Reducing Total Fulfillment Costs at Amazon EU Through Network Design Optimization

By
Ken Merriam

B.S. Civil Engineering Systems
University of Pennsylvania, 2002

Submitted to the MIT Sloan School of Management
and the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degrees of

Master of Business Administration
and
Master of Science in Civil and Environmental Engineering

In conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology
June 2007

Signature of Author: ______________________
Department of Civil and Environmental Engineering
MIT Sloan School of Management
May 11, 2007

Certified by: ________
Donald Rosenfield, Thesis Supervisor
Senior Lecturer, MIT Sloan School of Management
Director, Leaders for Manufacturing Fellows Program

Certified by: ______________________
David Simchi-Levi, Thesis Supervisor
Professor of Civil and Environmental Engineering
Co-Director, Leaders for Manufacturing Fellows Program

Accepted by: ______________________
Debbie Berechman, Executive Director of the MBA Program
MIT Sloan School of Management

Accepted by: ______________________
Daniele Veneziano, Chairman Departmental Committee for Graduate Students
Department of Civil and Environmental Engineering

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Reducing Total Fulfillment Costs at Amazon EU Through Network Design Optimization

By
Ken Merriam

Submitted to the MIT Sloan School of Management
and the Department of Civil and Environmental Engineering on May 11, 2007
in partial fulfillment of the requirements for the degrees of
Master of Business Administration and
Master of Science in Civil and Environmental Engineering

Abstract

A key supply chain management issue encountered by any business requiring a distribution system is in designing its distribution network. A distribution network configuration has both direct and indirect ongoing effects on how a firm operates influencing everything from supplier relationships and contracts to customer interface. A configuration affects both day-to-day operational and longer range strategic and tactical decision-making. From a pure cost perspective, a configuration has a significant impact on total fulfillment costs.

The effects of network configuration as well as the challenges and value behind the application of network design optimization techniques are well-illustrated by my 6-month experience working for the largest online retail distributor, Amazon.com, in their European (EU) operations. This paper further documents the process followed in identifying areas for improvement in Amazon's current EU fulfillment networks for the purpose of enabling total fulfillment cost reduction. The challenges and results from my experience are similarly included.

Two main projects were ultimately selected and documented in this paper: The first was aimed at minimizing transportation costs around the existing UK network configuration, while the second was targeted at minimizing total fulfillment costs through the alteration of EU network designs through the focused adjustment of product and inventory distributions. The first project has to date enabled significant minimization of UK transportation costs. The second project dealt with two complicated mathematical formulations ultimately intended for optimization, one of which is not yet covered in literature. In this case, further research and investigation is required for its practical implementation; nonetheless, the developed formulation was applied to a simplified scenario for the purpose of future study including validation and extension.

The ultimate objective of this paper however is to demonstrate the hidden potential and value behind the application of underutilized analytical techniques to network design through the tailored development and implementation of practical decision-support systems.

Thesis Supervisor: Donald Rosenfield
Title: Senior Lecturer, MIT Sloan School of Management

Thesis Supervisor: David Simchi-Levi
Title: Professor of Civil and Environmental Engineering

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Acknowledgements

I would like to acknowledge everyone who has supported and advised me throughout my studies at MIT which now culminates with this thesis. As such, I would like to start by thanking everyone at the Leaders for Manufacturing (LFM) Program. For almost 20 years, the program has offered LFM students the unique opportunity to put the latest tools and research in Operations Management as taught at MIT into practice. This experiential learning is made possible through LFM's unique partnership with leading industry firms such as Amazon that enable students such as myself to experience Operations Management first-hand.

On this point, I would like to thank Amazon for their support throughout the 6-month internship. Amazon's impressive growth rate along with its entrepreneurial and data-driven culture offered me a rich array of learning experiences while providing me with a great set of opportunities to apply those tools taught at LFM in a supportive context. At Amazon, I would like to especially acknowledge my internship supervisor Steven Harman, LFM Class of 1997, who always ensured I had access to the necessary information along with the required visibility to implement and present my results. More importantly, he provided me with the guidance and insight that enabled me to focus on those issues with most potential.

In addition, I would like to thank the rest of the Amazon EU Operations Management team including Allan Lyall, Nicholas Ng, Chris Peyre, Ingrid Ebner, Gerd Meyer-Taborsky, Eric Mack, Susan Griesse, Peter Shea and Mark Zeisler. They provided me with critical feedback, guidance, support and inside knowledge on how things operate. On the US front, I would like to especially thank Russell Allgor and his Operations Research team for their feedback and continual support. I would also like to acknowledge Jeff Wilke for his overall support of the LFM and his feedback on my work while at Amazon.

My management and engineering faculty advisors, Don Rosenfield and David Simchi-Levi, have both been of great importance in the development of this thesis. They provided me with many of the analytical fundamentals and procedures utilized in the development of both projects. In particular, I would like to thank Don Rosenfield for having led me through several seemingly "dead-ends".

More importantly, I would like to thank God for his blessings and my entire family for their support. I would like to especially acknowledge my beautiful wife for supporting me throughout all the long hours. She never faltered in patiently listening to my concerns and "theories" while always providing me with insightful advice. Looking back into these two amazing years at LFM, it would only be fair to say that she shares in the merit of my degrees.

Ken Merriam
Cambridge, MA May 2007

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Table of Contents

Table of Acronyms

LFM	Leaders for Manufacturing	RM	Regal Mail
EU	European	PiP	Proportional-in-Pricing
UK	United Kingdom	PF	PostalForce
DE	Germany	PN	PostalNet
FC	Fulfillment Center	DNL	Delivery Network
EOM	End of Month	SSD	Super Saver Delivery
MTD	Month to Date	AWPS	Average Weight per Shipment
S&OP	Sales & Operations Planning	WB	Weight Break
SPOT	Strategic Planning & Optimization	FAS	Factor Allocation System
IT	Information Technology	SOP	Standard Operating Procedure
Peak	Peak ordering season - Q4	PL	Product Line
SS	Safety Stock	PE	Product Entity (PL-Sortability-Velocity)
LP	Linear Program	OM	Order Mix (Mix of PLs in an order)
NLP	Non-linear Program	BMVD	Books, Music, Video & DVD

Table of Figures

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

PART I: INTRODUCTION

1.1 Purpose

The primary objective of this paper is to demonstrate examples of the practical implementation of analytical techniques and frameworks to network design optimization. Implementation is carried out through pragmatic decision-support systems aimed at total fulfillment cost reduction. The intent is to do so through a case study approach where two main types of projects relating to logistics network optimization are fully described. One project relates to optimizing logistics around an existing network configuration, while the other relates to directly altering a network's configuration specifically as applied to product and inventory allocation across warehouses.

Both projects are included in this paper due to their difference in nature and relevance to network design. In an attempt to reduce total fulfillment costs, there are two basic approaches: either by tailoring operational and tactical processes to better match an existing network's configuration or by altogether altering a network's design to better match the demands and nature of the business. The projects chosen for this paper are well suited to exemplify each basic approach.

A secondary objective of this paper is to outline through a case study approach the process used during the internship in identifying, selecting and improving upon aspects of a firm's logistic network with the purpose of reducing total fulfillment costs. More importantly though, the process is generalized and described to offer general guidance in the improvement of virtually any firm's logistics network.

1.2 Background & Basis for Thesis

The basis of this paper primarily lies on the experiential learning from a 6-month internship with the online retail distributor–Amazon in their European (EU) operations. In addition, the work, tools and frameworks are based from my studies and research at MIT as part of my dual-degree graduate studies with the Leaders for Manufacturing (LFM) Fellows program.

The extended internship experience, made possible through LFM's unique partnership with companies such as Amazon, is used as a case study in both developing a framework for investigating and optimizing a logistics network as well as demonstrating practical approaches at reducing total fulfillment costs through network design optimization.

The main objective of the internship, as proposed by Amazon, was to first analyze their current UK fulfillment network with the purpose of identifying areas for improvement to enable total fulfillment cost reduction. Upon identification, the final objective was then to validate and propose network-level changes in the UK fulfillment network along with their applicability to other Amazon EU fulfillment networks.

The specific projects were chosen after several steps including a rigorous due diligence process including several site visits and interviews with various managers for both the US and EU fulfillment networks. With the approval and support of my direct supervisor, various process owners were interviewed from almost every major department of Amazon's US and EU supply chain. This was critical in mapping out the network as well as in identifying cost drivers and Amazon's processes and systems used to support their networks.

1.3 Thesis Overview

The research and development of this paper mainly took place from June 2006 to December 2006 and is in large part due to the cooperation and collaboration of MIT faculty and various Amazon employees in both the US and EU.

The thesis is broken down into eight parts. The first two parts are introductory: Part I includes a brief outline of the research and support behind the thesis. Part II defines logistics networks from a general perspective. It also includes the key issues typically accounted for in planning and optimally designing a logistics network.

The next two parts are the bulk of the thesis. Both cover the internship experience at Amazon as pertaining to network design optimization. The intent is to not only cover the quantitative aspects behind both projects but also the organizational aspects behind the change initiatives. Part III firstly serves as an introduction to Amazon's EU operations while providing the reader with a description of the project selection process. It also outlines the background, stakeholders and improvement opportunities for each of the selected projects. Part IV deals with the solution for each project along with a description of challenges and results from validation and implementation.

Part V is intended to draw from the experience of both projects and provide the reader with a generalized framework for identifying, selecting and improving upon a firm's logistics network. Parts VI through VIII include a summary of conclusions, additional figures and references.

PART II: LOGISTICS NETWORK PLANNING & DESIGN

2.1 Objectives

At a fundamental level, all firms requiring the flow of products among facilities in the process of developing and delivering finished goods to a customer requires a distribution network. Designing a firm's distribution network should fit the nature of the business the firm caters to. Examples of factors influencing network design include but are not limited to transportation costs, space constraints, local labor costs, product properties, supplier lead times, raw material availability, customer dispersion, demand fluctuations, customer service levels and so forth. As such, distribution network designs are tailored to each business and are difficult to model.

An elemental part of any firm requiring a distribution system is its network configuration. Being the skeletal structure of a firm's distribution operations, a network configuration is ever-present in operational, tactical and strategic decision-making. Its structure is both strongly influenced by and influential on a firm's fulfillment cost drivers. For this reason, many firms seek to utilize the effectiveness of their distribution systems as a competitive differentiator. Companies such as WalMart are well-known for implementing innovative distribution networks (i.e. cross-docking) to remain competitive.

In essence, the end-goal behind improving a firm's logistics network configuration is profit maximization through total fulfillment cost reduction. In order to improve upon a network configuration it is important to recognize its determining components[1]:

1. The number, size and location of warehouses
2. The chosen transportation mode (e.g. outsourced? truck? rail?)
3. The allocation of product lines and inventory across warehouses (i.e. Inventory flow)
4. The sourcing of inventory from warehouses to fulfill orders (i.e. Order flow)

When implementing change, it is important to note that in most cases the further down the list, the easier change implementation can be. For example, the first component typically pertains to long-term decisions and is thus highly strategic. The second and third component can be generally characterized as tactical ranging from quarterly to yearly decisions. The last component can often be described as operational considering that a firm may manipulate this component in the short-term. Of course, this is not always the case and much depends on the firm's industry, products and IT systems. For example, in Amazon's case, the last component is actually considered both strategic and operational. Finally, it is important to note that these components can be highly interdependent; changing one requires re-evaluating and likely revising another component to better fit the revised configuration. As a result of this relationship, fulfillment costs are similarly driven by more than one component.

Reducing fulfillment costs while maintaining or improving service levels is at the core of establishing and continuously improving upon a network configuration along with those systems supporting the network. Given the complex and highly interdependent nature of a logistics network, it is important to maintain a holistic approach at reducing total fulfillment costs. Focusing cost reduction efforts for one aspect of fulfillment may increase the costs of another. Beyond costs, other non-monetary aspects (e.g. quality, service levels) need to be considered since these might also be influenced.

[1] Simchi-Levi, David et al. Designing & Managing the Supply Chain. 2nd Ed (2003) Chapter 1

Cost and service-related fulfillment components vary in significance between firms and industries. The following components are typically considered in varying importance when improving upon configurations and network-based activities:

1. Property, plant and equipment costs
2. Procurement costs
3. Labor costs
4. Inventory holding costs
5. Transportation costs
6. Service level requirements
7. Organizational behavior & reaction

Some important property, plant and equipment costs include rent, handling, maintenance and depreciation. Examples of service level requirements include product availability, fulfillment lead times, reliable quoted shipment times, reliable order fulfillment and quality of handled products.

2.2 Challenges

As previously alluded to, simply attempting to improve upon network configurations and network-based activities is highly challenging and complex. There are many reasons why this is the case.

Firstly, the interdependent nature of a logistics network requires a systems approach when mapping, analyzing and implementing change. Many times, even veterans at an organization are unable to foresee the consequences of changing any or a set of the aforementioned network components. Examples of such unforeseen effects are mentioned in Part IV.

