocates adding more *efficient* and *versatile* planes. The meaning of *efficiency* for planes in a MOG-constrained environment is a high ratio of cargo-delivered-per-plane to MOG-hours-consumed-per-plane. The more efficient a plane is in this sense, the more cargo it can deliver per day to a MOG-limited destination. According to the data furnished by USAF/SAA and the evaluation of MOG-hours consumed per plane at the most congested airfields in the model, the C17 is the most efficient airframe for a MOG-constrained environment. The meaning of *versatility* in the present context is having the ability to carry all three types of cargo (bulk, over-size and out-size), as only the C17 and C5 can. The optimization determines that the mobility system would perform better on the entire airlift if some more efficient and versatile air-frames were made available during the middle phase.

## 5.4 Sensitivity to Time Discretization

The cumulative aircraft balance constraints were added to lessen the effects of time discretization, as discussed in Section 3.6.2. The kinds of problems they are intended to remedy arise, for example, if the cycle time of a route is less than half the length of a time period. Without these constraints, such a cycle time would be rounded to zero and cause unrealistic results.

To test the effectiveness of the cumulative aircraft balance constraints, the model was run with time period lengths of 12, 24 and 48 hours. The resulting delivery profiles are displayed in Figure 2. The idea of the test is that in the absence of discretization error abatement measures, the error would increase as the time-step of the model gets larger. Figure 2, however, shows close agreement among the delivery profiles, regardless of time period length.

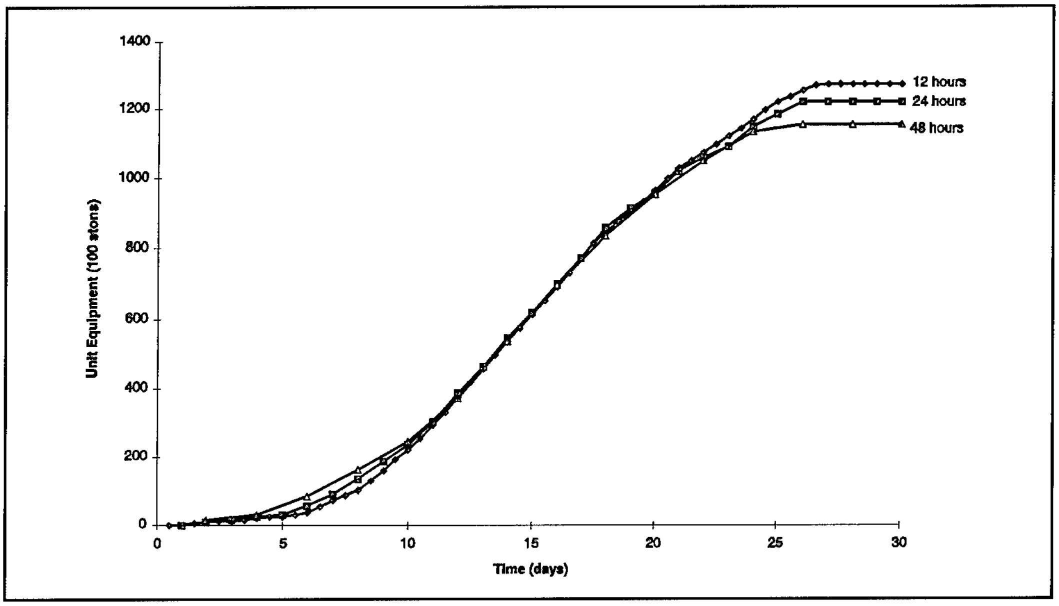



**Figure 2. Agreement among delivery profiles when time periods have length 12, 24 or 48 hours. Larger time-steps in linear programming yield smaller, easier-to-solve models, but usually cause greater discretization errors. In this model, however, the cumulative aircraft balance constraints effectively reduce discretization error.**

## 6. CONCLUSIONS AND FUTURE EXTENSIONS

The preceding analytic insights are typical of what that can be obtained through optimization, but not from simulation. They represent but a small sample of the kinds of questions that can be addressed with the optimization model. The model can give relatively rapid response to questions relating to major mobility issues such as: 1) Are the given aircraft and airfield assets adequate for the deployment scenario? 2) What are the impacts of shortfalls in airlift capability? 3) Where are the system bottlenecks and when will they become noticeable? This type of analysis can be used to help answer questions about selecting airlift assets and about investing or divesting in airfield infrastructure.

The optimization model has some limiting assumptions which must be taken into account when evaluating its results. As noted, they are: the approximation of airfield capacity by a uni-dimensional MOG factor, deterministic ground times, the absence of aerial refueling, and the rounding problems that are inevitably caused by the discretization of time. The cumulative aircraft balance constraints help address the last difficulty, by preventing overly optimistic or pessimistic results. Nevertheless, the one-day time scale of the model that typically has been used to date cannot accurately represent what happens at airfields during smaller time intervals.

In the Air Force analysis community, simulation has more acceptance than optimization. The advantage of simulation over optimization is that it can more readily accommodate uncertainty and it can handle a higher level of detail, such as tracking individual airplanes by tail number. The disadvantage is that it can only answer *what-if* questions, not *what's-best* questions. Simulations also usually take longer to run. Air mobility simulations used by the Air Force have had such long run times that the stochastic elements are sometimes left out in order to make them run faster.

Ideally, optimization and simulation should be used in concert, with the optimization being used to suggest mobility system configurations and modes of operation that are then analyzed in detail by the simulation. Simulation runs, in turn, would suggest new scenarios to be investigated by the optimization.

The optimization model described here is capable of being used in concert with other Air Force planning models, or it can stand alone to provide rapid and realistic responses in emerging conflict situations. Ongoing research is attempting to enhance the model in the following ways:

- Currently the routes made available to the optimization model are entered manually, based on USAF/SAA analysts' judgement. An auxiliary model is under development for generating routes [Turker, 1995]. Turker's research is also addressing the issue of decreasing the effects of airfield aggregation (and associated unit aggregation).
- Stochastic programming methods are under investigation for incorporating random ground times [Goggins, 1995].
- The Air Force is currently studying the formation and transportation of *global reach laydown packages.* The idea is to bring these packages to remote airfields to quickly create or augment airfield capacity. A related optimization model is addressing the optimal deployment of these mobile assets [Chapates, 1995].

## ACKNOWLEDGEMENTS

It is a pleasure to acknowledge the assistance and support of several Air Force officers involved with this project: LtCol Daniel Briand, LtCol James Hill, LtCol Robert Taft, MAJ Kevin Smith, MAJ Kirk Yost, and Capt David Horton. We have also greatly benefited from the participation of the following Naval Postgraduate School students in our research project: MAJ Steven Baker (USAF), LT David Fuller (USN), LT David Goggins (USN) and LTJG Yasin Turker (Turkish Navy). David Morton was partially supported by the National Research Council of the National Academy of Sciences under a research associateship at the Naval Postgraduate School.

## REFERENCES

Brooke, A., Kendrick, D., and Meeraus, A., *GAMS A User's Guide*, The Scientific Press, 1992.

Chapates, D., *Optimal Deployment of Global Reach Laydown Packages*, Master's Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, September 1995.

Gearing, CPT Rich, MAJ Jim Hill and MAJ Dave Wilson, "UTE Rates Revisited," *Airlift*, Spring 1988.

Goggins, D., *Stochastic Modeling for Airlift Mobility*, Master's Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, September 1995.

Killingsworth, P. and L. Melody, "CONOP Air Mobility Optimization Model," RAND, Santa Monica, CA, July 1994.

Lim, T.W., *Strategic Airlift Assets Optimization Model*, Master's Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, September 1994.

Morton, D.P and Rosenthal, R.E., "Minutes of Air Mobility Mini-Conference," Operations Research Department, Naval Postgraduate School, Monterey, CA, July 1994.

Turker, Y., *Route and Column Generation Methods for Airlift Mobility Optimization*, Master's Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, December 1995.

Wilson, MAJ Dave, "UTE: Utilization Rate: What Is It, How Is It Derived and How Should It Be Used?", *Airlift*, Winter, 1985.

Wing, V.F., Rice, R.E., Sherwood, R.W. and Rosenthal, R.E., "Determining the Optimal Mobility Mix," Joint Staff(J8), Force Design Division, The Pentagon, Washington, DC, 1 October 1991.

Yost, K.A., "The THRUPUT Strategic Airlift Flow Optimization Model," Air Force Studies and Analyses Agency, 30 June 1994.

# ENDNOTES

[1] Submitted July 1995; In revised form December, 1995

[2] *Note added in final revision:* More recent runs of the model have been with a larger data set corresponding to a two-MRC scenario. This instance of the model contains 200 units, 7 aircraft types, 155 routes, and 47 time periods. The linear program has 161,000 constraints, 183,000 variables and 1.9 million nonzero coefficients. It took 30 minutes to generate and 3 hours to solve with GAMS/CPLEX on the RS6000/590. On the advice of CPLEX Optimization, Inc., the model was solved with the "barrier" and "nocrossover" options; their assistance is gratefully acknowledged.

# Tutorials at the 64th MORS Symposium 18-20 June 1996 Ft. Leavenworth, KS

Tutorials Coordinator Roy Rice, Teledyne Brown Engineering, says to be sure to take advantage of the lunchtime educational opportunities at the symposium by attending some of the following outstanding tutorials. Abstracts of these tutorials, along with other symposium information, are in the *Preliminary Program.* Call the MORS office for more information.

- **Using Values to Generate Alternatives**
  *Dr. Gregory S. Parnell, Virginia Commonwealth University, Department of Mathematical Sciences*

- **Value-Focused Thinking**
  *Lt Col Jack A. Jackson, Maj Lee Lehmkuhl, AFIT/ENS and Maj Jeffrey S. Stonebraker, HQ USAFA/DFMS*

- **Modeling for Campaign Analysis: Lessons for the Next Generation of Models**
  *Richard Hillestad, Louis Moore, Bart Bennett, RAND*

- **Using DTIC to Publish MORS Papers**
  *Frank Scott, Defense Technical Information Center (DTIC)*

- **MASTR (Modeling, Analysis, Simulation and Training), A New Look**
  *Steve Boyd, AFSAA/SAGD*

- **Lanchester on Lanchester Intelligence**
  *Michael W. Garrambone, Veda, Inc.*

- **Operational Effectiveness Analyses for Systems That Don't Shoot**
  *Charles R. Hall III, MITRE*

- **Modeling Joint Mobility Problems: A Tutorial**
  *Dr. Yupo Chan, AFIT/ENS*

- **Determining the Force Structure Trade Space, Specifically Addressing Intelligence, Surveillance, and Reconnaissance (ISR) and Dominant Battlespace Awareness (DBA)**
  *Roy Rice, Teledyne Brown Engineering*

# Toward a Unified Modeling Framework for Real-Time Logistics Control[1]

**Warren B. Powell**
*Department of Civil Engineering and Operations Research Princeton University*

## ABSTRACT

Increasingly, the real-time management of complex operations are yielding to on-line support of mathematical optimization models. More than simple decision support systems that display data, these systems actually make decisions. In this article, we review actual applications of such systems, highlighting the opportunities that await us. The article develops a general taxonomy for dynamic resource scheduling problems, and provides a flexible notation system that synthesizes optimization and simulation. The goal is a flexible optimization approach that bridges the gap between optimization and simulation.

## INTRODUCTION

Operations research has finally reached its golden era: the demand to solve complex problems efficiently is being met with the data and computing power needed to feed and solve our complex models.

We live in an age where we need to do more with less. We need to manage ever more complex systems with fewer resources. The pressure for this comes from many sources: global competition, soaring costs for fixed facilities (roads, airports), demand for human resources, and pressure to reduce federal deficits. An enabling technology is our increasing ability to store and manipulate information. Centralized control of complex systems is replacing decentralized operations. Modern sensing technologies (e.g. bar-coding) and transmission technologies (fiber optics, satellites, cellular phone networks) have dramatically expanded our ability to "know what is going on." At the same time, these technologies have presented a new challenge: What do we do with all this data?

Operations research offers the technologies to transform data into decisions. However, actual implementations of operations research have rarely survived the test of time. The excitement of initial successes, often accompanied by great fanfare, is often replaced with the eventual realization that mathematical models still cannot compete with the multidimensional skills of human decision makers. A theme of this paper is that very different applications can be described with a common vocabulary.

In the past, we have tried to apply optimization methods by translating problems into the vocabulary of linear programming. As users have struggled with this exercise, many have turned to simple simulation methods, without any sense of optimality. By developing a common vocabulary, it becomes possible to identify properties of these problems that can be studied scientifically, and tackled using powerful mathematical solvers.

The paper begins in section 1 with a review of real applications of on-line optimization models for control of logistics operations. These applications highlight the opportunities that await those who undertake the challenge, but also help to focus our energies on the challenges that continue to face the field of operations research, and the companies that wish to use these technologies. Section 2 provides a taxonomy of resource scheduling problems that arise in logistics. Section 3 presents a flexible notation system for putting these problems into a mathematical setting. Finally, section 4 discusses formulation issues, and argues that an approximate solution of an accurate model is better than an optimal solution of an approximate model.

## 1. EXAMPLES OF REAL APPLICATIONS

CASTLE Laboratory at Princeton University has been developing and implementating real-time optimization models for a variety of logistics applications. A brief summary of some recent projects provides an indication of what is now possible (names of companies are withheld):

**Short-haul routing and scheduling**—A real-time scheduling system plans driver tours for loads into and out of rail yards. The system is updated with each transaction, reoptimizing complete driver tours within five seconds (including data input and output) for over 400 drivers. The tours must obey driver work rules, driver and customer preferences, and pickup and delivery time windows on the loads. Because loads are short, individual tours might cover as many as six or eight different loads within a work shift. However, customer orders are constantly changing, as is the status of each driver, requiring us to constantly reoptimize complete tours. The problem is a real-time equivalent to the crew scheduling problem.

**Load matching for long-haul trucking**—A real-time system matches drivers with loads for a fleet of over 2,000 drivers handling 5,000 loads per week. Assignments of drivers to loads must consider distance, time windows (service commitments), load priority, driver hours, and driver preferences (in particular, the need to assign a driver to a load that returns him home). An optimization model responds to each transaction (new loads, changes in loads, changes in driver status) within five seconds. With long-haul loads, most drivers are assigned to a single load, but some are assigned to tours of two loads. Users see not just the driver tour as recommended by the model, but a list of alternatives ranked in order of their impact on all driver assignments.

**Driver/load management over a linehaul relay network**—A tactical planning system manages over 6000 drivers and 10,000 loads per week over a national linehaul relay network. Using hourly updates, the system forecasts driver movements and loads over a four day planning horizon, allowing planners to make decisions on what drivers to assign to what loads, and when the loads should be moved so that drivers and loads remain balanced at intermediate relays.

**Tactical management of rail flatcars**—A major railroad must optimize the repositioning of its fleet of 10,000 flatcars to move intermodal trailers and containers over its rail network. There are over 50 types of flatcars, each holding between one and eight of the over 40 types of trailers and containers. The use of specific types of trailers and containers is not uniform, and there is a benefit to getting certain types of flatcars to specific locations to maximize the utilization of the flatcar (there is no point in sending in a flatcar that can hold four containers to a location that has a low container volume but ships a lot of trailers). The optimization model does not run in batch, but optimizes adaptively, adjusting to real-time data transactions, but not necessarily solving to optimality between transactions.

