

**Technical Report No. 1308:**  
**Real-time Capacity and Inventory Allocation Decisions**  
**for Repairables in a Two-Echelon System**  
**with Emergency Shipments \***

Kathryn E. Caggiano <sup>†</sup>

John A. Muckstadt <sup>‡</sup>

James A. Rappold <sup>†</sup>

August 2003

---

\*This work was supported in part by Xelus, Inc., and by the National Science Foundation (Grant DMI0075627)

<sup>†</sup>School of Business, University of Wisconsin, Madison, WI 53706

<sup>‡</sup>School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853

Copyright ©2003 by all authors  
All rights reserved.

## Abstract

Two critical decisions must be made daily when managing multi-echelon repair and inventory systems for reparable service parts: decisions for allocating available repair capacity among different items, and decisions for allocating available inventories to field stocking locations.

In many such environments, procurement lead times for service parts are lengthy and variable, operational requirements change frequently – resulting in demand processes that are highly uncertain and non-stationary, and repair capacity is limited. Moreover, it is common to have many items in short supply while others are in long supply. Such environments require integrated real-time decision-support systems that reduce the impact of inventory imbalances and respond appropriately to the volatile nature of the demand processes. By integrated, we mean employing models that simultaneously consider key aspects of the current state of the operating environment in determining what items to repair, where to ship available units, and by what mode to ship them.

In this paper, we develop an integrated model for making repair and inventory allocation decisions in a two-echelon reparable service parts system. We formulate the decision problem as a finite-horizon, periodic-review mathematical program, we show it can be formulated as a large-scale linear program, and we develop practical heuristic methods for solving the problem.

We demonstrate the value of employing this type of integrated real-time decision model over using separate repair and inventory allocation rules for a range of environments that are commonly encountered in practice. We also demonstrate the quality and computational efficiency of the heuristic approaches.

# 1 Introduction

When managing distribution systems for reparable service parts inventories, three levels of planning must be conducted: strategic, tactical, and operational. At the strategic level, supply chain design issues concerning the location and capacity of repair and stocking locations are addressed. Tactical planning focuses on establishing policies within the supply chain infrastructure, including setting inventory levels for different items, assigning field stocking locations to address various needs, and procuring long lead time items. Trade-offs among service levels, capacity usage, transportation costs, inventory costs, and other operating costs are considered at a high level when making these strategic and tactical decisions. Stationary models are commonly used to make inventory decisions for tactical planning purposes.

By contrast, operational planning focuses on day-to-day decision-making. The operational planning process in a service parts repair and distribution system, such as a computer hardware, commercial airline maintenance, or military logistics system, involves determining the best way to allocate available repair capacity among different items and determining the best way to allocate available inventories to field stocking locations. These allocation decisions, which frequently must be made in highly dynamic environments, are largely driven by the information available about the current state of the system, as well as the information available about the demands that are likely to occur in the time periods that the decisions will affect. Since the level and reliability of such information is limited, the operational planning horizon for such a system is finite. The usual goal of the operational planning activity is to satisfy operational requirements at minimum expected cost over the time periods impacted by the current decisions.

Our objectives in this paper are to formulate this operational planning problem in a manner that meaningfully measures the economic consequences of the current operating decisions and to present the problem in a form that can be readily solved on a rolling horizon basis for large-scale systems. We demonstrate the value of employing such integrated decision models over decentralized allocation rules in a range of commonly encountered operating environments.

## 1.1 System Description

The system we examine consists of a central repair facility, a central warehouse, and a set of field stocking locations that service customers. Figure 1 shows the cyclic flow of materials in this system. Each customer receives service from a designated field stocking location. When a part being used by a customer fails, the defective unit is removed, and a replacement part of the same type is dispatched from the customer's designated field stocking location. The defective unit is shipped back to the central repair facility, where it joins a queue of parts awaiting repair. After the part is repaired, it is shipped to the central warehouse, where it is then available for redeployment to any field stocking location. If a customer need arises and the required part is not on-hand at the customer's designated field stocking location, a backorder occurs. The backordered part will be subsequently supplied to the field stocking location from the central warehouse. Parts may be shipped from the central warehouse to the field stocking locations via a regular or an expedited shipment mode.

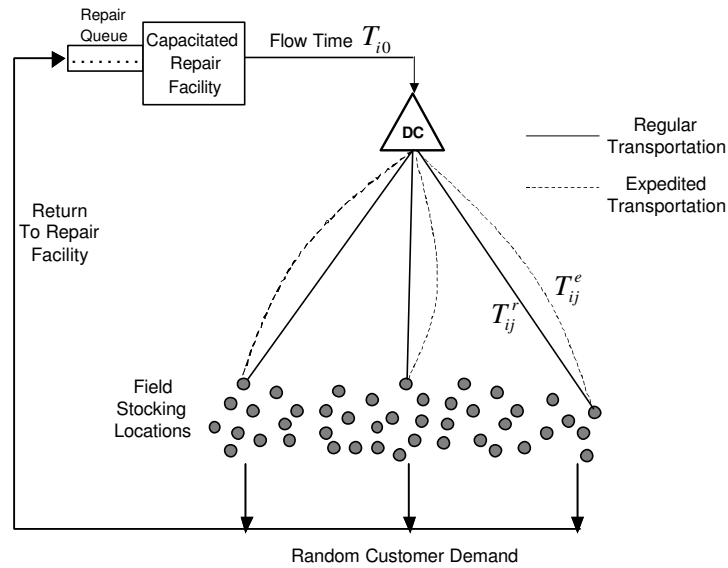


Figure 1: The System Under Study

At any given point in time, there are inventories of each part type in various places in the system. Specifically, a part may be:

1. In-transit from a field stocking location to the repair facility;

2. Awaiting repair at the repair facility;
3. In-repair at the repair facility;
4. In-transit from the repair facility to the central warehouse;
5. On-hand at the central warehouse;
6. In-transit from the central warehouse to a field stocking location; or
7. On-hand at a field stocking location.

Segments 1-4 constitute what we will call the *repair subsystem*, and segments 5-7 constitute the *distribution subsystem*. Given the quantities of each part type in each segment of the system, along with the knowledge of when parts in-transit will arrive at their destinations, decisions must be made each period regarding what parts to repair and what parts to ship from the central warehouse to the field stocking locations via each transportation mode.

## 1.2 Overview of the Paper

For ease of explication, we develop our model in three stages. The first model, although rather basic in its assumptions, provides both the framework and the insight for dealing with the second and third models, in which an additional shipment option and control over repair decisions, respectively, become available to the decision-maker. It is the third model for which we are ultimately interested in developing practical solution algorithms.

In describing each operational model, we take care to distinguish between the *planning horizon* of the model and the *effective horizon* of the model. The *planning horizon* of a model encompasses the time periods in which allocation decisions will be made. The *effective horizon* of a model encompasses the time periods in which the effects of the allocation decisions will be measured.

The first model, which we call the *stock allocation model*, or **SAM**, addresses only the inventory allocation problem and allows only one mode of shipment from the central warehouse to the field stocking locations (i.e, the regular shipment mode). The **SAM** determines how many parts to ship from the central warehouse to each field stocking location in each

period of the planning horizon. This model assumes that the decision-maker has complete visibility of parts in the distribution subsystem as well as those that are currently in the repair process and en route to the central warehouse, but has *no* knowledge of or control over parts that are in the repair queue or en route to the repair facility (i.e., segments 1 and 2). Accordingly, the length of the planning horizon for the **SAM** (i.e., the number of periods for which allocation decisions will be made) is the repair lead time plus the transport time from the repair facility to the central warehouse. Since the current contents of this pipeline are visible to the decision-maker, this is the horizon over which the supply to the central warehouse is known with certainty.

The *extended stock allocation model*, or **ESAM**, also addresses only the inventory allocation problem; however, in this model the decision-maker has the option of using an expedited shipment mode to transport parts from the central warehouse to the field stocking locations. In making allocation decisions, the incremental benefit gained by having parts arrive at a field stocking location earlier than they would have using the regular shipment mode must be weighed against the incremental cost of expedited shipment. The visibility of the decision-maker and the length of the planning horizon for the **ESAM** are the same as for the **SAM**.

In the *extended stock allocation model with repair*, or **ESAMR**, capacitated repair decisions as well as inventory allocation decisions must be made. The repair decisions to be made are which parts (of those awaiting repair) should have their repair commenced in the first period of the planning horizon. The **ESAMR** assumes that the decision-maker has visibility of all parts in the distribution subsystem as well as the repair subsystem, with the exception of parts that are in-transit to the repair facility (i.e., segment 1). The reason for this final restriction is that in practice, the decision-maker may have limited control over the timing and methods used by customers and/or field stocking locations to return defective parts for repair. The length of the planning horizon for the **ESAMR** is the same as for the **SAM** and the **ESAM**; however, the repair decisions to be made in this model affect the availability of parts at the depot warehouse in the last period of the planning horizon, and hence, are integrated with the inventory allocation decisions.

The objective of all three models is to minimize the *relevant* total expected system cost over the effective horizon. For the **SAM**, this cost includes the *incremental* expected cost of holding parts at the field stocking locations (instead of holding them at the central warehouse) and the expected backorder costs at the field stocking locations. For the **ESAM**, we also consider the *incremental* cost of shipping parts to field stocking locations via the expedited mode (instead of the regular shipment mode). For the **ESAMR**, in addition to the incremental shipping costs, we also allow for an *incremental* cost of holding parts at the central warehouse (instead of holding them in the repair queue). Repair costs, repair facility holding costs, and regular shipment costs are not captured in these models since they are *not relevant*. That is, assuming that all demand placed on the system is to be satisfied eventually, these costs will be incurred *regardless* of the current capacity and inventory allocation decisions. The only relevant costs in the operating models are those associated with the *timing* with which the parts are made to flow through the repair process to the central warehouse to the various field stocking locations.

In practice, these models would be implemented in a rolling horizon manner. That is, given the system's status at the beginning of a period, the model's recommendations would be followed for the current period only, recognizing that the model solution considers the consequences of the current period decisions on future period decisions and costs. Although the problems we describe could conceptually be formulated as dynamic programs, they could not be solved for realistically-sized problem instances due to the size of the state space. The models and solution approaches we will discuss are designed for making operational decisions in large-scale service parts repair and distribution systems.

The remainder of the paper is organized as follows. In Section 2 we review the relevant literature in the field. In Section 3 we define our notation and discuss the timing properties of the models. In Section 4 we formulate the **SAM** as a convex program, provide bounds for the optimal decision variable values, and show how **SAM** can be solved optimally using linear programming or approximately using simple greedy algorithms. In Section 5 we formulate the **ESAM**, provide bounds for the optimal decision variable values, and modify the **SAM** solution techniques to handle the **ESAM**. In Section 6 we present the **ESAMR**, an extension of **ESAM** that includes repair decisions, and describe how the overall problem can be solved



using solutions to the **ESAM** subproblems. We present a numerical study in Section 7 and summarize our contributions in Section 8.

