View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Calhoun, Institutional Archive of the Naval Postgraduate School

Calhoun
Institutional Archive of the Naval Postgraduate School

Calhoun: The NPS Institutional Archive

Faculty and Researcher Publications Faculty and Researcher Publications

1997

Application and Extension of the Thruput II Optimization Model for Airlift Mobility

Rosenthal, Richard E.

Military Operations Research

Military Operations Research, V3 N2 1997

DUDLEY KNOX LIBRARY

Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

Application and Extension of the Thruput II Optimization Model for Airlift Mobility

Richard E. Rosenthal
Steven F. Baker
Lim Teo Weng
David F. Fuller
David Goggins
Ayhan O. Toy
Yasin Turker

Operations Research Department
Naval Postgraduate School
1411 Cunningham Rd.
Monterey, California 93943-5219

David Horton
Daniel Briand

U.S. Air Force Studies and Analyses Agency

David P. Morton

Graduate Program in Operations Research
University of Texas at Austin

APPLICATION AREA:
Mobility & Transport of Forces

OR METHODOLOGY:
Linear Programming

ABSTRACT

THRUPUT II is a linear programming model developed at the Naval Postgraduate School for the U.S. Air Force Studies and Analyses Agency (AFSAA) to help improve the efficiency of the airlift mobility system. 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. THRUPUT II was used in the analysis provided by AFSAA to the C-17 Defense Acquisition Board in November, 1995. This paper reviews the model's formulation, describes its use in the C-17 analysis, and reports extensions that have been developed since the model's first appearance.

1 INTRODUCTION

This paper is a status report on a multi-year research effort to apply optimization modeling technology to the analysis of strategic airlift mobility. The purpose of the research is to help the U.S. Air Force improve logistical efficiency. Optimization is used to determine 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.

The primary model discussed in this paper is called THRUPUT II, which was introduced in a Naval Postgraduate School (NPS) Masters thesis [Lim, 1994] and further developed in a *Military Operations Research* article [Morton, Rosenthal, and Lim, 1996]. Since those earlier publications were written, THRUPUT II provided inputs to the C-17 Defense Acquisition Board decision of November 1995. This experience and other subsequent developments are covered here. A new model is currently under joint development between NPS and the RAND Corporation [Melody *et al*, 1996]. The distinguishing features of this new model are discussed in the conclusion.

The progenitors of THRUPUT II were the first THRUPUT, developed at the Air Force Studies and Analyses Agency [Yost, 1994]; and the Mobility Optimization Model (MOM), developed at the Joint Staff's Force Structure Resource, and Assessment Directorate (J8) [Wing *et al*, 1991]. All of these models are implemented with the General Algebraic Modeling System (GAMS) [Brooke, Kendrick and Meeraus, 1992].

Examples of the types of mobility questions that can be analyzed with optimization are: For a given fleet and a given network,

- Are the aircraft and airfield assets adequate for the deployment scenario?
- What are the impacts of shortfalls in airlift capability?
- 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. Such analyses 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 optimal values for the number of missions flown, and the amounts of cargo and passengers carried, for each unit, by each aircraft type, via each route, in each time period.

After describing the optimization model in Sections 2 and 3, Sections 4 and 5 discuss analyses performed in the recent non-developmental airlift aircraft (NDAA)/C-17 study. A special-purpose algorithm for solving large problem instances and a modeling extension to incorporate aircraft reliability are described in Section 6. Section 7 presents conclusions and ongoing research.

2 OVERVIEW OF MODEL

In this section we give a conceptual overview of the airlift optimization model. Then, Section 3 provides a detailed mathematical formulation. Sections 2 and 3 can be skipped by readers familiar with [Morton, Rosenthal, and Lim, 1996].

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 the maximum leg distance of a mission (shorter mission legs allow greater payloads), and aircraft with heavy loads may be required to make frequent 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, incorporating crews, and modeling intra-theater shuttles and ground transportation are the subject of the follow-on model mentioned earlier. 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 in 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

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 additional contingencies.

2.2.2 Constraints

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

- 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., C-17, C-5, C-141, 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.

- Air Force planners use a measure called Maximum-on-Ground (MOG) to represent airfield capacity. 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 available would yield a more accurate model. Ongoing projects at AMC [Schubert, Whisman, and Steppe, 1996] and RAND [Stucker, 1996] involve stochastic and deterministic simulations, respectively, whose purpose is to determine appropriate, and possibly multidimensional, MOG values. The model presented here will benefit from these investigations.
- 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. Then, in Section 6, we describe a stochastic optimization formulation that explicitly models stochastic ground times and indicate how this optimization model has been linked with a discrete-event simulation.

Other approximations of reality employed in the model for computational tractability are aggregation of airfields, discretization 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 model outlined above. The airlift optimization model is formulated as a multi-period, multi-commodity network-based linear program (LP) 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 closure (i.e., the time when all of a unit's deliveries are complete). 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. This is discussed in greater detail in Section 5.

3.1 Indices

u	indexes units, e.g., 82nd Airborne
a	indexes aircraft types, e.g., C-17, C-5, C-141
t,t'	indexes 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 × 108 inch platforms (84 × 104 usable) and can fit on any military aircraft (as well as cargo-configured CRAF) [Merrill, 1997]. Over-sized cargo is non-palletized rolling stock; it is larger than bulk cargo and can fit on a C-141, C-5 or C-17. Out-sized cargo is very large non-palletized cargo that can fit into a C-5 or C-17 but not a C-141.

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
$R_{au} \subseteq R_a$	permissible routes for aircraft type a carrying cargo or troops for unit u
$R_{ai} \subseteq R_a$	permissible routes for aircraft type a that use origin airfield i
$R_{ak} \subseteq R_a$	permissible routes for aircraft type a that use destination airfield 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 start times for aircraft of type a flying a mission for unit u on 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

$MovePAX_u$ Troop movement requirement for unit u

$MoveUE_u$ Equipment movement requirement in short tons (stons) for unit u

$ProBulk_u$ 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

$MaxLate$ 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

$PAXWt$ Average weight of a 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

$PAXSqFt_a$ Average cargo space (in sq. ft.) consumed by a 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 day) for an aircraft of type a

Airfield Data

$MOGCap_{bt}$ Aircraft capacity (in narrow-body equivalents) at airfield b in time t

$MOGReq_{ab}$ Conversion factor to a narrow-body equivalent for an aircraft of type a at airfield b

$MOGEff$ 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

$DaysLate_{uart}$ 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_{uart} Number of aircraft of type *a* that airlift unit *u* via route *r* with mission start time *t*

Y_{art} Number of aircraft of type *a* that recover from a destination airfield via route *r* with start time *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

$UENoGo_u$ Total stons of unit *u* equipment not delivered in the prescribed time frame

$PAXNoGo_u$ Number of unit *u* troops not delivered in the prescribed time frame

Each of the decision variables is constrained to be non-negative.