**Routing and scheduling for chemical distribution**—A system has been implementing for designing driver routes and schedules to deliver chemicals to customers which use the product at varying rates. Some customers have small tanks, which need to be clustered with other tanks to fully utilize the vehicle. The size of the load depends on the time the vehicle arrives, which reflects current inventory and customer usage patterns. The routing and scheduling system must form clusters, assign drivers, tractors and trailers to produce a complete schedule. In addition, these tours must be formed over a two-three day planning horizon. The optimization system receives real-time updates of all resources (drivers, tractors and trailers) and customer usage patterns.

All of these applications share certain common themes. In each case, the project was motivated by management's desire to improve operations and/or cut overhead. Improvements were sought through lower operating costs, better service, and better management/utilization of resources. In addition, all of the projects were initiated through a combination of competitive pressures and the availability of data using new or recently acquired information technologies.

All of the problems can be described as optimizing the use of resources to accomplish tasks. In fleet management applications (truckload, rail) there may be large numbers of resources with only a few types (three or four trailer types) or many types (50 types of flatcars), but with relatively simple attributes. In others, resources are people with very complex sets of attributes, and where often no two are alike. These resources must be managed to satisfy tasks, and tasks can vary based on attributes (origin, destination, time windows, equipment required) but also on the nature of the tasks (can one resource handle one task, or can one resource, such as a flatcar, handle multiple tasks, such as containers).

While it is possible to develop application-specific software, ultimately such systems are difficult to maintain, and tend to be plagued by heuristics that do not evolve as the problem changes. Our experience is that it is much more effective to start with a general statement of the problem, and provide general tools through which application-specific rules can be coded.

## 2. A TAXONOMY OF RESOURCE MANAGEMENT PROBLEMS

It is common when solving problems that arise in complex operational settings to develop a set of procedures that are unique to a particular setting. Such an approach tends to produce solution methods that can take advantage of the structure of a particular application. At the same time, the resulting code is often difficult to maintain, and may be hard coded to operational practices that may change in the future. The down side of this approach is two-fold: First, it is often hard to see the "forest through the trees" in the sense that while it is easy to design simple heuristics, it is harder to identify the important trade-offs that lead to improved solutions. Second, as advances in optimization algorithms progress, it is hard to incorporate these advances into the solution method.

We view our problems within the general framework of assigning resources to handle tasks (in a way, this describes the entire field of operations research). Resources might be drivers, tractors, trailers, flatcars, containers, locomotives and aircraft. Tasks represent the movement of goods over space and time (by contrast, manufacturing represents the transformation of goods from one type to another).

Our taxonomy, then, involves dividing problems in terms of the types of resources, and types of tasks.

## RESOURCE CHARACTERISTICS:

**Homogeneous resources**—All resources are identical. Two resources at the same point in space and time can be equally used on the same task. Tractors and, in some cases, trailers, are often modeled as homogeneous resources.

**Heterogeneous resources**—There are different types of resources which can handle different tasks, with some cross substitution (if there is complete cross substitution, we can model them as homogeneous resources—if there is no cross substitution, the problem can be decomposed into a series of homogeneous resource problems). Heterogeneous resources are labeled by a resource type (for example, a 45′ trailer or a 53′ trailer) which does not change over time. A complex problem might be railroad rolling stock, of which there are hundreds of different types.

**Multiattribute resources**—Some complex resources are not satisfactorily described by a static categorization, but rather by a complex set of attributes, that may evolve over time. The most common example is people (driver/crew scheduling) but other examples might include aircraft, ships and locomotives (due to their complex maintenance needs).

## RESOURCE LAYERING:

**Single layer**—Single layer resource problems assign a single resource to a task (for example, a driver is assigned to pull a load, a taxi driver serves a passenger).

**Multiple layers**—It might be that different types of resources are needed to handle a task. A pilot needs an aircraft to serve a flight. A driver needs a tractor and a trailer to pull a load of freight.

Resource layering is an issue that is largely ignored in most mathematical models. Of course it takes a pilot and a plane to move cargo around the world, but the routing and scheduling literature tends to focus on a single resource, which might be just the vehicle operator (driver, pilot), or just the vehicle, with suitable approximations to capture the presence of a human operator.

The other side of the equation is the tasks. The dimensions of a task include size, length (distance covered/time required), spatial density and booking profile. More specifically:

## TASK SIZE:

**Pickup and delivery**—Tasks represent small shipments, with typically 10 or more fitting on a single vehicle.

**Partial truckload**—A single shipment might fill a vehicle, or it might be possible to fit up to five on a single vehicle.

**Truckload**—A single task will fill a single vehicle.

**Bulk**—A single task might require two or more vehicles to handle all the freight. This arises frequently in rail and global air logistics.

## TASK LENGTH:

**Short**—A single driver can handle two or more tasks in a single shift.

**Medium**—Most tasks will consume an entire work shift for a driver.

**Long**—A task will span more than one work shift for completion.

## SPATIAL DENSITY:

**Many to many**—Tasks go from many locations to many locations, with typically only a few tasks originating or terminating in the same locations at a given point in time. This arises commonly in truckload trucking, where loads are highly dispersed.

**Few to many/many to few**—Tasks originate or terminate in a few locations (such as ports, air terminals or rail terminals), but the other end of the load might be at any location.

**Few to few**—Tasks might move between highly concentrated locations, creating large flows of tasks moving between specific pairs of terminals. These situations arise in rail, as well as the flow of drivers between hubs for less-than-truckload motor carriers.

## BOOKING PROFILE:

**Large prebooking times**—The time between when a task is known and when it needs to be served is large, implying that when a plan is put together, all the demands are known in advance. These problems arise most frequently in pickup and delivery operations, fixed "milk-run" operations and scheduled services.

**No prebooking**—Here, customers call in requests and expect immediate service. These situations arise in some areas of transportation, but more commonly in emergency situations where advance planning is not possible.

**Mixed prebooking**—The time from when a load is known to when it must be served might be very long or very short, requiring the ability to plan quickly. At the same time, customers who make their requests known well in advance expect a high level of service. Many carriers struggle to provide guaranteed service for requests known well in advance, when they also have to do a substantial amount of short term planning in response to last minute requests.

The last dimensions of resource scheduling problems in logistics cover how the tasks are serviced operationally.

## SERVICE RESPONSIVENESS:

**Demand responsive**—Transportation service is provided as demands arise. Examples are taxis and truckload motor carriers.

**Scheduled service**—Service schedules are fixed, and demands have to adjust their schedule to meet that of the transportation service. Most airlines and shipping companies work on this format.

**Mixed service**—Often, demands are fairly predictable, providing an incentive to plan some services in advance. However, last minute requests and sudden surges are better served by providing a degree of demand responsive service. Railroads fall in this category, as does the Airlift Mobility Command.

## CONSOLIDATION STRATEGY:

**Point to point**—Transportation service (full truckload) moves directly from origin to destination. Personal taxi and limousine services work this way, as does most truckload trucking.

**Relay operations**—A load may move from origin to destination through a sequence of relays, often with a change of resources (most commonly drivers or crews, but this might apply to railroads with a change in locomotives).

**Pickup and delivery**—Smaller shipments are moved from point to point as a vehicle performs in-vehicle consolidation.

**Transshipment**—Here, goods are moved from one vehicle to the next to achieve additional economies of consolidation.

This taxonomy is quite broad, and encompasses problems ranging from classical pickup and delivery, to airline crew scheduling, to dynamic fleet management.

## 3. AN ALGEBRA FOR RESOURCE MANAGEMENT PROBLEMS

The problem faced by operations research specialists in logistics is one of translating very complex problems, with a lot of application specific details, into the very general language of linear programming. Thus, if we are faced with the problem of managing aircraft and crews to move loads globally in response to national emergencies, with all the issues and constraints that arise, we have to write code that translates this problem into a cost vector **c**, constraint matrix **A**, and constraint vector **b**, producing an optimization model of the form:

$$\min_{\mathbf{x}} \mathbf{cx}$$

subject to:

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{x} \geq 0$$

Once the problem is in this form, we can turn it over to powerful solvers that recognize this type of problem statement. It is the thesis of this paper that we need a richer vocabulary for logistics settings, that allows problems in logistics to be described in a more natural setting, but is still general enough so that a wide range of problems in logistics can be tackled using a common set of solvers. Our view is that problems in logistics consist of the following basic components: resources (drivers or crews, tractors, trailers, locomotives, aircraft), tasks (goods to be moved from one location to the next) and facilities (terminals, ports, relays, yards).

In this section we introduce the notation which is used to describe the problem throughout the paper. Because the model is dynamic, we need to make decisions over a number of time periods. Let T be the number of planning periods in our planning horizon. We allow decisions to be made at T+1 points in our planning horizon, which we refer to as time indices t = 0, 1, ...,T. Time period t for t = 0,...,T−1 represents the interval [t,t+1) and is of uniform length. Time indices and time periods are represented by the letters s and t.

## THE NETWORK

We define the following network data:

$C$=the set of all physical terminals in the transportation network.

$\tau_{ij}$=travel time (in integer time periods) from city i to city j

$c_{ij}$=the cost of moving a resource from i to j

We use the letters i and j to index terminals in $C$.

Solving the DRSP means deciding which resources should move which tasks over the transportation network. In making a sequence of such decisions over time, we are solving the DRSP on a discrete-time, dynamic network which is generated by replicating the physical network at each time index. A node in this network is represented by (i,t) in space time, and link (i,j,t) represents the movement from (i,t) to $(j,t+\tau_{ij})$, where $\tau_{ij}$ is the travel time from i to j. Let

$$\tau^{\max} = \max_{i,j \in C}\{\tau_{ij}\}$$

as the maximum travel time in the transportation network; by assumption.

## RESOURCES

To account for the resources moving through our dynamic network we define the following set for each $i, j \in C$, $t \in T$ and $s = 0,\ldots,\tau^{\max}-1$:

$R_{ijt}(s)$ = the set of all resources that are currently scheduled to arrive to terminal j from terminal i at time t+s.

A resource's attribute vector completely describes its state at any point on our space-time grid. These attributes are used to determine the cost and feasibility of a potential resource-to-task assignment. To simplify our notation we define for each $t \in T$, $j \in C$ and $s = 0,\ldots,\tau^{\max}-1$,

$$R_{jt}(s) = \bigcup_{i \in C} R_{ijt}(s)$$

as the set of resources currently scheduled to arrive at terminal j at time t+s. We can then define:

$$R_t(s) = \bigcup_{s=0}^{\tau^{\max}-1} \bigcup_{i \in C} R_{ijt}(s)$$

as the set of all resources in the network at time t. The set $R_t$ is the set of resources r defined over the entire population of resources.

Because of the heterogeneity of resources, we must track each individual resource over time. To this end, we define the following resource state vector:

$$\mathbf{a} = \{\ldots, \mathbf{a}_{rt}, \ldots\}$$

where

$$\mathbf{a}_{rt} = \text{attribute vector of resource } r \in R_t \text{ at time } t \in T.$$

We refer to the k[th] element of a resource's attribute vector as $a_{rt}(k)$. The minimum elements in the vector **a** is location, time of availability, and a unique label or identifier. Other elements might include hours of service, recent pay history, training or skill level and domicile. It is also useful to group resources by common attributes, using an aggregation function. Let $A$ be the space of all possible attribute vectors, and let G be an aggregation mapping:

$$G: A \rightarrow \hat{A}$$

where $|A| << |\hat{A}|$. Now let:

$R_a$=the number of resources with attribute vector $a \in \hat{A}$

## TASKS

Similarly, to account for the tasks that need to be moved about our dynamic network we define the following set for each $i, j \in C$ and $t \in T$.

$L^0_{ijt}$ = the set of all tasks with origin terminal i and destination terminal j that first become available at time t.

$L_{ijt}$ = the set of all uncovered tasks with origin terminal i and destination terminal j at time t. Similarly, we define the set of all uncovered tasks in the network at time $t \in T$ as

$$L_t = \bigcup_{i \in C} \bigcup_{j \in C} L_{ijt}$$

As with resources, we assign an attribute vector for each task:

$\mathbf{b}_t = \{\ldots, \mathbf{b}_{lt}, \ldots\}$

= attribute vector of task $l \in L_t$ at time $t \in T$ where $\mathbf{b}_{lt}$ is the vector of attributes for task l at time t.

## DECISION VARIABLES

We can now present the notation that is used in the recursive formulation of the DRSP. First we define our decision variables. For each $t \in T$, $i, j \in C$, $r \in \mathbf{R}_{it}(0)$ and $l \in \mathbf{L}_t$, we define

$$x_{rjt} = \begin{cases} 1 & \text{if resource r is assigned to cover task l at time t} \\ 0 & \text{otherwise} \end{cases}$$

Notice that we have restricted resource assignments to tasks available at the resource's current terminal. The movements of unassigned resources (i.e., repositioning movements) arise naturally out of scheduled resource assignments. Hence, we define for each $t \in T$, $i$, $j \in C$ and $r \in \mathbf{R}_{it}(0)$

$$y_{rjt} = \begin{cases} 1 & \text{if resource r is assigned to move empty to j at time t} \\ 0 & \text{otherwise} \end{cases}$$

For notational clarity we define the vectors of resource assignments and repositioning movements as

$$\mathbf{x}_t = \{\ldots, x_{rlt}, \ldots\}$$

and

$$\mathbf{y}_t = \{\ldots, y_{rjt}, \ldots\},$$

respectively.

To record the coverage of tasks we define for each $t \in T$, $i, j \in C$ and $l \in L_{ijt}$ the integer variable $z_{lt}$ by

$$z_{lt} = \sum x_{rlt}$$

Clearly $z_{lt}=1$ if task l is first covered during period t and $z_{lt}=0$ otherwise. For notational convenience we define the variables $\tilde{z}_{lt}$ by

$$\tilde{z}_{lt} = \sum_{\tau=0}^{t-1} z_{l\tau}$$

which record whether or not a task has been covered in any time period preceding the current.

## SYSTEM DYNAMICS

It is convenient to describe our collection of resources and tasks as an abstract "system". We denote the "state" of the system at time $t \in T$ immediately prior to the knowledge of the decisions in $\mathbf{x}_t$ and $\mathbf{y}_t$ by $\mathbf{S}_t$. The state of the system is described by the sets of resources

$$\mathbf{S}_t = \{(R_t, \mathbf{a}_t), (L_t, \mathbf{b}_t)\}$$

Our state variable consists of the set of resources $R_t$, the attributes of these resources $\mathbf{a}_{rt}$, $r \in R_t$, the set of uncovered tasks $L_t$, and the attributes of these tasks, $\mathbf{b}_{lt}$, $l \in L_t$. $\mathbf{S}_0$, $\mathbf{L}_0$ and $\mathbf{R}_0$ are required as input data.

To keep our system state variable current, we must update it after each time period has passed and the accompanying set of decisions have been implemented. We can describe a system state variable update in more detail using the algebra which we have developed. We define operators A and B to update the resource and task attribute vectors, respectively. An update of $\mathbf{S}_t$ may then be phrased in terms of these operators, where we suppose that we have just fixed $\mathbf{x}_t$ and $\mathbf{y}_t$ at their optimal values. That is,

$$\mathbf{a}_{t+1} = A(\mathbf{x}_t, \mathbf{y}_t, \mathbf{a}_t, \mathbf{b}_t)$$

and

$$\mathbf{b}_{t+1} = B(\mathbf{b}_t, \mathbf{x}_t, \tilde{z}_t)$$

The A operator is used to update the resource attribute vectors. The operation of A on $\mathbf{x}_t$, $\mathbf{y}_t$, $\mathbf{a}_t$ and $\mathbf{b}_t$ to produce $\mathbf{a}_{t+1}$ is really a sequence of actions that depends on the specifics of the individual problem. The B operator similarly updates the task attribute vector b.