## 2 Literature Review

Steady-state models for determining optimal inventory levels in multi-echelon systems have proven to be very useful in practice. In the seminal works by Feeney and Sherbrooke (1966) and Sherbrooke (1968), the METRIC model establishes base stock levels that minimize expected backorders in a two-echelon system with compound Poisson demands. The METRIC model and its numerous extensions make an infinite server assumption for the repair process, implying that the repair lead times are statistically independent over time. Moreover, these models commonly assume that repair capacity is allocated on a first-come, first-served (FCFS) basis. Despite their utility in many environments, these assumptions are often not valid and can consequently result in less-than-expected system performance.

Dynamic real-time allocation rules have been shown to impact operating performance significantly under certain conditions. The first to examine real-time allocation decisions in these environments was Miller (1968). In this work, a bidding-type system is developed in which material is allocated to the location with the greatest need, provided that the central warehouse is willing to ship the unit. For the scheduling of more complex repair processes, Maxwell (1969) examined the value of obtaining current system status information when making the job dispatching decision.

Muckstadt (1980) constructed a model and conducted a simulation study to assess the impact of non-stationary demand processes on inventory levels and system performance. The study examines the impact of using models that assume a stationary demand process when demand is, in fact, non-stationary. Hillstead and Carrillo (1980) and Hillstead (1982) also examine environments with non-stationary demand processes and develop the Dyna-METRIC model. Another model that addresses many real-world complexities is Vari-METRIC, developed by Slay (1984).

In Scudder and Hausman (1982), an inventory stocking model is developed that implicitly considers the presence of finite production (or repair) capacity and dispatching rules through

a characterization of dependent lead times. Hausman and Scudder (1982) and Scudder (1984) examine several scheduling heuristics to prioritize repair capacity. They conclude that, under most circumstances, the infinite server assumption is robust. Verrijdt et al. (1998) construct a model in which items awaiting repair can be expedited if their stock on-hand falls below a critical level. Hoadley and Heyman (1977) develop a single period two-echelon model that can be used to determine an initial amount of system inventory and its optimal placement among locations to minimize periodic adjustments and inventory rebalancing across locations and between echelons. Gross et al. (1977) consider the joint problem of setting system inventory levels for a single item and the number of repair servers. They apply their queuing-based approach to solve a multi-year planning problem in which each year is modeled in steady state. They are keen to observe that in highly dynamic environments with low demand rate items in which population sizes change, usage rates change, and engineering changes occur, using steady state models may be problematic. They offer an approximation to a dynamic program to make decisions across successive periods.

The impact of dynamic scheduling rules has also been addressed in the queueing literature. Wein (1992) uses a linear program imbedded in a queueing system to determine sequencing priorities for capacity constrained workstations. See Duenyas (1994) and Wein and Chevalier (1992) for model extensions and performance comparisons with static scheduling rules. Peña Perez and Zipkin (1997) model a multi-item production-inventory system with stationary Poisson demand and argue that inventories should be stored in those items with the lowest holding cost rate. Kim et al. (1996) apply dynamic dispatching rules to control the flow of released orders into a semiconductor facility.

Another large body of literature exists to explore the performance impact of expediting in supply chains. The models developed in Moinzadeh and Schmidt (1991), Aggarwal and Moinzadeh (1994), and Moinzadeh and Aggarwal (1997) offer two modes of transporting material between echelons in a two-echelon system. They develop the form of the optimal control policies and offer solution approximations. They demonstrate the value of having multiple modes of transportation.

The work done in Pyke (1990) is most relevant to ours in that he jointly considers the repair capacity allocation decision and the inventory allocation of items to bases. However,

our work differs from past research in several ways. First, we construct a computationally tractable multi-period approximation to a dynamic program to determine both optimal capacity allocation and inventory allocation decisions. Second, we develop a heuristic approach that is appropriate for very large systems. We prove several properties of the solution approach. Finally, our framework is quite general and permits the incorporation of additional system complexities, including multiple modes of transportation, non-stationary stochastic demand, and finite repair capacity.

### 3 Notation and Assumptions

In this section we define notation for the first two models we will discuss, **SAM** and **ESAM**. Additional notation for the third model, **ESAMR**, will be defined in Section 6.

The key to the model formulations lies in appropriately constraining the inventory allocation decisions by the part availability at the central warehouse, as well as accurately capturing the relevant cost consequences of the inventory allocation decisions. Namely, for each part type, we must capture:

- the current stock level at the central warehouse and the quantities due to arrive at the central warehouse over the planning horizon;
- the current stock levels at the field stocking locations and the quantities due to arrive at these locations as a consequence of the inventory allocation decisions; and
- the costs incurred each period as a consequence of the inventory allocation decisions; namely, the incremental expedited shipment costs to the field stocking locations, the incremental expected holding costs at the field stocking locations, and expected back-order costs at the field stocking locations.

We reiterate that repair costs, central warehouse holding costs, and regular shipment costs are not relevant to the **SAM** or the **ESAM** since these costs will be incurred regardless of the current inventory allocation decisions. Since the flow of parts from the repair facility to the central warehouse is predetermined in these models, the only relevant costs are those

associated with the *timing* with which the parts are made to flow from the central warehouse to the field stocking locations.

We will use the following notation throughout the paper:

### Network Parameters

- $I$  - the set of items, indexed by  $i$ .
- $J$  - the set of field stocking locations, indexed by  $j$ .

### Time Parameters (All are assumed to be an integer number of periods.)

- $T_{i0}$  - the repair lead time for item  $i$ , including transport to the central warehouse from the repair facility. (The index 0 denotes the central warehouse.)
- $T_{ij}^r$  - the regular transportation lead time for item  $i$  to location  $j$  from the central warehouse.
- $T_{ij}^e$  - the expedited transportation lead time for item  $i$  to location  $j$  from the central warehouse. We assume that  $1 \leq T_{ij}^e < T_{ij}^r$ .

### Decision Variables

- $y_{ijt}^r$  - the number of units of item  $i$  to ship via regular transport from the central warehouse to field location  $j$  in time period  $t$ ,  $t = 0, \dots, T_{i0}$ .
- $y_{ijt}^e$  - the number of units of item  $i$  to ship via expedited transport from the central warehouse to field location  $j$  in time period  $t$ ,  $t = 0, \dots, T_{i0}$ .

### Supply and Demand Parameters

- $\tilde{S}_{i0t}$  - the *known* cumulative supply of item  $i$  available at the central warehouse through period  $t$  (i.e., inventory on-hand at the beginning of the planning horizon plus stock arriving through period  $t$ ). These parameters define the pipeline stock profile coming into the central warehouse at the beginning of the planning horizon and are unaffected by current allocation decisions. They are defined for periods  $t = 0, \dots, T_{i0}$ .
- $\tilde{S}_{ijt}$  - the *known* cumulative supply of item  $i$  available at field location  $j$  through period  $t$  (i.e., net inventory at the beginning of the planning horizon

plus stock arriving through period  $t$ ). These parameters define the pipeline stock profile coming into location  $j$  from the central warehouse at the beginning of the planning horizon and are unaffected by current allocation decisions. They are defined for periods  $t = 0, \dots, T_{ij}^r + T_{i0}$ . Note, however, that  $\tilde{S}_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)}$  for all  $t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}$ .

- $S_{ijt}$  - the cumulative supply of item  $i$  available at field location  $j$  through period  $t$ . These parameters *are* affected by current central warehouse allocation decisions, and are defined for periods  $t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}$ .
- $X_{ijt}$  - the cumulative demand of item  $i$  at field location  $j$  through period  $t$ , a random variable, defined for periods  $t = 0, \dots, T_{ij}^r + T_{i0}$ .

### Cost Parameters and Functions

- $h_{ij}$  - the incremental cost per period associated with holding a unit of item  $i$  at field location  $j$  instead of at the central warehouse.
- $\pi_{ij}$  - the unit shortage cost per period for item  $i$  at field location  $j$ .
- $e_{ij}$  - the incremental cost of shipping a unit of item  $i$  via expedited mode from the central warehouse to field location  $j$ . This cost is assumed to include any incremental holding costs while in transit.
- $G_{ijt}(\cdot)$ - the function describing the expected incremental holding costs and backorder costs incurred in period  $t$  for item  $i$  at field stocking location  $j$ , defined for periods  $t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}$ . The function argument is  $S_{ijt}$ , the cumulative supply of item  $i$  at field location  $j$  through period  $t$ . That is,  $G_{ijt}(S_{ijt}) = h_{ij}E[S_{ijt} - X_{ijt}]^+ + \pi_{ij}E[X_{ijt} - S_{ijt}]^+$ .
- $Q_{ij}(\cdot)$ - the function describing the expected incremental holding costs incurred beyond the end of the effective horizon for item  $i$  at field stocking location  $j$ . The function argument is  $S_{ij(T_{ij}^r+T_{i0})}$ , the cumulative supply of item  $i$  at field location  $j$  at the end of the effective horizon. That is,  $Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) = h_{ij} \sum_{t=T_{ij}^r+T_{i0}+1}^{\infty} E[S_{ij(T_{ij}^r+T_{i0})} - X_{ijt}]^+$ .

Some explanation should be given for the item-specific time periods over which our parameters and variables are defined. As we mentioned earlier, our goal is to capture all of the incremental expected costs that are a *direct consequence* of the allocation decisions made during the planning horizon (i.e., in periods  $t = 0, \dots, T_{i0}$  for item  $i$ ). Note that since different items may have different repair lead times, the number of periods for which we have supply information (i.e.,  $T_{i0}$ , the pipeline length) may differ from item to item. Also, since the shipment lead times  $T_{ij}^r$  and  $T_{ij}^e$  to the various field stocking locations may differ across items and locations, the consequences of the allocation decisions made at the central warehouse may be realized within time windows that vary by item and by location. Specifically, for a given item  $i$  and a given field location  $j$ :

- Incremental expected holding costs are realized in periods  $t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}$ .
- Expected backorder costs are realized in periods  $t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}$ .
- Incremental costs for expedited shipments from the central warehouse are realized in periods  $t = 0, \dots, T_{i0}$ .

In addition, we capture the end-of-horizon implications of carrying inventory at each field stocking location for each item. Note that although the supply may make it possible to send extra inventory from the central warehouse to field stocking locations, carrying inventory at a field stocking location at the end of the effective horizon is appropriate only if the expected future demand at that location warrants it. To address this, we include in the model a cost term that reflects the expected incremental holding costs that would be incurred in periods *following* the end of the effective horizon. This cost function, first introduced in Chan et al. (1999), measures the expected number of future periods worth of inventory that are on-hand at the end of the effective horizon and multiplies this figure by the item's incremental holding cost per period. That is, the expected future holding cost associated with having  $[S_{ij(T_{ij}^r+T_{i0})} - X_{ij(T_{ij}^r+T_{i0})}]^+$  units of item  $i$  on-hand at field location  $j$  at the end of the effective planning horizon is expressed as:

$$Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) = h_{ij} \sum_{t=T_{ij}^r+T_{i0}+1}^{\infty} E[S_{ij(T_{ij}^r+T_{i0})} - X_{ijt}]^+. \quad (3.1)$$

It is easily shown that  $Q_{ij}$  is a convex function of its argument.

## 4 The Stock Allocation Model

In this section, we formulate the *stock allocation model*, or **SAM**, as a convex program and show that the problem is separable by item. We then propose two approaches for solving the item subproblems. In the first approach, we derive bounds on the optimal values of the decision variables and use these bounds to formulate and solve the subproblems as linear programs. Although the LPs can be quite large, this method is exact, and it provides a benchmark against which we can measure the performance of faster, greedy heuristics. Two such greedy algorithms are presented here.