3.5 Formulation of the Objective Function

minimize:

$$\sum_u \sum_a \sum_{r \in R_a} \sum_{t \in T_{uar}} LatePenUE_u$$
$$\cdot DaysLate_{uart} \cdot TonsUE_{uart}$$
$$+\sum_u \sum_a \sum_{r \in R_a} \sum_{t \in T_{uar}} LatePenPAX_u$$
$$\cdot DaysLate_{uart} \cdot TPAX_{uart}$$
$$+\sum_u (NoGoPenUE_u \cdot UENoGo_u$$
$$+NoGoPenPAX_u \cdot PAXNoGo_u)$$
$$+\sum_a \sum_t Preserve_{at} \cdot NPlanes_{at}$$

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.

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:

$$\sum_{a \in A_{bulk}} \sum_{r \in R_{au}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_u$$
$$= MoveUE_u \quad \forall u \text{ with } MoveUE_u > 0$$

Demand Satisfaction Constraints for Out-Sized Cargo:

$$\sum_{a \in A_{out}} \sum_{r \in R_{au}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_u$$
$$\geq ProOut_u \cdot MoveUE_u$$
$$\forall u \text{ with } MoveUE_u > 0$$

Demand Satisfaction Constraints for Over-Sized Cargo:

$$\sum_{a \in A_{ovr}} \sum_{r \in R_{au}} \sum_{t \in T_{uar}} TonsUE_{uart} + UENoGo_u$$
$$\geq (ProOver_u + ProOut_u) \cdot MoveUE_u$$
$$\forall u \text{ with } MoveUE_u > 0$$

Demand Satisfaction Constraints for Troops:

$$\sum_{a} \sum_{r \in R_{au}} \sum_{t \in T_{uar}} TPAX_{uart} + PAXNoGo_u$$
$$= MovePAX_u \quad \forall u \text{ with } MovePAX_u > 0$$

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 pre-assign 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.

In the following constraints we use the notation $[Ctime_{ar}]$ to denote $Ctime_{ar}$ rounded to the nearest integer.

Aircraft Balance Constraints at Origin Airfields:

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

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 Origin Airfields:

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

The above constraint is in the cumulative form, rather than in the simpler form $\Sigma i \; Allot_{ait} \leq Supply_{at}$ 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:

$$\sum_{r \in R_a} \sum_{t'=1}^{t} \sum_u K_{artt'} X_{uart'}$$
$$+ \sum_{r \in R_a} \sum_{t'=1}^{t} K_{artt'} Y_{art'} + \sum_i \sum_{t'=1}^{t} H_{ait'}$$
$$+ \sum_k \sum_{t'=1}^{t} HP_{akt'} \leq \sum_{t'=1}^{t} NPlanes_{at'} \quad \forall a, t$$

$$\text{where} K_{artt'} =$$
$$\begin{cases} t - t' + 1 & \text{if } t' \leq t < t' + CTime_{ar} - 1 \\ CTime_{ar} & \text{if } t \geq t' + CTime_{ar} - 1 \end{cases}$$

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 reported in [Morton, Rosenthal, and Lim, 1996]. 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:

$$TPAX_{uart} \leq MaxPAX_a \cdot X_{uart} \quad \forall u, a, r, t : t \in T_{uar}$$

Maximum Payload Constraints:

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

Cargo Floor-Space Constraints:

$$PAXSqFt_a \cdot TPAX_{uart} + UESqFt_u \cdot TonsUE_{uart} \leq ACSqFt_a \cdot LoadEff_a \cdot X_{uart} \quad \forall u, a, r, t : t \in T_{uar}$$

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.

As an illustration of the above constraint, consider a fleet of five 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 flight time 1000 hours (10 hours/plane-day × 20 days × 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.

$$\sum_u \sum_{r \in R_a} \sum_{t \in T_{uar}} FltTime_{ar} \cdot X_{uart} + \sum_{r \in R_a} \sum_t FltTime_{ar} \cdot Y_{art} \leq \sum_t URate_a \cdot NPlanes_{at} \quad \forall a$$

3.6.5 Aircraft Handling Capacity at 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. The MOG efficiency factor *MOGEff* is introduced to cushion the effect of not explicitly modeling uncertainty. In Section 6.2, we describe a stochastic programming model that more directly handles aircraft reliability and its effect on airfield capacity. The MOG constraints are formulated for each airfield and time period (as before, we use the notation $[Dtime_{abr}]$ to denote $Dtime_{abr}$ rounded to the nearest integer).

$$\sum_u \sum_a \sum_{r \in R_a} \sum_{\substack{t' \in T_{uar} \\ t' + [DTime_{abr}] = t}} (MOGReq_{ab} \cdot GTime_{abr}/24) \cdot X_{uart'} + \sum_a \sum_{r \in R_a} \sum_{t' + [DTime_{abr}] = t} (MOGReq_{ab} \cdot GTime_{abr}/24) \cdot Y_{art'} \leq MOGEff \cdot MOGCap_{bt} \quad \forall b, t$$

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* 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 $Y_{art'}$ 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 MOGs" become available.

3.6.6 Initial Conditions

$$H_{ait} \equiv 0 \quad \forall a, i, t : t \leq 0$$
$$HP_{akt} \equiv 0 \quad \forall a, k, t : t \leq 0$$
$$X_{uart} \equiv 0 \quad \forall u, a, r, t : t \leq 0$$
$$Y_{art} \equiv 0 \quad \forall a, r, t : t \leq 0$$

4 FLEET-MIX TRADEOFF ANALYSIS

Prior to the C-17 Defense Acquisition Board (DAB) decision in November, 1995, there were a number of fleet options being considered as replacements for the aging C-141 fleet. These included "pure" C-17 fleets, as well as mixed fleets that included not only C-17s, but also a number of Non-Developmental Airlift Aircraft (NDAA), a Boeing 747-400F assigned the USAF designation C-33. THRUPUT II's first "operational" test supported the analysis required by the C-17 DAB.