Models are helpful tools for providing insight and guidance to streamlining processes and improving upon logistics. However, it is important to note their limitations. For example, in an attempt to try to capture the complexity of a system, models can quickly become intractable and impractical. This modeling tradeoff between accuracy and practicality is often hard to manage and expectations for different stakeholders are seldom met. Models are often considered too *simplistic yet tractable* or too *complex yet accurate*[2]. Complexity can often be overcome through the appropriate application of decision-support systems with practical user interfaces; however, there will always be those that remain wary of the "black box" nature behind its results.

There are two basic techniques used to model networks: Optimization and simulation. Both types of models exhibit the aforementioned modeling tradeoff. Beyond this, an important limitation of optimization models is that they do not directly capture either the variability of forecast input or the sensitivity to changes in assumptions. On the other hand, a limitation of simulation models is that they are built to test only one network configuration at a time without any mathematical optimization. Simulation models however track system behavior better and stochastic simulations can even capture variability in both forecast input and assumptions. Determining which model to use depends on the nature of the network and is in itself yet another

[2] Accuracy is considered herein as how close results reflect reality (i.e. high validity).
Rardin, Ronald. Optimization in Operations Research. Chapter 1 refers to a tradeoff between tractability & validity

challenge in network design optimization. For example, the larger the planning time horizon and expected variability in forecasts, the less dependable an optimization model will be.

Another challenge in developing and maintaining decision-support systems for the purpose of guiding network improvement is the changing internal and external environment. Anything from reorganizations, mergers and acquisitions to demand trends and government policies can quickly invalidate a model's assumptions. Thus, incorporating flexibility into a model is important; however, it brings us back to square one: the modeling tradeoff. Counter-intuitively, the more flexible, the less practical a model can actually become.

Finally, a commonly overlooked challenge in improving a network is organizational resistance to change. It is not uncommon to have these efforts result in controversial actions. Changing any of the four network configuration components can have a significant controversy and ripple effect throughout an organization. Anything from closing a warehouse to changing what product line allocation across warehouses has an effect on jobs and roles and will therefore meet organizational resistance by certain stakeholders. Recognizing and keeping all stakeholders involved in proposing and implementing change is a key determinant of success.

Several of these challenges and my approach at overcoming them are reflected in the selection, proposal and implementation process of both Amazon EU network improvement projects further described in Part IV.

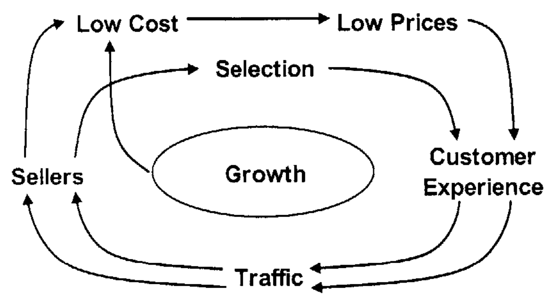
[THIS PAGE IS INTENTIONALLY LEFT BLANK]

PART III: IMPROVEMENT OPPORTUNITIES AT AMAZON EU

3.1 Amazon Background

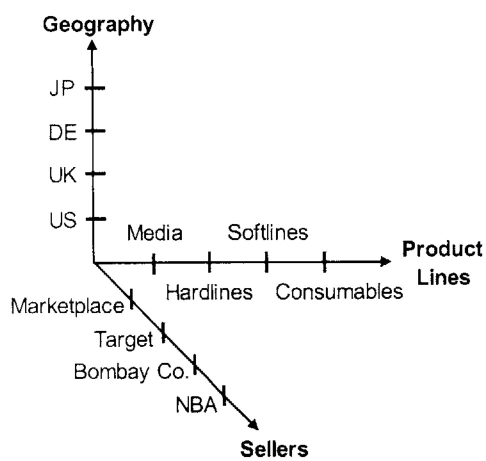
Amazon is a Fortune 500 company and is considered the largest online retailer. Amazon's main goal is to offer customers the world's largest online selection of products. Their objective is to have customers look first to Amazon for any item they would ever like to purchase. Amazon prides itself with their low prices, wide selection and fulfillment immediacy. Improvement on all three fronts can best be described by their growth model as shown in Figure III-1.

Figure III-1: Amazon's Growth Model



The idea is that growth is fueled by a positive feedback loop. For example, the more sellers, the higher the product selection and the better the customer experience. Good customer experience drives customer traffic which in turn attracts more sellers. This interior loop fosters growth which enables economies of scale leading to a lower cost structure. Amazon then passes on the lower costs to customers in the form of lower prices thus improving the customer experience. Amazon's growth can be translated across any of three different directions: Geography, Products and Sellers as shown in Figure III-2.

Figure III-2: Amazon's Growth Axes[3]



Amazon's worldwide operations consist of country-based fulfillment networks each with its associated website. See Figure III-3 for a list of these networks and their respective product

[3] Growth analogy by Amazon's US Director of Supply Chain, Nader Kabbani

lines.[4] A customer, anywhere in the world, can log in to any of these websites and place an order. The ordered product(s) are then fulfilled from one or several fulfillment centers located in the country associated with the customer-chosen website. For example, if a customer from the US were to order two books and a DVD from the Amazon.co.uk website then that customer's order would be fulfilled from one or several fulfillment centers in the UK even though it might be likely that the customer's products are available in a fulfillment center in the US.

Figure III-3: Amazon's Worldwide Fulfillment Networks

	Fulfillment Network & Product Line Expansion						
Product Line Rollout	**U.S.** Amazon.com 07/1995	**U.K.** Amazon.co.uk 10/1998	**Germany** Amazon.de 10/1998	**Japan** Amazon.co.jp 11/1999	**France** Amazon.fr 08/2000	**Canada** Amazon.ca 06/2002	**China** Joyo.com 08/2004
Books	1995	1998	1998	2000	2000	2002	2004
Music/Video	1998	1999	1999	2001	2000	2002	2004
DVD	1998	1999	1999	2001	2000	2002	2004
Video Games	1999	2000	2000	2001	2001	2003	2004
Software	1999	2000	2000	2001	2001	2003	2004
Electronics	1999	2001	2001	2003	2005		
Toys & Baby	1999	2001	2004	2004			
Home Improve	1999	2004	2004	2005			
Kitchen	2000	2004	2004	2003			
Apparel	2002						
Sports & Outdoors	2003			2005			
Jewelry	2003						
Health & Personal	2003						
Beauty	2004						
DVD Rental		2004	2005				
Groceries	2006						

Notes:
1. Years indicate launch dates for product lines by fulfillment network
2. Table relevant as of start of 2006

Each fulfillment network is composed of one or more fulfillment centers that carry either sortable, non-sortable or both types of inventory[5]. A fulfillment center is referred to as an FC at Amazon. An FC is a warehouse where products are delivered by suppliers and then stowed, picked, packed and labeled for shipping to customers by Amazon. Shipment is then carried out by non-affiliated mail carriers. The number, capability and capacity of FCs varies across networks and is tailored to meet the network's demands.

Amazon's IT systems and support are for the most part centralized in their Seattle worldwide headquarters. These systems enable both order and inventory flow. Each fulfillment network's infrastructure is built to accommodate these systems. Amazon's processes are either centralized or decentralized depending on the nature of the activities related to the process. For example, operations research, finance and labor planning function at a global-, network- and FC-level, respectively. More insight on this point is provided in the next section where Amazon's EU supply chain is mapped for the purpose of identifying and selecting improvement opportunities.

Amazon can be characterized as a process-driven company. They are highly focused on continuously improving those processes and systems affecting both order and inventory flow. The relative efficiency of these processes and resulting network flow differentiates Amazon from its competitors. Competitors include other online retailers (e.g. Overstock.com, eBay.com) and brick-and-mortar retailers (e.g. Barnes & Noble, Circuit City). Amazon's low fulfillment costs,

[4] Mendelson, Haim. Amazon.Com: Marching to Profitability (Case Study EC-25). Graduate School of Business Stanford University, 2006. Exhibit 3: Amazon's Expansion Across Countries and Product Categories.
[5] "Sortable" relates to those goods having all of its dimensions and weight below Amazon's sortability criteria.

wide selection, reliability and product immediacy are accomplished due to two main reasons: (1) Their unique business platform on which they effectively remove brick-and-mortar retailers from the value chain and (2) their highly tailored and continuously improving systems and processes built to support their logistics network.

3.2 Project Selection

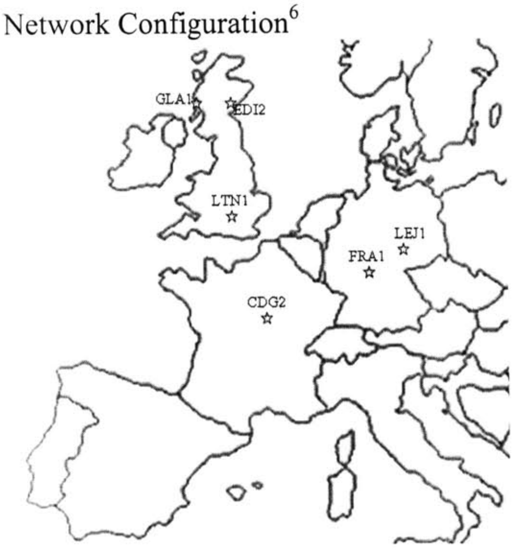
The process of identifying and selecting a project can be as important as the proposal and implementation process for a project. This section describes those first steps taken to identify and select two improvement opportunities which were at the centerpiece of the Amazon EU internship experience. These steps ultimately led to two clearly relevant and burning network design-related issues.

3.2.1 Amazon EU Network Configuration

Mapping Amazon's EU supply chain and network configuration was fundamental. It also served as a way to introduce myself and gain credibility within the organization. This was accomplished through interviews with different process owners, FC training, customer service training, meeting attendance and internal document research. The mapping effort was broken down into the four main components mentioned in Part II.

Investigation began with the first two components: (1) The number, size and location of warehouses and (2) the chosen inbound and outbound transportation modes. Similarly to the rest of Amazon's worldwide operations, the EU fulfillment network is composed of three country-based networks: United Kingdom (i.e. Amazon.co.uk), Germany (i.e. Amazon.de) and France (i.e. Amazon.fr). Each fulfillment network has a different number, capacity and capability arrangement of FCs built to fit the country's demand. Figure III-4 shows the number and location of these FCs. With respect to transportation mode, consistently with worldwide operations, both inbound and outbound transportation is managed by independent mail carriers.

Figure III-4: Amazon EU Network Configuration[6]

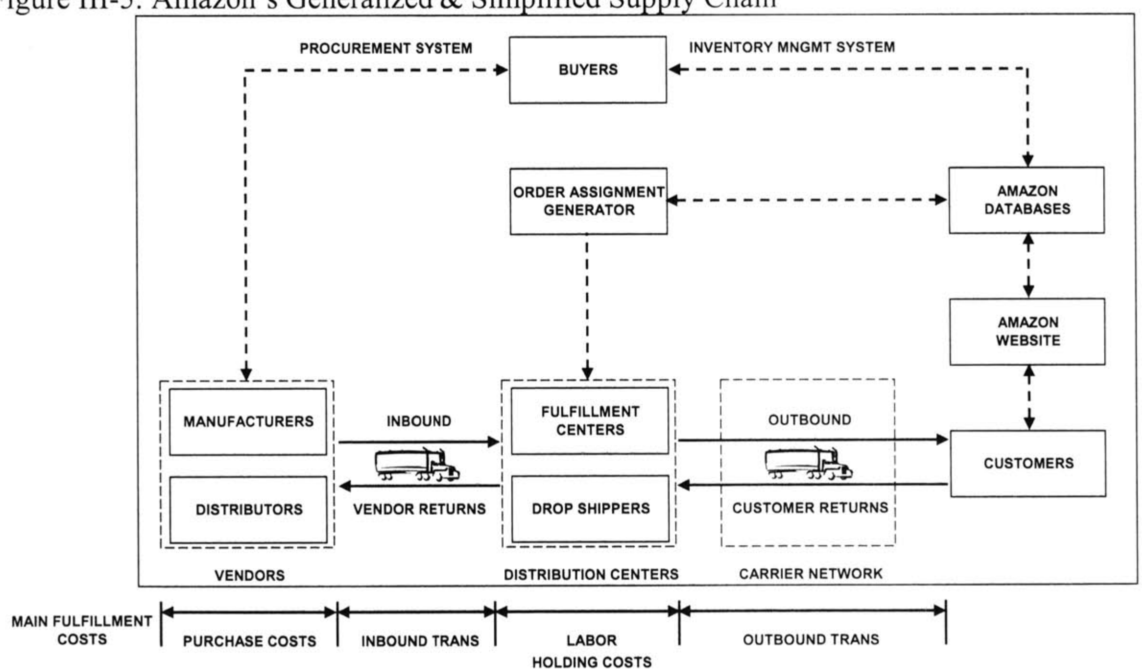


[6] Map Source: www.travel-eu.com, August 2006

It is important to note the major differences between the US and EU networks. First of all, given differences in demand volume and geography coverage, each EU network requires less FCs than the US network. Two main differences further arise resulting from the disparity in geography coverage. The first is that EU mail carriers have different fee structures which for domestic shipments are independent of distance. This is further discussed in Section 3.3. The second difference is that the US sortable FC network is setup and stocked with product lines such that customers are typically able to be fully served from an FC within their given region. In doing so, long-distance shipments throughout the US are minimized. In the EU, this is not an issue, as neither the shipping cost nor shipping times are distance-dependent.