### 4.1 Model Definition

Recall that the **SAM** addresses only the inventory allocation problem and allows only one mode of shipment (i.e., regular shipment) from the central warehouse to the field stocking locations. Given the notation of the previous section, we can describe the **SAM** as follows:

$$\text{(SAM) minimize} \quad \sum_{i \in I} \sum_{j \in J} \left\{ \sum_{t=T_{ij}^r}^{T_{ij}^r+T_{i0}} G_{ijt}(S_{ijt}) + Q_{ij}(S_{ij}(T_{ij}^r+T_{i0})) \right\} \quad (4.1)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t y_{ijt'}^r, \quad \forall i \in I, t = 0, \dots, T_{i0}, \quad (4.2)$$

$$S_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)} + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r, \quad \forall i \in I, j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}, \quad (4.3)$$

$$y_{ijt}^r \geq 0 \text{ and integer} \quad \forall i \in I, j \in J, t = 0, \dots, T_{i0}. \quad (4.4)$$

Constraints (4.2) ensure that stock is available at the central warehouse before it is allocated, and constraints (4.3) relate the quantities received at the field locations with the corresponding quantities shipped from the central warehouse.

Observe that no more than one item type occurs in any one constraint in this formulation. Hence, the problem is separable by item. Letting  $Z^*$  denote the optimal objective function value to **SAM**, this means that we can write:

$$Z^* = \sum_{i \in I} Z_i^*,$$

where  $Z_i^*$  denotes the optimal objective function of the subproblem  $\mathbf{SAM}_i$ , given by:

$$(\mathbf{SAM}_i) \text{ minimize } \sum_{j \in J} \left\{ \sum_{t=T_{ij}^r}^{T_{ij}^r+T_{i0}} G_{ijt}(S_{ijt}) + Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) \right\} \quad (4.5)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t y_{ijt'}^r, \quad \forall t = 0, \dots, T_{i0}, \quad (4.6)$$

$$S_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)} + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r, \quad \forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}, \quad (4.7)$$

$$y_{ijt}^r \geq 0 \text{ and integer } \quad \forall j \in J, t = 0, \dots, T_{i0}. \quad (4.8)$$

The problem of solving  $\mathbf{SAM}$  thus reduces to solving  $\mathbf{SAM}_i$  for each  $i \in I$ .

## 4.2 LP Formulation of $\mathbf{SAM}_i$

The objective function of  $\mathbf{SAM}_i$  given in (4.5) has two types of terms, those involving single-period expected costs, and those involving end-of-horizon expected holding costs. For a given field location  $j$  and a given time period  $t \in [T_{ij}^r, \dots, (T_{ij}^r + T_{i0})]$ , let us focus on the single-period cost function  $G_{ijt}$ . Since  $G_{ijt}$  is convex in its argument, it is a simple matter to find the solution to the constrained newsvendor problem  $\mathbf{CN}_{ijt}$ :

$$(\mathbf{CN}_{ijt}) \text{ minimize } G_{ijt}(S_{ijt}) \quad (4.9)$$

subject to  $S_{ijt} \geq \tilde{S}_{ijt}$  and integer.

Let  $\hat{S}_{ijt}$  denote the *largest* optimal solution to this problem, so that

$$\hat{S}_{ijt} = \max(\tilde{S}_{ijt}, \arg \min_{S \in \mathcal{S}}(G_{ijt}(S))), \quad (4.10)$$

where  $\mathcal{S} = \{ \lceil F_{X_{ijt}}^{-1}(\frac{\pi_{ij}}{\pi_{ij} + h_{ij}}) \rceil, \lfloor F_{X_{ijt}}^{-1}(\frac{\pi_{ij}}{\pi_{ij} + h_{ij}}) \rfloor \}$ .

Recall that  $\tilde{S}_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)}$  for all  $t \in [T_{ij}^r, \dots, T_{ij}^r + T_{i0}]$ . Thus, in order for the solutions  $\hat{S}_{ijt}, t \in [T_{ij}^r, \dots, (T_{ij}^r + T_{i0})]$  to be nondecreasing in  $t$ , it suffices for the corresponding distribution functions  $F_{X_{ijt}}(x)$  to be nonincreasing in  $t$  for all values  $x$ . Since  $X_{ijt}$  denotes the *cumulative* demand of item  $i$  at field location  $j$  through period  $t$ , this condition must hold.

Hence,

$$\hat{S}_{ij(t-1)} \leq \hat{S}_{ijt} \quad \forall j \in J, t \in [T_{ij}^r, \dots, (T_{ij}^r + T_{i0})], \quad (4.11)$$



and we have the following theorem to bound the optimal cumulative stock levels:

**Theorem 1** *For all field locations  $j \in J$  and all time periods  $t \in [T_{ij}^r, \dots, (T_{ij}^r + T_{i0})]$ , let  $\hat{S}_{ijt}$  denote the largest optimal solution to  $\mathbf{CN}_{ijt}$ , and let  $S_{ijt}^*$  denote the corresponding term within an optimal solution to  $\mathbf{SAM}_i$ . Then:*

$$\tilde{S}_{ij(T_{ij}^r-1)} \leq S_{ijt}^* \leq \hat{S}_{ijt} \quad \forall j \in J, t \in [T_{ij}^r, \dots, (T_{ij}^r + T_{i0})]. \quad (4.12)$$

*Proof:* The first inequality must hold for any feasible solution. Only the second inequality requires proof. Suppose it does not hold for some field location  $j$ , and let  $k$  be the smallest index (i.e., the earliest time period) for which the location violates this condition. That is,  $S_{ijk}^* > \hat{S}_{ijk}$ , and  $S_{ijt}^* \leq \hat{S}_{ijt}$  for all  $t \in [T_{ij}^r, \dots, (k-1)]$ . (If  $k = T_{ij}^r$ , this means that  $S_{ij(k-1)}^* = S_{ij(T_{ij}^r-1)}^* = \tilde{S}_{ij(T_{ij}^r-1)} \leq \hat{S}_{ij(T_{ij}^r-1)}$ .) By (4.11), we have that  $\hat{S}_{ij(k-1)} \leq \hat{S}_{ijk}$ . Thus,  $S_{ijk}^* > \hat{S}_{ijk} \geq \hat{S}_{ij(k-1)} \geq S_{ij(k-1)}^*$ , which implies that  $S_{ijk}^* - S_{ij(k-1)}^* > 0$ . Recall, however, that  $\tilde{S}_{ijk} - \tilde{S}_{ij(k-1)} = \tilde{S}_{ij(T_{ij}^r-1)} - \tilde{S}_{ij(T_{ij}^r-1)} = 0$ . Thus, the optimal solution must allocate at least one unit of item  $i$  that arrives at location  $j$  in time period  $k$ . That is, we must have  $y_{ij(k-T_{ij}^r)}^{r*} > 0$ . Consider the following minor changes to the optimal solution to  $\mathbf{SAM}_i$ :

$$\begin{aligned} y_{ij(k-T_{ij}^r)}^r &\leftarrow y_{ij(k-T_{ij}^r)}^{r*} - 1 \quad \text{and} \\ y_{ij(k-T_{ij}^r+1)}^r &\leftarrow y_{ij(k-T_{ij}^r+1)}^{r*} + 1. \end{aligned}$$

The resulting solution is feasible. We are simply delaying the shipment of one unit from the central warehouse to location  $j$  by one period, so that one *less* unit arrives in period  $k$  and one *more* unit arrives in period  $k+1$ . Moreover, this modified solution will have  $S_{ijk} = S_{ijk}^* - 1$ , but for all  $t \neq k$ ,  $S_{ijt} = S_{ijt}^*$ . If  $k \neq T_{ij}^r + T_{i0}$ , then the only change to the objective function is that  $G_{ijk}(S_{ijk})$  replaces  $G_{ijk}(S_{ijk}^*)$ . But since  $G_{ijk}$  is convex in its argument and  $\hat{S}_{ijk}$  is the *largest* optimal solution to  $\mathbf{CN}_{ijt}$ , we have that  $G_{ijk}(\hat{S}_{ijk}) \leq G_{ijk}(S_{ijk}) = G_{ijk}(S_{ijk}^* - 1) < G_{ijk}(S_{ijk}^*)$ . If  $k = T_{ij}^r + T_{i0}$ , then the second term in the objective function will also change; but, since  $Q_{ij}$  is a strictly increasing function of its argument,  $Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) = Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}^* - 1) < Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}^*)$ . In either case, the modified solution will have an objective function value that is strictly less than the value

achieved by the original solution. Hence, the original solution cannot be optimal, and in any optimal solution we must have  $S_{ijt}^* \leq \hat{S}_{ijt}$  for all  $t = T_{ij}^r, \dots, (T_{ij}^r + T_{i0})$ .  $\square$

Theorem 1 provides upper and lower bounds on the cumulative stock levels in any optimal solution to  $\mathbf{SAM}_i$ . We now use this result to construct a linear programming formulation of  $\mathbf{SAM}_i$ . Letting

$$\delta_{ijtk} = \begin{cases} 1 & \text{if } S_{ijt} = k, \\ 0 & \text{otherwise,} \end{cases} \quad (4.13)$$

we can reformulate  $\mathbf{SAM}_i$  as follows:

$$\text{minimize } \sum_{j \in J} \left\{ \sum_{t=T_{ij}^r}^{T_{ij}^r+T_{i0}} \sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\hat{S}_{ijt}} \delta_{ijtk} G_{ijt}(k) + \sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\hat{S}_{ij}(T_{ij}^r+T_{i0})} \delta_{ij(T_{ij}^r+T_{i0})k} Q_{ij}(k) \right\} \quad (4.14)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t y_{ijt'}^r, \quad \forall t = 0, \dots, T_{i0}, \quad (4.15)$$

$$\sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\hat{S}_{ijt}} \delta_{ijtk} \cdot k = \tilde{S}_{ij}(T_{ij}^r-1) + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r, \quad \forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}, \quad (4.16)$$

$$\sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\hat{S}_{ijt}} \delta_{ijtk} = 1 \quad \forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}, \quad (4.17)$$

$$\delta_{ijtk} \in \{0, 1\} \quad \forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}, k = \tilde{S}_{ij}(T_{ij}^r-1), \dots, \hat{S}_{ijt}, \quad (4.18)$$

$$y_{ijt}^r \geq 0 \text{ and integer } \quad \forall j \in J, t = 0, \dots, T_{i0}. \quad (4.19)$$

Since the cost functions  $G_{ijt}$  and  $Q_{ij}$  are convex in their arguments, and since the cumulative stock levels  $S_{ijt}$  and the parameters  $\tilde{S}_{i0t}$  and  $\tilde{S}_{ijt}$  only assume integer values, the integer restrictions on  $\delta_{ijtk}$  and  $y_{ijt}^r$  are unnecessary. That is, solving the LP relaxation of the preceding ILP will result in an integer optimal solution. Hence, the integer restrictions on  $\delta_{ijtk}$  and  $y_{ijt}^r$  in (4.18) and (4.19), respectively, can be dropped, and the solution to the resulting linear program will be an optimal solution to  $\mathbf{SAM}_i$ . Details can be found in Dantzig (1962).