Although many criteria must be considered when designing a fleet mix, a principal consideration is the ability to deliver the U.S. mobility requirements in support of our National Defense Strategy—currently two nearly-simultaneous Major Regional Contingencies (2-MRCs). Since THRUPUT II was designed to study strategic airlift, a two theater mobility study was a natural application of the model.

In the 2-MRC scenario, much of the cargo being flown from CONUS to the theaters is considered "out-sized" equipment, such as tanks or helicopters. Out-sized cargo is problematic, since it can only fit on certain wide-body aircraft, such as the C-5 or C-17. The C-33 is a hybrid in this regard; it can carry some, but not all types of out-sized cargo. THRUPUT II's features are well suited to contrast the capabilities of the long range, high payload C-33, with the more versatile, but smaller C-17. It was conceivable that THRUPUT II would show the lifting capability of a modest C-33 fleet could move most of the bulk and over-size cargo, allowing C-5s to satisfy the out-size requirement. Alternatively, the results might show that the demand for out-sized cargo movement dominates, and that the C-5 must be supplemented with C-17s to meet that requirement.

An additional fleet mix tradeoff involves the consumption of ground resources. The C-17 is designed to onload and offload quickly in an austere environment, while the C-33 is principally an airliner, and requires longer runways and a more robust support infrastructure. However, unless refueled in flight, the C-17 needs to stop more frequently than a C-33, which could offset any advantage derived from its reduced ground requirements. These two contrasting aspects of C-17 and C-33 resource utilization could interplay so as to give one aircraft considerable advantage in a contingency.

Cargo loading and airfield utilization are just two of a myriad of issues surrounding the procurement of any new airlifter. Without detailed modeling and simulation, the C-17 DAB could not hope to make an informed choice based on objective criteria. However, unlike previous boards, this time the analysis included results provided by a detailed optimization model.

4.1 Input

THRUPUT II's input requirements are generalized into four categories: 1) unit, 2) airfield, 3) aircraft, and 4) route data. The source of the unit movement requirements is called the Time-Phased Force Deployment Data, or TPFDD. This highly detailed list of equipment and personnel requirements is intended to identify everything necessary to carry out our national strategy. Consequently, it can be quite detailed and extremely long. In fact, the TPFDD used in this analysis initially consisted of more than 21,000 entries. Modeling each of these entries as a THRUPUT II unit was unthinkable, given current computational limitations. Through careful screening and consolidation (see Section 5), the TPFDD was reduced to just over 200 entries, each of which was read into THRUPUT II as a unit. From the pared TPFDD, we examined the origins (Aerial Ports of Embarkation—APOEs), and destinations (Aerial Ports of Debarkation—APODs) and attempted to set up a realistic enroute basing scheme that could support the movement.

The primary guidance for this airfield and TPFDD information was the Joint Chiefs of Staff, J8, force structure analysis called the Mobility Requirements Study, Bottom-Up Review Update, MRS-BURU [Joint Chiefs of Staff, 1995]. Its results were driven by a specific 2-MRC TPFDD, considered to be the most widely accepted requirements listing in existence. This report not only identified the who, what, when, and where of every movement requirement, but also listed the available airfields, including their relative capacities for air cargo traffic flow. MRS-BURU is credited for providing the motivation for upgrading the U.S. airlift fleet.

Compared with unit and airfield information, aircraft and route data were relatively straightforward to gather. Although an aircraft's effect on the airlift system is contentious, its performance characteristics are largely objective and easily derived. Route data presented a more difficult, yet not insurmountable challenge. Relying only on currently established AMC routing condemns the model to favor aircraft whose payload-range characteristics resemble the current fleet. Allowing THRUPUT II the latitude to choose new routes based on an aircraft's unique capabilities was preferable, so we offered many more route-aircraft combinations than might seem necessary at first glance. The tradeoff between making sufficient routes available and model tractability is discussed in [Toy, 1996].

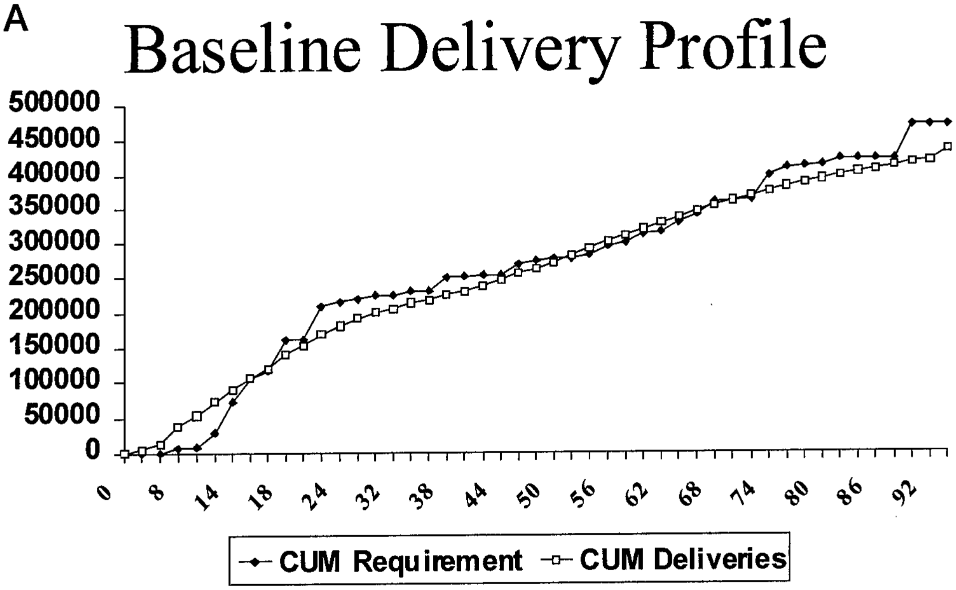
In addition to the airlift system parameters, there were several subjective factors to consider when setting up the scenario. One such factor was the *MaxLate* parameter, which establishes how late cargo and passengers can arrive before incurring an extremely large nondelivery (no-go) penalty. Increasing *MaxLate* naturally allows more overall cargo to be delivered, but has the unfortunate effect of dramatically increasing the size of the model, since there are more feasible movement options. However, the need to keep the model small must be balanced with a reasonable estimate of when "late" becomes "too late" from an operational standpoint. For the purposes of this work, *MaxLate* was set at eight days, meaning any cargo or passengers that could not be moved by the Required Delivery Date (*RDD*) + 8 would be considered not delivered and cause the maximum penalty to be charged. Fortunately, the eight-day maximum affected all excursions similarly, thus mitigating any relative advantage attributable to this subjectivity when comparing fleet mixes.