Amazon's supply chain can be generalized for all of its global networks. Figure III-5 provides a simplified graphical interpretation of the processes, systems and high-level stakeholders involved in both the inventory and order flow at any given network. As noted earlier, Amazon's unique value proposition includes removing brick and mortar retail outlets from the value chain. Drop shippers have not been mentioned earlier; these independently owned warehouses are part of Amazon's external fulfillment. Any ordered item available in a drop shipper location but not in an FC will get fulfilled from the former. The intention is to improve product immediacy and availability.

Figure III-5: Amazon's Generalized & Simplified Supply Chain[7]



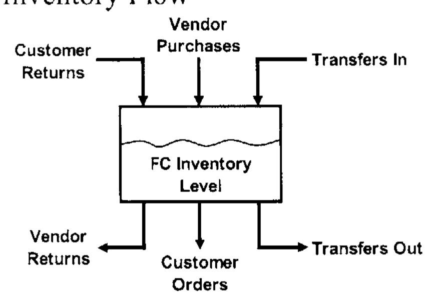
The third component defining a network configuration relates to the method behind product and inventory allocation between warehouses (i.e. inventory flow). This component addresses questions such as: How should the latest UK vendor shipment of Harry Potter books be allocated between LTN and GLA? Are inventory transfers required between LTN

[7] Allgor, Russell et al. *Company Report: "Supply Chain Design, Management & Optimization"*. 9th International Symposium on Process Systems Engineering, Volume 21, 2006

and GLA to address any inventory imbalances? Inventory flow is represented in Figure III-5 as the outer loop. At the center of this component are the inventory planning and management processes and systems. In order to understand these, it is important to firstly get a better sense of inventory flow from an FC perspective.

For the most part, Figure III-6 can be generalized as applicable to all FCs and is essentially a more detailed look into the lower, center piece of Figure III-5. Transfers in and out refer to the transfer of inventory from one FC to another within a network. Vendor purchases are done based on both reactive (i.e. pull) and predictive (i.e. push) methods. For example, if a customer were to order a rare book from Amazon that is currently not in-stock, Amazon would promise a ship date of sufficient length (e.g. 4 to 6 weeks) to allow for vendor lead time and then reactively place a purchase order for the book. On the other hand (and for the most part), other more frequently ordered titles are predictively purchased by Amazon from vendors using a technique described in Section 4.2.

Figure III-6: Generalized FC Inventory Flow[8]



While the figure shown above provides insight into the flow of inventory, it does not clearly outline how vendor purchases are allocated among FCs. This process is important because the allocation of product lines and inventory has a significant and lasting effect on fulfillment costs. The allocation of inventory purchases is a function of the Factor Allocation System (FAS) factors. These factors are currently determined on a seasonal basis (i.e. peak and non-peak) mainly by the collaborative work of the US Operations Research team and the EU Supply Chain Operations, Sales & Operations Planning (S&OP) and Capacity Planning teams. The US Strategic Planning and Optimization team (SPOT) developed an inventory allocation optimization model which proposes optimal FAS factor to the EU teams. A more precise description of the allocation process and the model is offered in Section 3.4.

The fourth component defining a network configuration relates to the method in which inventory from different FCs is matched to orders (i.e. order flow). This flow is represented in Figure III-5 as the inner loop. At the center of this component are customer assignment processes and systems. These are mainly centralized in Amazon's worldwide headquarters. Each network assigns orders using the same algorithm that essentially optimizes the allocation of items in an order to FCs while keeping to promise ship dates and FC inventory availability. Orders are currently optimized on an individual basis upon order creation. Optimality is defined as the assignment of all items within a single order to an FC or set of FCs within a network such that an order's total fulfillment costs are minimized while keeping

[8] FC inventory flow analogy by Amazon's US Director of Supply Chain, Nader Kabbani

to constraints. Once an order gets assigned, items are naturally picked, assembled and packed on a pull basis.

The fulfillment costs considered in order assignment include labor processing, transportation and load balancing costs. Load balancing is a process used by Amazon to influence new order assignment and thus the allocation of workload between FCs to help reach weekly FC workload forecasts. This is done through the dynamic assignment of virtual fulfillment costs to an FC called load balancing costs. This process allows FCs to plan labor and other resources in accordance to the stated forecasts with greater reassurance.

3.2.2 Main Fulfillment Cost Drivers

The next step, after mapping Amazon's main network components, was in identifying its main cost drivers. Amazon, as any other firm in the online retail industry, has the following main drivers of fulfillment costs (listed in the same order as they are relevant in magnitude): Material, transportation and labor processing costs. As such, Amazon's processes and continuous improvement efforts are geared towards ensuring that each of these costs is maintained as low as possible. This information was used as a sounding board in selecting the projects for the internship.

3.2.3 Considered Processes & Systems

Linking Amazon's processes and respective network components to Amazon's main cost drivers provides insight as to how improvement in any single process could potentially translate into cost savings. Provided that material, transportation and labor costs weighed so heavily on total fulfillment costs, it was clear that any improvement in the procurement, inventory assignment or order assignment process could have significant economic returns potential given that each had a significant impact on some or all of the 3 main cost drivers.

In my research process, special attention was provided in looking for opportunities in all three of these processes. For example, in procurement, the accuracy and timeliness of predictive buying methods was considered. The mix of predictive (i.e. push) and reactive (i.e. pull) purchasing was also of interest. In terms of inventory assignment, the area of most interest was that of optimal inventory and product line allocation given that these in a way predetermine the network distribution of future transportation and labor costs. Finally, in terms of order assignment, load balancing and the nature of the order assignment algorithm were investigated. Most of the focus however turned out to take place in the determination of the transportation cost component used by the order assignment algorithm.

3.2.4 Improvement Opportunities

Having mapped the EU network configuration, determined Amazon's main fulfillment cost drivers and linked these to several of Amazon's processes, the next step involved gathering the research data and identifying the areas with most potential for improvement. In doing so, several identifiers were used, more specifically any opportunities to:

1. Implement risk management techniques
2. React to internal or external change
3. Implement decision-support systems
4. Increase the accuracy, scope and/or flexibility of existing models

Five main filters were used in project selection: (1) Supervisor and stakeholder feedback, (2) cost driver relation, (3) available resources, (4) organizational "pull" and (5) time constraints. The first was the most important in getting a clear expectation of challenges and project potential. The last filter was also a critical factor in selecting projects with realistic scope and more importantly realistic development and implementation horizon for the 6-month internship. Related to ensuring a realistic scope was the organizational pull filter. Only those projects where most stakeholders were in approval and supportive were considered. The available resources filter was not a limiting factor due to Amazon and LFM's support.

Based on the identification criteria and selection filters, only two projects were chosen. Time constraints enforced these projects to deal only within the realm of network-defining components 3 (i.e. inventory flow) and 4 (i.e. order flow).

3.3 Project 1: UK Transportation Cost Minimization

The first project titled UK Transportation Cost Minimization dealt with the fourth network-defining component (i.e. order flow). This project was a result of a change in the external environment coupled with the opportunity and need to implement a risk management technique through the use of a decision-support system.

3.3.1 Background

Since its inception, Amazon has strategically chosen to outsource both inbound and outbound transportation in order to maintain focus in their competitive strengths: Efficient inventory and order flow. Across networks, Amazon works and negotiates with suppliers and mail carriers to obtain competitive inbound and outbound rates. For example, in order to receive lower rates Amazon has agreed to pre-sort packages per region for certain high volume FCs.

Forging new relationships and integrating mail carriers to Amazon's network is a challenge that Amazon has had to overcome with every new network. Each country has a small set (i.e. oligopoly) of well-established mail carriers and surely enough their rate systems are just as varied. It then comes to no surprise that Amazon's global transportation resources are decentralized. A network-based Transportation team deals with these challenges and work with US-based IT teams to incorporate unique mail carrier rate systems into Amazon's order assignment system. This project is a result of the challenges that arise in dealing with mail carriers and their unique rate structures in this case specifically with the UK.

At the time of the internship, Amazon offered customers two types of shipments: Standard and Super Saver Delivery (referred to as SSD) shipping. Standard shipping offers quicker delivery times than SSD. In the UK, all mail carrier rates are independent of domestic shipping distance. There are two basic types of shipment rate structures: flat and weight-based rates. Amazon deals with both rate structures in each shipment type. Each mail carrier primarily identifies itself with one type of rate structure. Figure III-7 shows Amazon's main UK mail carriers[9] by shipment type and their respective rate structure.

[9] For confidentiality, mail carrier names have been masked.

Figure III-7: Amazon UK Mail Carriers by Shipment Type & Method

Typ Package Weight	Rate Structure	Amazon Shipment Type	
		Standard	SSD
Light	Weight-Based Rates	Regal Mail (RM High)	Regal Mail (RM 2nd)
Medium	Flat Rate	PostalForce (PF)	PostalNet (PN)
Heavy	Flat Rate		Delivery Network (DNL)

The mail industry in the UK is highly concentrated within weight segments. Mail carriers have strategically based their rates to locate themselves competitively within a specific weight range in a focused attempt to dominate that range. For example, low weight packages (up to approximately 1.5kg) are mainly controlled by Regal Mail and as such their rate cards are structured such that heavy packages can only be shipped by them at uncompetitive rates and up to a maximum weight limit. PostalForce and PostalNet on the other hand have based their rates to competitively focus on medium weight packages and in the case of PostalForce heavy packages as well. Other carriers such as DHL and Delivery Network have rates competitively focused in the heavier end of the spectrum.

3.3.2 Problem Statement

Regal Mail (RM) carries the bulk of all of Amazon's UK shipments. This is mainly due to the fact that most of Amazon's shipments are low weight (e.g. books, DVDs, music) and as mentioned earlier, Regal Mail is the only competitive mail carrier in the UK for this weight range. This has effectively bound Amazon UK to RM in both Standard and SSD shipments.

Regal Mail has for some time now imposed a unique and complex charging mechanism that most likely frustrates most of their client's accounting and planning departments. Regal Mail charges a single shipment fee for every package shipped through Regal Mail High (i.e. Amazon's Standard shipping) and another single fee for every package shipped through Regal Mail 2nd class (i.e. Amazon's SSD shipping) on the basis of a shipping source's monthly average weight per shipment (AWPS). In order to calculate the average weight per shipment for any locale, Regal Mail divides the total weight of all packages they shipped from a location in a given month by the total count of packages they shipped from that location and in that month. This process is followed for both RM High and 2nd Class shipments. A month's AWPS is then plotted in their semi-linear rate cards for each shipment type to determine the single shipment fee to be charged to every single shipment.