Similarly, we define operators R and L to update the sets of resources and tasks. Given $\mathbf{a}_t$, $\mathbf{b}_t$, $\mathbf{x}_t$, $\mathbf{y}_t$, we can define:

R: $R_t \rightarrow R_{r+1}$

L: $L_t \rightarrow L_{r+1}$

The operators A, B, R and L are intended to encompass completely general rules governing the evolution of the state of the system. They are highly application specific and do not, in a direct way, affect the optimization procedures. Though these operations, we are trying to merge the strengths of simulation and optimization.

# THE OBJECTIVE FUNCTION

We complete our notation system by defining the cost and reward parameters which are used to evaluate each resource assignment:

$\mathbf{c} = \{\ldots, c_{ij}, \ldots\}$

= the cost of repositioning a resource from location i to location j (this can be made time dependent if necessary)

$h(\mathbf{a},\mathbf{b}) = \{\ldots, h(\mathbf{a}_{rt}, \mathbf{b}_{lt}), \ldots\}$

= the vector of costs of assigning a resource with attributes $\mathbf{a}_{rt}$ to cover task $l \in L_t$ beginning during period t, $i, j \in C$.

$\mathbf{r} = \{\ldots, r_{lt}, \ldots\}$

= the vector of rewards received for begining task $l \in L_t$ during time period t, $i, j \in C$.

The cost vector **c** captures traditional transportation costs from moving over space and time (this includes holding in inventory). The function **h** captures application-specific costs that depend on the specifics of the resource and task attribute vectors. Some of these costs might be real, and others might represent artificial bonuses and penalties designed to encourage or discourage specific behavior. While not very rigorous, these "soft costs" are a part of every application.

To develop an objective function, first define:

$g_t(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{S}_t) = \mathbf{r}_t \mathbf{x}_t - \mathbf{h}_t(\mathbf{a}_t, \mathbf{b}_t)\mathbf{x}_t - \mathbf{c}_t \mathbf{y}_t$

= the net 'profit' from the decisions made at time t.

Our problem is to determine x, y and z to solve:

$$\max_{\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t} \sum_{t \in T} g_t(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{S}_t) \tag{1}$$

It is possible, in principle, to solve this as a large scale optimization problem. If the problem is relatively simple, it might reduce to a network or a multicommodity network flow problem. For many applications, the attribute vector a and the updating process A is sufficiently complex that a column generation method is needed. Complex work rules and other operational concerns can be built into the logic for generating possible schedules for a resource. These schedules become the columns of a set partitioning model, which can be solved with linear programming. Such techniques have been effectively applied to airline crew scheduling (an excellent review of these techniques can be found in Desrosiers *et al.* [2]). There are several problems with these approaches. First, they are not well suited to handling forecasting uncertainties. Second, they perform very poorly in the face of long planning horizons.

An alternative approach is to express the problem using a recursive formulation.

$$G_t(\mathbf{S}_t) = \max_{\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t} g_t(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{S}_t) + G_{t+1}(\mathbf{S}_{t+1})$$

= the best net profit from all decisions made in the interval [t,T].

In the traditional language of dynamic programming, $G_{t+1}$ is the value function of our optimality recursion. We define $G_{T+1} = 0$. We may then break the DRSP into a sequence of easier subproblems. By Bellman's "principle of optimality", we know that solving such a sequence of subproblems will provide us with the same solution in the end as solving the original problem. Specifically, the subproblem to be solved at each time index t, is as follows:

$$\max_{\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t} g_t(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{S}_t) + G_{t+1}(\mathbf{S}_{t+1})$$

with system dynamics given by:

$$\begin{aligned} \mathbf{a}_{t+1} &= A(\mathbf{x}_t, \mathbf{y}_t, \mathbf{a}_t, \mathbf{b}_t) \\ \mathbf{b}_{t+1} &= B(\mathbf{b}_t, \mathbf{x}_t, \tilde{\mathbf{z}}_t) \\ R_{t+1} &= R(R_t, a_{t+1}) \\ L_{t+1} &= L(L, \mathbf{x}_t, \mathbf{a}_{t+1}, \mathbf{b}_{t+1}) \end{aligned}$$

It is well known that in practice this system is intractable. However, if we replace $G_t(\mathbf{S}_t)$ with a careful approximation, the problem can become quite easy to solve. Furthermore, it is quite easy to incorporate a very high level of detail. While the resulting solution may not be optimal, in the sense of solving an approximate problem, we will be better able to solve a more accurate model of the real problem.

# 4. FORMULATION STRATEGIES

The problem with equation (1) is that for most applications, the problem is intractably large. Furthermore, uncertainty in the data brings into question the meaning of such a global "optimum." In our experience, it is not uncommon that a mathematical, global optimum solution is unimplementable (because it assumes a level of data accuracy that is not achievable). Furthermore, such global optima are only possible for deterministic models. Experimental results in Frantzeskakis and Powell (1990) and Powell and Cheung (1994) show that optimal solutions of deterministic models can provide significantly worse results than an approximate solution of a stochastic model. The results of this research provides supporting evidence to the following fundamental hypothesis of modeling:

*An approximate solution to a more accurate model is better than an exact solution to an approximate model*

Practitioners have realized this for many years. Not uncommonly, highly heuristic simulation models, such as the MASS program used by the Air Mobility Command, supplant more powerful but less flexible linear programs. Simulation offers the analyst much greater flexibility, but gives up in the process any possibility of finding even near optimal solutions.

What is often missing from the basic modeling paradigm is the realization that simple rules can provide near-optimal solutions. Furthermore, it is possible, using hierarchical control strategies, to guide simple strategies toward a global optimum. Recently, for example, Powell *et al.* (1995) developed a method that can be applied to large fleet management problems with multiple equipment types and time windows. Classically formulated as a large scale, multicommodity network flow problem with GUB constraints to handle time windows (see Powell *et al.* (1995) and Magnanti and Simpson (1978)), it was shown that the same problem could be solved much more easily from the perspective of optimal control of queues. It is important in large, complex problems not to overreach the problem in the pursuit of global optimality. Much more effective strategies can be developed by decomposing the problem and solving sequences of smaller problems.

There are three dimensions to global formulations of complex problems that can each be exploited to provide more efficient solution algorithms:

**Spatial**—It is possible to optimize over all decisions over all points in space, coordinating the activities in location i with activities in location j.

**Temporal**—It is possible to optimize decisions over all points in time, so that decisions at time t are coordinated with decisions at time t' > t.

**Organizational**—Real decisions take place at several levels of an organization. It is possible to combine these organizational levels to achieve a more global optimum.

Optimizing globally over space requires a level of coordination that is not always warranted. Many decisions can be made at one location without worrying too much about decisions at other locations. At the same time, some coordination is required, and obvious mistakes can be made if every location is allowed to work independently. In Powell *et al.* (1995), a control variable plus dual information was communicated to each location. These variables are adjusted globally based on up-dated information. This research produced a spatial decomposition allowing each location to work independently, and yet still achieve a solution close to a global optimum.

Temporal decomposition arises naturally in stochastic formulations of problems. The ability to optimize across time periods is an artifact of deterministic models. Temporal decomposition requires optimizing today using only approximate information about what will happen tomorrow. The result of this approach is a much easier optimization problem.

Organizational decomposition provides explicit recognition to multiple levels of decision making. Global optimization formulations assume a single decision-maker.

Organization decomposition implies a hierarchical structure. For example, decisions to reposition aircraft from location to the next might occur at a central command level, while the decision of what freight to put on an aircraft, or what pilot to assign, might be made locally. Again, organizational decomposition can, in a theoretical sense, produce lower quality solutions (in a theoretical sense), but hierarchical decisions make better use of information available to different levels of decision makers, and which might not be available to the computer.

## ACKNOWLEDGEMENT

This research was supported in part by grant DDM-9102134 from the National Science Foundation, and by grant AFOSR-F49629-93-1-0098 from the Air Force Office of Scientific Research.

## REFERENCES

[1] R. K.-M. Cheung and W.B. Powell, "An Algorithm for Multistage Dynamic Networks with Random Arc Capacities, with an Application to Dynamic Fleet Management," *Operations Research*, 1994 (to appear).

[2] J. Desrosiers, M. Solomon and F. Soumis, "Time Constrained Routing and Scheduling," *Handbook in Operations Research and Management Science: Networks"* (C. Monma, T. Magnanti and M. Ball, eds.) North Holland, 1995.

[3] L. Frantzeskakis and W. B. Powell, "A Successive Linear Approximation Procedure for Stochastic, Dynamic Vehicle Allocation Problems," *Transportation Science*, Vol. 24, No. 1, pp. 40-57 (1990).

[4] T. Magnanti and R. Simpson, "Transportation Network Analysis and Decomposition Methods," Report No. DOT-TSC-RSPD-78-6, U.S. Department of Transportation, 1978.

[5] W.B. Powell, T. Carvalho, G. Godfrey and H. Simao, "Dynamic Fleet Management as a Logistics Queueing Network," *Annals of Operations Research on Transportation*, 1995.

[6] W.B. Powell, P. Jaillet and A. Odoni, "Stochastic and Dynamic Networks and Routing," in *Handbook in Operations Research and Management Science*, (C. Monma, T. Magnanti and M. Ball, eds.), North Holland, 1995.

## ENDNOTE

1 Submitted June 1995; In final form September, 1995

# Publications Subscription Rates

***PHALANX* — 4 Issues Annually**

| Domestic Rates | | Foreign Rates | |
|---|---|---|---|
| 1-Year | $20 | 1-Year | $40 |
| 2-Year | $35 | 2-Year | $70 |

***Military Operations Research* — 4 Issues Annually**

| Domestic Rates | | Foreign Rates | |
|---|---|---|---|
| 1-Year | $40 | 1-Year | $80 |
| 2-Year | $75 | 2-Year | $150 |

**Combination Subscription — The Best Value!**

| Domestic Rates | | Foreign Rates | |
|---|---|---|---|
| 1-Year | $55 | 1-Year | $110 |
| 2-Year | $100 | 2-Year | $200 |

**Foreign rates are based on checks drafted on US banks.**

## Please Start My Subscription Today!

*PHALANX* ☐ MOR ☐ Combination Subscription ☐

1-Year ☐ 2-Year ☐ Total Enclosed ____________

Name: ____________________________________________

Address: __________________________________________

City ______________________ State ________ Zip Code ____________

# Modeling and Optimization of Mobility Analysis: Optimal Requirement Studies[1]

**Fan Yang,**
**Ervin Y. Rodin,**
**S. Massoud Amin**

*Center for Optimization and Semantic Control[2]*
*Department of Systems Science and Mathematics*
*Washington University*

## ABSTRACT

In this report, we describe our work in developing models, methodologies and simulations for network optimization problems in the planning, analyzing and optimizing of large scale (air) transportation networks with time window constrained routing and scheduling. Our research is motivated by certain problems encountered in the United States military's strategic mobility analysis, in general, and specifically in Mobility Analysis Support System (MASS) of the USAF's Air Mobility Command (AMC).

This work is performed within the framework of Semantic Control paradigm, a three-layer supervisory hierarchical structure. In this context a new mathematical programming model, called Network Optimization Mobility Analysis (NETO), for the mobility analysis system is formulated as a pickup-delivery vehicle routing and scheduling problem with time-window constraints (PDPTW). In order to cope with the computational complexity inherent in the PDPTW formulation, we have developed and implemented a novel algorithm called SP-CGCE (set-partitioning formulation, column generation and column elimination). The computational results indicate a promising and robust performance by this solution algorithm. The problems tested/solved here involve many more nodes than similar problems previously attempted. The test results indicate that the SP-CGCE algorithm is at least twice as fast as currently available column generation-branch and bound schemes; this increase in speed is due to the effectiveness of the column elimination process used after the completion of the linear programming phase to obtain integer solutions.

In particular, the focus of this report is the optimal *requirement studies problem,* where the following question is addressed: *"How many of what types of transportation assets are necessary to move cargo to the specified destinations, satisfying a particular desired closure schedule?"*

## 1. INTRODUCTION

The Center for Optimization and Semantic Control at Washington University in St. Louis has been conducting research jointly with the Air Mobility Command of the United States Air Force with respect to the Mobility Analysis Support System. In the past, we have solved several large-scale, time-dependent, mixed variable, uncertain and complex problems encountered in aerospace and decision support domains [1-5,8,9] using the Semantic Control paradigm (see below). The Center researchers approach the solution of such problems using a judicious combination of classical mathematical methodologies (mathematical programming, computational geometry, control theory, game theory, stochastic, etc.), together with Artificial Intelligence paradigms such as Planning, Search, Fuzzy System Theory, Neural Networks, Rule Based Systems, and Logic Programming [8-9]. Our approach is based on the Semantic Control paradigm—a three-level hierarchical structure (Figure 1-1) consisting of:

- an Identifier, which processes the list of requirements, known as the time-phased force deployment data/document (TPFDDs[3]), and interprets the available information;
- a Goal Selector, which generates and evaluates several plans; and
- an Adapter, which implements the optimal plan.

For example, the Identifier module consists of neural networks for processing, pattern recognition and optimization of TPFDDs. Once trained, neural networks identify requirements and consequently recommends assignment and allocation of aircraft in order to deliver those requirements. Currently, given a requirement containing:

i) a commodity code (such as outsize, oversize, bulk, passengers),
ii) an onload-offload region, and
iii) the percent of the total requirement to be moved,

the neural network recommends the appropriate assignment and allocation of aircraft to deliver that requirement.[4] The neural network module serves as a "pattern recognizer" in order to reduce the complexity; in addition to this, we are currently developing a "fuzzy model" of the air transportation which incorporates leeways in the constraints and goal. This should prove very useful since several quantities such as MOG (maximum on

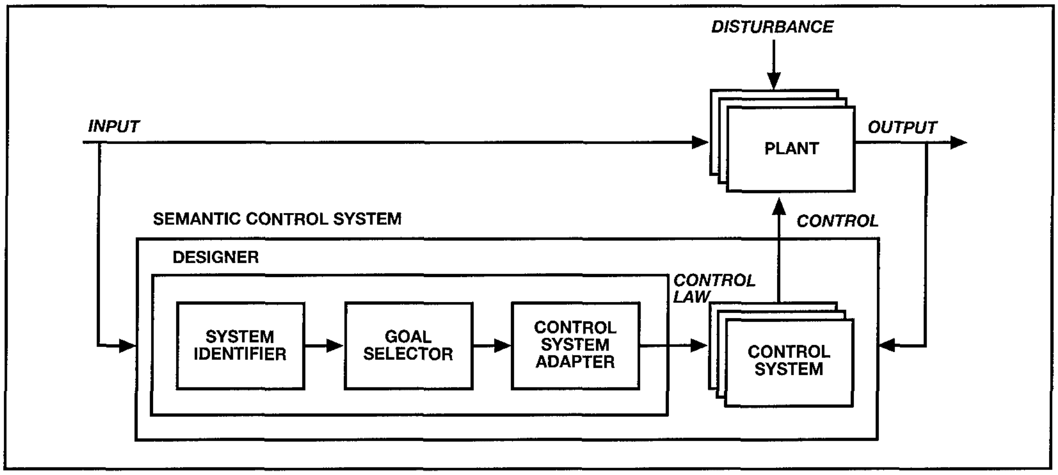



**Figure 1-1: A semantic control system consists of a System Identifier, a Goal Selector, a Control System Adapter, and one or more control systems/laws.**

ground) are not crisp variables. This approach admits such uncertainties as part of the model, thus reducing labor-intensive post-optimality sensitivity analysis. These issues will not be discussed further in this report. We refer the interested reader to [8,9]. This report deals mainly with algorithm development and simulation of exact mathematical programming and optimization methodologies (cf. [6] and [10]) for the Goal Selector module of the Semantic Controller. More specifically our objectives in this report are:

1) reviewing existing mobility analysis models and addressing their various limitations,
2) presenting the new mobility analysis model NETO formulated as a PDPTW problem,
3) discussing our solution algorithm (SP-CGCE) and comparing its performance to other published results,
4) providing a brief overview of the system implementation and related issues,
5) giving an example of the optimal requirements studies problem, and
6) concluding with a discussion of other relevant problems addressed by this approach as well as related open problems.