As with all analyses, preparing the above inputs took vast amounts of time. However, we believe many of these inputs are not scenario specific, and can be re-used with little adjustment in a variety of studies.

4.2 Analysis

THRUPUT II's C-17/NDAA analysis was conducted parametrically by running each proposed fleet mix as a separate excursion. The performance of each fleet was evaluated primarily by how much of the movement requirement (cargo and passengers) was delivered in a timely manner. Examining unit "closure" in this way, we were able to identify several significant differences between fleets. One such difference is illustrated by Figures 1A and 1B. In these figures, days from the "kickoff" of the first contingency are given on the horizontal

A



B

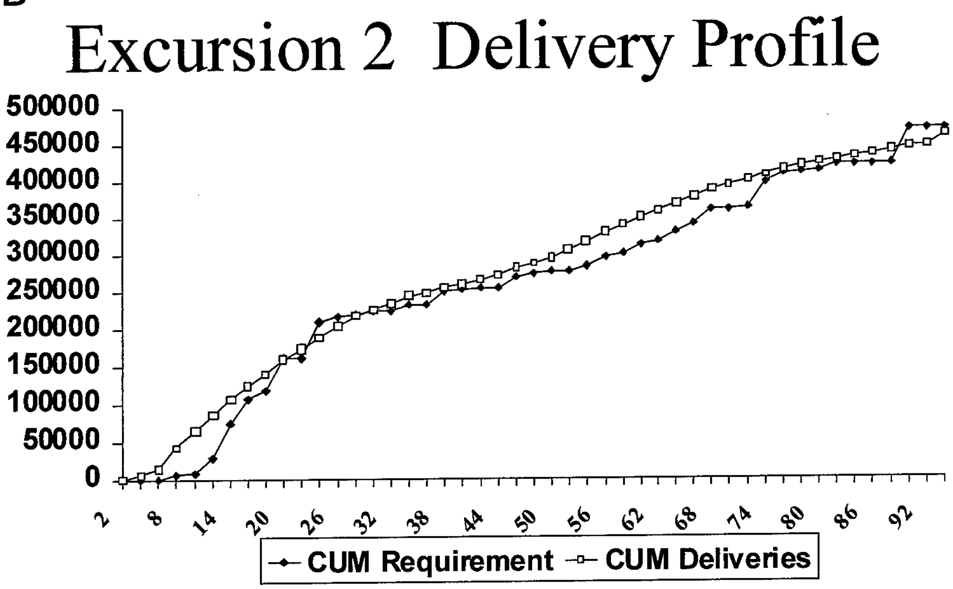


Figure 1. Tons delivered (vertical axis) versus contingency day (horizontal axis). The difference between the requirement and delivery lines in each scenario indicates that the aircraft fleet used in Excursion 2 performs better than the baseline.

axis. The two plots show the cumulative amount of cargo moved to date (stons), contrasted with the cumulative amount of cargo required to date. Figure 1A corresponds to the baseline fleet mix for the study. Figure 1B corresponds to one of the alternative fleet options under consideration.

Ideally, the airlift fleet should be able to accommodate the entire demand on time. However, given the non-uniform nature of the TPFDD requirements, all of the fleets examined had difficulty recovering from extremely large spikes in demand. The fleet used in the baseline case falls behind early and experiences great difficulty catching up with the requirement. However, the fleet used in Excursion 2 experiences only brief lags in the cargo delivery. It was more able to move cargo early, and hence stayed ahead of the imminent demand surges.

Another key metric used to evaluate the different fleets was the relative proportions of on-time, late, and undelivered cargo and passengers. Figures 2A and 2B detail the results for the five cases examined. While all of the cases delivered similar amounts of on-time cargo, total cargo delivered (including late deliveries) varied significantly—notably between Excursion 1, and Excursions 2, 3, and 4. Interestingly, a comparison of the two figures shows that as the airlift fleet was tailored to improve cargo delivery, the number of passengers delivered went down. As a result, it appears that none of the proposed fleets dominate with respect to both cargo and passenger delivery.

During the course of our output analysis, we were unexpectedly enlightened by what began as a casual look at the marginals, or "shadow prices" associated with an optimal solution. Although not a key aspect of the study, airfield size played an enormous role in the overall performance of each of the fleets. The output revealed that relatively small changes in the airfield's capacity at key enroute and destination airfields would yield disproportionate changes in system performance. Moreover, these key bases differed depending on which fleet mix was under consideration. For example, fleets with many C-17s stressed enroute airfields considerably more than fleets with many C-33s. Conversely, excursions with a large number of C-33s relied heavily on destination airfield size, but did not require as extensive an enroute infrastructure due to their longer range. Given the performance characteristics of the two aircraft, this insight is not surprising, but the quantification of an airfield's marginal value is a great benefit of optimization that is unavailable in simulation. Moreover, this discovery clearly emphasized that a fleet mix decision is not one to be made in isolation. All aspects of the airlift system, including such factors as airfield infrastructure must be considered when choosing a mix of aircraft. They are not independent; treating them as such risks providing decision makers with skewed information about a critical piece of our nation's mobility force.

The analysis described here, performed by the THRUPUT II team in support of the 1995 C-17/NDAA DAB decision, is indicative of the type of insight that can be provided by this optimization model to a decision maker. We have elected to emphasize this theme rather than delve into the scenario and excursion specific details such as fleet composition and basing structure. However, one aspect of this project that does demand closer description involves the methods used to reduce such a large (indeed initially intractable) model to a manageable size.

5 MODEL REFINEMENTS FOR IMPLEMENTATION

5.1 Model Reduction

One of the key issues regarding implementation of optimization modeling, particularly in military applications, is the balancing of realism vs. tractability.

No mathematical model can ever be totally realistic. The optimization modeling process is itself a constrained optimization problem. The objective of the process is to maximize the amount of realism achieved, subject to the limitations on computational tractability. Regardless of the rapid rate of advances in computing, we will always be faced with finite limits on tractability and hence never achieve total realism. The question is: how much realism can one achieve with the resources at hand?

THRUPUT II has decision variables with as many as four indices, such as X_{uart}, so the crux of the balancing problem is the number of (u, a, r, t)-tuples included in any real instance of the model. In general, the more tuples allowed, the more realistic the model, but the more difficult it is to solve. The number of tuples depends on, first, how many of each index type exists (how many units, aircraft types, routes and time periods are modeled), and, second, what rules are used for allowing or prohibiting any (u, a, r, t) combination from being considered. These two aspects of model reduction are discussed next.