For example, assuming an FC in the UK shipped 100,000 Standard shipments through Regal Mail High last month weighing a total of 51,000kg, then that FC's RM High end of month average weight per shipment (EOM AWPS) would be equal to 510 grams:

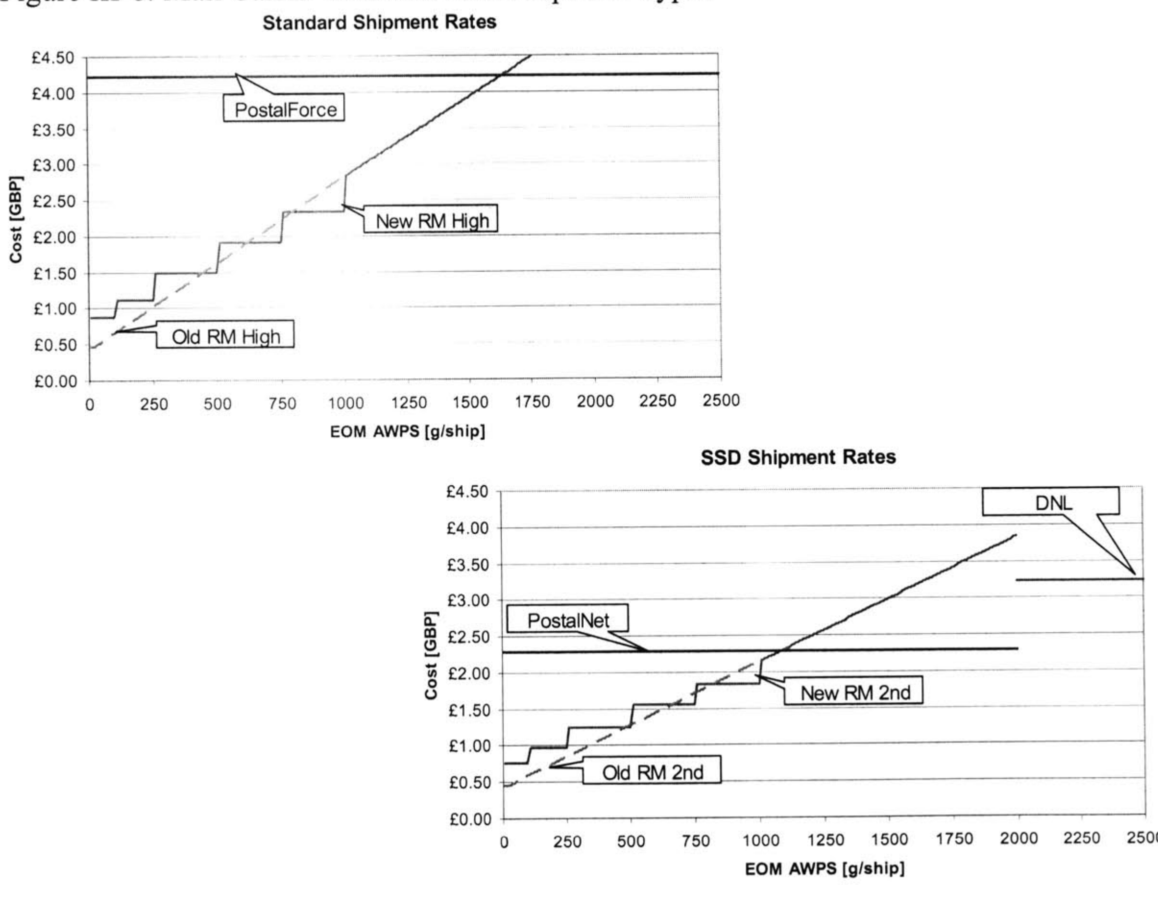
$$EOM\,AWPS_{RMHigh} = 51{,}000{,}000g/100{,}000shipments$$

According to their standard semi-linear rate card shown in Figure III-8, the resulting shipment fee would then equal to about GBP 1.65. Therefore, Amazon's FC would be charged GBP 165,000 (= 1.65*100,000) for last month's RM High shipments.

A major problem with this charging mechanism is that it is a backward-looking fee structure. As a result, this makes it difficult for Amazon to estimate its transportation costs throughout a month in progress. This poses a difficulty for the EU Transportation team in attempting to minimize outbound transportation costs. For example, for Standard shipments, the team has to determine what shipments are to be labeled and sent through RM High as opposed to PostalForce (PF) even before the end of month AWPS for that FC is known. In other words, they have to determine ex-ante what shipments are to be classified as "light" and "heavy" and thus get shipped through RM High and PF, respectively. This fee structure also poses a challenge for optimizing order assignment since the assignment algorithm needs to know immediately after order creation how much an order's transportation costs will be.

To complicate things further, near the start of the internship, Regal Mail announced that they would change their rate structure to include size starting August 2006. Their initiative was called Proportional-in-Pricing (PiP). They now added an extra shape dimension into their pricing scheme whereby a package was now characterized as a Packet, a Large Letter or a Letter. According to their criteria, all of Amazon's packages were to be now considered Packets. Within Packets, the same charging mechanism and rate cards were maintained with one exception: The transition from semi-linear to step function rate cards. This transition along with the rates from other mail carriers is shown in Figure III-8 for each shipment type.

Figure III-8: Mail Carrier Rates for Both Shipment Types



The EU Transportation team immediately recognized the added challenge in this rate structure change. The Regal Mail EOM AWPS for Standard and SSD shipments not only had to fall below PostalForce and PostalNet fixed rate fees, respectively, but it now also had to optimally fall to the rightmost edge of a rate step. Figure III-9 expands upon the previous example and demonstrates the change in level of sensitivity resulting from the transition to the new Regal Mail rate cards.

Figure III-9: Simplified Sensitivity Demo with New Regal Mail Rate Cards

Regal Mail High	Scenario 1: Heavy		Scenario 2: Light	
EOM Total Wgt [kg]	51,000		49,000	
EOM Total Shipped	100,000		100,000	
EOM AWPS [g/ship]	510		490	
Rate Structure	**Pre-PiP**	**Post-PiP**	**Pre-PiP**	**Post-PiP**
Respective Ship Rate	£1.65	£1.93	£1.60	£1.49
EOM RM High Costs	£165,000	£193,000	£160,000	£149,000

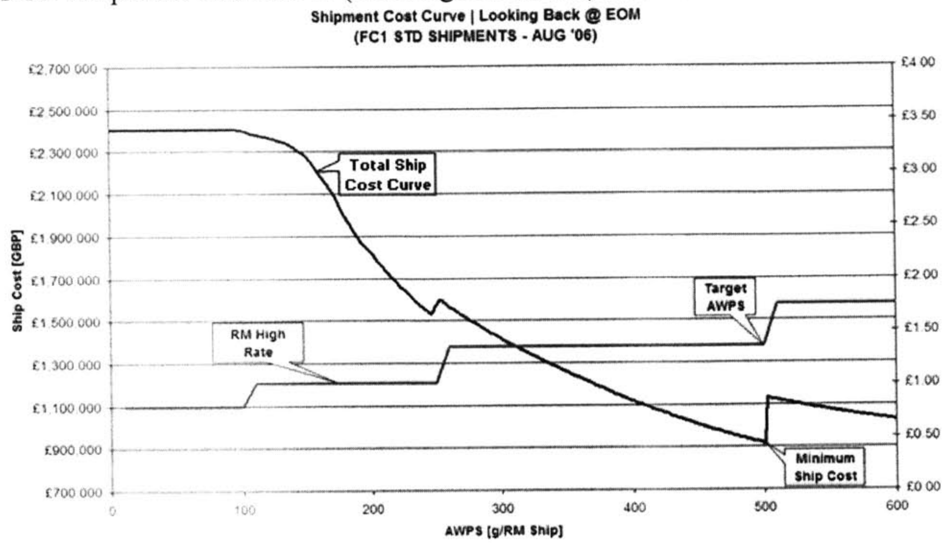
The figure above demonstrates that a 2,000 kg difference prior to PiP would have resulted in a reduction of GBP 5,000 whereas with PiP, it now results in a reduction of GBP 44,000. The difference in variation is over 8 times larger and is a result of shifting between rate steps. Clearly, this hypothetical example is an oversimplification of the actual total EOM change in costs between both scenarios primarily since only RM High shipping costs were considered. The exercise nonetheless underscores the much higher level of risk that is now involved.

The example above also demonstrates the new-found importance of ensuring that Amazon's EOM AWPS for each shipment type in each FC lie to the rightmost edge of rate steps. Moreover, upon incorporating other mail carriers for a given shipment type into the problem, it begins to make intuitive sense that there should only be one optimal rate step whereby having the EOM AWPS at its rightmost edge would minimize total shipping costs.

This is actually demonstrated later, but the point is that the EU Transportation team was now in need of a new process that would enable them to minimize total outbound transportation costs. The objective would be to first identify an optimal rate step for each FC-shipment type combination and to then have the means to ensure that the EOM AWPS actually reaches the target. The latter would involve the periodic manipulation of weight breaks (WB). A WB is simply a threshold defined by Amazon whereby any shipment weighing below it gets shipped through RM and anything over it gets shipped by a flat rate carrier. The challenge lies in determining the WB most likely to reach the predetermined EOM AWPS target.

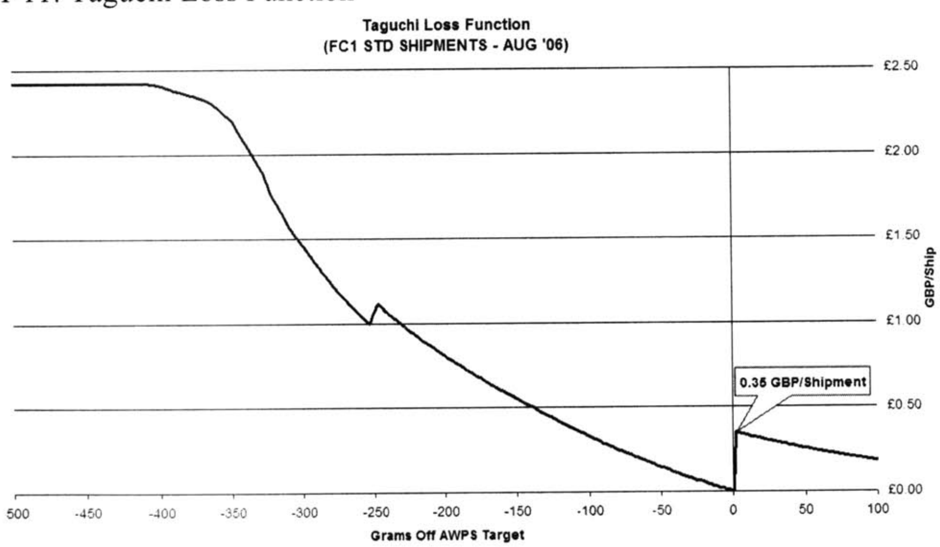
The optimal rate step is easily determined looking backwards into a month by developing that month's cost curve. These curves can be developed by running simulations of a previous month's shipments for a given FC for each possible WB and recording each simulation's resulting total EOM shipment cost (i.e. the y-axis) and RM EOM AWPS (i.e. the x-axis). For example, the cost curve shown in Figure III-10 clearly identifies for this masked case scenario that the rightmost edge of the 250-500g rate step is the optimal EOM AWPS since this is where total Standard ship costs are minimized.

Figure III-10: Shipment Cost Curve (Looking Backward) as a Function of AWPS[10]



The importance of setting and reaching optimal EOM AWPS targets can be further stressed through these cost curves. Note how costs increase drastically by missing the target by only 1 additional gram over 500g. In order to clearly demonstrate the marginal cost of error, a Taguchi Loss function is provided in Figure III-11 for the masked case shown above. The plot shows that one gram above target would cost an additional 0.35 GBP per shipment, whereas one gram below target would cost only an additional 0.005 GBP per shipment.

Figure III-11: Taguchi Loss Function



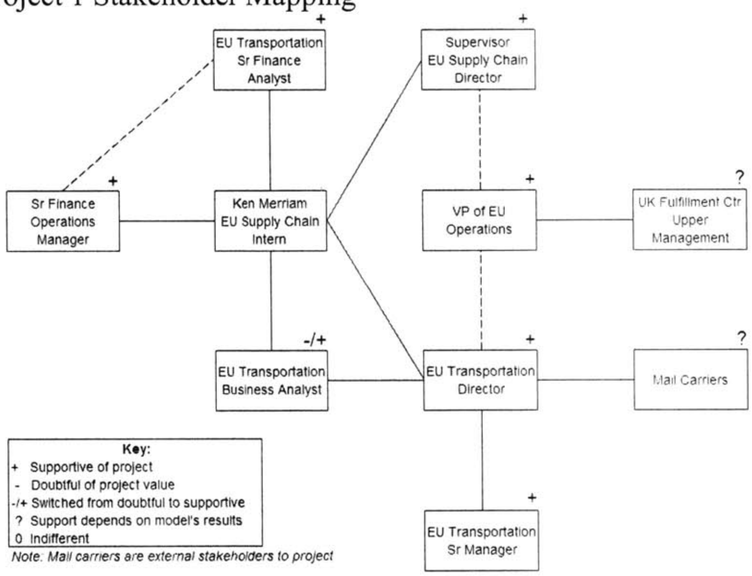
The sensitivity of these cost curves clearly demonstrate the need for a practical decision-support tool which would allow the EU Transportation team to effectively manage such a high monetary risk between straying to the right as opposed to the left of a target.

[10] For confidentiality, the cost curve figures, month and FC's name have been shifted and masked.

3.3.3 Stakeholder Mapping

As part of the project selection process, all significant stakeholders were mapped for this project. Figure III-12 provides stakeholder stances and their inter-relationship. For example, solid lines indicate a direct relationship between the project and a stakeholder, whereas dashed lines indicate a stakeholder's indirect influence through a direct report(s).

Figure III-12: Project 1 Stakeholder Mapping



The purpose of this exercise was to gauge the level of organizational support and difficulty in implementing any solution. As demonstrated in the figure, there was a high level of clearly defined support behind this project from all levels.

3.3.4 Project Justification

The project fit well under the criteria outlined in Section 3.2. The project was directly linked with a major cost driver: transportation costs. Key stakeholders recognized the urgency to this problem thus alluding to the level of future support in implementing a solution. The EU Transportation team was at that point under pressure to develop a process to enable cost minimization, yet only had about two months until the August 21st implementation of Regal Mail's new rate structure. Beyond this, the team was also preparing for the opening of a second FC in Germany. All this meant that the project would have a realistic time horizon for a 6-month internship and the EU Transportation team would welcome the aid.