The major drawback to solving  $\mathbf{SAM}_i$  using the above LP formulation is that the number of variables can be very large. For practical purposes, it is possible to reduce the number of variables by limiting the number of values each  $y_{ijt}^r$  can assume (e.g., multiples of 5, instead of 1). By making such restrictions, the LP will yield a solution that is only approximately optimal, but for large problems, this type of rescaling can be a very useful technique. Furthermore, since the second derivatives of the  $G$  functions are not large in the neighborhood of their minimizers, the expected cost of the optimal solution to a scaled problem will, in most realistic instances, be very close to the expected cost of the optimal solution to the unscaled problem.

### 4.3 Greedy Algorithms for $\mathbf{SAM}_i$

Instead of solving  $\mathbf{SAM}_i$  as a linear program, we have developed two alternative greedy algorithms that myopically exploit the convexity of the objective function. Solutions resulting from these algorithms may be used in two ways. They may be used directly, or, since these algorithms provide basic feasible solutions for the preceding LP, they can be used to seed the LP with a near-optimal solution.

Before describing these algorithms, we define the following terms to capture incremental changes in the objective function. For all  $t = T_{ij}^r, \dots, T_{ij}^r + T_{i0}$ , let:

$$\Delta G_{ijt}(S_{ijt}) = G_{ijt}(S_{ijt} + 1) - G_{ijt}(S_{ijt}) \quad (4.20)$$

be the incremental change in expected holding and backorder costs realized in period  $t$  when the allocation of item  $i$  to location  $j$  increases by one unit prior to period  $t$  or in period  $t$  (i.e., when one more unit of item  $i$  arrives at location  $j$  prior to period  $t$  or in period  $t$ ).

Next, define

$$\Delta C_{ijt} = \sum_{k=t}^{T_{ij}^r + T_{i0}} \Delta G_{ijk}(S_{ijk}) \quad (4.21)$$

to be the incremental change in the expected holding and backorder costs over the entire effective horizon when the allocation of item  $i$  to location  $j$  increases by one unit in period  $t$ . Finally, let

$$\Delta Q_{ij} = Q_{ij}(S_{ij(T_{ij}^r + T_{i0})} + 1) - Q_{ij}(S_{ij(T_{ij}^r + T_{i0})}) \quad (4.22)$$

be the incremental change in the end-of-horizon expected holding costs when the allocation of item  $i$  to location  $j$  increases by one unit in any planning horizon period.

Given these definitions, if in period  $t$  we *send* an additional unit of item  $i$  from the central warehouse to location  $j$ , the total incremental change in the objective function is as follows:

$$\Delta Z_i(j, t) = \Delta C_{ij(t+T_{ij}^r)} + \Delta Q_{ij} \quad \text{when } y_{ijt}^r \leftarrow y_{ijt}^r + 1. \quad (4.23)$$

Each of the algorithms begins with  $y_{ijt}^r$  equal to zero for all  $t \in [0, \dots, T_{i0}]$  and iteratively assigns available stock at the central warehouse to field stocking locations according to a greedy rule. For all  $t \in [0, \dots, T_{i0}]$ , we use  $A_{i0t}$  to measure the cumulative number of units of item  $i$  that have been allocated up through period  $t$  (i.e., sent to field service locations from the central warehouse).

In the first algorithm, **GA**, the greedy rule is simple: Over all periods  $t$  in which stock is available for allocation, find the  $(j, t)$  combination that produces the largest objective function reduction  $\Delta Z_i(j, t)$ , and assign a unit of stock to be sent to location  $j$  in time period  $t$ . We now formally state this heuristic:

**(GA) A Greedy Algorithm for SAM<sub>i</sub>:**

**Step 0:** Set  $y_{ijt}^r \leftarrow 0$  for all  $j \in J, t \in [0, \dots, T_{i0}]$ . (Note that this implicitly sets  $S_{ijt} \leftarrow \tilde{S}_{ijt}$  for all  $t \in [T_{ij}^r, \dots, T_{ij}^r + T_{i0}]$ .) Set  $A_{i0t} \leftarrow 0$  for all  $t \in [0, \dots, T_{i0}]$ .

**Step 1:** If  $A_{i0T_{i0}} = \tilde{S}_{i0T_{i0}}$ , then **STOP** – no more allocations can be made. Otherwise, determine  $t^* = \min\{t : \forall t' \geq t, A_{i0t'} < \tilde{S}_{i0t'}\}$ , the earliest period in which a unit of stock is still available for allocation.

**Step 2:** For all  $j \in J, k \in [t^*, \dots, T_{i0}]$ , compute  $\Delta Z_i(j, k)$ , and determine  $(j^*, k^*) = \arg \min_{(j,k)} (\Delta Z_i(j, k))$ .

**Step 3:** If  $\Delta Z_i(j^*, k^*) \geq 0$ , then **STOP** – no further objective function reductions are possible. Otherwise, set  $y_{ij^*k^*}^r \leftarrow y_{ij^*k^*}^r + 1$  (so that implicitly  $S_{ij^*(T_{ij^*}^r+t)} \leftarrow S_{ij^*(T_{ij^*}^r+t)} + 1$  for all  $t \in [k^*, \dots, T_{i0}]$ ). For all  $t \in [k^*, \dots, T_{i0}]$ , set  $A_{i0t} \leftarrow A_{i0t} + 1$ . Go to Step 1.

The primary benefit of **GA** is that it only requires  $O(j)$  computation at each iteration. This

is because in Step 2, it can be shown that for each  $j \in J$ , the largest reduction  $\Delta Z_i(j, k)$  will be achieved in the first period  $k \geq t^*$  such that  $S_{ij(k+T_{ij}^r)} < \hat{S}_{ij(k+T_{ij}^r)}$ . Hence, not all periods  $k \in [t^*, \dots, T_{i0}]$  need to be checked. The total time required for the algorithm is  $O(jT_{i0} + j\tilde{S}_{i0T_{i0}})$ .

**GA** will not necessarily find the optimal solution to **SAM<sub>i</sub>**. The reason is that the availability of the central warehouse pipeline stock is spread out over time periods  $0, \dots, T_{i0}$ , but the algorithm allocates each unit of stock as if it were the very last unit available. That is, **GA** does not consider the fact that stock that will be available next period may be able to provide almost as much benefit to a location  $j$  as stock that is available now, but for another location  $j'$  waiting until next period may reduce the benefit significantly. Under certain conditions, however, **GA** will find the optimal solution to **SAM<sub>i</sub>**. Namely, if for some integer  $N$  we have  $\tilde{S}_{i0t} = N$  for all  $t \in [0, \dots, T_{i0}]$ , or if for some period  $k \in [0, \dots, T_{i0}]$  we have  $\tilde{S}_{i0t} = 0$  for all  $t \in [0, \dots, (k-1)]$  and  $\tilde{S}_{i0t} = N$  for all  $t \in [k, \dots, T_{i0}]$ , then **GA** will find the optimal solution to **SAM<sub>i</sub>**.

The second algorithm, **LGA**, is similar to the first, except that a look-ahead step is performed before an allocation decision is made. That is, the algorithm checks to see what the objective function reduction would be if the next *two* units of available stock were the last two units available.

**(LGA) A Look-Ahead Greedy Algorithm for SAM<sub>i</sub>:**

**Step 0:** Set  $y_{ijt}^r \leftarrow 0$  for all  $j \in J, t \in [0, \dots, T_{i0}]$ . (Note that this implicitly sets  $S_{ijt} \leftarrow \tilde{S}_{ijt}$  for all  $t \in [T_{ij}^r, \dots, T_{ij}^r + T_{i0}]$ .) Set  $A_{i0t} \leftarrow 0$  for all  $t \in [0, \dots, T_{i0}]$ .

**Step 1:** If  $A_{i0T_{i0}} = \tilde{S}_{i0T_{i0}}$ , then **STOP** – no more allocations can be made. Otherwise, determine  $t_1^* = \min\{t : \forall t' \geq t, A_{i0t'} < \tilde{S}_{i0t'}\}$ , the earliest period in which a unit of stock is still available for allocation, and  $t_2^* = \inf\{t \geq t_1^* : \forall t' \geq t, \tilde{S}_{i0t'} - A_{i0t'} \geq 2\}$ , the earliest period in which a *second* unit of stock is available for allocation. (Note that it is possible to have  $t_2^* = \infty$  if  $t_1^*$  is the last period in which stock is available for allocation and only one unit remains.)

**Step 2:** For all  $j \in J$ ,  $k \in [t_1^*, \dots, T_{i0}]$ , compute  $\Delta Z_i(j, k)$ . If  $\min_{(j,k)} \Delta Z_i(j, k) \geq 0$ , then **STOP** – no further objective function reductions are possible. Otherwise, go to step 3.

**Step 3:** If  $t_2^* = \infty$ , then for each  $j \in J$ , determine  $k_1^j = \arg \min_{k \in [t_1^*, \dots, T_{i0}]} (\Delta Z_i(j, k))$ , the locally best period in which to send another unit to location  $j$ , and set  $Change(j) = \Delta Z_i(j, k_1^j)$ . Otherwise, for each  $j \in J$ , determine  $k_1^j = \arg \min_{k \in [t_1^*, \dots, T_{i0}]} (\Delta Z_i(j, k))$  and  $(j_2, k_2) = \arg \min_{(j,k \in [t_2^*, \dots, T_{i0}])} (\Delta Z_i((j, k_1^j), (j_2, k_2)))$ , where  $\Delta Z_i((j, k_1^j), (j_2, k_2))$  represents the change in the objective function  $\Delta Z_i(j_2, k_2)$  after setting  $y_{j(k_1^j)}^r \leftarrow y_{j(k_1^j)}^r + 1$  (this is only a look-ahead check – we are not actually augmenting  $y_{j(k_1^j)}^r$  in this step); then set  $Change(j) = \Delta Z_i(j, k_1^j) + \min(0, \Delta Z_i((j, k_1^j), (j_2, k_2)))$ .

**Step 4:** Determine  $j^* = \arg \min_{j \in J} (Change(j))$ . Set  $y_{ij^*k_1^{j^*}}^r \leftarrow y_{ij^*k_1^{j^*}}^r + 1$  (so that implicitly  $S_{ij^*(T_{ij}^r+t)} \leftarrow S_{ij^*(T_{ij}^r+t)} + 1$  for all  $t \in [k_1^{j^*}, \dots, T_{i0}]$ ). For all  $k \in [k_1^{j^*}, \dots, T_{i0}]$ , set  $A_{i0k} \leftarrow A_{i0k} + 1$ . Go to Step 1.

**LGA** requires  $O(j^2)$  computation time for each iteration, and the total time required for the algorithm is  $O(jT_{i0} + j^2\tilde{S}_{i0T_{i0}})$ . Clearly, **LGA** can be generalized to be an  $n$ -step look-ahead algorithm for any  $n$ , but since the computation time for each iteration is  $O(j^{(n+1)})$ , such an algorithm is likely to be very computationally intensive for  $n \geq 3$ . We note, however, that such an  $n$ -step look-ahead algorithm will find the optimal solution to **SAM** <sub>$i$</sub>  if  $\tilde{S}_{i0T_{i0}} \leq (n+1)$ .