5.1.1 Aggregation

The number of units, aircraft types, routes and time periods in the instances we ran of the model were chosen with a great deal of attention to the issues raised above. Distasteful as it may seem, a certain amount of aggregation of

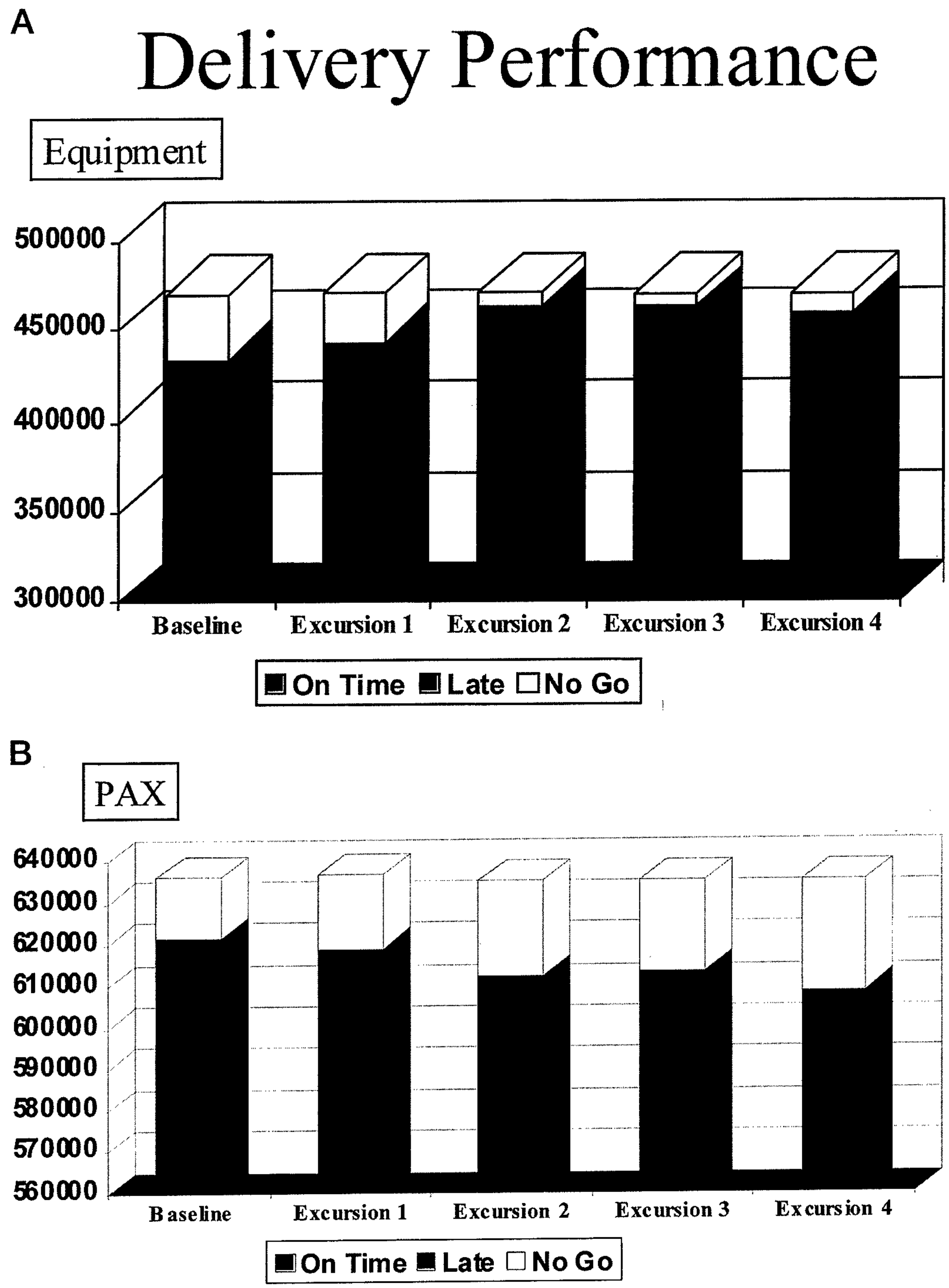


Figure 2. On time, late, and undelivered cargo for various airlift fleets.

entities is needed in any real-world modeling project. In this case, the most significant aggregation took place in the generation of airfields, units and time periods. Aggregation of airfields implies, in turn, a limit on the number of available routes.

We developed a location-theoretic optimization model for airfield aggregation, but in the case of the C-17/NDAA study, the USAF aviators on our team just used military judgement to decide which airfields to aggregate [Turker, 1995]. In the end, the infrastructure of the 2-MRC scenario was represented by 29 airfields.

The routes generated over the selected airfields were the product of a combination of the location-theoretic optimization model and the oversight of the aviators. The program used a tree structure to consider possible routes and screened them for inclusion based on various rules. The rules included: critical leg length of the aircraft, required crew rest or crew change, deviation of route length from great circle distance, aircraft/airfield compatibility (the civilian reserve fleet has landing restrictions not imposed on military aircraft and vice versa), and others. The oversight step was particularly intriguing, because the location and routing model was developed by a Turkish naval officer, who had no access to the real names and locations of the airfields during development and testing. Some routes had to be added or deleted based on understanding of the actual situation. The final result of this process was the inclusion of 313 routes for the entire scenario.

As stated in Section 4, the TPFDD file we were given for the C-17/NDAA analysis had over 21,000 movement requirements. This data set was first screened for the deletion of extremely small requirements. Then it was aggregated by assuming two movement requirements could be merged if the following conditions held: they had the same type of cargo (or passengers) to be moved, they had the same origin and destination (after airfield aggregations), and they had nearly simultaneous RDDs. The definition of "nearly simultaneous RDDs" was governed by a set of user-supplied parameters, which enforced simultaneity less rigorously as we went further out towards the horizon.

Aggregation of time is always a delicate issue in optimization modeling. Time has to be discretized, and nothing has a more direct effect on model size than the choice of time discretization units. For the C-17/NDAA study, we chose to divide time into two-day time periods, of which there were 47. Using 94 one-day time periods was intractable. See [Morton, Rosenthal, Lim, 1996] for an experiment on a single-MRC scenario, which showed that the cumulative aircraft balance constraints in Section 3.6.2 help lessen the effect of time discretization.