More importantly, the project had a significant effect on order flow and thus network cost-efficiency. While, the solution most likely would not involve any network configuration alteration, it would require working around the existing UK network configuration and processes while ultimately affecting order assignment. The reader is referred to Section 4.1 for this project's proposal, validation and implementation results.

3.4 Project 2: EU Product & Inventory Allocation Optimization

The second project titled EU Product and Inventory Allocation Optimization dealt with the third network-defining component (i.e. inventory flow). This project was a result of a reaction to internal change in the form of growth coupled with the opportunity to increase the scope, accuracy and flexibility of an existing optimization model.

3.4.1 Background

Up until August 2006, the UK was Amazon's only multi-node network in the EU. In August 2006, Germany transitioned to a bi-nodal network with the successful opening of the Leipzig (LEJ1) FC. Operating a network with a single node as is the case for France is naturally simpler than a multi-node network especially with respect to order and inventory flow. This is the case because order, product line and inventory allocation is not of any relevance.

In terms of order assignment, the UK and DE multi-node networks are successfully using virtually the same order assignment algorithm utilized in the US multi-node network. This being that the cheapest possible fulfillment plan for an order actually gets enacted. Unfortunately, the inventory assignment logic used in the US is not applicable to the UK or DE. The main reason is that, as mentioned earlier, the US sortable FC network is setup and stocked with PLs such that customers are typically able to be fully served from an FC within their region. In this manner, long-distance shipments are minimized. In the EU however, this is not an issue, as neither the shipping cost nor times are distance-dependent due to the smaller geographical coverage. As a result of this, FC fulfillment overlap among US regions is minimized whereas this is not the case for EU FC fulfillment; as such, the same inventory assignment logic is inapplicable to EU networks.

3.4.2 Problem Statement

As the UK, DE and FR networks increase in size, both in terms of nodes and capacity, the importance of setting a consistent and tailored EU framework for allocating product lines and inventory becomes more critical. Currently, there is no theoretical framework for determining the optimal placement of product lines across FCs in EU multi-node networks. Allocation decisions have been mostly the result of organic growth and on the basis of supplier constraints. For example, the UK started off with FC1 which was originally built with the capability to hold all product lines. After the transition to a dual-node network this actually remained the case and FC2 now simply shares in carrying the load for FC1's high demand product lines. While this logic was sufficient during the early stages of expansion, it should not be the case going forward.

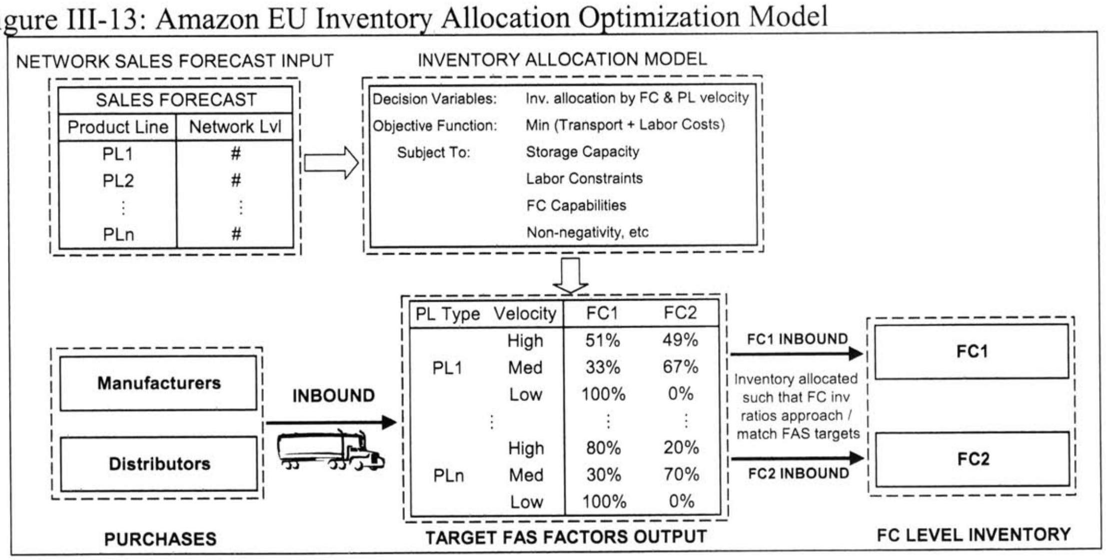
With all of the Amazon EU networks growing at double digit rates in terms of order volume, an analytical and scalable framework is required to determine the optimal allocation of product lines and inventory across FCs. Optimality in this case is defined as the allocation of product lines and inventory across FCs such that total expected fulfillment costs are minimized subject to FC storage capacity, labor capacity and supplier delivery constraints. The following questions need to be answered by such a framework:

1. What product lines should be carried by each FC?
2. If a product line is carried by more than one FC, then how should the total inventory for that product line get allocated?

Amazon is aware of the importance in developing an optimal inventory allocation framework for the EU and efforts have been underway to address this issue since 2005. The main challenge has been in developing a practical yet sufficiently accurate model that reflects the complex aspects of an EU network across time. The Strategic Planning & Optimization (SPOT) team developed a network model for the UK in 2005 that provided inventory allocation guidance for a seasonal period assuming a given product line distribution across FCs. For example, given that books were already stored in both FCs in the UK, the model would propose Factor Allocation System (FAS) factors such as 70% and 30% for FC1 and FC2, respectively. These factors would be based on the model's attempt at minimizing transportation and labor costs. Figure III-13 offers more detail on the development and application of these factors. If accepted by the EU Capacity Planning, SC Operations and S&OP teams, these FAS factors would be input into the procurement FAS system to determine the automated allocation of newly procured inventory across FCs.

It is important to note that the network optimization model was built as a linear program (LP) to ensure tractability and low processing time. Classical LP theory indicates that polarized solutions, such as 100% and 0%, should be commonly expected of LP programs. For practical purposes which are outlined in further sections, these polarized solutions were mostly eliminated by manually imposing allowable range constraints on all FAS factors.

Figure III-13: Amazon EU Inventory Allocation Optimization Model



The problem with the existing inventory allocation model is that it does not offer guidance for the optimal allocation of product lines across FCs. In other words, it optimizes the allocation of inventory across FCs for a product line (PL) if more than one FC carries it, but it does not indicate whether each of those FCs should carry the PL to begin with. Another issue was that the model was built in 2005 and the SPOT team had since then developed new and improved micro models for FC labor and storage that superseded those from the 2005 network model. Finally, it was not clear whether the existing network model neglected other relevant fulfillment costs that were a function of product and inventory allocation.

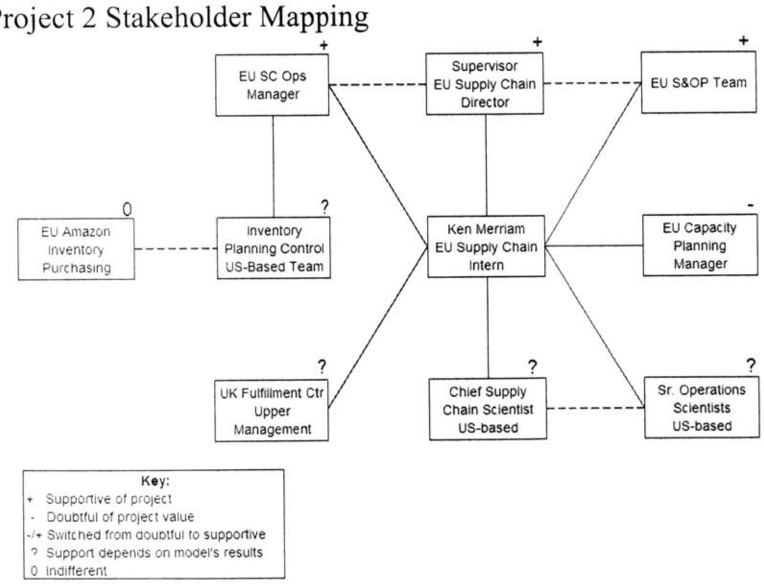
At a much deeper level, there were concerns by several managers that there were currently no clear links between order assignment and inventory assignment or similarly between order behavior and inventory assignment. In the former case, there was concern that inventory might be being allocated by the model in different proportions relative to how inventory was being assigned to fulfill orders by the order assignment algorithm. In the latter case, the inventory allocation model did not consider customer's product line ordering patterns in its decision-making logic. As a result, the likelihood of sourcing multi item orders from different FCs (i.e. order splits) and its respective additional transportation costs were not being directly considered by the model.

The possibility of having a disassociation between demand and supply was of concern since this would lead to inventory imbalances, more specifically shortages in one FC and excess in another. Similarly, this would also mean that order assignments are sub-optimal due to sub-optimal inventory allocation. In fact, a Six Sigma project by an EU S&OP member actually demonstrated the occurrence and severity of these imbalances. Projects were underway to solve this issue reactively through transshipments (i.e. inventory transfers across FCs). However, it was clear to everyone that improving upon the network model such that the model's resulting FAS factors better reflected demand patterns would help alleviate this problem proactively.

3.4.3 Stakeholder Mappings

Once again, as part of the project selection process, all significant stakeholders were mapped for this project. Figure III-14 provides stakeholder stances and their inter-relationship. As in the previous stakeholder map, solid lines indicate a direct relationship between the project and a stakeholder, whereas dashed lines indicate a stakeholder's indirect influence through a direct report.

Figure III-14: Project 2 Stakeholder Mapping



As can be noted from the figure, this project cuts across more departments and has a larger variety in stakeholders than the first project. Another difficulty behind this project lies in the varied support across stakeholders. For the most part, support depended upon project results.

3.4.4 Project Justification

This project is based on the opportunity to increase the accuracy, flexibility and scope of an existing model. The need for such an improvement is a result of internal change in the form of network growth. While determining whether adding further accuracy and scope would be outweighed by the loss in practicality was challenging, the project's subject matter was definitely well aligned with the internship's objective. Dealing with inventory flow which is such an important component that is definitive of a network's configuration was an alluring aspect of the project.

Furthermore, any improvements upon the model, and thus the network configuration, would provide a relatively quick impact upon fulfillment costs with a low potential for disruption. Affected costs included labor, transportation and inventory holding costs. The need for such an extended model was clear and accepted by many in the organization. The project had high visibility and support. The main concern was the project's time horizon. Given the high level of impact in changing the model, validation and buy-in would be a central part of the project and this would require time. More than half-way into the internship this was of concern, yet it was determined that sufficient value would be gained if at least a running solution were to be proposed for this project even if it meant not being able to validate and fully implement. The reader is referred to Section 4.2 for this project's proposal.

PART IV: THEORY TO PRACTICE AT AMAZON EU

4.1 UK Transportation Cost Minimization

Regal Mail, a main Amazon carrier in the UK, announced that it would change its rate structure on Aug 21st to a rate system which they labeled PiP. In order to minimize UK outbound transportation costs, a new process was now required to ensure that the EOM AWPS for every FC-shipment type combination be as close as possible to the rightmost edge of an optimal rate step. This involved setting and achieving monthly AWPS targets through the periodic manipulation of the Regal Mail weight breaks (WBs).

4.1.1 Project Proposal

The EU Transportation team had clearly determined that changing the current shipment assignment processes was required as a result of PiP. They did so by running a simulation in which the new PiP rates were applied to actual shipment volumes and profiles from 2005 in order to gauge what the cost impact of "doing nothing" could be in the future. The hypothetical cost increase for 2005 shipments was estimated at GBP 2.1 million.[11]

There was obviously a clear need for a practical decision-support tool to (1) define EOM AWPS targets and (2) identify the associated weight breaks required to most likely reach those targets for each FC-shipment type combination. Furthermore, the model's proposed EOM AWPS targets would have to be located sufficiently close to the rightmost edge of a rate step, yet not so much as to have variation push the average onto the next higher rate step.

Such a decision-making tool was ultimately developed for this project working alongside the EU Transportation team. A tool was required immediately though given the proximity of the rate structure change (August 21st). Thus, the project was broken down into two phases. The first phase involved the development of a simplified version of the end model to be implemented starting August, whereas the second phase involved the development and implementation of the final version model.

Both models were developed to propose the most likely WB to reach an EOM AWPS target at any given point in time within a month in progress using the following main input:

1. Actual – Month to date (MTD) total shipment quantity and weight by ship method for Standard and SSD shipments
2. Forecast – Total remaining shipment volume for the remainder of the month for Standard and SSD shipments by FC
3. Forecast – Shipment weight profile shape for the remainder of the month

Both models served a dual purpose to meet the needs of the EU Transportation team: As forward planning tools (at the onset of each month) and as dynamic control decision-making tools (at any point during a month-in-progress). Thus, as Amazon gains more insight on a month's shipment weight profile and volume, the team can dynamically and proactively change WBs through a month in progress to increase the likelihood of reaching the EOM AWPS targets.