This report is divided into six sections:

- Section 1 and subsections 1.1 through 1.3 present background information on the strategic mobility analysis and limitations of current simulation and mathematical models.
- Section 2 discusses the system architecture and the components of our model NETO.
- Section 3 focuses on the mathematical formulation, algorithmic details, and performance analysis. The mathematical model for our formulation is given in more detail in Appendix B.
- Section 4 presents system implementation and gives an example of the optimal requirements studies problem.
- Section 5 discusses other related problems and defines future work.
- Section 6 concludes with a brief summary.

## 1.1. STRATEGIC MOBILITY ANALYSIS: BACKGROUND [7]

Various objectives of strategic mobility analysis are grouped into three broad planning categories:

- Resource Planning: long-range deployment planning and programming.

- Deliberate Planning: mid-range deployment planning that encompasses the development and analysis of operational plans.
- Execution Planning: including both the short-range crisis action planning before an engagement begins and the continuing planning and replanning as execution proceeds.

There are two fundamental questions involved in the above planning categories as well as in all other planning activities:

1) How to accomplish the objectives (and to what degree) given the resources?
2) Given the objectives, what are the minimum resources required to accomplish them and how to do so?

In particular *Resource Planning* encompasses program development and related policy research that is conducted in the planning, programming, and budgeting system (PPBS). Although mobility studies in resource planning may presume specific theater scenarios, the analyses are collectively meant to conduct coordinated long-range resource planning for total forces. These studies are generally of two types: capability assessments, which determine the force closure that can be supported by a given set of lift assets, and requirements studies, which estimate the lift assets necessary to support a given force closure.

In *capability assessments* (the forward problem), a strategic mobility model is used to assess how soon a particular set of transportation assets can effect theater closure of a particular set of forces, support resource, and resupply, given the constraints of scenario and cargo priorities. Although capability assessments theoretically are one-shot uses of the model, more runs are almost always needed to assess the implications of uncertainty. To explore degrees of risk with a given force structure and operational objectives, the model may be exercised numerous times with different versions of scenario assumptions.

In *requirements studies* (the backward problem) the following question is asked, "How many of what types of transportation assets are necessary to move cargo to the specified destinations, satisfying a particular desired closure schedule?" The results of the analysis describe a set of transportation assets , or perhaps the required increments to a baseline set of assets. Conducting this type of study with the currently available mobility models is necessarily a tedious iterative process. At the Joint Staff, an important recent example of a requirements study is the RIMS (Revised Intertheater Mobility Studies), which required over 400 MIDAS runs (Model for Intertheater Deployment by Air and Sea) between October 1986 and April 1989.

## 1.2. ANALYSIS PROCESS OF THE CURRENT MOBILITY MODELS

The models that are currently being used in the defense communities [7], such as MIDAS (Model for Intertheater Deployment by Air and Sea, a Joint Deployment System model, 1980), RAPIDSIM (Rapid Intertheather Deployment Simulator, 1974), TFE (Transportation Feasibility Estimator, a Joint Operation Planning System), FLOGEN (Flow Generator, an Air Mobility Command model), SEACOP (Strategic Sealift Contingency Planning System, a Military Sealift Command model), MASS (Mobility Analysis Support System, an Air Mobility Command model, 1980's), etc., all process data in a similar way. Each model uses several inputs in the form of data files; all use similar algorithms to simulate the transportation system, and all produce similar outputs, e.g. delivery dates, utilization rates, and delays/queues.

Typically, four files provide input for the simulation models: a requirement file, a PREPO (prepositioning) file, a transportation resources file, and a scenario file. The model then assigns cargoes to transportation assets according to certain rules, and simulates cargo movement through the transportation system. All of the current models use the same solu-

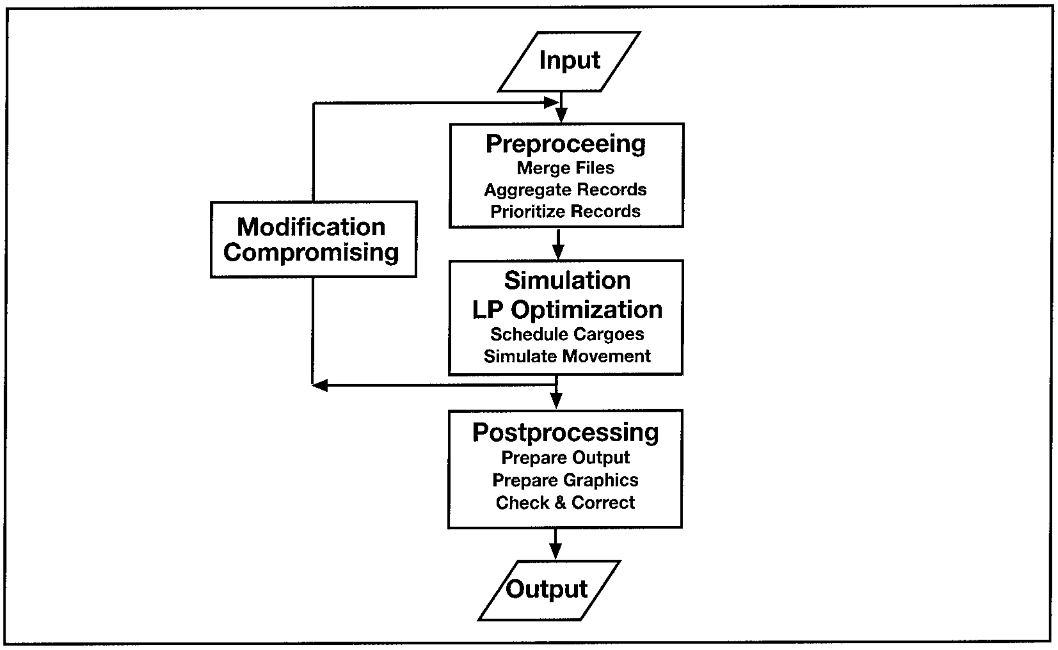



**Figure 1-2. The Analysis Process of the Current Mobility Models**

tion technique (deterministic simulation) and basically follow the same steps to arrive at a delivery profile. A model may undergo some or all of the following analysis tasks (Figure 1-2): merge files; aggregate records (by ports, route, ships, or cargo); prioritize records; select models (i.e., air or sea for those with no chosen mode); schedule cargoes; simulate movement; prepare textual output; prepare graphical output; check and correct.

As mentioned earlier, we were motivated by problems encountered in the Mobility Analysis Support System (MASS). MASS is a family of analytical tools developed originally by the Command Analysis Group at the headquarters, Military Airlift Command, Scott AFB, Illinois, from the mid-1980's to the early 1990's. The Command Analysis Group (Studies and Analysis Flight now) is presently working under the Plans and Analysis Directorate at the headquarters, Air Mobility Command. The MASS family consists of a variety of models to aid in the analyses of the full spectrum of airlift operations from daily peacetime cargo movement to full-scale global wartime movements such as Desert Shield/Storm.

MASS is a deterministic simulation model which directs aircraft through a network of onload, enroute, offload, and recovery bases in order to deliver a set of requirements needed to achieve some predefined scenario goal. MASS is capable of handling many diverse scenarios. An enroute base is an intermediate stop, normally for fuel or to change crews, between an offload and onload base. A recovery base is visited after the offload for fuel and/or crew change, in order to relieve congestion at the offload base. The recovery base is where aircraft wait to be scheduled for their next mission.

The cargo requirements (TPFDD) contain cargo information or requirements such as onload(origin), offload(destination), available date, required delivery date, size, weight and nature of the cargo, etc. The set of cargo requirements given by TPFDD are taken as input, based on the availability of aircraft, airfields, parking space, crew members and routes etc., MASS works through the entire airlifting operation, simulating onloading, offloading, scheduling, routing, refueling, crew changing processes, generating a multitude of step by step aircraft activities, cargo movement and delivery information. MASS also simulates the impacts made by some anticipated/unanticipated changes in the airlifting system, such as increased number of aircraft, base closure, etc. It is an effective tool in that it offers a feasible solution to the airlift problem; however, it does not address the backward problem directly nor does it guarantee an optimal solution. In subsection 1.3 we describe the limitations of current models in more detail.

## 1.3. LIMITATIONS OF CURRENT MODELS

Schank et al. ([7], pp. 39-50) from the RAND corporation provide a very comprehensive review of strategic mobility models and analyses. The following limitations of the current models are identified:

- All work in only one direction, accepting similar types of input data and producing the same general information.
- None are optimal.

**Current Models All work in One Direction:** All major current mobility models are simulations or have major simulation components. They accept data on what has to be moved (cargoes), what is prepositioned (PREPO), what transportation assets are available, and what the assumptions are regarding timing and available infrastructure. They then assign cargoes to transportation assets according to specific rules and simulate their movement through the transportation system. Finally, all produce estimates of when units are delivered into the theater and utilization rates of the transportation assets and facilities. All existing models use this process, regardless of the decisions and objectives being addressed. All models basically provide the closure profile for these forces, support units and resupply given these transportation assets (forward problem).

This question may be appropriate for deliberate planning or execution planning analysis, but it does not directly address the concerns of how many transportation assets are required (important for resource requirement studies).

Strategic mobility analysis that addresses transportation asset requirements seeks the best mix of transportation assets for achieving a desired closure profile for a given set of forces, support units and resupply (backward problem). The unknown values in requirements determination are required inputs to existing models.

At present, therefore, analysis cannot directly answer the question of how many of each type of transportation assets are required. They can only obtain an approximate solution by multiple runs and trial and error.

**The Solution Is Not Optimal:** This laborious process, more art than science, certainly does not provide "optimal" answers. In fact, a good deal of expertise is typically needed to develop even a "good" answer to transportation force structure issues. This suggests that a different modeling approach is warranted, one that moves away from simulations, or at least from current simulation methods, to an approach that directly addresses force requirements questions.

Our mobility analysis system, NETO, accepts the same input file information as the above models; however, it differs from them and other newer approaches (such as ADANS) in its optimization and analysis capabilities. For example, all other models are geared toward addressing the capability assessment (the forward problem) while unable to solve the optimal requirements studies (the backward problem); NETO is capable of solving both forward and backward problems. In the remainder of this report we focus precisely on the commonly ignored optimal requirement studies problem. In what follows, we discuss our approach to address the above limitations and to solve the optimal requirement studies problem without the need for repeated runs.

## 2. NETWORK OPTIMIZATION MOBILITY ANALYSIS SYSTEM

NETO consists of two interrelated components: a network optimization engine with time window-constrained routing and scheduling based on integer and combinatorial optimization methodology; and an analysis system with a information management system built upon RDBMS and multimedia technology. RDBMS is not presented here; it falls outside of the focus of this report. In this section, we will give an overview of the NETO system architecture, describe the underlying labeled digraph and PDPTW. More detailed issues as well as the forward problem, selection of cutting plane, column generation, column elimination, numerical results and comparisons are reported in [6].

## NETO SYSTEM ARCHITECTURE

The diagram and system hierarchy of NETO system architecture are shown in Figure 2-1. The input information is the same as the original input information (cf. Figure 1-1), except that it might come from a database system. We will substitute the original simulation process with our optimization system.

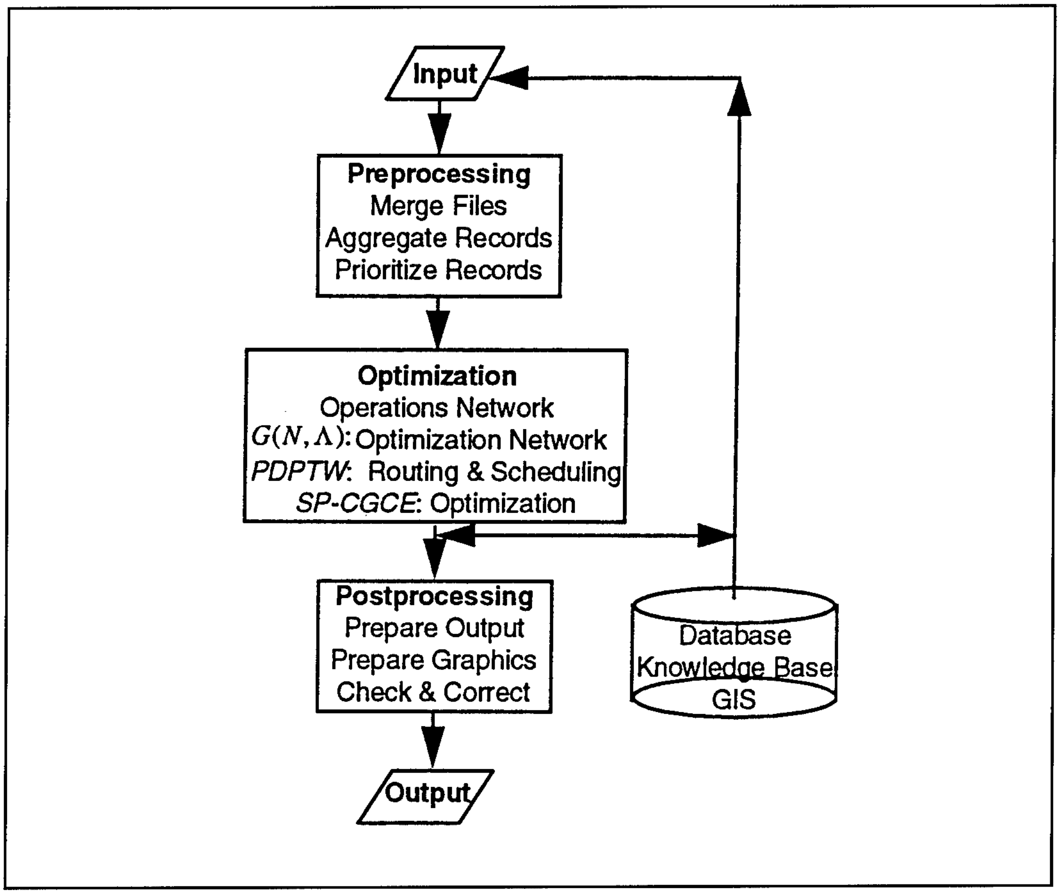



**Figure 2-1. NETO System Architecture**

## NETO SYSTEM COMPONENTS

The functions of the various components are described as follows:

*Operations Network:* The Operations Network is the original mobility operation information represented in the form of transportation network consisting of all relevant data, such as air bases, seaports, air routes, sea routes, onloads, offloads, enroutes, cargoes, transportation vehicles, weather, scenario, movement requirements, logistics factors, and so forth. The operations network is more than just a geographic network such as a map; it is a model of the concept of operations, a database of the operation information.