5.1.2 Variable Elimination and Sparsity

An algebraic modeling language such as GAMS is very conducive to implementing rules for limiting the number of admissible combinations of indices. In the case of the (u, a, r, t) tuples mentioned above, an X_{uart} variable is allowed to exist if all the following conditions hold:

- Route r flies from unit u's origin to its destination
- Aircraft type a can fly on route r with an acceptable payload
- The start time, t, of the mission is after unit u's available-to-load date
- The arrival time, if the mission starts at time t, is on or before $RDD(u) + MaxLate$
- There is a match between some cargo type (or passengers) that aircraft type a can carry and unit u's movement requirement
- Aircraft of type a must be available at the origin of route r at time t

These rules are evaluated once for each tuple and stored in a GAMS dynamic set. This set is referenced when the constraints are generated in order to achieve as much model sparsity as possible. The set is also used to eliminate other variables. For example, the aircraft inventory variables at destination airfields, HP_{akt}, cannot exist unless some aircraft of type a can potentially arrive at airfield k prior to time t.

There were other dynamic sets in THRUPUT II. In our experience with real-world optimization models, a serious investment of development time in the fine-tuning of dynamic sets for implementing model reduction rules can have a big payoff in tractability.

5.2 Computational Experience

After using the aggregations and reductions noted above, the model runs required for the C-17/NDAA study had the following problem dimensions: 200 units, seven (or fewer) aircraft types, 313 routes, 47 time periods and 29 airfields. The resulting model sizes and solution times are given below:

Scenario	Rows (000)	Cols (000)	Non-zeros	Solve Time
Baseline	161	183	1.9 mil	2.98 hrs
Ex 1	124	142	1.5 mil	1.95 hrs
Ex 2	154	177	1.9 mil	2.57 hrs
Ex 3	154	177	1.9 mil	3.14 hrs
Ex 4	154	177	1.9 mil	2.48 hrs

These runs were performed with GAMS as the problem generator and CPLEX 3.0 [CPLEX Optimization Inc., 1994] as the solver on an IBM RS6000/590 workstation. These rather large-scale optimization models presented a challenge. There were, in fact, several unsuccessful early attempts. We were very fortunate to be able to get advice from CPLEX Optimization, Inc. [Lowe, 1995] on solver settings. We sent them, via FTP over the Internet, a file containing a 3 million non-zero instance of the model, which they were able to solve. (This was when most but not all of the variable eliminations and sparsity refinements were implemented, so they have solved an even larger problem than the ones reported above.) The key advice from the CPLEX people was to use the barrier (interior point) algorithm, with tolerances and options tuned for this particular model.

5.3 Output

Each run of THRUPUT II for the C-17/NDAA study produced large amounts of output data. This profusion of information was too much for an analyst to absorb, so it had to be organized in relevant summary reports of the optimal solution. Since each case took a long time to solve, we made sure that all the optimal solution information was stored in readily accessed files. Then, a separate GAMS reporting program could be run many times against the same optimal solution. This proved to be useful because the AFSAA analysts often thought of new ideas for interesting summary reports while analyzing the old ones. Among the most widely used reports were:

- Total number of delivery missions flown, by aircraft type.
- Total number of missions flown, by route.
- For each unit, the closure date (date of last delivery if unit is fully delivered) displayed next to the ALD and RDD, along with the total amounts delivered on-time, late and not at all.
- For each MRC and time period, cumulative deliveries vs. requirements, separated by bulk, over-size, out-size and passengers.
- For each airfield, a report on MOG utilization, summarized as the number of days when MOG use exceeds P percent of capacity for P = 10, 25, 50, 75, 90, 95, and 100.

6 MODEL EXTENSIONS

6.1 Solution by Cascade

Although the difficulties associated with solving a large THRUPUT II model can be partially redressed by the model reduction techniques just described, ongoing research at the Naval Postgraduate School demonstrates that THRUPUT II may be solved in a piecemeal manner, thus greatly increasing the allowable problem size. This section describes that effort.

Consider how a scheduler would approach the 2-MRC scenario. Meticulously optimizing all aircraft, loading, and route decisions over the entire scenario length is impossible for at least two reasons: 1) future uncertainty makes gathering accurate data for the latter periods of a scenario problematic, and 2) a sufficiently long contingency overloads the scheduler's ability to reconcile the myriad of decisions. A modeler formulating a linear program faces the same difficulties, namely incorporating the increasing problem size with decreasing certainty as the length of the scenario grows. For either scheduler or modeler, perhaps the most straightforward way of dealing with the difficulties incurred by a large scenario is to focus sequentially on a subset of the scenario's periods, then move forward in time to a new subset. This temporal "myopia" degrades the solution quality, but makes the problem simpler to solve. Moreover THRUPUT II, which is used to mimic scheduling but does not produce schedules, is more "accurate" if it can incorporate the realism of nearsighted scheduling. For example, when choosing fleet size or infrastructure for use in future mobility contingencies, THRUPUT II ideally wishes to optimize given the *current scheduling capabilities*, instead of a utopian capability. A truly optimal schedule generated by THRUPUT II might alter decisions made at the outset of a contingency based on specific delivery requirements several weeks later. This is unrealistic, and can be avoided by reducing the ability of the formulation to look so far ahead.

The *proximal cascade* heuristic applied to THRUPUT II proceeds by solving for all variables and constraints whose domain is defined for the first 20 (for example) periods. Thus missions are flown so as to minimize delivery penalties in the first 20 periods, subject to the constraints applicable in those periods. Then, the process is cascaded forward in time to solve for a later set of periods. Mathematically, this implies generating a feasible solution by successively solving for only a subset of rows and columns, then moving to a set of rows and columns corresponding to later time periods. Each of these subproblems should overlap the previously solved subproblem in order to minimize the end effects caused by the former's temporal limitation. Fortunately, this methodology is facilitated by the structure of THRUPUT II. Variables and constraints in this model directly affect only nearby time periods. For example, missions flown on day 5 of a scenario have a large impact on the missions that can be flown on day 7, but only a minor impact on the missions that can be flown on day 25. This characteristic manifests itself as an overlapping "staircase" along the main diagonal of an LP's constraint coefficient matrix. The width of the overlap gives the number of time periods directly affected by the decisions (variable levels) made in a given time period. The rest of the coefficient matrix is relatively sparse, since variables (columns) associated with the early time periods rarely appear in constraints (rows) corresponding to the later time periods. This well known methodology is known as either the rolling horizon, or *proximal cascade* heuristic. However, the heuristic is sparsely documented, and is theoretically incomplete, since no scheme to bound the solution quality has been offered.