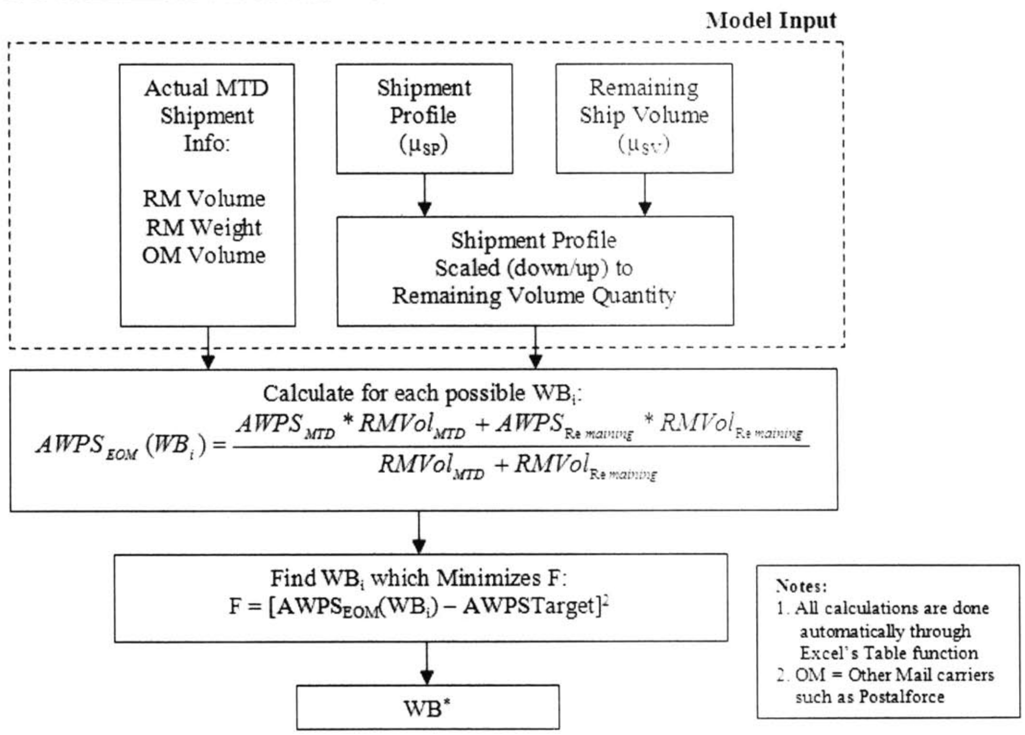
[11] Shea, Peter et al. *Company Report: "UK Domestic Ship Method Modeling"*. EU Transportation Team. Nov 2006

The main difference between both models was that the first assumed all forecast inputs as 100% certain (i.e. a deterministic approach) whereas the second one accounted for any uncertainty/variability behind the forecast input (i.e. a stochastic approach).

Phase I: Deterministic Model

This model was quickly developed and implemented in time for Regal Mail's PiP launch. In order to implement the model within time though, it was built as simple as possible; the model even used a rudimentary exact numerical search method utilizing Excel's Table function as opposed to a mathematical program. Based on MTD shipment data and deterministic shipment weight profile and volume forecast input, the Excel model simply calculated the EOM AWPS for each possible weight break (WB) within the full range of 0 to 10kg in steps of 10g. The model then proposed a WB whose respective EOM AWPS best matched a targeted AWPS. The model's logic is described in the process map shown in Figure IV-1.

Figure IV-1: Deterministic Model Flow Chart



There were three main flaws with this model. The first was that it did not prescribe an optimal EOM AWPS target. Instead the user had to provide this as input so that the model could provide a weight break (WB) whose resulting AWPS best approximated the target. The target was fixed for every month and was based on a study by the Transportation team which minimized the *previous* year's total shipping costs. Therefore, with this model, there was no real way of determining whether the provided target was on an optimal rate step for each future month. This would have not been the case however if the model would have initially been built with a total EOM shipment cost function as opposed to a virtual cost function. This was the case due to time constraints and the model's short term purpose.

The second flaw similarly related to the model's limited scope. The model's prescriptive output was completely independent of how mail was allocated between Regal Mail and the other carriers. For example, some mail carriers had volume limitations and the output did not account for this. Similarly, the output offered no guidance or expectations relevant to total EOM shipment costs. Finally, and most importantly, it was naïve to expect that the forecasted weight profile and volume input would actually take place; variability, as later demonstrated, can not be neglected.

Despite the model's limitations, the EU Transportation team was in fact successful in managing PiP as shown in Figure IV-2, which shows the marked difference in EOM AWPS between the months of June through July (when WB's were not being manipulated to reach AWPS targets) and August through November (during the model's implementation). In doing so, WB's were typically reviewed and/or manipulated twice per month by the team to ensure targets were achieved.

Figure IV-2: WB Manipulation Effect on EOM Average Weight per Shipment (AWPS)

FC	Regal Mail Ship Method	BEFORE		AFTER			
		June 2006	July 2006	Aug 2006	Sep 2006	Oct 2006	Nov 2006 MTD
FC2	Standard	503	512	481	482	445	411
	SSD	554	597	494	486	481	464
FC1	Standard	505	513	488	489	474	463
	SSD[3]	558	587	503	474	461	455

Note:

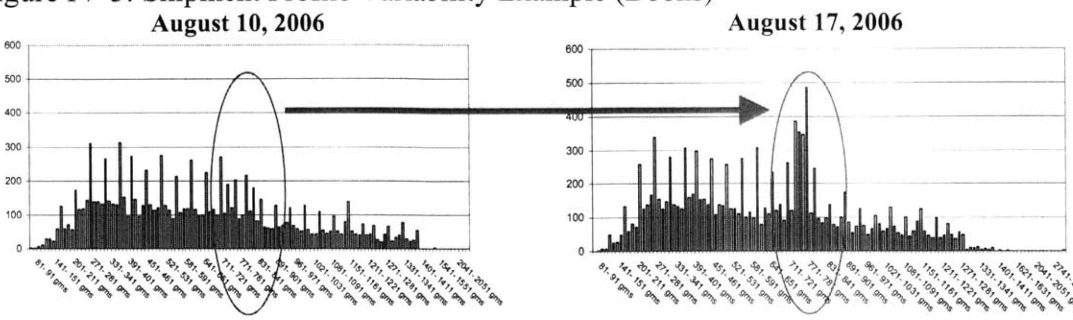
1. *Weight Break manipulation first took place in August 2006*
2. *Cells are shaded RED if EOM AWPS was over target AWPS*
3. *From the introduction of PiP on Aug 21st through the rest of August, the actual AWPS fell below 500 g/ship for FC1 SSD*

Figure IV-2 shows that from August to November, with the aid of the deterministic model, the imposed 500 g/shipment EOM AWPS target (i.e. the rightmost edge of the 250-500g rate step) was in fact closely approximated without jumping to the next higher rate step.

However, two main issues remained. Firstly, it was not clear whether the 250-500g rate step was in fact the optimal rate step. Secondly, this being a deterministic decision-making process there was a lack of any risk management: By aiming for the rightmost edge of a chosen rate step, such as 500 g/ship, the high monetary risk of any slight variation in forecasts that could push the average onto the next higher rate step was being neglected. Yet, even if a buffer were to be included between the edge and the target, which was actually the case during this period, then how large should that buffer be?

Figure IV-3 clearly demonstrates an example of the type of variance contributing to the total monthly shipment profile forecast uncertainty. The y-axis represents frequency and the x-axis weight. This figure shows the large difference in weight profile that can exist from one week to the next. The following simplified example illustrates the potential severity of such variance: Say the left weight profile shape was used as representative for an entire month's weight profile shape. In this case, the model would prescribe a WB of say 1,200g to reach the EOM AWPS target of 500g. However, during the last week of the month, the profile suddenly drastically changes to that shown on the right. Without any modification, the chosen WB would have resulted in a high EOM AWPS (e.g. 510g) corresponding to the next higher RM rate step. The costs of such a rate jump are highly taxing.

Figure IV-3: Shipment Profile Variability Example (Books)[12]



Up until November, the EU Transportation team had been able to overcome this model's weakness through close monitoring and good judgment in interpreting and applying the model's output.

Phase II: Stochastic Model

The stochastic model quantitatively answers the question posed above: How far is too far, and how close is too close to a rate step's rightmost edge when selecting an EOM AWPS target? The answer lies in the level of forecast variability.

Back in Part II, two basic techniques used to model networks were mentioned: Optimization and simulation. The EU Transportation team was looking for a tool that would offer them WB selection guidance that accounted for risk. This signaled a need for both prescriptive (i.e. optimization) and descriptive (i.e. simulation) output[13]: Prescriptive to enable AWPS target and WB selection guidance and descriptive to enable the simulated evaluation of any given WB under the effect of forecast variability. The latter could only be obtained through stochastic simulation.

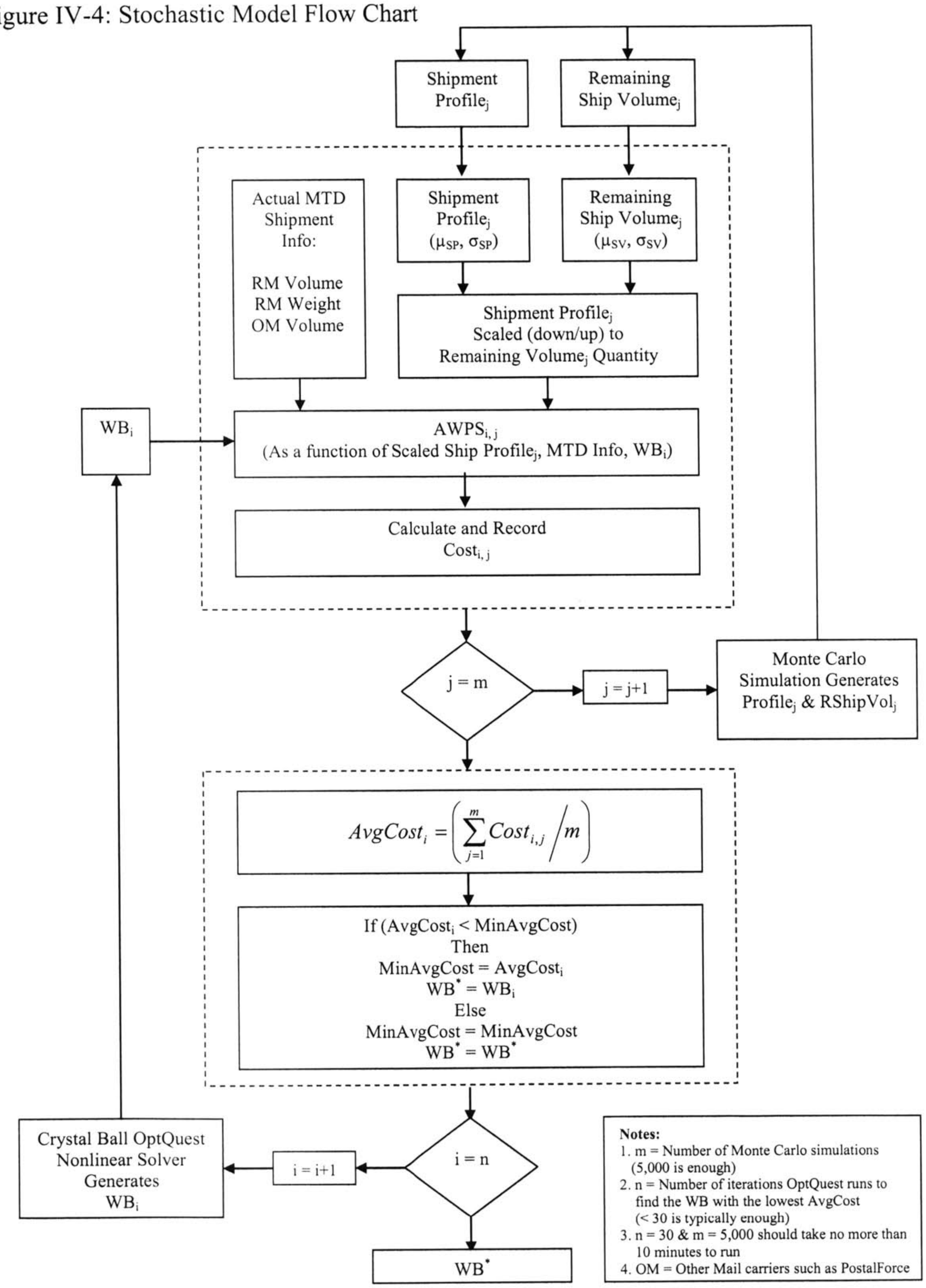
The logic behind this model is similar to that of the deterministic model with three main exceptions: (1) the incorporation of stochastic simulation to account for forecasting error, (2) the inclusion of all mail carriers as opposed to only Regal Mail and (3) the use of actual total shipment costs as opposed to virtual costs for the objective function. These enable the model to propose a WB that maximizes the savings of closely approximating the rightmost edge of an optimal rate step, while minimizing the chances of jumping to the next higher rate step.