*Optimization Network $G(N,\Lambda)$:* The Optimization Network is the Operations Network represented in the form of a *labeled digraph* suitable for mathematical optimization purposes. It has four types of nodes: starting nodes $S$, terminating nodes $T$, pickup nodes $P^+$and delivery nodes $P^-$. $P^+$ and $P^-$ forms a *complete digraph* $P^+ \times P^-$. For $S$ arcs only go from $S$ to $P^+$. For $T$ arcs only go from $P^-$ to $T$. Denote $P=P^+ \cup P^-$, $N=S \cup P \cup T$, and $\Lambda = S x P^+ \cup P^+ x P^- \cup P^- x T$. Then we can write the digraph as $G(N,\Lambda)$.

In the mobility analysis system, the set of nodes $S$ could be the home depots. The set of nodes $T$ could correspond to the recovery bases. $P^+$ depicts the onload bases of requirements and $P^-$ describes their offload bases. A node in the optimization network may correspond to many physical nodes in the operations network or vice versa; an arc in the optimization network may correspond to several arcs/paths in the operations network and vice versa. The labels contain various relevant information derived from the operations network. Among these data are the cost of arc $(i,j)$, time window constraints $[a_i,b_i],[a_j,b_j]$ (time intervals during which service is required; "service" meaning either pickup or delivery), the physical nodes that make up the arc $(i,j)$, etc. Some of the variables used in the optimization network are:

| | |
|---|---|
| $\overline{d}_i$: | load vector(volume, weight...) of cargo $i$ at node $i$ |
| $[a_i,b_i]$: | pickup time window at node i for movement/cargo $i$ |
| $[a_0,b_0]$: | time window for vehicle leaving the depot $S$ |
| $[a_{2n+1},b_{2n+1}]$: | time window for vehicle returning to the depot $T$ |
| $\overline{D}$: | capacity of vehicle (load weight limit, volume,...) |
| $t_{ij}$: | travel time from node $i \in N$ to node $j \in N$ |
| $s_i$: | service time(pickup time or delivery time) at node $i \in N$ |
| $\overline{Y}_i$: | the total load on the vehicle just after it leaves node $i \in N$ |
| $T_i$: | time of start service at node $i \in N$ |
| $T_0$: | arrival time at node $i$ or time vehicle leaves the depot $S$ |
| $T_{2n+1}$: | time vehicle returns to the depot $T$ |
| $RT$: | feasible route defining formulation |

A complete list of terminology, definitions, notation, and symbols is given in Appendix A at the end of this report. If the labels which store the transformed mobility information are oriented, all mobility analysis will result in the same kind of optimization network. Therefore if labels are not considered, the optimization network $G(N,\Lambda)$ is a topological representation of the mobility system. An operations network is converted to an optimization network through *Network Construction.*

*Network Construction:* This transforms an operations network into an optimization network, taking into consideration such factors as routes, enroutes, cargoes, fuel, and other information in association with the operations network. There are basically two tasks: building up the digraph topology $G$ and computing the labels. For example, to construct the arc from a pickup node $i \in P^+$ to a delivery node $n+i \in P^-$, we may select the shortest path $P$ with the maximum length of any segment in $P$ not exceeding a certain quantity in the operations network. This could mean that a certain type of aircraft can make a sustained flight with supported available refueling along the route.

*Reduced Optimization Network:* The Reduced Optimization Network is an Optimization Network reconfigured by tightening some excessively wide time windows and by eliminating as many as possible inadmissible arcs. By excessively wide time windows, we mean those time windows that can be narrowed without changing the problem under consideration. By an inadmissible arc we mean arcs which violate the constraints imposed upon the mobility system, such as time window and vehicle capacity, etc. PDPTW [16,17,20,21] is the underlying model for the optimization. The PDPTW model represents a vehicle routing and scheduling problem where cargoes are to be picked up in specified origins (sources) within given pickup time periods and to be delivered to desired destinations (sinks) within given delivery time periods.

The PDPTW was first formulated based on a vehicle flow/multicommodity flow-based nonlinear model (cf. [6]) and then reformulated into set-partitioning formulation and solved by the Column Generation Column Elimination Algorithm (SP-CGCE). The Column Generation Technique is based on the primal-simplex method to efficiently solve LP problems with a very large number of columns. It decomposes the original LP program into a master problem and a subproblem. In our research we have decomposed the linear relaxation of the set partitioning formulation of the PDPTW into a shortest-path subproblem with constraints. After the LP optimal is obtained by the column generation process, the Column Elimination Technique has been employed to obtain integer optimality.

The optimization result has been utilized for various output analysis purposes, according to the specific needs of the operation. In particular, the output data have been stored in the database system for further analysis.

## 3. SP-CGCE SOLUTION ALGORITHM AND PERFORMANCE

The use of a set-partitioning formulation, with the column generation scheme for solving vehicle routing problem (VRP) and PDPTW problems has recently become more frequent [11-21]. This is mainly because good alternative formulations for PDPTW problems are not known and the linear programming relaxation of the set partitioning formulation often yields a strong bound.

Other algorithms for solving PDPTW problems found in the literature [17,20] generally take the following *set-partitioning formulation, column generation, branch-and-bound* approach: these algorithms use a set-partitioning formulation and solve the relaxed set-partitioning problem by column generation, where columns are generated when necessary by solving a constrained Shortest Path Problem. Often the linear optimal solution is also an integer solution. If it is not, the linear optimal solution offers a good lower bound for the original set-partitioning problem, especially if some heuristic cutting planes are used. Then this scheme resorts to branch-and-bound to find the integer optimal. Since the original integer formulation is a set-partitioning formulation, branch-and-bound can not take place on the decision variables directly, but rather, on the arcs/paths in the network. This creates a tremendous number of subproblems/new nodes in the branch-and-bound process and each of them corresponds to a subgraph of the original graph $G(N,A)$. Again, column generation with the shortest path problem can be used to solve the subproblem on the subgraph to linear optimality in order to get the lower bound for a further branch-and-bound process.

In set-partitioning literature, the concept of identifying columns that would not contribute to the optimal solution and thus be excluded from the optimizing process was first mentioned in 1963 by Balinski [22]. Agarwal [23] applied this in 1989 for a VRP problem based on a well known result in combinatorial optimization by Pierce in 1973, [24]; we call this general concept column elimination.

In this work, the column generation technique is used to solve the linear relaxation of the SP to its linear optimal. The generating algorithm of the column generation is a constrained shortest path problem which is solved by dynamic programming. Based on the information of the reduced costs of the SP linear relaxation and its linear optimal value and an integer optimal upper bound, a column elimination technique is developed to eliminate many non-promising columns, thus reducing the size of the SP. The reduced SP then can be solved directly. By combining column generation with column elimination, we developed a solution algorithm for NETO; furthermore, it is mathematically guaranteed that the reduced SP will yield the integer optimal solution for the original problem (cf. section 3 of chapter 4 in [6]). The set-partitioning formulation of NETO is provided in Appendix B of this report. In subsection 3.1 we present a performance comparison of our SP-CGCE algorithm and previously published results.

## COMPARISON OF PERFORMANCE BETWEEN SP-CGCE AND OTHER ALGORITHMS

The computational experiment was conducted on 99 different test problems; the problem size varied from ten pickup nodes to 120 pickup nodes, the number of feasible arcs ranged from 180 to 23,198, the feasible routes range from 16 to 32,375. Numerical results [6] indicate robust performance of the algorithm, especially the column elimination technique which generally reduces the SP problem size by an order of 2. The test results indicate an at least 100% speed increase over currently available column generation, branch-and-bound scheme; this is due to the effectiveness of the column elimination process. Additionally, in return for the sacrifice of some optimality , larger and more difficult problems can be solved several times faster. The gap between the LP bound and the integer optimum for the 99 problems tested range from 0% to 3.7% with an average of 0.1%.

As discussed earlier, most other algorithms for solving the PDPTW problem solve the LP to optimality and then utilize a branch-and-bound scheme to find the integer optimum. Since some of the subproblems on the subgraph can be almost as difficult as the original graph, solving one such subproblem might as well double the total solution time, and solving two might triple the time. This is evident in the numerical results given by Dumas[17] as recompiled here in Table 3-1.

**Table 3-1. Time Required to Solve LP and ILP to Optimality [17]**

| Problem | A19 | A30 | B30 | C20 | C30 | D40 | D50 | D55 |
|---|---|---|---|---|---|---|---|---|
| Z(LP) cpu time(sec) | 92 | 47 | 112 | 28 | 111 | 66 | 95 | 204 |
| Z(ILP) cpu time(sec) | 95 | 51 | 114 | 51 | 169 | 172 | 215 | 313 |

In the above table, Z(LP) cpu time is the time required to solve LP relaxation and to obtain the LP optimal solution Z(LP) via column generation. Z(ILP) cpu time is the time required to solve the Integer Linear Program optimal solution Z(ILP) by branch-and-bound using the LP optimal as a lower bound. The average ratio of Z(ILP) cpu time over Z(LP) cpu time is 1.56.

The *SP-CGCE* algorithm developed for NETO, however, does not use branch-and-bound to solve the problem to integer optimality after the LP optimal is obtained. It uses column elimination. From the computational results presented previously in this report and in [6], some of which are recompiled here in Table 3-2, we can see that the column elimination (*TCE*) is a fraction of the time required to solve the LP optimal (*TRts+TCG*). The only overhead involved is *Tzu*, the time required to find an upper bound of the integer problem (*ASP*), which is also a fraction of the time needed to solve the LP optimal.

**Table 3-2. Time Required to Solve LP and ILP Optimal(SP-CGCE)**

| Problem | A36 | B45 | D60 | D70 | E40 | E45 |
|---|---|---|---|---|---|---|
| Z(LP) cpu time(sec) | 257 | 1420 | 43 | 182 | 80 | 404 |
| Z(ILP) cpu time(sec) | 0 | 1 | 2 | 1 | 6 | 3 |

Z(LP) cpu time= *TRts+TCG;* (ILP) cpu time=TZu+TCE

In conclusion, the numerical experiments show that the column generation/column elimination algorithm is indeed a powerful, flexible, stable and efficient one. The column elimination procedure, in particular, is remarkably efficient. For more details on performance evaluation and theoretical issues such as network reduction (time-window tightening, arc elimination) and SP-CGCE algorithm development as well as an up-to-date literature review of PDPTW and related subjects, we refer the reader to [6].

# 4. NETO SYSTEM IMPLEMENTATION AND DEMONSTRATION

The first part of this section discusses the implementation of NETO, and the second part provides an example for solution of the backward problem.

## 4.1 SYSTEM IMPLEMENTATION

To test the SP-CGCE algorithm and demonstrate the new model *NETO*, a prototype system is implemented on a SunSparc Server 670MP workstation in a total 8181 lines of source code in C. The Linear Programming and Integer Programming solver for the *SP-CGCE* algorithm is built upon Cplex 3.0 Callable Library.® The user interface is implemented using Xt Tool Kit Intrinsics and Xlib. Technically speaking, the GUI interface class hierarchy based on object oriented programming is as in Figure 4.1.

In Figure 4.1 the "inputButton" controls the user input interface; the "outputButton" takes care of the output of optimization results and statistics; the "generateButton" generates a test problem; the "optimizeButton" activates the column generation column elimination program to solve the problem; the mapBox Widget Class allows a programmer to draw geographic maps in an X window. It is designed to give the application programmer the ability to work entirely in world (latitude, longitude) coordinates and frees him/her from thinking about the projection, scale, and display of the data. It has a 'zoom box' built in. The user can drag out a zoom box with the first mouse button (changeable through added translations). He/she can then zoom by clicking the first button within the zoom box, or cancel it by clicking outside its boundaries.

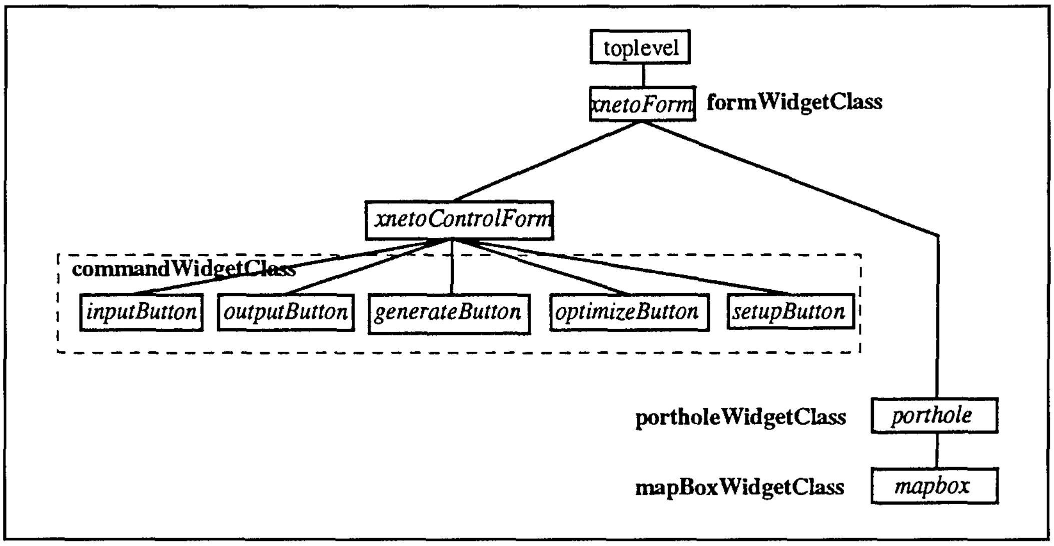



**Figure 4.1 X GUI Interface Class Hierarchy**

## 4.2 SYSTEM DEMONSTRATION

An example of the optimal requirement studies (the backward problem) is given in this sub-section. Five additional examples are given in [6] as demonstrations of the NETO and the SP-CGCE algorithm. In [6], the first example is a backward requirement analysis type problem; the second is a forward capability analysis-type problem; the third is a full-load problem; the fourth is a multidepot problem; and the last is a split type of problem.

### 4.2.1 EXAMPLE: OPTIMAL REQUIREMENTS STUDIES (BACKWARD PROBLEM)

In this example the problem is to find how many aircraft are necessary to move cargoes to the specified destination, while satisfying the closure schedule specified by the TPFDD. An illustration is provided of how the system and the algorithm function.

#### 4.2.1.1 INPUT AND PREPROCESSING

Raw input information stored in the database system will first be preprocessed by the tasks listed previously in this report, such as by merging files and aggregating records into a correct and efficient form. Here we start from a regular and simplified TPFDD format and proceed to the optimization process described below.

#### 4.2.1.2 OPTIMIZATION

*Operations Network:* The operations network constitutes the original mobility operation information represented in a form of transportation network consisting of all relevant data, such as air bases, seaports, air routes, sea routes, onloads, offloads, enroutes, cargoes, transportation vehicles, weather, scenarios, movement requirements, logistics factors, etc. The operations network in this example is outlined in terms of cargoes, transportation resource and operation scenarios:

*Cargo:* Cargo information described by the simplified TPFDD as movement requirement is specified in Table 4-1. For a graphical representation of the TPFDD, please refer to Figure 4-2.

**Table 4-1 TPFDD for the Backward Problem**

| APOE | APOD | EAD(min) | LAD(min) | TONNAGE |
|---|---|---|---|---|
| ALLEGHENY CO | MYRTLE BEACH AFB | 303 | 1685 | 256 |
| HAWTHORNE MUNI | GADSDEN MUNI AFB | 144 | 1825 | 267 |
| DULUTH INT | MYRTLE BEACH AFB | 636 | 2851 | 138 |

*Transportation Resource:*

- Aircraft Capacity: 500 tons
- Aircraft Speed: 120 mph.
- Aircraft Berth:
  - Starting Depot: SAN FRANCISCO INTL.
  - Returning Depot: SAN FRANCISCO INTL.