## 5 The Extended Stock Allocation Model

The *extended stock allocation model*, or **ESAM**, is a modification of the **SAM** in which the decision-maker has the option of using an expedited shipment mode to transport parts from the central warehouse to the field stocking locations, in addition to the regular shipment mode. In this section, we extend the formulation of the **SAM** to allow for an expedited mode of shipment and modify the solution methods presented in the previous section to accommodate this extension. Specifically, since Theorem 1 does not apply to the subproblems of the **ESAM**, we derive new bounds on the optimal values of the decision variables. Using

these new bounds, the subproblems can still be formulated and solved as linear programs. We also describe how the two greedy algorithms presented in the previous section can be modified for the **ESAM**.

## 5.1 Model Definition

Given the previously defined notation, we formulate the **ESAM** as follows:

$$\text{(ESAM) minimize} \quad \sum_{i \in I} \sum_{j \in J} \left\{ \sum_{t=T_{ij}^e}^{T_{ij}^r+T_{i0}} G_{ijt}(S_{ijt}) + Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) + \sum_{t=0}^{T_{i0}} e_{ij} y_{ijt}^e \right\} \quad (5.1)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t (y_{ijt'}^r + y_{ijt'}^e), \quad \forall i \in I, t = 0, \dots, T_{i0}, \quad (5.2)$$

$$S_{ijt} = \tilde{S}_{ijt} + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad \forall i \in I, j \in J, t = T_{ij}^e, \dots, T_{ij}^r - 1, \quad (5.3)$$

$$S_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)} + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r, \quad (5.4)$$

$$\forall i \in I, j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0},$$

$$y_{ijt}^e, y_{ijt}^r \geq 0 \text{ and integer} \quad \forall i \in I, j \in J, t = 0, \dots, T_{i0}. \quad (5.5)$$

Note that the objective function (5.1) contains a new term to capture the cost of expedited shipment. Also, note that we now use two sets of constraints, (5.3) and (5.4), to relate the quantities received at the field locations with the corresponding quantities shipped from the central warehouse. When there was only a regular shipment mode,  $t = T_{ij}^r$  was the earliest time period in which the allocation decisions could affect the cumulative stock level  $S_{ijt}$  of item  $i$  at location  $j$  received through period  $t$ . However, with the option of using expedited shipment, it is possible to affect  $S_{ijt}$  for time periods  $t = T_{ij}^e, \dots, T_{ij}^r - 1$  as well.

As with **SAM**, **ESAM** is separable by item, and the problem reduces to solving **ESAM<sub>i</sub>**, for each  $i \in I$ , given by:

$$\text{(ESAM}_i\text{) minimize} \quad \sum_{j \in J} \left\{ \sum_{t=T_{ij}^e}^{T_{ij}^r+T_{i0}} G_{ijt}(S_{ijt}) + Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) + \sum_{t=0}^{T_{i0}} e_{ij} y_{ijt}^e \right\} \quad (5.6)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t (y_{ijt'}^r + y_{ijt'}^e), \quad \forall t = 0, \dots, T_{i0}, \quad (5.7)$$

$$S_{ijt} = \tilde{S}_{ijt} + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad \forall j \in J, t = T_{ij}^e, \dots, T_{ij}^r - 1, \quad (5.8)$$

$$S_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)} + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad (5.9)$$

$$\forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0},$$

$$y_{ijt}^e, y_{ijt}^r \geq 0 \text{ and integer } \quad \forall j \in J, t = 0, \dots, T_{i0}. \quad (5.10)$$

## 5.2 LP Formulation of $\mathbf{ESAM}_i$

Unfortunately, Theorem 1 does not hold for  $\mathbf{ESAM}_i$ , and hence we must derive new bounds in order to construct a meaningful linear program. The reason that Theorem 1 does not hold for  $\mathbf{ESAM}_i$  is that while we can affect the cumulative stock levels  $S_{ijt}$  for  $t \in [T_{ij}^e, \dots, T_{ij}^r - 1]$  by sending expedited shipments in periods 0 through  $(T_{ij}^r - 1) - T_{ij}^e$ , we cannot *completely* control the cumulative stock pattern in these time periods. That is, the cumulative stock levels  $S_{ijt}$  for  $t \in [T_{ij}^e, \dots, T_{ij}^r - 1]$  have lower bounds of  $\tilde{S}_{ijt}$  because of what is in the pipeline at time 0, and while we have the ability to *augment* the pipeline with extra stock, we cannot *remove* stock that is already in the pipeline. This means that in addition to  $S_{ijt} \geq \tilde{S}_{ijt}$  for every  $t \in [T_{ij}^e, \dots, T_{ij}^r - 1]$ , every feasible solution also must have  $S_{ij(t+1)} - S_{ijt} \geq \tilde{S}_{ij(t+1)} - \tilde{S}_{ijt}$ . Because of this constraint, we will *not* necessarily have  $S_{ijt}^* \leq \hat{S}_{ijt}$ .

In order to derive new bounds for the decisions variables of  $\mathbf{ESAM}_i$ , we once again focus on the single-period expected cost function  $G_{ijt}$  and the constrained newsvendor problem  $\mathbf{CN}_{ijt}$  given in (4.9). For each  $j \in J$  and each  $t \in [T_{ij}^e, \dots, T_{ij}^r + T_{i0}]$ , let  $\hat{S}_{ijt}$  denote the *largest* optimal solution to  $\mathbf{CN}_{ijt}$ , and define:

$$M_{jt} = \max_{k \in [T_{ij}^e, \dots, t]} \{\hat{S}_{ijk} - \tilde{S}_{ijk}\}. \quad (5.11)$$

We then have the following theorem:

**Theorem 2** *For all field locations  $j \in J$  and all time periods  $t \in [T_{ij}^e, \dots, (T_{ij}^r + T_{i0})]$ , let  $\hat{S}_{ijt}$  denote the largest optimal solution to  $\mathbf{CN}_{ijt}$ , and let  $S_{ijt}^*$  denote the corresponding term*



within an optimal solution to **ESAM**<sub>*i*</sub>, and let  $M_{jt} = \max_{k \in [T_{ij}^e, \dots, t]} \{\hat{S}_{ijk} - \tilde{S}_{ijk}\}$ . Then:

$$\tilde{S}_{ijt} \leq S_{ijt}^* \leq \tilde{S}_{ijt} + M_{jt} \quad \forall j \in J, t \in [T_{ij}^e, \dots, (T_{ij}^r + T_{i0})]. \quad (5.12)$$

*Proof:* The first inequality must hold for any feasible solution. Only the second inequality requires proof. Suppose it does not hold for some field location  $j$ , and let  $k$  be the smallest index (i.e., the earliest time period) for which the location violates this condition. That is,  $S_{ijk}^* > \tilde{S}_{ijk} + M_{jk}$ , and  $S_{ijt}^* \leq \tilde{S}_{ijt} + M_{jt}$  for all  $t \in [T_{ij}^e, \dots, (k-1)]$ . (If  $k = T_{ij}^e$ , then note that  $S_{ij(k-1)}^* = S_{ij(T_{ij}^e-1)}^* = \tilde{S}_{ij(T_{ij}^e-1)} \leq \tilde{S}_{ij(T_{ij}^e-1)}$ .) We have then that

$$\begin{aligned} S_{ijk}^* - S_{ij(k-1)}^* &> (\tilde{S}_{ijk} + M_{jk}) - S_{ij(k-1)}^* \\ &\geq (\tilde{S}_{ijk} + M_{jk}) - (\tilde{S}_{ij(k-1)} + M_{j(k-1)}) \\ &\geq (\tilde{S}_{ijk} + M_{jk}) - (\tilde{S}_{ij(k-1)} + M_{jk}) \\ &= \tilde{S}_{ijk} - \tilde{S}_{ij(k-1)}. \end{aligned}$$

Since the  $\tilde{S}_{ijt}$  are nondecreasing, the optimal solution must allocate at least one unit of item  $i$  that arrives at location  $j$  in time period  $k$ . That is, we must have  $y_{ij(k-T_{ij}^e)}^{e*} > 0$  or  $y_{ij(k-T_{ij}^e)}^{r*} > 0$ . Without loss of generality, assume the latter, and consider the following minor changes to the optimal solution to **ESAM**<sub>*i*</sub>:

$$\begin{aligned} y_{ij(k-T_{ij}^r)}^r &\leftarrow y_{ij(k-T_{ij}^r)}^{r*} - 1 \quad \text{and} \\ y_{ij(k-T_{ij}^r+1)}^r &\leftarrow y_{ij(k-T_{ij}^r+1)}^{r*} + 1. \end{aligned}$$

The resulting solution is feasible. We are simply delaying the shipment of one unit from the central warehouse to location  $j$  by one period, so that one *less* unit arrives in period  $k$  and one *more* unit arrives in period  $k+1$ . Moreover, this modified solution will have  $S_{ijk} = S_{ijk}^* - 1$ , but for all  $t \neq k$ ,  $S_{ijt} = S_{ijt}^*$ . If  $k \neq T_{ij}^r + T_{i0}$ , then the only change to the objective function is that  $G_{ijk}(S_{ijk})$  replaces  $G_{ijk}(S_{ijk}^*)$ . But since  $G_{ijk}$  is convex in its argument and  $\hat{S}_{ijk}$  is the *largest* optimal solution to **CN**<sub>*ijt*</sub>, we have that  $G_{ijk}(\hat{S}_{ijk}) \leq G_{ijk}(\tilde{S}_{ijk} + M_{jk}) \leq G_{ijk}(S_{ijk}) = G_{ijk}(S_{ijk}^* - 1) < G_{ijk}(S_{ijk}^*)$ . If  $k = T_{ij}^r + T_{i0}$ , then the second term in the objective function will also change; but, since  $Q_{ij}$  is a strictly increasing function of its argument,  $Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) = Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}^* - 1) < Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}^*)$ . In either case, the

modified solution will have an objective function value that is strictly less than the value achieved by the original solution. Hence, the original solution cannot be optimal, and in any optimal solution we must have  $S_{ijt}^* \leq \tilde{S}_{ijt} + M_{jt}$  for all  $t = T_{ij}^e, \dots, (T_{ij}^r + T_{i0})$ .  $\square$