Using Excel and Crystal Ball, an Excel add-in which provides Monte Carlo analysis and a non-linear optimization package, thousands (= m, as shown on Figure IV-4) of different possible ship volume and weight profile scenarios are evaluated for every selected WB. The model then determines, through Crystal Ball's non-linear optimization package called OptQuest, which WB would offer the lowest total shipment cost across all scenarios. Figure

[12] Shea, Peter et al. *Company Report: "UK Domestic Ship Method Modeling"*. EU Transportation Team. Nov 2006
[13] Rardin, Ronald. Optimization in Operations Research. (1998) Chapter 1

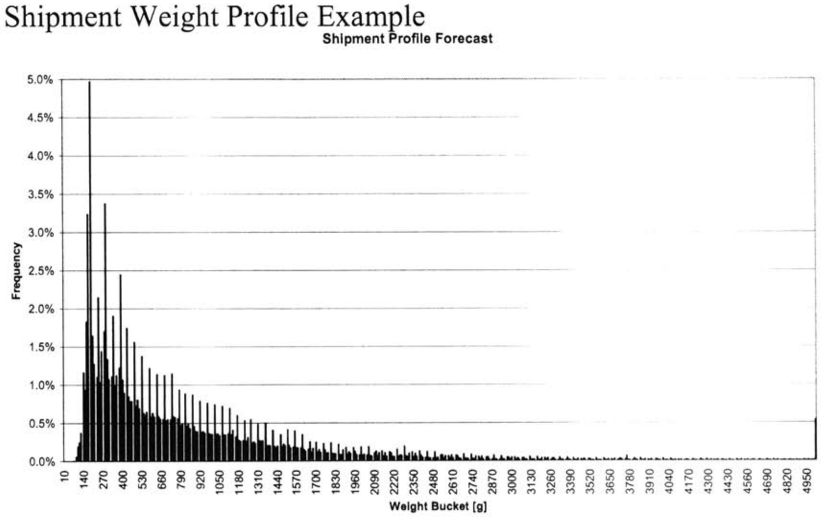
IV-4 demonstrates the exact logic used in order to capture the forecast variability into the model's optimal WB and AWPS target output.

Figure IV-4: Stochastic Model Flow Chart



In terms of changes to input format, in addition to Regal Mail's MTD shipment volume and weight data, the MTD shipment volume data for other mail carriers is now also required. This enables total monthly shipment cost estimation. With respect to forecasts, both the average and variance for the remaining Regal Mail shipment volume is required. This data is easily obtained from the EU Transportation team's forecasting tools. With the deterministic model, only the average was considered. In order to capture both an average and variance in shipping weight profile, four different weight profiles each of identical time span and similar demand profile to that of the period under investigation is required. For example, assuming today were August 8th, then there would be a total of 24 days left in the month, including the 8th. Thus, for this case, four representative weight profiles each of 24 days length could be (1) Aug07-Jul15, (2) Jul14-Jun21, (3) Jun20-May28 and (4) May27-May04. In the previous model, only one weight profile was used. Figure IV-5 provides an example of a weight profile for a randomly chosen time period.

Figure IV-5: Shipment Weight Profile Example



Based on the four selected weight profiles, an average and standard deviation is determined for each 10g weight bucket using exponential weighted averages. The exponential weights are intended to place more relevance to the most recent data set and the least relevance to the oldest. Based on the EU Transportation team's experience, with a few exceptions mostly based on seasonality or new product releases, the latest shipment profile generally best reflects the forthcoming shipment profile. The following general exponential weighted moving average (EWMA[14]) formula was used with a slight modification to that of textbooks:

$$w_p = \frac{(1-\alpha)^{p-1}}{\sum_p (1-\alpha)^{p-1}} \quad \text{where } \alpha = \frac{2}{N+1}$$

p = 1 (i.e. newest set), 2, 3, 4 (i.e. oldest set)

However, the application of these formulas was slightly modified. As stated earlier, the length of time each profile covers is a function of how many days are left in the month for the month in question. In the case given earlier, the profiles are 14 days long. If the present

[14] Diebold, Francis. Elements of Forecasting. 2nd Ed (2001) p. 355

day would have actually instead been Aug 28th, then the length would have changed to 4 days. The chosen data sets could have then been (1) Aug27-Aug24, (2) Aug23-Aug20, (3) Aug19-Aug16 and (4) Aug15-Aug12. The closer one is to the start of the month, the farther away in time the fourth data set will be. Thus, it would be unfair to penalize the fourth data set for the Aug 28th case (i.e. Aug15-Aug12) equally to that of the fourth data set for the Aug 8th case (i.e. May27-May04). In order to account for the length of time periods and the imposed effect on the age and thus relevance of the data, the following changes were made:

$$w_p' = w_p + \left(\frac{1}{P} - w_p\right) * \sqrt{\frac{DaysPassed}{TotalDays - 1}} \quad \text{where } P = p_{max} = 4$$

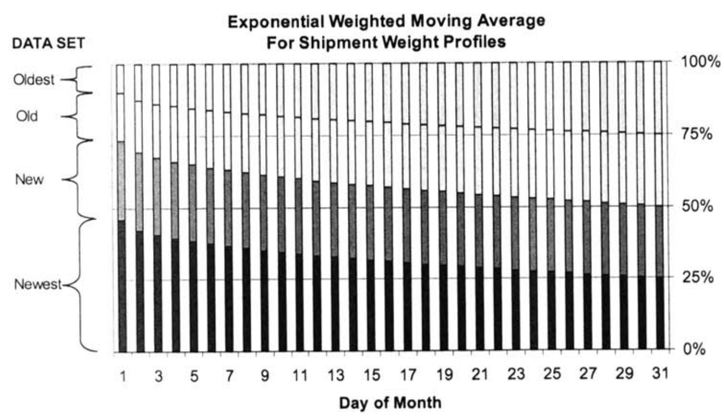
Using these equations along with N set equal to 4[15], the weights shown in Figure IV-6 would apply to each of the above examples: Aug18 vs. Aug 30. Note how the closer "today" is to the start of the month, the closer the weights are to an application of pure exponential weights (i.e. 46%, 28%, 16% and 10%). Similarly, the closer "today" is to the end of a month, the closer the weights are to an even distribution (i.e. 25%, 25%, 25% and 25%). This makes sense considering that the last period for Aug 28 is only 17 days away as compared to Aug 8 being 97 days away.

Figure IV-6: Modified Exponential Weighted Average Example

Scenario "Today"	Data Set Newest	New	Old	Oldest
August 8	Aug07-Jul15	Jul14-Jun21	Jun20-May28	May27-May04
Exp Wgts	**36%**	**26%**	**21%**	**17%**
August 28	Aug27-Aug24	Aug23-Aug20	Aug19-Aug16	Aug15-Aug12
Exp Wgts	**26%**	**25%**	**25%**	**24%**

The above example demonstrates the modified application of exponential weights for 24 and 4 day long periods. Figure IV-7 provides a look into the weight allocation for any period length as a function of the day in the month in which the study is made. It is important to notice the exponential decrease (increase) in weight for the newest (oldest) data set as the month comes to an end.

Figure IV-7: Modified Exponential Weighted Average Application



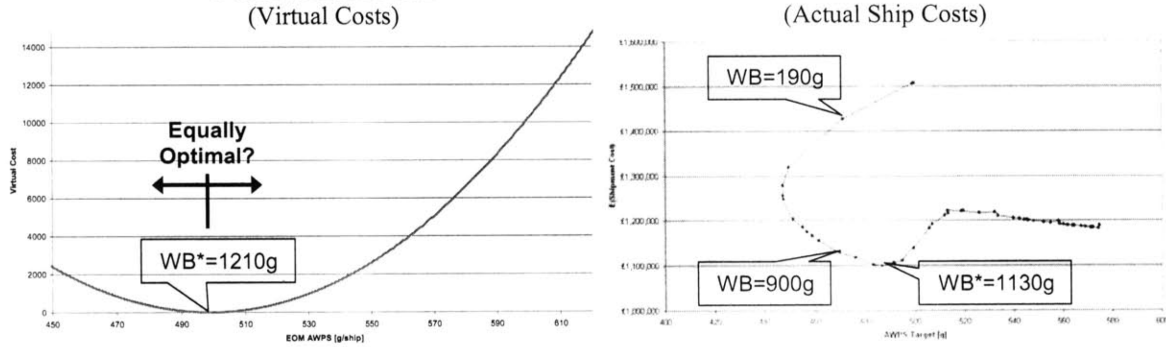
15 N is set to 4 since there are 4 periods being studied. However, the user is allowed to modify N. For example, the less confident the user is with the older data sets then the lower N should be set (i.e. 2) to reduce their relevance.

The stochastic model was developed to address all three flaws previously mentioned of the deterministic model. Firstly, through the improved logic, the model now incorporates variability into its output. Secondly, the model now provides an optimal AWPS target. Lastly, the model now offers descriptive information regarding all mail carriers, including Regal Mail. This allows users to study the optimization effects on all mail carriers. For example, the user is now able to set constraints on carrier monthly volume limits, evaluate how close these limits are approached in any given solution and determine the probability (i.e. risk) of such an event.

The last two improvements are a result of changing the objective function from a virtual to an actual shipment cost function which includes MTD and remaining expected shipment costs for all mail carriers within each FC-shipment type combination. Moreover, including all carriers enabled total outbound transportation cost optimization. Minimizing only Regal Mail shipment costs would most likely have led to higher total transportation costs or infeasible "solutions" when incorporating the effects upon other carriers. Figure IV-8 demonstrates the drastic change in objective functions between both models. For confidentiality, dates and actual cost data have been masked for the stochastic model plot.

Figure IV-8: Objective Function Difference

Deterministic Model (Virtual Costs) | **Stochastic Model**[16] (Actual Ship Costs)



In theory, the stochastic model is a clear improvement over the deterministic model as it includes all costs and accounts for variability. In order to prove this in actuality though, a validation study was performed.

4.1.2 Validation Process & Results

The best way to demonstrate the efficacy of both models was through simulation using the actual shipment volume and profile for a previous month and then comparing resulting EOM total shipping costs[17]. The following validation study was preformed mainly to determine the difference in total shipment costs between having had run the stochastic model as opposed to the deterministic model for a previous month and FC-Shipment type combination. In addition, as a benchmark to both models, the results were compared to the lowest possible costs that could have been obtained if one had known the exact ship profile and volume in advance of the start of the month.

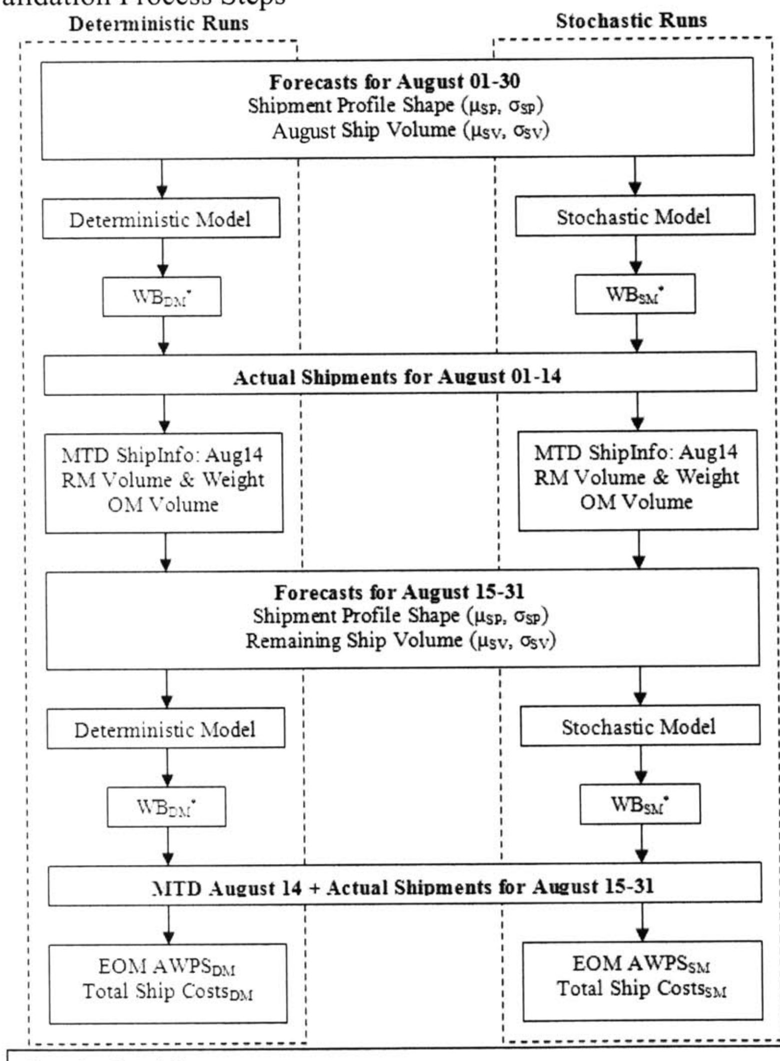
[16] The peculiar shape of this total shipment cost curve is later explained within this section.