The quality of the solution produced by the *proximal cascade* heuristic is dependent on many scenario specific factors, and cannot be stated theoretically for most problems. However, a bound on the solution quality may be derived by exploiting information derived from this heuristic solution. Since a given time period is only directly linked to a few adjacent time periods, relaxing the rows associated with these nearby periods can separate subproblems out of an otherwise linked model. As with most decompositions however, the success of this scheme is dependent on the ability to compute accurate prices for resource consumption of the relaxed constraints. With such prices, a Lagrangian penalty can be applied to the subproblems, and a lower bound can be derived. Often, price selection is computationally intensive, which makes Lagrangian methods undesirable. However, in this case, reasonable prices are readily available from the *proximal cascade* heuristic just computed.

The *proximal cascade* heuristic offers a way to produce a more realistic schedule than an unencumbered optimization model. It also greatly reduces the tractability problems associated with the large models demanded by mobility planners. Finally, the cost of scheduling myopia may be estimated by solving a series of relaxed subproblems. For these reasons, the method shows great promise for use with THRUPUT II [Baker, 1997].

6.2 Incorporating Aircraft Reliability

Aircraft reliability is an important factor in the ability of the airlift system to deliver troops and materiel in a timely fashion. The current fleet has a mix of planes with differing reliability characteristics. For example, the C-5 requires unscheduled maintenance on approximately 15% of its landings while the rate for the newer (and smaller) C-17 fleet is under 7% (1994 peace-time data, AMC). Broken aircraft reduce the lift capability of the system by reducing the size of the effective fleet. In addition, aircraft requiring unscheduled maintenance and repairs reduce throughput by consuming scarce resources (e.g., maintenance, crew-duty hours, ramp space) that might otherwise contribute to on-time deliveries.

Simulation models for airlift systems are attractive because they can incorporate high levels of detail such as tracking individual aircraft and incorporating unscheduled maintenance and repairs. However, simulation models typically use naive aircraft routing and scheduling rules; as a result, it is possible to provide a simulation model with additional resources (e.g., more aircraft or routing options) and yet have system performance degrade. Linear programming models use more aggregate representations of the airlift fleet and infrastructure and do not incorporate uncertainty. However, due to optimal scheduling and routing, linear programming models better lend themselves to analysis of system bottlenecks by providing marginal values on specific resources, and in some cases, LP models may be more appropriate for comparing system performance under different sets of resources.

A *stochastic optimization* model for strategic airlift combines the ability of a simulation to include uncertain aircraft ground times with an LP's ability to optimally schedule and route aircraft. However, the resulting stochastic optimization model is typically very large and requires special-purpose optimization software. We have extended the LP model of Section 3 to incorporate aircraft reliability [Goggins, 1995]. The model is identical to the deterministic model except that the ground time $GTime_{abr}$ which appears in the airfield capacity constraint (Section 3.6.5) is replaced with a discrete random variable and the modified constraint includes an elastic decision variable which allows the constraint to be violated at a certain cost.

Mathematically, these modifications can be summarized as follows. Let ω denote a specific ground-time scenario (such as a scenario where a C-5A breaks, and its repair time is seven hours), and let p^{ω}_{bt} be the probability of observing scenario ω for a particular base b and time t. $GTime^{\omega}_{abr}$ represents the "effective ground-time" spent by aircraft a at base b when flying route r under ground time scenario ω, and $MOGPen_{bt}$ is the unit penalty for violating the airfield capacity at base b in time t. The elastic decision variable $R^{\omega}_{bt} \geq 0$ denotes the amount by which capacity is exceeded at airfield b in time t under scenario ω. The new airfield capacity constraints and the additional objective function term are specified in the following two equations.

The stochastic optimization model has been solved for the modest-sized data set in [Lim, 1994] which has 20 units, seven aircraft types, 17 airfields, and 30 time periods. Three of the seven aircraft were modeled as having random ground times (C-5, C-17, and C-141) and we assumed each aircraft type breaks independently. Ground times were approximated by discrete distributions with 9 realizations for each aircraft type, resulting in $9^3 = 729$ realizations for each base b and time t combination. The resulting stochastic model increases the number of airfield capacity constraints by a factor of 729 over the deterministic model from $30 \times 17 = 510$ to 371790. There are an equal number of additional decision variables of type R^{ω}_{bt}. We solved the stochastic model with a Benders' decomposition algorithm. While the total number of constraints in the stochastic model is greater by a factor of more than 50, the increase in running time over the deterministic model is a factor of 12 (20 minutes to 100 seconds on an IBM RS6000 590 workstation) [Goggins, 1995].

Since the linear and stochastic programming models contain more aggregate representations of the airlift system, we examined whether the schedules proposed by these optimization models were "flyable" in a more detailed simulation. We developed a discrete-event stochastic simulation model that took as input the output of these mathematical programs; specifically, the simulation model attempts to execute a proposed aircraft routing schedule. The strategy of coupling optimization and simulation models in this way is very attractive. Confidence can be gained in the optimization model as certain parameters are tuned (as we describe below) and the performance of the simulation can be improved since naive

Airfield Capacity Constraint when Aircraft have Random Ground Times:

$$\sum_{u}\sum_{a}\sum_{r\in R_a}\sum_{t'+[DTime_{abr}]=t} (MOGReq_{ab}\cdot GTime^{\omega}_{abr}/24)\cdot X_{uart'} + \sum_{a}\sum_{r\in R_a}\sum_{t'+[DTime_{abr}]=t} (MOGReq_{ab}\cdot GTime^{\omega}_{abr}/24)\cdot Y_{art'} - R^{\omega}_{bt} \leq MOGCap_{bt} \quad \forall b,t,\omega$$

Additional Objective Function Term to Discourage Capacity Violations:

$$\sum_{b}\sum_{t}\sum_{\omega} p^{\omega}_{bt}\cdot MOGPen_{bt}\cdot R^{\omega}_{bt}$$

scheduling rules are replaced with those proposed by an optimization model.

Our experimental results on the modest-sized data set of [Lim, 1994] compared schedules proposed by the linear and stochastic programming models. During the peak demand periods, we observed a 10% increase in cargo and troop deliveries when the simulation model executed schedules proposed by the stochastic program. Because the stochastic optimization model is larger and more difficult to solve (and cannot currently be solved within algebraic modeling languages such as GAMS), it is desirable to "tune" the deterministic optimization model so that it yields delivery schedules that are achievable in the simulation model. As described in Section 3.6.5, controlling the "MOG efficiency value" *MOGEff* is one way to achieve this. We empirically determined that a *MOGEff* value of 0.80 gave deterministic optimization schedules that were "flyable" in the stochastic simulation.