Theorem 2 provides upper and lower bounds on the cumulative stock levels in any optimal solution to **ESAM**<sub>*i*</sub>. As before, we can use these bounds to construct a linear programming formulation of **ESAM**<sub>*i*</sub>. Letting

$$\delta_{ijtk} = \begin{cases} 1 & \text{if } S_{ijt} = k, \\ 0 & \text{otherwise,} \end{cases} \quad (5.13)$$

we can reformulate **ESAM**<sub>*i*</sub> as follows:

$$\text{minimize } \sum_{j \in J} \left\{ \sum_{t=T_{ij}^e}^{T_{ij}^r+T_{i0}} \sum_{k=\tilde{S}_{ijt}}^{\tilde{S}_{ijt}+M_{jt}} \delta_{ijtk} G_{ijt}(k) + \sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\tilde{S}_{ij}(T_{ij}^r+T_{i0})+M_j(T_{ij}^r+T_{i0})} \delta_{ij(T_{ij}^r+T_{i0})k} Q_{ij}(k) + \sum_{t=0}^{T_{i0}} e_{ij} y_{ijt}^e \right\} \quad (5.14)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t (y_{ijt'}^r + y_{ijt'}^e), \quad \forall t = 0, \dots, T_{i0}, \quad (5.15)$$

$$\sum_{k=\tilde{S}_{ijt}}^{\tilde{S}_{ijt}+M_{jt}} \delta_{ijtk} \cdot k = \tilde{S}_{ijt} + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad \forall j \in J, t = T_{ij}^e, \dots, (T_{ij}^r - 1), \quad (5.16)$$

$$\sum_{k=\tilde{S}_{ij}(T_{ij}^r-1)}^{\tilde{S}_{ijt}+M_{jt}} \delta_{ijtk} \cdot k = \tilde{S}_{ij}(T_{ij}^r-1) + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad (5.17)$$

$$\forall j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0},$$

$$\sum_{k=\tilde{S}_{ijt}}^{\tilde{S}_{ijt}+M_{jt}} \delta_{ijtk} = 1 \quad \forall j \in J, t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}, \quad (5.18)$$

$$\delta_{ijtk} \in \{0, 1\} \quad \forall j \in J, t = T_{ij}^e, \dots, T_{ij}^r + T_{i0}, k = \tilde{S}_{ijt}, \dots, \tilde{S}_{ijt} + M_{jt}, \quad (5.19)$$

$$y_{ijt}^e, y_{ijt}^r \geq 0 \text{ and integer} \quad \forall j \in J, t = 0, \dots, T_{i0}. \quad (5.20)$$

As with the **SAM**<sub>*i*</sub>, solving the LP relaxation of the preceding ILP will result in an integer optimal solution. Hence, the integer restrictions in (5.19) and (5.20) can be dropped, and the solution to the resulting linear program will be an optimal solution to **ESAM**<sub>*i*</sub>.

### 5.3 Greedy Algorithms for ESAM<sub>i</sub>

Both of the greedy algorithms presented in Section 4.3 are easily modified to accommodate the ESAM<sub>i</sub>. The incremental changes in costs  $\Delta G_{ijt}(S_{ijt})$ ,  $\Delta C_{ijt}$ , and  $\Delta Q_{ij}$  are defined as before, but now to capture the total change in the objective function we must take into account the mode of shipment used in the allocation. That is, we now define:

$$\Delta Z_i(j, t) = \begin{cases} \Delta Z_i^r(j, t) = \Delta C_{ij(t+T_{ij}^r)} + \Delta Q_{ij} & \text{when } y_{ijt}^r \leftarrow y_{ijt}^r + 1, \\ \Delta Z_i^e(j, t) = \Delta C_{ij(t+T_{ij}^e)} + \Delta Q_{ij} + e_{ij} & \text{when } y_{ijt}^e \leftarrow y_{ijt}^e + 1. \end{cases} \quad (5.21)$$

The modified version of **GA**, called **EGA**, is given below. Similar changes are required to tailor the look-ahead algorithm **LGA** for the ESAM<sub>i</sub>, although in the interest of time and space we omit the details of **ELGA**. As with the original algorithms, the solutions resulting from the modified greedy algorithms **EGA** and **ELGA** may be used directly, or they can be used to quickly seed the ESAM<sub>i</sub> LP with a near-optimal solution.

#### (EGA) A Greedy Algorithm for ESAM<sub>i</sub>:

**Step 0:** Set  $y_{ijt}^m \leftarrow 0$  for all  $j \in J, t \in [0, \dots, T_{i0}], m \in [r, e]$ . (Note that this implicitly sets  $S_{ijt} \leftarrow \tilde{S}_{ijt}$  for all  $t \in [T_{ij}^e, \dots, T_{ij}^r + T_{i0}]$ .) Set  $A_{i0t} \leftarrow 0$  for all  $t \in [0, \dots, T_{i0}]$ .

**Step 1:** If  $A_{i0T_{i0}} = \tilde{S}_{i0T_{i0}}$ , then **STOP** – no more allocations can be made. Otherwise, set  $t^* = \min\{t : \forall t' \geq t, A_{i0t'} < \tilde{S}_{i0t'}\}$ , the earliest period in which a unit of stock is still available for allocation.

**Step 2:** For all  $j \in J, k \in [t^*, \dots, T_{i0}]$ , compute  $\Delta Z_i^r(j, k)$  and  $\Delta Z_i^e(j, k)$ , and determine  $(j^*, k^*, m^*) = \arg \min_{(j, k, m)} (\Delta Z_i^m(j, k))$ .

**Step 3:** If  $\Delta Z_i^{m^*}(j^*, k^*) \geq 0$ , then **STOP** – no further objective function reductions are possible. Otherwise, set  $y_{ij^*k^*}^{m^*} \leftarrow y_{ij^*k^*}^{m^*} + 1$  (so that implicitly  $S_{ij^*(T_{ij^*}^{m^*} + t)} \leftarrow S_{ij^*(T_{ij^*}^{m^*} + t)} + 1$  for all  $t \in [k^*, \dots, T_{i0}]$ ). For all  $k \in [k^*, \dots, T_{i0}]$ , set  $A_{i0k} \leftarrow A_{i0k} + 1$ . Go to Step 1.

We now turn our attention to an operating environment in which repair decisions as well as inventory allocation decisions must be made.

## 6 The Extended Stock Allocation Model with Repair

The premise for the *extended stock allocation model with repair*, or **ESAMR**, is that a repair process, shared by all items, is the supply source for the central warehouse. The repair allocation decisions to be made are what items should be entered into repair in period  $t = 0$ .

Each item type  $i$  is assumed to have a fixed repair time from the time the repair commences. Units of item type  $i$  that are selected to enter repair in period 0 will reach the central warehouse in time period  $T_{i0}$ , the last period of the planning horizon for item  $i$ ; hence, this supply will affect the costs realized over the effective horizon for item  $i$ .

The modeling extension is fairly straightforward. First, we introduce the following additional notation:

- $R_i$  - the number of units of item  $i$  available for repair in time period 0.
- $C_{max}$  - the maximum number of units (over all item types) that can be entered into repair in time period 0 (i.e., the repair facility capacity).
- $C_{min}$  - the minimum number of units (over all item types) that must be entered into repair in time period 0. We assume that  $C_{min} \leq \sum_{i \in I} R_i$ . In many practical applications, this parameter would likely be set to  $\min(\sum_{i \in I} R_i, C_{max})$  so that as many units as possible are repaired.
- $C_i$  -  $\min(R_i, C_{max})$ , the maximum number of units of item  $i$  that can enter the repair process in time period 0.
- $h_{i0}$  - the incremental cost per period associated with holding a unit of item  $i$  at the central warehouse instead of in the repair queue.
- $Q_{i0}(\cdot)$  - the function approximating the expected incremental holding costs incurred beyond the end of the planning horizon for item  $i$  at the central warehouse. The function argument is  $\tilde{S}_{i0T_{i0}}$ , the cumulative supply of item  $i$  at the central warehouse at the end of the planning horizon. (See below.)
- $v_i$  - the number of units of item  $i$  to enter into repair in time period 0.

The end-of-horizon expected incremental holding cost function in **ESAMR** is given by:

$$Q_{i0}(\tilde{S}_{i0T_{i0}}) = h_{i0} \sum_{t=T_{i0}+1}^{\infty} E[(\sum_{j \in J} \tilde{S}_{ij(T_{ij}^r-1)} + \tilde{S}_{i0T_{i0}}) - \sum_{j \in J} X_{ijt}]^+. \quad (6.1)$$

As with  $Q_{ij}$ ,  $Q_{i0}$  is a convex function of its argument. By incorporating  $Q_{i0}$  into the objective function of **ESAMR**, we provide a mechanism for determining the most cost-effective use of repair capacity. That is, we will prioritize items to be repaired that are likely to be needed in the near future instead of repairing items that may not be needed for some time. Note that for each period beyond the end of the planning horizon, the expression (6.1) charges an incremental holding cost on the expected excess of the cumulative total supply (in the subsystem rooted at the central warehouse) over the cumulative total demand. As such, it is only an *approximation* of the true expected incremental holding costs since it implicitly assumes that the cumulative total supply is fungible in satisfying the cumulative total demand.

Having defined these additional parameters, the *extended stock allocation model with repair*, or (**ESAMR**), is:

(**ESAMR**)

$$\text{minimize } \sum_{i \in I} \left[ \sum_{j \in J} \left\{ \sum_{t=T_{ij}^e}^{T_{ij}^r+T_{i0}} G_{ijt}(S_{ijt}) + Q_{ij}(S_{ij(T_{ij}^r+T_{i0})}) + \sum_{t=0}^{T_{i0}} e_{ij} y_{ijt}^e \right\} + Q_{i0}(\tilde{S}_{i0T_{i0}}) \right] \quad (6.2)$$

subject to

$$\tilde{S}_{i0t} \geq \sum_{j \in J} \sum_{t'=0}^t (y_{ijt'}^r + y_{ijt'}^e), \quad \forall i \in I, t = 0, \dots, T_{i0}, \quad (6.3)$$

$$S_{ijt} = \tilde{S}_{ijt} + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad (6.4)$$

$$\forall i \in I, j \in J, t = T_{ij}^e, \dots, T_{ij}^r - 1,$$

$$S_{ijt} = \tilde{S}_{ij(T_{ij}^r-1)} + \sum_{t'=0}^{t-T_{ij}^r} y_{ijt'}^r + \sum_{t'=0}^{\min(t-T_{ij}^e, T_{i0})} y_{ijt'}^e, \quad (6.5)$$

$$\forall i \in I, j \in J, t = T_{ij}^r, \dots, T_{ij}^r + T_{i0},$$

$$\tilde{S}_{i0(T_{i0})} = \tilde{S}_{i0(T_{i0}-1)} + v_i, \quad \forall i \in I, \quad (6.6)$$

$$C_{min} \leq \sum_{i \in I} v_i \leq C_{max}, \quad (6.7)$$

$$0 \leq v_i \leq C_i \text{ and integer } \forall i \in I, \quad (6.8)$$

$$y_{ijt}^e, y_{ijt}^r \geq 0 \text{ and integer } \forall i \in I, j \in J, t = 0, \dots, T_{i0}. \quad (6.9)$$

Since  $Q_{i0}$  is a convex function of its argument, the objective function remains convex in each  $S_{ijt}$ . Constraints (6.6), (6.7), and (6.8) are new. Constraints (6.6) replace the previously deterministic values  $\tilde{S}_{i0(T_{i0})}$  with  $\tilde{S}_{i0(T_{i0}-1)}$  plus whatever gets entered into repair in period 0. Constraint (6.7) ensures that we meet the minimum repair requirements without exceeding repair capacity, and constraints (6.8) ensure that we only repair what is available to repair. Unlike the **SAM** and **ESAM**, **ESAMR** is *not* separable by item due to the capacity constraint (6.7). However, with a little more work we can represent **ESAMR** in a form that is easily solvable.