For simplicity, cargo loading/unloading time is converted into vehicle speed thus the service time s is set to zero. Aircraft capacities and cargoes are modeled here with only one dimension (weight); other dimensions such as load size and passenger/cargo type can also be incorporated.

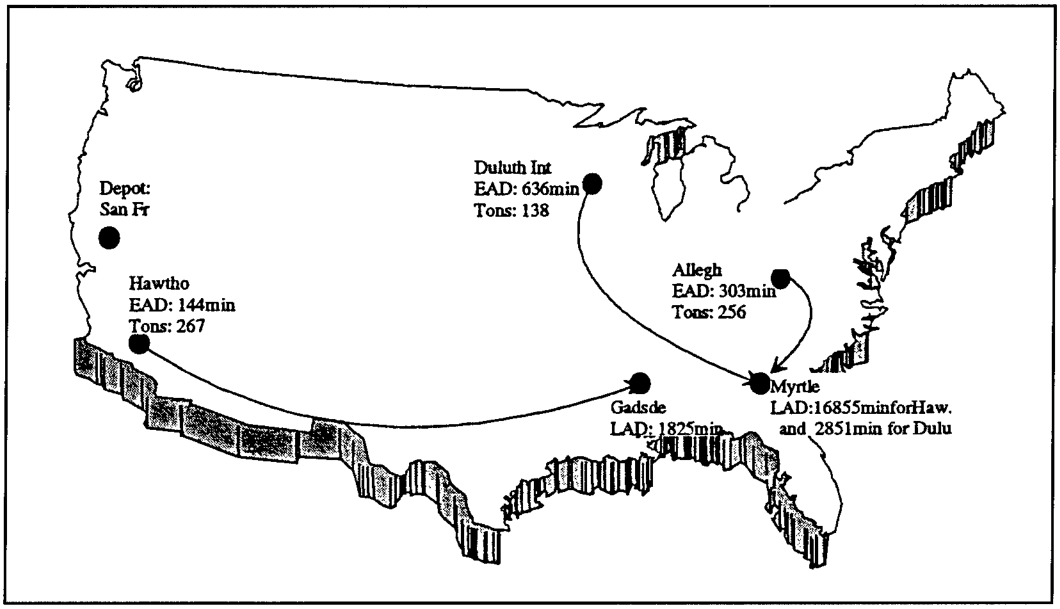



**Figure 4.2 Graphical Representation of the TPFDD in the Backward Problem**

Table 4-2 Operations Network Information

a. Distance Among Bases: Great Circle Distance

| AIR BASES | SAN FRANC ISCO INT | ALLEGHE NYCO | HAWTHOR NE MUNI | DULUT H INT | MYRTLE BEACH AFB | GADSDEN MUNI AFB |
|---|---|---|---|---|---|---|
| SAN FRANCISCO INT | 0 | 2282 | 347 | 1657 | 2451 | 2038 |
| ALLEGHENYCO | 2282 | 0 | 2165 | 764 | 473 | 581 |
| HAWTHORNE MUNI | 347 | 2165 | 0 | 1632 | 2269 | 1842 |
| DULUTH INT | 1657 | 764 | 1632 | 0 | 1151 | 947 |
| MYRTLE BEACH AFB | 2451 | 473 | 2269 | 1151 | 0 | 431 |
| GADSDEN MUNI AFB | 2038 | 581 | 1842 | 947 | 431 | 0 |

b. Flying Time Among Bases: Using t[i,j]= GCD[i,j]/AifcraftSpeed

| AIR BASES | SAN FRANC ISCO INT | ALLEGHE NYCO | HAWTHOR NE MUNI | DULUT H INT | MYRTLE BEACH AFB | GADSDEN MUNI AFB |
|---|---|---|---|---|---|---|
| SAN FRANCISCO INT | 0 | 1141 | 173 | 828 | 1270 | 1019 |
| ALLEGHENYCO | 1141 | 0 | 1082 | 382 | 236 | 290 |
| HAWTHORNE MUNI | 173 | 1082 | 0 | 816 | 1134 | 921 |
| DULUTH INT | 828 | 382 | 816 | 0 | 575 | 473 |
| MYRTLE BEACH AFB | 1270 | NA | 1134 | 575 | 0 | 215 |
| GADSDEN MUNI AFB | 1019 | 290 | 921 | 473 | 215 | 0 |

*Operation Scenarios:*

- Operation Time: Starting at 00.00 hr, ending at 48:00 hr (i.e. Time Duration = 48.00 hrs)
- Infrastructure:
  - Availability of Aircraft: to be decided optimally
  - Transportation Network /Operations Network: Table 4-2.

**Network Construction & Optimization Network:** The network construction transforms the above operations network into the optimization network , taking into consideration factors such as routes, enroutes, cargoes, fuel and other information in association with the operations network. TPFDD and transportation network information is transformed into the labeled digraph $G(N,A)$. After the network construction process, the graph and labels information are shown in Table 4-3.

**Table 4-3. Optimization Network Information**

a. Node $N$, *labels* a[i], b[i] and d[i] in $G(N,\Lambda)$

| Node N | 0 | 1 | 2 | 3 | -1 | -2 | -3 | 7 |
|---|---|---|---|---|---|---|---|---|
| Air Base | SAN FR | ALLEGH | HAWTHO | DULUTH | MYRTLE | GADSDE | MYRTLE | SAN FR |
| a[i] | 0 | 303 | 144 | 636 | 0 | 0 | 0 | 0 |
| b[i] | 2880 | 2880 | 2880 | 2880 | 1685 | 1825 | 2851 | 2880 |
| d[i] | 0 | 256 | 267 | 138 | -256 | -267 | -138 | 0 |

b. Cost $c_{ij}$ for $G(N,\Lambda)$

| Node | 0 | 1 | 2 | 3 | -1 | -2 | -3 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 3722 | 1787 | 3097 | NA | NA | NA | NA |
| 1 | NA | 0 | 2165 | 764 | 473 | 581 | 473 | NA |
| 2 | NA | 2165 | 0 | 1632 | 2269 | 1842 | 2269 | NA |
| 3 | NA | 764 | 1632 | 0 | 1151 | 947 | 1151 | NA |
| -1 | NA | NA | 2269 | 1151 | 0 | 431 | 0 | 2451 |
| -2 | NA | 581 | NA | 947 | 431 | 0 | 431 | 2038 |
| -3 | NA | 473 | 2269 | NA | 0 | 431 | 0 | 2451 |
| 7 | NA | NA | NA | NA | NA | NA | NA | 0 |

c: Flying Time $t_{ij}$ for $G(N,\Lambda)$

| Node | 0 | 1 | 2 | 3 | -1 | -2 | -3 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1141 | 173 | 828 | NA | NA | NA | NA |
| 1 | NA | 0 | 1082 | 382 | 236 | 290 | 236 | NA |
| 2 | NA | 1082 | 0 | 816 | 1134 | 921 | 1134 | NA |
| 3 | NA | 382 | 816 | 0 | 575 | 473 | 575 | NA |
| -1 | NA | NA | 1134 | 575 | 0 | 215 | 0 | 1225 |
| -2 | NA | 290 | NA | 473 | 215 | 0 | 215 | 1019 |
| -3 | NA | 236 | 1134 | NA | 0 | 215 | 0 | 1225 |
| 7 | NA | NA | NA | NA | NA | NA | NA | 0 |

Please note that $-i$ is equivalent to $n+i$. So either $(i,n+i)$ or $(i,-i)$ denote the same pickup-delivery pair. NA means not applicable.

To construct the optimization network, two tasks arise: building up the digraph topology $G$, and computing the labels. For simplicity, we take the direct physical route $(i,j)$ as the arc $(i,j)$ of $G(N,\Lambda)$. With the arcs available, other parts of the network can be built very easily. In particular, the cost of arc $(i,j)$ is defined as

$$c_{ij} = \begin{cases} GCD(i,j), & \text{if } i \neq 0 \\ K + GCD(i,j), & \text{if } i = 0 \end{cases}$$

**, and $K$=1441, here $K$ represents the "fixed cost" of** utilizing a vehicle. In reality, arc construction is a complicated procedure which can be done in various ways according to the actual operational situation. Information concerning crew scheduling, traffic congestion, aircraft mechanical limitations, weather situation, closed air bases, hostile regions and so forth might all be included in the optimization network construction process. For example, to construct the arc from a pickup node $i \in P^+$ to a delivery node $n+i \in P^-$, we may select the shortest path $P$ with the maximum length of any segment in path $P$ not exceeding certain number in the operations network, which may mean that a certain type of aircraft can make a sustained flight with supported available refueling along the route.

**Network Reduction & Reduced Optimization Network:** The above optimization network can be further reconfigured by tightening some time windows and some inadmissible arcs through the process of network reduction, which reduces the size of the problem. There are nine rules for time window tightening and inadmissible arc elimination (cf. Chapter 4 of [6]); these rules identify infeasible/inadmissible arcs and reduce the network size. Table 4-4 shows the result of network reduction.

**Table 4-4 Reduced Optimization Network Information**

a: Node, labels a[i], b[i] and d[i]

| Node | 0 | 1 | 2 | 3 | -1 | -2 | -3 | 7 |
|---|---|---|---|---|---|---|---|---|
| Air Base | SAN FR | ALLEGH | HAWTHO | DULUTH | MYRTLE | GADSDE | MYRTLE | SAN FR |
| a[i] | 0 | 1141 | 173 | 828 | 1377 | 1094 | 1403 | 0 |
| b[i] | 2880 | 1419 | 904 | 1080 | 1655 | 1825 | 1655 | 2880 |
| d[i] | 0 | 256 | 267 | 138 | -256 | -267 | -138 | 0 |

b. Cost $c_{ij}$ for $G(N,\Lambda)$

| Node | S | 1 | 2 | 3 | -1 | -2 | -3 | T |
|---|---|---|---|---|---|---|---|---|
| S | 0 | 3722 | 1787 | 3097 | NA | NA | NA | NA |
| 1 | NA | 0 | - | - | 473 | 581 | 473 | NA |
| 2 | NA | 2165 | 0 | 1632 | - | 1842 | - | NA |
| 3 | NA | 764 | - | 0 | - | 947 | 1151 | NA |
| -1 | NA | NA | - | - | 0 | 431 | 0 | 2451 |
| -2 | NA | 581 | NA | - | 431 | 0 | 431 | 2038 |
| -3 | NA | - | - | NA | 0 | 431 | 0 | 2451 |
| T | NA | NA | NA | NA | NA | NA | NA | 0 |

Note: "-" entry in above table means the arc is eliminated.

Comparing 4-3(a) with 4-4(a), it can be seen that 6 out of 8 time windows are tightened, e.g. the original time-window in Table 4-3(a) for node 1 was [303, 2880], in the reduced version it has been tightened to [1141, 1419]; Comparing 4-3(b) with 4-4(b), it can be seen that 11 out of 33 arcs in the original optimization network are eliminated, e.g. in the entry for the arc connecting node 1 to node 2 has been eliminated. These window tightening and arc reductions result in a reduced network with less computational complexity.

## PDPTW & SP-CGCE OPTIMIZATION:

With $G(N,\Lambda)$ available, the *SP* formulation for the PDPTW problem can be carried out as discussed in Appendix B. The *SP* formulation is an implicit one, because it offers the structure, but does not explicitly express the parameter values. These values, such as the cost coefficients and columns, will be generated along with the solution of the formulation. The Column Generation part of the SP-CGCE algorithm solves the LP relaxation of the SP formulation to LP optimal after generating 4 columns (Table 4-5). During the first iteration of column generation, the column which is generated corresponds to the feasible route of (0,3,1,-3,-1,7) (also see Table 4-3):

1) the vehicle leaves its home base at node 0 (referring to SAN FR),
2) picks up cargo at node 3 (referring to DULUTH, picking up 138 tons there),
3) the vehicle then goes to node 1 (referring to ALLEGH, picking up 256 tons),
4) the vehicle delivers the cargo from DULUTH (labeled node "3") to node "-3" which is the drop-off at MYRTLE,

5) the vehicle then delivers the cargo from ALLEGH (labeled node "1") to node "-1" which here also refers to MYRTLE.

The reduced cost corresponding to this column/route is -6560. By column generation technique, we know it is the minimum reduced cost among all other columns which are not in the base of the simplex algorithm. Since this is a negative value, the LP solution is not yet an optimum one; therefore, more column generations are needed. As shown in Table 4-5, after the 4th column generation iteration, the optimal solution is obtained.

**Table 4-5 Column Generation Process**

| Column Generations | min Reduced Cost | Corresponding Route | LP Optimal? | Add This Column? |
|---|---|---|---|---|
| 1st | -6560 | (0,3,1,-3,-1,7) | N | Y |
| 2nd | -5179 | (0,2,-2,1,-1,7) | N | Y |
| 3rd | -3946 | (0,2,3,-3,-2,7) | N | Y |
| 4th | 0 | All 7 Feasible Routes | Y | N |

The Column Elimination part of the algorithm solves the SP problem to integer optimal using 5 out 7 total feasible columns/routes (Table 4-6), i.e. two columns are eliminated.

**Table 4-6 Column Elimination Process**

| Route | Reduced Cost | Zupper -Zlower | Eliminated? | The Path |
|---|---|---|---|---|
| 1 | 3206 | 1973 | Y | 0(T:0), 1(T:1141), -1(T:1377), 7(T:26 02) |
| 2 | 1973 | 1973 | N | 0(T:0), 2(T:173), -2(T:1094), 7(T:211 3) |
| 3 | 3354 | 1973 | Y | 0(T:0), 3(T:828), -3(T:1403), 7(T:262 8) |
| 4 | 50000000 | 1973 | N* | 0(T:0), 2(T:173), -2(T:1094), 1(T :1384), -1(T:1620), 7(T:2845) |
| 5 | 0 | 1973 | N | 0(T:0), 3(T:828), 1(T:1210), -3(T:1446,-1(T:1446), 7(T:2671) |
| 6 | 0 | 1973 | N | 0(T:0), 2(T:173), 3(T:989), -3(T:1564), -2(T:1779), 7(T:2798) |
| 7 | 0 | 1973 | N | 0(T:0), 3(T:828), 1(T:1210), -1(T:1446), -3 (T:1446), 7(T:2671) |
| 8 | not PD PTW feasible | | | 0(T:0), 2(T:173), 3(T:989), 1(T:1371), -1(T:1607), -3(T:1607), -2(T:1822), 7(T:2841) |
| 9 | | | | 0(T:0), 2(T:173), 1(T:1255), -1(T:1491), -2(T:1706), 7(T:2725) |
| 10 | | | | 0(T:0), 2(T:173), 3(T:989), 1(T:1371), -3(T:1607), -1(T:1607), -2(T:1822), 7(T:2841) |

*The 50000000 value is an internal flag of the implementation for code optimization

The optimization statistics and optimal solution are shown in Table 4-7, from which we know that the minimum number of vehicles used is 2 and the optimal routes and schedule are 0(T: 0), 2(T:173), -2(T:1094), 7(T:2113) and 0(T: 0), 3(T: 828), 1(T:1210), -1(T:1446), -3(T:1446), 7(T:2671).