7 CONCLUSIONS AND ONGOING RESEARCH

THRUPUT II is an optimization model of the airlift mobility system that has proven useful to Air Force analysts in an important acquisition study. The Air Force analytical community has in the past put much more reliance on simulation than on optimization. This is in contrast to civilian industries, such as petroleum, electronics, airlines, forestry and many others, where optimization is very widely used.

While we were developing THRUPUT II, a similar and concurrent effort was under way at the RAND Corporation. The CONOP model of [Killingsworth and Melody, 1994] is also a GAMS-based, multi-period linear programming model for airlift optimization. It has some features not found in THRUPUT II. In May, 1996, the NPS and RAND groups started a joint effort to develop a new optimization model with the best features of both THRUPUT II and CONOP. The new model is called the NPS/RAND Mobility Optimizer (NRMO). Among NRMO's features that are not modeled in THRUPUT II are:

- The use of tankers for aerial refueling, and the facility for some tankers to change roles between refueling and cargo hauling.
- The modeling of shuttle flights and ground transportation in theater: some units have the option of direct delivery vs. transshipment, and some aircraft have the option of changing roles between strategic carriers and shuttlers.
- Detailed flow balance and utilization constraints for crews.
- The modeling of recovery bases, so that aircraft arriving in theater have the option of receiving services and crew changes at some other airfield besides the MRC's main port of debarkation.

The NRMO model is currently in use in a study of airfield infrastructure and in a large Pacific scenario. The detailed formulation of NRMO and the results of these studies will be given in a future report [Melody *et al.*, 1996].

REFERENCES

Baker, S., *A Cascade Approach for Staircase Linear Programs with an Application to Air Force Mobility Optimization*, PhD Dissertation, Operations Research Department, Naval Postgraduate School, Monterey, CA., June 1997.

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

CPLEX Optimization Inc., *CPLEX*, version 3.0 for *GAMS*, Incline Village NV, 1994.

Gearing, R., Hill, J., and Wilson, D., "UTE Rates Revisited," *Airlift*, Spring 1988.

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

Joint Chiefs of Staff, "Mobility Requirements Study—Bottom-Up Review Update" (U) SECRET, The Pentagon, Washington, DC., March 1995.

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

Lim Teo Weng, *Strategic Airlift Assets Optimization Model*, M.S. Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA., September 1994.

Lowe, T., Telephone interviews and electronic mail, CPLEX Optimization Inc., Incline Village NV, September 16–18, 1995.

Melody, L., Rosenthal, R., Morton, D., and Baker, S., "NPS/RAND Mobility Optimizer (NRMO)," Fifth Air Force Mobility M&S Users' Group, 1–2 August 1996, Naval Postgraduate School, Monterey, CA.

Merrill, D., Electronic mail, HQ Air Mobility Command Studies and Analyses Group, Scott AFB IL, September 29, 1997.

Morton, D., Rosenthal, R., and Lim Teo Weng, "Optimization Modeling for Airlift Mobility," *Military Operations Research*, 1, No. 4, Winter 1996.

Schubert, K., Whisman, A., and Steppe, J., "Airfield Capability Modeling," 64th MORS Symposium, 18–20 June 1996, Fort Leavenworth, KS.

Stucker, J., "Airfield Capacity Estimator (ACE)," Fifth Air Force Mobility M\&S Users' Group, 1–2 August 1996, Naval Postgraduate School, Monterey, CA.

Toy, A., *Route Prioritization for an Airlift Mobility Optimization Model*, M.S. Thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, September 1996.

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

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

Wing, V., Rice, R., Sherwood, R., and Rosenthal, R., "Determining the Optimal Mobility Mix," Joint Chiefs of Staff (J8), The Pentagon, Washington, DC, October 1991.

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

ERRATUM: BENCHMARKING AND EFFICIENT PROGRAM DELIVERY FOR THE DEPARTMENT OF DEFENSE'S BUSINESS-LIKE ACTIVITIES

In the article "Benchmarking and Efficient Program Delivery for the Department of Defense's Business-Like Activities" that appeared in the Spring 1996 issue of *Military Operations Research* there is an error in the DEA model (2) presented on page 24. The formulation as it appears in the article is as follows:

$$\text{maximize } h_0 = \frac{u_s S_0}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0}$$

$$\text{subject to: } \frac{u_s S_1}{v_{wy} WY_1 + v_{sqft} SQFT_1 + v_{apf} APF_1} + \frac{u_s S_2}{v_{wy} WY_2 + v_{sqft} SQFT_2 + v_{apf} APF_2} + \ldots \frac{u_s S_{237}}{v_{wy} WY_{237} + v_{sqft} SQFT_{237} + v_{apf} APF_{237}} \leq 1$$

$$\frac{u_s}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{wy}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{sqft}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{apf}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

This formulation shows each ratio of weighted outputs to weighted inputs to be summed and that the sum must be less than or equal to one. This is incorrect. The correct formulation is to have each ratio to be less than or equal to one. Following is the corrected model (2) as it should have appeared in the article.

$$\text{maximize } h_0 = \frac{u_s S_0}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0}$$

$$\text{subject to: } \frac{u_s S_1}{v_{wy} WY_1 + v_{sqft} SQFT_1 + v_{apf} APF_1} \leq 1$$

$$\frac{u_s S_2}{v_{wy} WY_2 + v_{sqft} SQFT_2 + v_{apf} APF_2} \leq 1 \vdots \frac{u_s S_{237}}{v_{wy} WY_{237} + v_{sqft} SQFT_{237} + v_{apf} APF_{237}} \leq 1$$

$$\frac{u_s}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{wy}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{sqft}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

$$\frac{v_{apf}}{v_{wy} WY_0 + v_{sqft} SQFT_0 + v_{apf} APF_0} \geq \epsilon$$

In the corrected model there would be 241 constraints—one for each commissary (237 commissaries) plus one for each output and input variable. In the incorrect model that appeared in the article there would have been only five constraints.

I would like to note that the error did not affect the results of the research presented in the article. The model as used in the actual analysis was correct. Only the presentation of the model in the paper was incorrect.

I would like to thank Major Greg Hoscheit of the Army Recruiting Command for noticing the error and pointing it out.