Given an instance of **ESAMR**, let  $Z_i^*(s)$  denote the optimal objective function value of the subproblem **ESAM** $_i$ , subject to the condition that  $\tilde{S}_{i0T_{i0}} = s$ . Suppose that we can determine  $Z_i^*(s)$  for each  $i \in I$  and for each  $s \in \{\tilde{S}_{i0(T_{i0}-1)}, \dots, \tilde{S}_{i0(T_{i0}-1)} + C_i\}$ . Letting

$$\delta_{ik} = \begin{cases} 1 & \text{if } v_i = k, \\ 0 & \text{otherwise,} \end{cases} \quad (6.10)$$

we can write **ESAMR** as follows:

$$\text{minimize} \quad \sum_{i \in I} \left[ \sum_{k=0}^{C_i} Z_i^*(\tilde{S}_{i0(T_{i0}-1)} + k) + Q_{i0}(\tilde{S}_{i0(T_{i0}-1)} + k) \right] \delta_{ik} \quad (6.11)$$

subject to

$$\sum_{k=0}^{C_i} \delta_{ik} = 1, \quad \forall i \in I, \quad (6.12)$$

$$C_{min} \leq \sum_{i \in I} \sum_{k=0}^{C_i} k \delta_{ik} \leq C_{max}, \quad (6.13)$$

$$\delta_{ik} \in \{0, 1\}, \quad \forall i \in I, k = 0, \dots, C_i. \quad (6.14)$$

It is not hard to show that for each item  $i$ , the optimal **ESAM** $_i$  value  $Z_i^*(\tilde{S}_{i0(T_{i0}-1)} + k)$  is non-increasing and (discretely) convex in  $k$ . (This convexity depends critically on the fact that the  $k$  additional units of available inventory all arrive at the central warehouse in the *same* planning horizon period  $T_{i0}$ . If the inventory were to arrive in different periods, then we could not guarantee convexity.) Moreover,  $Q_{i0}(\tilde{S}_{i0(T_{i0}-1)} + k)$  is increasing and convex in  $k$ . Hence, the objective function (6.11) captures the tradeoffs among the item types of

using repair capacity to make more inventory available for use at the central warehouse and potentially incurring incremental holding costs on this additional inventory. Because of the convexity of (6.11), the following greedy marginal analysis algorithm, **EGAR**, can be used to solve **ESAMR** to optimality, provided that the optimal  $Z_i^*$  values are used.

**(EGAR) A Greedy Algorithm for ESAMR:**

**Step 0:** For all  $i \in I$ ,  $k = 0, \dots, C_i$ , determine

$$W_i(k) = Z_i^*(\tilde{S}_{i0(T_{i0}-1)} + k) + Q_{i0}(\tilde{S}_{i0(T_{i0}-1)} + k).$$

Set  $v_i \leftarrow 0$  for all  $i \in I$ . Set  $A \leftarrow 0$ .

**Step 1:** If  $A = C_{max}$ , then **STOP** – no more repair capacity is available. Otherwise, for all  $i \in I$  with  $v_i < C_i$ , compute

$$\Delta W_i(v_i) = W_i(v_i + 1) - W_i(v_i)$$

and determine  $i^* = \arg \min_{i \in I} (\Delta W_i(v_i))$ .

**Step 2:** If  $\Delta W_{i^*}(v_{i^*}) \geq 0$  and  $A \geq C_{min}$ , then **STOP** – no further objective function reductions are possible. Otherwise, set  $v_{i^*} \leftarrow v_{i^*} + 1$ , and  $A \leftarrow A + 1$ . Go to Step 1.

Even if the greedy algorithm **EGA** is used to solve the **ESAM<sub>i</sub>** subproblems, so that the resulting values  $Z_i^*(\tilde{S}_{i0(T_{i0}-1)} + k)$  for  $k = 0, \dots, C_i$  are only approximate, it can be shown that these values will remain non-increasing and convex in  $k$ . Hence, **EGAR** can be used to make real-time repair and inventory allocation decisions using an exact or approximate method for solving the **ESAM<sub>i</sub>** subproblems. Note also that since this approach is largely separable by item, the computational effort required to perform sensitivity analysis is minimal. As part of our computational experiments, we implement **EGAR** with **EGA** embedded and compare the quality of the solutions with the exact approach.

## 7 Numerical Study

We had two primary objectives in conducting numerical experiments. First, we wanted to compare the quality of the solution resulting from our heuristic allocation techniques with the optimal solution to the **ESAMR** under a variety of different operating conditions.

Second, and more important, we wanted to test the conjecture that integrated real-time decision models are of significant operational value in dynamic environments where inventory imbalances exist, as is often the case in actual service parts supply chains. As we noted earlier, inventory levels found in service parts supply chains are often too high or too low for the current operating situation. This can occur because an item's inventory position often must be decided far in advance of its availability. When demands eventually arise for the item, the available quantity may not be appropriate for the current demand processes. See Ramey (1999), for a detailed description and example of this phenomenon for the C-5 Galaxy aircraft.

To address the first objective, we implemented the **EGAR** algorithm outlined in Section 6, where the item subproblems were solved approximately using the **EGA** heuristic from Section 5.3. We compared the solution from this approach with the optimal solution to the **ESAMR**, which we obtained by solving a linear programming version of the **ESAMR** formulation given in (6.2)-(6.9). A simulator was used to randomly generate problem instances for this comparison. Details are given in Section 7.2.

To address the second objective, we simulated the continuous operation of a service parts supply chain using two different methods for making allocation decisions, and compared the long-run performance of these methods. The first method used the **EGAR** algorithm (with the **EGA** heuristic embedded) to jointly make repair and inventory allocation decisions in a rolling-horizon manner. The second method employed a decentralized approach in which the **EGAR** algorithm was used for making inventory allocation decisions, but a first-come, first-served rule was used to manage the repair queue. Details of the experiment are given in Section 7.3.



## 7.1 Testing Environment

To facilitate our numerical study, a large-scale periodic-review service parts supply chain simulator was developed and used to create operating system states for different supply system configuration. A supply system configuration is defined by the following factors:

- the number of items and locations;
- the lead times for repair, regular transport, and expedited transport for each item and location;
- the mean and variance of demand each period for each item and location;
- the incremental holding, backorder, and transportation costs for each item and location;
- the maximum repair capacity each period; and
- the total system inventory level of each item.

In the numerical experiments for testing the value of the integrated approach, the following factors were held constant: There were 12 items demanded at 5 locations. Repair, expedited transport, and regular transport lead times were 3, 1, and 2 periods, respectively. The return transportation lead time for failed items to the repair facility was 5 periods. The incremental holding costs at the central warehouse ranged between \$0.14 and \$0.27 per unit per period across the items. The incremental holding costs at the demand locations were twice that of the central warehouse and ranged between \$0.28 and \$0.54 per unit per period across the items. The backorder cost was 9 times the holding cost rate for each item at each location. The incremental unit cost of expedited transport was \$0.80. The factors that were varied include the repair capacity utilization (80%, 90% and 95%) and the demand process variance-to-mean ratios (1 and 10).

In the numerical experiments for testing the quality of the **EGAR** heuristic approach, we used the same factor values just described, expect that two values of the regular transport time were tested (2 and 5), and two backorder-to-holding-cost ratios were tested (9 and 20).

For both experiments, the demand processes for each item at each location were taken to be stationary and independent over time and location. While our methodology can handle

	<b>Item Demand Rates</b>											
<b>Location</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>1</b>	6.2	5.0	4.0	3.0	1.2	0.6	6.2	5.0	4.0	3.0	1.2	0.6
<b>2</b>	6.2	5.0	4.0	3.0	1.2	0.6	6.2	5.0	4.0	3.0	1.2	0.6
<b>3</b>	6.2	5.0	4.0	3.0	1.2	0.6	6.2	5.0	4.0	3.0	1.2	0.6
<b>4</b>	6.2	5.0	4.0	3.0	1.2	0.6	6.2	5.0	4.0	3.0	1.2	0.6
<b>5</b>	6.2	5.0	4.0	3.0	1.2	0.6	6.2	5.0	4.0	3.0	1.2	0.6
<b>Item System Inventory Levels</b>	515	410	325	240	115	60	255	220	155	85	40	15

Table 1: An example of the item demand rates and system inventory levels.

more complicated non-stationary demand processes, we chose to conduct our experiments to demonstrate the value of making integrated decisions even with relatively stable demand processes. Two types of demand distributions were used, Poisson (to simulate a variance-to-mean ratio of 1) and negative binomial (to simulate a variance-to-mean ratio of 10).

The demand rates for the 12 items were established as follows. For the first 6 items, a mixture of high, medium, and low demand rates were set to represent a composite set of items, as is typically found in practice. The demand rates for the second 6 items were identical to those of the first 6 items. The demand rates for each item were identical across the locations.

The system inventory levels for the first 6 items were set to values that were slightly above those required to minimize their steady-state expected holding and backorder costs. We did this to represent items that are in long supply. For the second 6 items, the system inventory levels were set to approximately one-half of the inventory levels of the first 6 items, representing items that are in short supply. An example is given in Table 1. (The inventory levels shown in Table 1 correspond to an instance in which the demand processes are Poisson processes.) Thus, the testing environment contains a mixture of high, medium, and low demand rate items, some of which have ample inventory and others of which do not have sufficient inventory in the supply chain.

## 7.2 Exact Approach vs. EGAR Heuristic

For this comparison, we chose to test supply system configurations that exhibit conditions under which the performance of the methods is likely to differ. Twenty-four different configurations were tested in all. For each configuration, one thousand operating system states were randomly generated, using the simulator, to create problem instances. We applied the **EGAR** algorithm and the exact approach to each instance. We used the simulator to create instances instead of creating them manually to avoid potentially biasing the experimental results with statistically unlikely system states.

Utilization	Demand VTMR	Regular Lead Time = 2		Lead Time 2 Average	Regular Lead Time = 5		Lead Time 5 Average	Average
		Backorder Cost Multiple = 9	Backorder Cost Multiple =20		Backorder Cost Multiple = 9	Backorder Cost Multiple =20		
80%	1	0.85%	0.95%	0.90%	0.68%	0.91%	0.79%	0.85%
	10	0.93%	0.79%	0.86%	0.85%	1.03%	0.94%	0.90%
80% Average		0.89%	0.87%	0.88%	0.77%	0.97%	0.87%	0.87%
90%	1	0.83%	0.95%	0.89%	0.78%	0.72%	0.75%	0.82%
	10	0.84%	1.02%	0.93%	1.11%	0.95%	1.03%	0.98%
90% Average		0.84%	0.98%	0.91%	0.95%	0.83%	0.89%	0.90%
95%	1	1.04%	0.75%	0.90%	0.89%	0.94%	0.91%	0.90%
	10	1.07%	0.85%	0.96%	1.29%	1.00%	1.15%	1.05%
95% Average		1.06%	0.80%	0.93%	1.09%	0.97%	1.03%	0.98%
<b>Average</b>		0.93%	0.88%	0.91%	0.93%	0.92%	0.93%	0.92%

Table 2: Average deviations of the EGAR cost above the ESAMR optimal cost.

Table 2 contains the results for each of the twenty-four configurations examined. The numbers in the table represent the average percentage deviation of the expected effective horizon cost incurred by employing the **EGAR** heuristic from the expected effective horizon cost incurred by using the optimal **ESAMR** solution.