Table 4-7 Optimization Results

a. Optimization Statistics

| Optimal Values | Vehicles Needed | Pickup-Delivery Pair | Nodes | Arcs | F_Arcs |
|---|---|---|---|---|---|
| 12452 | 2 | 3 | 7 | 33 | 22 |
| FRts | ColumGens | LPOptimal | Zupper | RtsElimd | Ttotal |
| 7 | 4 | 10479 | 12452 | 2 | 0 |

b. Optimal Routing & Scheduling

| Binary Variable | Route Cost | Routing & Scheduling Information Format: Node(T:Arrival/Departure Time) |
|---|---|---|
| x1 | 5667 | 0(T: 0), 2(T:173), -2(T:1094), 7(T:2113) |
| x6 | 6785 | 0(T: 0), 3(T: 828), 1(T:1210), -1(T:1446), -3(T:1446), 7(T:2671) |

## 4.2.1.3 POSTPROCESSING AND OUTPUT

Depending on the situation, various postprocessings could be done and user-friendly output could be generated. Here, we will give the GUI display of the optimal routes in Figure 4.3, the throughput in Figure 4.4, and the Vehicle-in-Use Information in Figure 4.5.

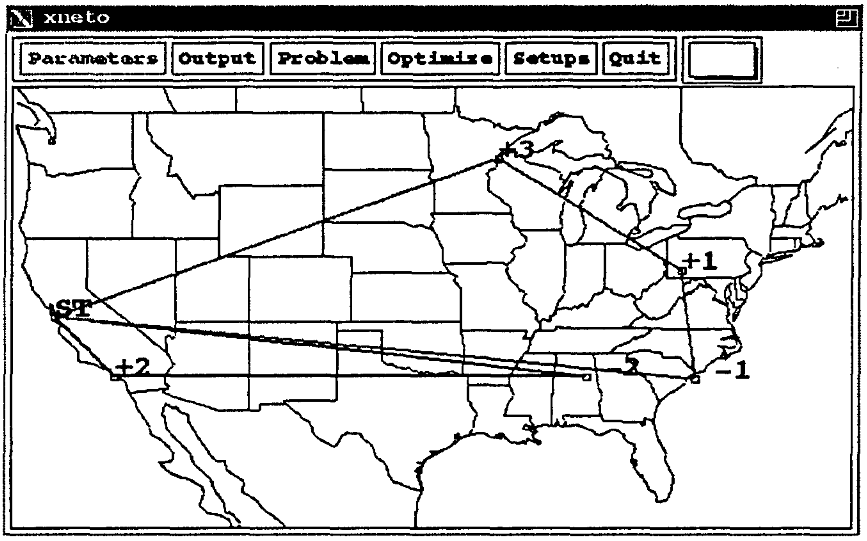



Figure 4.3 Optimal Routing in the Requirements Studies Example

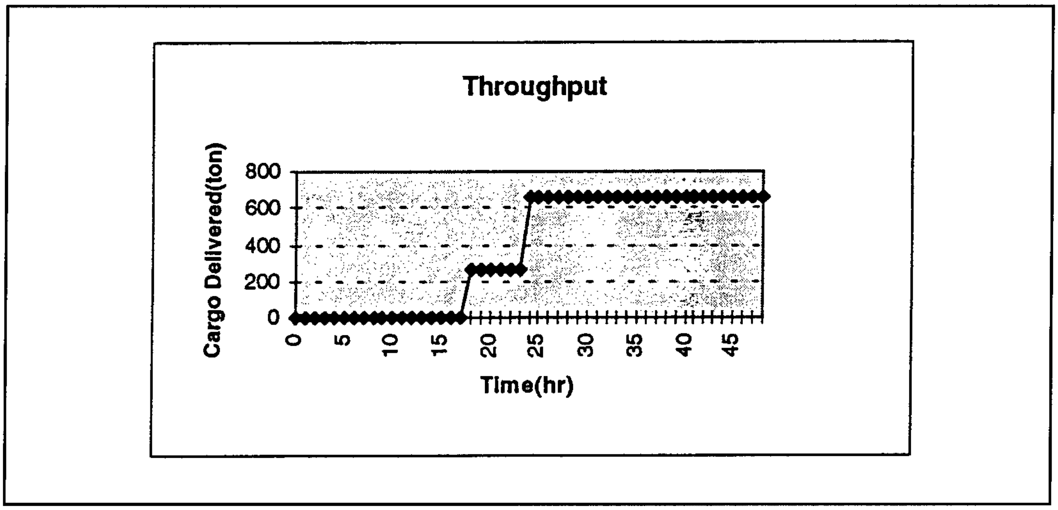



Figure 4.4 Throughput

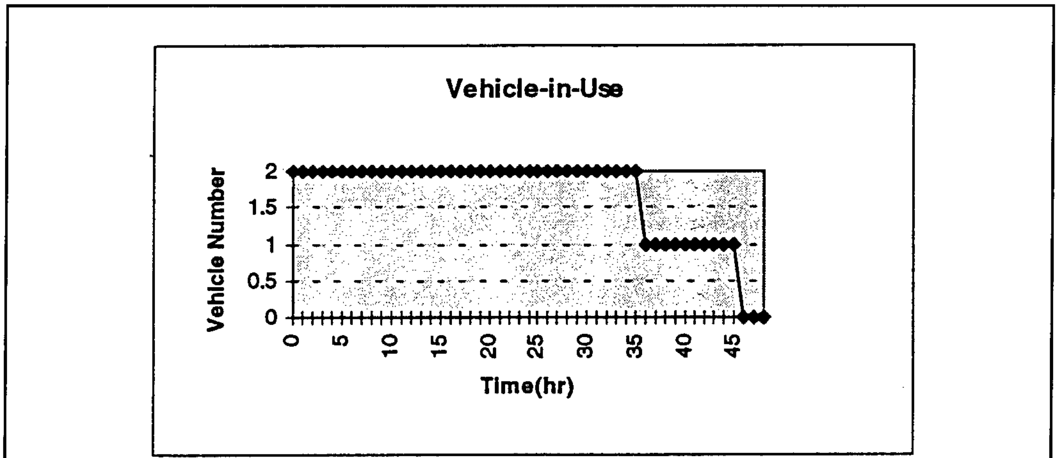



Figure 4.4 Vehicles-in-Use

**Figure 4.5 Vehicles-in-Use**

## 5. DISCUSSION OF RELATED ISSUES AND FUTURE WORK

Although the focus of this report has been our solution to the requirement studies problem, the above implemented model and solution scheme is powerful, flexible and extendible in dealing with many other real world issues. Here we mention, with some modification, those problems addressed by this approach and additional issues for further work.

1) *Vehicle Numbers:* Consideration of the number of vehicles is easily incorporated in this approach. For the minimum number of vehicles problem, what is needed is to take.

$c_{ij} = \begin{cases} K & if\, i = 0 \\ 0 & if\, i \neq 0 \end{cases}$. **For problems that are concerned with an exact number m of vehicles, the constraint:** $\sum_{r\in\Omega} x_r = m$ **must be included. For problems with a maximum number m vehicles the constraint:** $\sum_{r\in\Omega} x_r \leq m$ **must be included.**

2) *Multi-depot, Nonhomogenous-vehicle:* This approach can easily be extended to a multi-depot, nonhomogenous-vehicle situation in which the feasible routes would be obtained by applying the *Constrained Shortest Path* Algorithm to different depots and types of vehicles. The computation time/complexity increases linearly.

3) *General VRP Problems:* This scheme can solve general *VRP* problems, pickup only problem, delivery only problem, split, full load, with or without time windows. The only major modification is the constrained shortest path problem.

4) *Forward Problem:* The approach can also address the aforementioned forward problem. The forward problem can be formulated similarly by incorporating the penalty term $T_{n+i}$-$a_i$, $i \in P^+$ into the cost coefficient $c_r$.

5) *Soft Time Windows:* The approach can also address the so called soft time window problem. The penalties will be incorporated into the cost coefficient $c_r$ to include a route with violated time windows as a feasible route.

6) *Larger Problems:* The algorithm can be tailored to solve even larger problems by sacrificing optimality and settling for a sub-optimal solution. One way of doing this is to first divide the original problem into subproblems using the concept of clustering, and then use this scheme to solve each subproblem to optimality. It is also possible to not require optimality in the constrained shortest path and column elimination algorithm. From another point of view, a semantic control paradigm can be employed to deal with larger problems in which the higher and intelligent layer of the system will identify the problem situation and transfer control accordingly to the lower and actuating layer of the system which in our case would be the PDPTW algorithm.

7) *Full Load and Split Problem:* By adding the full load requirement into the constrained shortest path problem, the algorithm solves the full load problem; by assigning different nodes to the split loads, it solves the split problem. Also, the algorithm is easily adapted to deal with regular routing problems, pickup problems, delivery problems and TSP problems. Of course, it performs better with problems which are more tightly constrained.

The objective function is flexible, i.e. various objective functions can be included. As mentioned above, an objective function to address the forward problem, the backward problem, and soft time windows can be included in the scheme. Other factors, such as travel distance, travel time, vehicle utilization considerations etc., can also be easily incorporated.

The mobility system is a large-scale and complicated system; in order to address more realistic and larger problems in mobility analysis and other large scale transportation systems, much more research is needed in addition to the work presented in this report. The following outlines several open problems that need to be addressed:

1) *Crew Scheduling:* There are regulations/constraints on the working hours of crew members. The crew scheduling issues can also be incorporated into the scheme. One way is to do (vehicle) routing first and (crew) scheduling second, in which case crew scheduling will take place after the vehicle routes are settled. Crew scheduling could also be done along with feasible route generation, in which case each feasible route needs to meet crew scheduling constraints.

2) *More efficient parallel algorithm development for the constrained SPP problem:* The column generation-column elimination algorithm is efficient in finding the integer optimal when the relaxed LP optimal is achieved. Unfortunately, solving the constrained SPP problem for the column generation process to obtain the LP optimal is very time-consuming and computer memory-intensive. It is the major bottleneck of the algorithm; therefore a more efficient algorithm for the constrained SPP problem is desired. In recognition of the ever-increasing use of parallel computing, parallel algorithm development may be a worthwhile pursuit. The use of parallel algorithms will

be achievable in the near future since the process for solving the constrained SPP problem is based on dynamic programming and has very strong parallelism.

3) *Vehicle Concurrence Issues:* Vehicles might compete for common resources such as routes, crew members, air fields etc., which could affect operations. The effects of one vehicle on another make the system a time-dependent and vehicle-dependent dynamic system; these issues are not considered here. The set-partitioning formulation, column-generation solution approach might be inherently weak in modeling these factors. One possible way of handling this is by a route-first and concurrence-check-second approach. Since the optimal solution is often not unique, the check can be first conducted for all optimal solutions. If none satisfies the concurrence check criteria, certain modifications need to be performed.

4) *Nonlinear Loading Algorithm:* In this report, a multidimensional linear loading algorithm is used (constraint B-5 for load progression in the formulation given in Appendix B). In more complicated cases, nonlinear loading may be involved, and issues concerning nonlinear loading algorithms coupled with the optimality analysis should be explored. Generally, any loading algorithm could replace the existing loading in the NETO as long as it can be incorporated in the constrained shortest path problem.

5) *Dynamic Routing and Scheduling:* The situation studied in this paper is basically a static routing and scheduling problem with time windows and capacity constraints. The movement requirement is known in advance. In some situations, the movement requirement is dynamic and the routing and scheduling should be performed continuously.

6) *Probabilistic Considerations:* In this report we have assumed that the parameters of the operation, including the network, resources, etc. are all deterministic. But in real situations, various uncertainties could be involved in many aspects of the problem; therefore probabilistic studies are useful in addressing more realistic scenarios.

# 6. CONCLUSION

This report is based on a doctoral dissertation by the first author [6]; it is the first attempt to use column generation-column elimination scheme to solve VRP problems in general and VRPTW and PDPTW problems in particular; it is also the first attempt to model and solve the mobility analysis system problems using network optimization with time window constrained routing and scheduling.

The new model (NETO) not only offers optimal solutions but also solves both the forward problem and the backward problem. Above all, it is flexible and can be extended to include many additional practical and operational constraints and considerations. The SP-CGCE algorithm is an efficient and competitive approach to solve practical vehicle routing and scheduling problems. The computational results presented briefly in section 3.3 of this report indicate robust performance for the algorithm. All these characteristics discussed above make NETO, with the SP-CGCE algorithm powerful, flexible and practical.

In summary a new mobility analysis model named NETO [6] is proposed to address various limitations of the existing ones. The new model consists of a network optimization engine with time window constrained routing and scheduling that is based on integer and combinatorial optimization methodology, and an analysis system with a management information system built upon RDBMS and multimedia technology. It is our belief that NETO with the SP-CGCE algorithm can be, should be and will be utilized to solve practical mobility analysis problems as well as other transportation system related problems.

## ACKNOWLEDGMENTS

The authors wish to express their gratitude to the AFOSR for funding this research, to anonymous referees for their insightful suggestions, to Lt. Col. Anthony Waisanen of AMC for many fruitful discussions and his knowledge of military transportation problems and to Col. Craig M. Northrup of AMC for his support of and interest in this work.

## APPENDIX A: NOTATIONS AND DEFINITIONS

| | |
|---|---|
| AMC: | Air Mobility Command |
| APOE: | Aerial Port of Embarkation |
| APOD: | Aerial Port of Debarkation |
| CINC: | Commander-in-chief |
| COA: | Courses of Action |
| CONUS: | Continental United States |
| EAD: | Earliest Available Date |
| FLOGEN: | Flow Generator |
| GCD: | Great Circle Distance |
| GUI: | Graphical User Interface |
| IP: | Integer Programming |
| LAD: | Latest Arrival Date |
| LP: | Linear Programming |
| MASS: | Mobility Analysis Support System |
| MASS: | Mobility Analysis Simulation System |
| MIDAS: | Model for Intertheater Deployment By Air and Sea |
| MIP: | Mixed Integer Programming |
| MIS | Management Information System |
| MSC: | Military Sealift Command |
| NETO: | Network Optimization Mobility Analysis System |
| OSD: | Office of the Secretary of Defense |
| PDPTW: | Pickup and Delivery Problem with Time Window Constraint |
| RDBMS | Relational Database Management System |
| RIMS: | Revised Intertheater Mobility Study |
| SP-CGCE: | Set-partitioning Formulation , Column Generation Column Elimination |
| SPP: | Shortest Path Problem |
| TPFDD: | Time Phased Force Deployment Data |
| TSP: | Traveling Salesman Problem |
| VRP: | Vehicle Routing and Scheduling Problem |
| VRPTW: | Vehicle Routing and Scheduling Problem with Time Window Constraint |
| $R^n_+$: | $n$ dimensional real vector space |