There are several questions of interest to discuss:

1. *How frequently did the decisions resulting from the two approaches match?* Matches in every single allocation decision were observed for many of the problem instances (i.e., operating states) tested; however, complete matches were not a usual or frequent occurrence. Due to the combinatorial nature of the problem, small differences in repair decisions, shipment timing, and quantity were common, and we found it difficult to devise a meaningful measure of the “closeness” of two solutions directly. It is clear from Table 2, however, that these small differences did not greatly affect the resulting expected costs.

*2. How close are the expected costs over the effective horizon resulting from the two approaches?* As Table 2 shows, the average relative difference between the expected costs of the **EGAR** heuristic and the exact approach were very small over all supply system configurations tested. The largest overall cost differences occurred when the demand process variance-to-mean ratios were high, and repair capacity utilization was high. This combination caused large and frequent shortages to occur for long periods of time, which, in turn, magnified the small deviations from the optimal inventory allocation that were made by the myopic **EGAR** heuristic. However, at the item level, the most significant factor in determining the quality of the **EGAR** solution relative to the exact solution was the choice of system inventory level. The simple explanation for this is that the constrained newsvendor cost functions are relatively flat in the area of their minima. When there is sufficient inventory of an item to meet most demand requirements, small deviations (in terms of quantity shipped and shipment timing) from the optimal operating allocation will have relatively little impact on the overall system cost incurred for this item. In such cases, the performance of most reasonable allocation policies is likely to be good. Since the **EGAR** heuristic attempts to minimize expected costs, albeit myopically, it performs well when the system, as a whole, has enough stock. When the system inventory level for an item is far too low, small deviations from the optimal solution will produce much larger relative cost differences than when the system has sufficient inventory. It is in these cases that the logic and robustness of the allocation rule becomes extremely important. Our experiments show that the **EGAR** heuristic performs well even when there are significant inventory imbalances in the system. (While the method for choosing item inventory levels is not detailed explicitly in this paper, there is a body of relevant literature. See Chan et al. (1999) and Caggiano et al. (2001).)

*3. What is the magnitude of the computational advantage of using the EGAR heuristic (with EGA embedded) over the exact method?* On average, the exact method's solution time was approximately five times longer than that of the **EGAR** heuristic. On a 700MHz Intel-based computer, the exact algorithm and the **EGAR** heuristic averaged 10 seconds and 2 seconds, respectively. The **EGAR** heuristic is much faster than the exact method because the necessary expected cost computations are done on an as-needed

basis. The majority of the processing time for the exact method is devoted to computing all of the necessary cost coefficients found in the **ESAMR** model.

### 7.3 First-Come, First-Served vs. Integrated Repair Allocations

As stated earlier, to test the value of an integrated approach over a decentralized one, we implemented two different methods for making allocation decisions. The first method used the **EGAR** algorithm (with the **EGA** heuristic embedded) to jointly make repair and inventory allocation decisions. The second method employed a decentralized approach in which the **EGAR** algorithm was used for making inventory allocation decisions, but a first-come, first-served rule (FCFS) was used to manage the repair queue. That is, the repair decisions were made with no knowledge of downstream requirements and were based solely on the order in which items were returned for repair. We chose not to implement a FCFS rule for *both* inventory allocation and repair since such an approach will perform less well than FCFS for repair decisions and **EGAR** for inventory allocation decisions.

For each supply system configuration examined, the simulation was run for one thousand time periods, after a warmup phase. Five separate random seeds were employed, along with their antithetic random number streams, for a total of ten simulations runs for each configuration. The outcomes for different levels of repair capacity utilization and demand uncertainty are summarized in Table 3. The data in the table indicate the average per period cost observed across the ten runs and the standard deviation of these average per period cost values. These results show that by considering repair and inventory allocation decisions jointly, costs are reduced by 13.3% on average, with the reductions ranging between 9.2% and 16.0%.

Note that these results are based on fairly benign stationary demand processes. Real demand data are typically non-stationary and exhibit attributes for which our integrated approach will likely outperform simple and/or decentralized allocation rules by a wider margin. Furthermore, although it is outside the scope of this paper, our experiments revealed that by allowing lateral transshipments in our integrated model, costs were lowered by an additional 10.1%, on average. This demonstrates that there is potential for significant economic value in using an integrated repair and stock allocation decision process.

Repair Capacity Utilization	Demand Process VTMR	Allocation Rule		Relative Cost Improvement of ESAMR
		FCFS	ESAMR	
80%	1	\$1,712.71 ±\$11.70	\$1,443.95 ±\$13.05	15.7%
	10	\$1,763.01 ±\$26.89	\$1,554.13 ±\$26.50	11.8%
90%	1	\$1,713.15 ±\$11.73	\$1,443.80 ±\$13.03	15.7%
	10	\$1,773.05 ±\$29.25	\$1,569.92 ±\$29.61	11.5%
95%	1	\$1,716.65 ±\$11.71	\$1,442.60 ±\$12.92	16.0%
	10	\$1,802.50 ±\$40.32	\$1,636.45 ±\$56.32	9.2%
<b>Average</b>		<b>\$1,746.84 ±\$21.93</b>	<b>\$1,515.14 ±\$25.24</b>	<b>13.3%</b>

Table 3: Simulation results for FCFS vs. ESAMR.

## 8 Concluding Remarks

In this paper, we developed an integrated real-time model for making repair and inventory allocation decisions in a two-echelon repairable service parts system. Our objectives were to present this operational planning problem in a form that can be readily solved on a rolling horizon basis for large-scale systems and to demonstrate the value of using such models. To this end, we formulated the decision problem as a large-scale linear program, and we developed practical heuristic methods for solving the problem. Although we do not discuss the details here, we note that the model can easily be extended to include transshipments between field stocking locations as well as the pipeline of materials in-transit to the repair facility.

Our numerical analyses showed that for making inventory allocation decisions, the **EGAR** heuristic is highly competitive with the exact approach and is much more computationally efficient. Moreover, we observed that for operating situations in which inventory imbalances exist, such as in Ramey (1999), the **ESAMR** decision approach performed substantially better than the decentralized approach that uses **EGAR** for inventory allocation and a first-come, first-served rule for repair decisions. This demonstrates the importance of using an integrated decision approach in many dynamic operating environments.

## References

- Aggarwal, P. K. and Moinzadeh, K. (1994), "Order Expedition in Multi-Echelon Production / Distribution Systems," *IIE Transactions*, 26, 86–96.
- Caggiano, K. E., Jackson, P. L., Muckstadt, J. A., and Rappold, J. A. (2001), "Optimal Stocking in Repairable Parts Networks with Repair Capacity and Inventory Pooling," *Technical Report, School of OR&IE, Cornell University, Ithaca, NY 14853*.
- Chan, E. W., Rappold, J. A., and Muckstadt, J. A. (1999), "Determining and Allocating Capacity-Driven Safety Stock in Multi-Item, Multi-Echelon Systems," *School of Business, University of Wisconsin, Madison, WI*.
- Dantzig, G. (1962), *Linear Programming and Extensions*, Princeton, NJ: Princeton University Press.
- Duenyas, I. (1994), "A Simple Release Policy for Networks of Queues with Controllable Inputs," *Operations Research*, 42, 1162–1171.
- Feeney, G. J. and Sherbrooke, C. C. (1966), "The (S-1,S) Inventory Policy under Compound Poisson Demand," *Management Science*, 12, 391–411.
- Gross, D., Kahn, H., and Marsh, J. (1977), "Queueing Models for Spares Provisioning," *Naval Research Logistics Quarterly*, 24, 521–536.
- Hausman, W. H. and Scudder, G. D. (1982), "Priority Scheduling Rules for Repairable Inventory Systems," *Management Science*, 29, 1215–1232.
- Hillstead, R. J. (1982), *Dyna-METRIC: A Mathematical Model for Capability Assessment and Supply Requirements when Demand, Repair, and Resupply are Non-Stationary*, Santa Monica, California: The RAND Corporation R-2785-AF.
- Hillstead, R. J. and Carrillo, M. J. (1980), *Models and Techniques for Recoverable Item Stockage when Demand and the Repair Process are Nonstationary - Part I: Performance Measurement*, Santa Monica, California: The RAND Corporation N-1482-AF.

- Hoadley, B. and Heyman, D. P. (1977), “A Two-Echelon Inventory Model with Purchases, Dispositions, Shipments, Returns, and Transshipments,” *Naval Research Logistics Quarterly*, 24, 1–19.
- Kim, J., Leachman, R. C., and Suh, B. (1996), “Dynamic Release Control Policy for the Semiconductor Wafer Fabrication Lines,” *Journal of the Operational Research Society*, 47, 1516–1525.
- Maxwell, W. L. (1969), *Priority Dispatching and Assembly Operations in a Job Shop*, Santa Monica, California: The RAND Corporation RM-5370-PR.
- Miller, B. L. (1968), *Real Time METRIC for the Distribution of Serviceable Assets*, Santa Monica, California: The RAND Corporation RM-5687-PR.
- Moinzadeh, K. and Aggarwal, P. K. (1997), “An Information Based Multiechelon Inventory System with Emergency Orders,” *Operations Research*, 45, 694–701.
- Moinzadeh, K. and Schmidt, C. P. (1991), “An  $(S - 1, S)$  Inventory System with Emergency Orders,” *Operations Research*, 39, 308–321.
- Muckstadt, J. A. (1980), *Comparative Adequacy of Steady-State Versus Dynamic Models for Calculating Stockage Requirements*, Santa Monica, California: The RAND Corporation R-2636-AF.
- Peña Perez, A. and Zipkin, P. (1997), “Dynamic Scheduling Rules for a Multiproduct Make-to-Stock Queue,” *Operations Research*, 45, 919–930.
- Pyke, D. F. (1990), “Priority Repair and Dispatch Policies for Repairable-Item Logistics Systems,” *Naval Research Logistics*, 37, 1–30.
- Ramey, T. L. (1999), “Lean Logistics: High-Velocity Logistics Infrastructure and the C-5 Galaxy,” Tech. Rep. MR-581-AF, RAND Corporation, Santa Monica, CA.
- Scudder, G. D. (1984), “Priority Scheduling and Spares Stocking Policies for a Repair Shop: The Multiple Failure Case,” *Management Science*, 30, 739–749.



- Scudder, G. D. and Hausman, W. H. (1982), "Spares Stocking Policies for Repairable Items with Dependent Repair Times," *Naval Research Logistics Quarterly*, 29, 303–322.
- Sherbrooke, C. C. (1968), "METRIC: A Multi-Echelon Technique for Recoverable Item Control," *Operations Research*, 16, 122–141.
- Slay, F. M. (1984), "VARI-METRIC: An Approach to Modeling Multi-Echelon Resupply when the Demand Process is Poisson with Gamma Prior," *Working Note AF301-3, Logistics Management Institute*.
- Verrijdt, J., Adan, I., and de Kok, T. (1998), "A Trade-Off Between Emergency Repair and Inventory Investment," *IIE Transactions*, 30, 119–132.
- Wein, L. M. (1992), "Scheduling Networks of Queues: Heavy Traffic Analysis of a Multistation Network with Controllable Inputs," *Operations Research*, 40, S312–S334.
- Wein, L. M. and Chevalier, P. B. (1992), "A Broader View of the Job-Shop Scheduling Problem," *Management Science*, 38, 1018–1033.