| | |
|---|---|
| $Z_+^n$: | a set of non-negative integral $n$-dimensional vector space |
| $B_+^n$: | a set of non-negative binary $n$-dimensional vector space |
| $P^+$: | pickup node set, $P^+ = \{1, 2, \ldots, n\}$. The corresponding delivery node to $i \in P^+$ is $n+i$, also referred to as $-i$ |
| $P^-$: | delivery node set, $P^- = \{n+1, n+2, \ldots, 2n\} = \{-1, -2, \ldots -n\}$ |
| $P$: | operation node set, $P \equiv P^+ \cup P^-$. $P$ includes all pickup and delivery nodes |
| $S$: | Starting node set from which vehicles departure. For single depot case, $S = \{0\}$. $S$ is also used for the space $\{0,1\}^\Omega$ |
| $T$: | Terminating node set to which vehicles return. For single depot case $T = \{2n+1\}$ |
| $N$: | all nodes of the optimization network, $N = S \cup P \cup T = \{0, 1, \ldots, n, n+1, \ldots, 2n, 2N=1\}$ |
| $\Lambda$: | all arcs of the optimization network, $\Lambda = S \times P^+ \cup P^+ \times P^- \cup P^- \times T$ |
| $G(N, \Lambda)$: | the original underlying graph of the *optimization network* |
| $\rho_r$: | a feasible route in $G(N, \Lambda)$ |
| $\Omega$: | or $\Omega(N, \Lambda)$, the set of all feasible routes in $G(N, \Lambda)$. $\Omega = \{\rho_r\}$ |
| $\|\Omega\|$: | the cardinality of $\Omega$ |
| $X_B$: | the basic variables in the simplex method |
| $X_N$: | the non-basic variables |
| $c_B$: | the cost coefficient corresponding to $X_B$ |
| $c_N$: | the coefficient corresponding to $X_N$ |
| $B$: | the basis in the simplex method |
| $N$: | the non-basic columns. |
| $\Omega_B$: | i.e. $\{\rho_r : \delta_r \in B\}$, the set of the feasible routes that correspond to the columns in the feasible base $B$ |
| $\delta_r$: | column coefficient of the set partitioning formulation, corresponding to feasible route $\rho_r$, where $\delta_{ir} = \begin{cases} 0 & \text{if node } i \text{ is not on route } r \\ 1 & \text{if node } i \text{ is on route } r \end{cases}$ $i \in P^+,\ r \in \{ s \mid \rho_s \in \Omega\}$ |
| $x_r$: | $x_r = \begin{cases} 0 & \text{if feasible route } r \text{ is not selected in the solution} \\ 1 & \text{if feasible route } r \text{ is selected in the solution} \end{cases}$ $r = \{r \mid \rho_r \in \Omega\}$ binary variable |

| | |
|---|---|
| $SP$: | the original master problem with set partition formulation |
| $RSP$: | the linear relaxation of $SP$ |
| $ASP$: | the augmented set partition master problem |
| $RASP$: | the linear relaxation of $ASP$ |
| $\tilde{x}_r^R$: | the optimal solution variable for $RSP$ or $RASP$ |
| $\tilde{x}_i$: | a feasible solution for $SP$ (integer solution) |
| $x_{ij}$: | $x_{ij}$ where $(i,j) \in \Lambda$ is the vehicle flow variables of the feasible route $\rho_r \in \Omega$. |

$$x_{ij} = \begin{cases} 1 & \text{if feasible route } \rho_r \text{ goes directly from } i \text{ to } j \\ 0 & \text{if feasible route } \rho_r \text{ does not go directly form } i \text{ to } j \end{cases}$$

| | |
|---|---|
| $c_{ij}$: | the cost of arc $(i,j)$ |
| $\tilde{c}_{ij}$: | artificial cost of arc $(i,j)$ for the shortest path problem |
| $c_r$: | the cost of route $\rho_r$: $c_r = \sum_{(i,j)\in\Lambda} c_{ij} x_{ij}$ |
| $\bar{c}_r$: | reduced cost $\bar{c}_r = c_r - \pi\delta_r$, where $\pi$ is the dual variables vector/simplex multiplier |
| $S$: | the original problem space, $S = \{0,1\}^{\lvert\Omega\rvert}$. $S$ is also used for the starting nodes |
| $S_R$: | i.e. $S_{RSP}$, or $S_{RASP}$, the problem space of $RSP$ or $RASP$, $S_R \in R_+^{\lvert\Omega\rvert}$ and is the linear relaxation of S |
| $\bar{d}_i$: | load vector(volume, weight...) of cargo $i$ at node $i$ |
| $[a_i, b_i]$: | pickup time window at node i for movement/cargo $i$ |
| $[a_0, b_0]$: | time window for vehicle leaving the depot $S$ |
| $[a_{2n+1}, b_{2n+1}]$: | time window for vehicle returning to the depot $T$ |
| $\bar{D}$: | capacity of vehicle(load weight limit, volume,...) |
| $t_{ij}$: | travel time from node $i \in N$ to node $j \in N$ |
| $s_i$: | service time(pickup time or delivery time) at node $i \in N$ |
| $\bar{Y}_i$: | the total load on the vehicle just after it leaves node $i \in N$ |
| $T_i$: | time of start service at node $i \in N$ |
| $T_0$: | arrival time at node $i$ or time vehicle leaves the depot $S$ |
| $T_{2n+1}$: | time vehicle returns to the depot $T$ |
| $RT$: | feasible route defining formulation |

| | |
|---|---|
| *RTL:* | columns to be generated each time for the column generation process |
| $\delta^-(j)$: | the set of nodes that are connected to node $j$ |
| $\rho^k_\alpha(j)$: | the $\alpha^{th}$ route in all routes that starts at $S$ and ends at $j$ with $k$ arcs |
| $P^k(j)$: | the set of all routes that start at $S$ and end at $j$ with $k$ arcs, i.e. $P^k(j) = \bigcup_\alpha \rho^k_\alpha(j)$ |
| $h^k_\alpha(j)$: | the cost of route $\rho^k_\alpha(j)$ |
| $T^k_\alpha(j)$: | arrival time at node $j$ of route $\rho^k_\alpha(j)$ |
| $\overline{Y}^k_\alpha(j)$: | vehicle load at node $j$ along route $\rho^k_\alpha(j)$ |

# APPENDIX B: SET PARTITIONING FORMULATION OF THE PDPTW

*B.1 Route Defining Formulation*

The PDPTW problem is formulated as a set-partitioning model; the formulation is based on the concept of feasible routes:

A ***feasible route*** $\rho_r$ in $G(N,\Lambda)$ is a non-cyclic path that originates from $S$ and terminates at $T$, while satisfying pairing constraints, precedence constraints, capacity constraints and time window constraints. Introducing binary route flow variable $x_{ij}$ as

$$x_{ij} = \begin{cases} 1 & \text{if the feasible route r goes directly from } i \text{ to } j \\ 0 & \text{if the feasible route r does not go directly form } i \text{ to } j \end{cases} \quad (i,j) \in \Lambda.$$

Then $\rho_r$ can be defined as follows:

$$\sum_{j \in N} x_{ij} - \sum_{j \in N} x_{j,n+i} = 0, \quad i \in P^+ \qquad \text{(B-1) (pairing constraints)}$$

$$T_i + s_i + t_{i,n+i} \le T_{n+i}, \quad i \in P^+ ) \qquad \text{(B-2) (precedence constraints)}$$

$$\text{Route:} \quad \left. \begin{array}{l} x_{ij} = 1 \Rightarrow T_i + s_i + t_{ij} \le T_j, \quad i,j \in P \\ x_{0j} = 1 \Rightarrow T_0 + t_0 \le T_j, \quad j \in P^+ \\ x_{i,2n+1} = 1 \Rightarrow T_i + s_i + t_{i,2n+1} \le T_{2n+1}, \quad i,j \in P \end{array} \right\} \quad \text{(B-3) (time progression)}$$

$$\left. \begin{array}{l} a_i \le T_i \le b_i, \ i \in P \\ a_0 \le T_0 \le b_0 \\ a_{2n+1} \le T_{2n+1} \le b_{2n+1} \end{array} \right\} \quad \text{(B-4) (time window constraints)}$$

$$\left. \begin{array}{l} x_{ij} = 1 \Rightarrow \overline{Y}_i + \overline{d}_j = \overline{Y}_j, \quad i \in P, \ j \in P^+ \\ x_{ij} = 1 \Rightarrow \overline{Y}_i - \overline{d}_j = \overline{Y}_j, \quad i \in P, \ j \in P^- \\ x_{0j} = 1 \Rightarrow \overline{Y}_0 + \overline{d}_j = \overline{Y}_j, \quad j \in P^+ \end{array} \right\} \quad \text{(B-5) (load progression)}$$

$$0 \le \overline{Y}_i \le \overline{D}, \quad i \in P^+ \qquad \text{(B-6) (capacity constraint)}$$

In the above formulation, equation (B-1) ensures that both the pickup node $i$ and its correponding delivery node $n+i$ are on the same route $\rho_r$; (B-2) ensures that on the route $\rho_r$, pickup is performed before delivery; (B-3) represents the time progression in the network, while (B-4) are time window constraints. Contraints (B-5) express the compatibility requirements between routes and vehicle loads, while constraints (B-6) are the capacity constraints.

*B-2 Set Partitioning Formulation:*
In set-partitioning formulation, the column coefficients $\delta_r$ in the constraint matrix are defined by the *feasible route* $\rho_r$ in $G(N,\Lambda)$ in the following way:

***B-2 Set Partitioning Formulation:***
In set-partitioning formulation, the column coefficients $\delta_r$ in the constraint matrix are defined by the *feasible route* $\rho_r$ in $G(N,\Lambda)$ in the following way:

$$\delta_r = [\delta_{ir}]_{n\times 1}, \text{ where } \delta_{ir} = \begin{cases} 0 & \text{if node } i \text{ is not on route } \rho_r \\ 1 & \text{if node } i \text{ is on route } \rho_r \end{cases} \quad i \in P^+.$$

Next let's introduce the binary decision variable $x_r$:

$$x_r = \begin{cases} 0 & \text{if the feasible route } \rho_r \text{ is not selected in the solution} \\ 1 & \text{if the feasible route } \rho_r \text{ is selected in the solution} \end{cases}, \quad r \in \{r \mid \rho_r \in \Omega\}$$

and the cost coefficient $c_r$ associated with $x_r$ or within the feasible route $\rho_r$. Note that $x_r$ is defined through $\rho_r$, and $\rho_r$ is defined through $x_{ij}$. Then we are ready to give the set-partitioning formulation for the *PDPTW* problem:

$$z = \min \sum_{r\in\Omega} c_r x_r$$

$$SP: \quad st. \quad \sum_{r\in\Omega} \delta_{ir} x_r = 1, \quad i \in P^+$$

$$x_r \in \{0,1\}, \quad r \in \Omega. \text{ ie. } X \in S = \{0,1\}^{|\Omega|}$$

In the above *SP* formulation, the column coefficients $\delta_r$ and the cost coefficient $c_r$ are not explicitly available. They need to be obtained through corresponding feasible route $\rho_r$, which were defined previously. A flexible objective function can be obtained by varying the exact formulation of the $c_r$'s. In particular to solve the requirement studies problem, if $c_{ij} = \begin{cases} K \text{ if } i = 0 \\ 0 \text{ if } i \neq 0 \end{cases}$ then the objective is to minimize the number of vehicles used. In general, the number of feasible routes $\rho_r$ and the number of columns $\delta_r$ in $|\Omega|$ is huge, making it computationally prohibitive to enumerate all feasible routes/columns and solve the *SP* problem to integer optimality.

## REFERENCES

[1] E.Y. Rodin, "Semantic Control Theory," Appl. Math. Lett. Vol. 1, No., 1 (1988), pp. 73-78

[2] Y. Lirov, Artificial Intelligence Methods in Decision and Control Systems, doctoral dissertation, Washington University, St. Louis, MO, (August 1987)

[3] S.M. Amin, Intelligent Prediction Methodologies in the Navigation of Autonomous Vehicles, doctoral dissertation, Washington University, St. Louis, MO, (January 1990)

[4] R.D. Weil, AI Methods in Utilizing Low Dimensional Models of Differential Games, doctoral dissertation, Washington University, St. Louis, MO, (September 1990)

[5] D. Geist, Semantic Control in Continuous Systems Applications to Aerospace Problems, doctoral dissertation, Washington University, St. Louis, MO, (December 1990)

[6] F. Yang, Network Optimization with Time Window Constrained Routing and Scheduling, doctoral dissertation, Washington University, St. Louis, MO , (Aug. 1995)

[7] J. Schank, M. Mattock,G. Sumner, I . Greenberg, J. Rothenberg, and J. P. Stucker, "A review of Strategic Mobility Models and Analysis," RAND National defense research institute, report # R-3926-JS, (1991)

[8] E.Y. Rodin, et al, "Intelligent Aircraft Routing & scheduling Project," *Center for Optimization and Semantic Control*, Washington University, St. Louis, MO, (1993)

[9] S.M. Amin, "Artificial intelligence methodologies in airlift transportation networks", *Center of Optimization and Semantic Control,* Washington University, St. Louis, MO (1992)

[10] K. Ruland,"Polyhedral solution to Pick-up and Delivery problems", doctoral dissertation, Systems Science and Mathematics, Washington University,St. Louis,MO, (August 1995)

[11] S. Arunapuram "Vehicle routing and scheduling with full loads", doctoral dissertation, *Case Western Reserve University,* (1993)

[12] M. Desrochers, J. Lenstra, M. Savelsbergh, F. Soumis, Vehicle Routing with time windows: optimization and approximation. In *Vehicle routing: Methods and Studies.* B. Golden and A. Assad(eds.). Norht-Holland, Amsterdam, 65-84, (1988)

[13] M. Desrochers, J.K. Lenstra, M.W.P. Savelsbergh. "A classification scheme for vehicle routing and scheduling problems," European. J. of Opers. Res., Vol. 46, 322-332, (1990)

[14] M. Desrochers, J. Desrosiers, M. Solomon, "A new optimization algorithm for the vehicle routing problem with time windows," *Operations Research,* Vol. 40, No. 2, (March-April 1992)

[15] J. Desrosiers, Soumis,F., and Desrochers,M., "Routing with time windows by column generation," *Networks,* Vol.14, 545-565, (1984)

[16] Y. Dumas, " Confection d'itineraire de vehicules en vue du transport de plusieurs origines a plusieurs destinations, " Publication 434, centre de techerche sur les transports, Universite de Montreal, (1985)

[17] Y. Dumas, J. Desrosiers and F. Soumis, "The pickup and delivery problem with time windows," *European Journal of Operational Research,* Vol. 54, 7-22, (1991)

[18] Laporte, "The Vehicle routing problem: An overview of exact and approximate algorithms," *European Journal of Operational /research,* Vol. 9, 345-358, (1992)

[19] C.C. Ribeiro and Soumis, F., "A Column Generation Approach to the Multiple-depot vehicle Scheduling Problem", *Operations Research* Vol 42, No. 1, (January-February 1994)

[20] M. Sol and Savelsbergh, M.W.P "A Branch-and-Price algorithm for the Pickup and Delivery Problem with time windows", Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands, (1995)

[21] M. Solomon and Desrosiers,J., "Time window constrained routing and scheduling problems," *Transportation Science,* Vol. 22, 1-13, (1988)

[22] M.L. Balinski and Quandt, R.E., "On an integer program for a delivery problem," Oper. Res. Vol. 12, 300-304, (1964)

[23] Y. Agarwal, Mathur, K., and Salkin,H.M., "A set-partitioning-based exact algorithmfor the vehicle routing problem", Networks, Vol.19, 731-749, (1989)

[24] J.F. Pierce and J. S. Lasky, J.S., "Improved combinatorial programming algorithms for a class of all zero-one integer programming problems," Management Sci., Vol. 19, No. 5, 528-543, (1973)


## ENDNOTES

1 Submitted August 1995; In final form November 1995

2 This work was supported in part by AFOSR under grant number 890158.

3 A complete list of terminology, definitions, notation, and symbols is given in the Appendix A at the end of this report.

4 The networks were trained and validated on a declassified TPFDD file from Operation Just Cause in Panama.

# MILITARY OPERATIONS RESEARCH SOCIETY BOARD OF DIRECTORS