[17] For confidentiality, all cost curve figures, months and FC names have been shifted and/or masked.

The following table outlines the fact basis for the validation study. The validation steps are also provided in Figure IV-9.

Month	August
Fulfillment Center	FC1
Shipment Type	Standard
WB Review Periods	2 (Aug01 & Aug15)

Figure IV-9: Validation Process Steps



Narrative Description:

1. Data mined actual daily shipments for August 2006 FC1 Standard ship option
2. Determined WBs (as if it were Aug 01) using the Deterministic and then the Stochastic Model
3. Ran a simulation of actual shipments from Step 1 for all days between 01-Aug and 14-Aug using the Deterministic and then the Stochastic WBs from Step 2
4. Using MTD Regal Mail and Postal shipment results from the simulation in Step 3, the WBs were determined (as if it were Aug 15) using the Deterministic and then the Stochastic Model
5. Ran a simulation of actual shipments from Step 1 for all days between 15-Aug and 30-Aug using the Deterministic and then the Stochastic WBs from Step 4 along with the respective MTD shipment results from Step 3
6. Recorded the EOM Ship Costs and AWPS from Step 5 for each model
7. Determined the WB that achieved the lowest possible shipment costs for all of August had one known the exact actual ship profile in advance (optimal EOM AWPS 500g)
8. Collected the Best Possible Scenario EOM Ship Costs from Step 7 and recorded it

The validation results are shown in Figure IV-10. The study proves that the stochastic model is an improvement over the deterministic model. It also demonstrates how the stochastic model better allows for correction throughout a month in progress. The "Best Possible" scenario shows that an EOM AWPS of 500g would yield the lowest possible cost. At the start of the month, the stochastic model proposed a WB of 1,490g which led to a high AWPS of 506g at mid-month. At this point, the model recognized the costly overage and based on new and improved forecasts for the remainder of the month proposed a lower WB of 1,350g which led to an AWPS of 497.6g at the EOM.

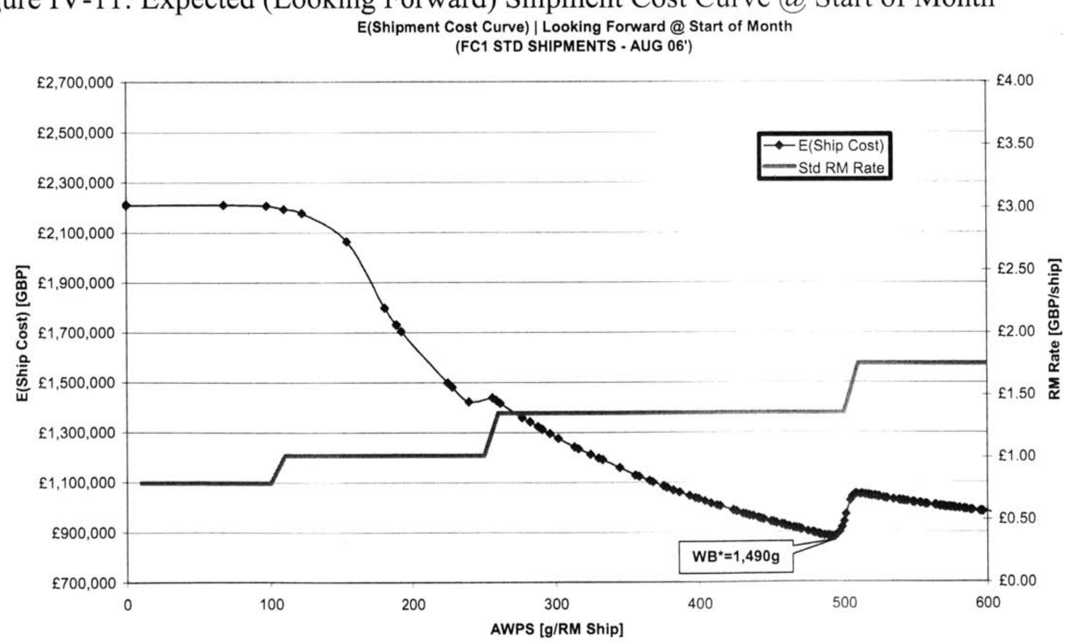
Figure IV-10: Validation Results

FC1 Standard Shipments
August 2006

Model	Weight Break Aug 01 - Aug 14	AWPS Aug 14	Weight Break Aug 15 - Aug 30	EOM AWPS August	Total Ship Costs August
Deterministic	1,510	509.3	1,440	507.5	**20.5% Higher**
Stochastic	1,490	506.4	1,350	497.6	**0.5% Higher**
Best Possible	1,420	496.9	1,420	500.0	**Benchmark**
					Stochastic WBs Results in **16.6% Lower Costs** than the Deterministic WBs

Using the validation study runs as an example, Figure IV-11 and Figure IV-12 show plots of the stochastic model's calculations for the run on Aug01 and Aug15, respectively. For each review period, these plots clearly answer how close EOM AWPS targets should be set to the rightmost edge of the rate step along with the required WB necessary to achieve it given the amount of forecast variability. Also note how the stochastic model's expected (looking forward) cost curve in Figure IV-11 resembles the actual (looking backward) cost curve shown in Figure III-10. The similarity in shape and magnitude further validate the model.

Figure IV-11: Expected (Looking Forward) Shipment Cost Curve @ Start of Month[18]

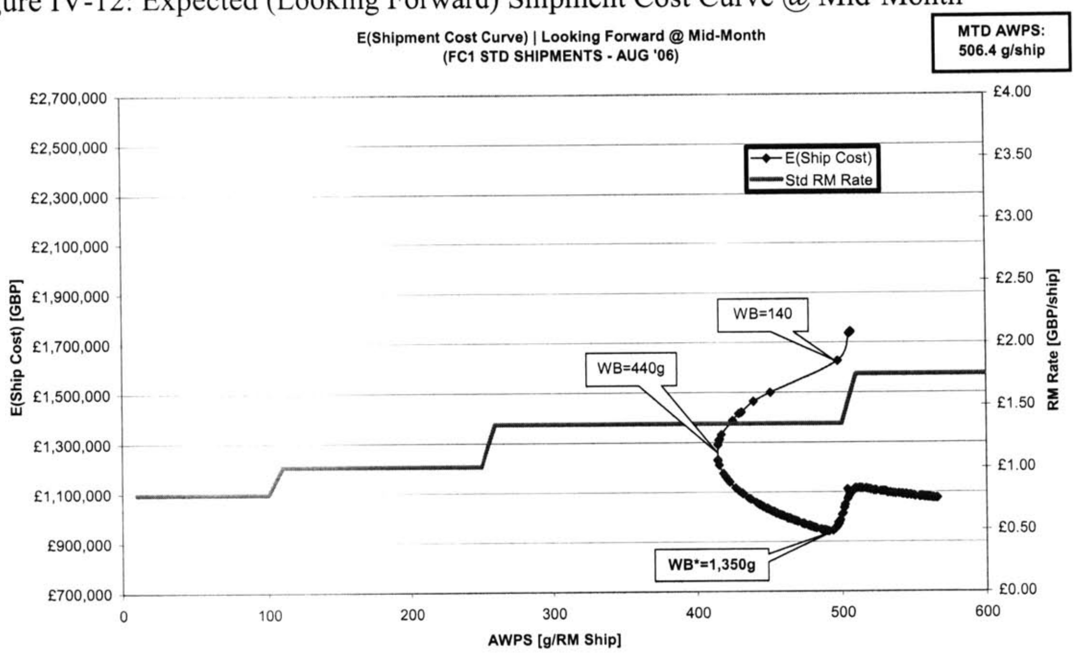


[18] For confidentiality, all cost curve figures, months and FC names have been shifted and/or masked.

It is important to note that each point of the total cost curve is defined by three dimensions: expected total EOM shipment cost, expected EOM AWPS and the respective WB. For simplicity, two dimensional plots were provided (cost vs. AWPS); however, WBs are labeled at key points of the cost curve. For example, implementing a WB of 1,490g at the start of the month would yield the lowest expected EOM total cost with an EOM AWPS below 500g.

Figure IV-12 shows the value of the model as a dynamic control decision-making tool by informing management of the range of possible choices (EOM AWPS targets) given MTD performance along with onward WB guidance to most likely minimize EOM total ship costs.

Figure IV-12: Expected (Looking Forward) Shipment Cost Curve @ Mid-Month



As shown in the top right corner of Figure IV-12, the MTD AWPS at mid-month was 506.4g. This however was higher than the targeted AWPS. As a result, two options were available at that point to bring the AWPS back down to target: (1) lowering the WB to 140g resulting in high total EOM ship costs as shown in the figure or (2) lowering the WB to 1,350g resulting in the lowest possible total EOM ship cost as also shown in the figure.

The first option involved a drastic change in WB at mid-month from 1,490 to 140g. This would have been a suboptimal approach at lowering the AWPS since it would have been expensive to ship only a few but very light shipments (below the 140g WB) through RM for the rest of the month while shipping the remainder (and majority) of shipments through the higher-rate, flat-rate carrier. On the other hand, the second option involved a more gradual change in WB at mid-month from 1,490 to 1,350g. This approach would optimally bring the AWPS back down by replicating the first half of the month with the exception that the WB would now be lowered just enough so that the remaining RM shipment volume would pull the total month's AWPS down to the lower rate step.

The above explanation clarifies the peculiar shape of the total ship cost curve during a month in progress. In fact, it is not surprising that as the end of the month is further approached, the total cost curve loop actually tightens since the impact of changing the WB on the month's total AWPS (or similarly total ship cost) reduces with time.

4.1.3 Implementation Process & Results

The project was broken down into two implementation phases for two main reasons: time and buy-in. The time factor was previously covered; however, the buy-in factor has not. The reasoning was that it would be much easier to implement a stochastic, non-linear optimization model if a simpler, deterministic version of it was first well-accepted. Once buy-in for the first model was acquired, demonstrating the old model's weaknesses and the method in which the new model addressed them all seemed an easier approach at proposing and eventually implementing seemingly complex analytical techniques (i.e. Monte Carlo simulation and non-linear optimization).

The EU Transportation team's willingness to improve upon decision-making for PiP management along with the phased buy-in strategy worked well since the stochastic model was in fact implemented starting December 2006. In order to make the transition smooth, the stochastic model was built as user friendly as possible; training and support on the use of the model and the Crystal Ball software was also provided to members of the team to whom ownership of the model was eventually transferred to by the end of the internship. At the time this paper was written, the model continued to be used on a bi-weekly basis. As shown by the validation results outlined in Figure IV-10, the stochastic model has been estimated to reduce outbound transportation costs by 15%.

In the end, the most valuable aspect of the stochastic model lies in that even though the model is an improvement in accuracy, scope and flexibility it is nonetheless still as practical as the deterministic model. During each WB review period, going through each FC and each shipment type takes a total of no more than 45 minutes. A Standard Operating Procedure for operating the model is shown in **Figure A-1**. Screenshots of the model's Excel user interface are shown in **Figure A-2**. Further screenshots of a Monte Carlo simulation and stochastic optimization process are shown in **Figures A-3** and **A-4**. As previously mentioned, the model is meant to be flexible; flexible enough to accommodate for events such as new product releases (e.g. Harry Potter releases) as well as any future changes to Regal Mail's rates and pricing strategy by adjusting specific input data accordingly.

4.1.4 Next Steps & Challenges

While the model was build as user friendly as possible within the allotted time frame of the internship, there is still room for improvement. Using Excel VBA and MSAccess it is possible to actually automate the entire process, including automation of input data transfer from Amazon's database, input data transfer from the EU Transportation team's daily ran shipment volume forecasts and the necessary Crystal Ball operations to obtain WB and AWPS target output. Automation could reduce the WB review process by up to 60%.

The purpose behind automation would be to reduce time spent on non-value added activities. On the long-run however, it would be of great value to investigate the possibility of completely automating the monitoring, review and adjustment of WBs. Having process controls setup so that alarms were to be activated in case MTD AWPS deviated from target beyond a certain predetermined percentage control limit. The alarm would itself trigger the model to be run to review MTD data and propose a new WB and/or AWPS target. The EU Transportation team could then review the model's results and implement changes as per their best judgment.

4.2 EU Product & Inventory Allocation Optimization

In August 2006, Amazon's DE network transitioned to a bi-nodal network with the opening of its second FC in Leipzig. This and the continued strong growth rate of all Amazon EU networks coupled with capacity constraints underscored the need for a scalable analytical framework for determining the optimal placement of product lines (PLs) and inventory across FCs in multi-node EU networks.

This project was intended to develop short-term guidance in this matter and propose a long-term, scalable framework. For the short-term, inventory allocation guidance based on PL placement was provided for the UK for both peak 2006 and non-peak 2007. For the long-term, an extension of the current EU network model shown on Figure III-13 was proposed to not only determine the optimal inventory allocation but also the optimal PL placement across FCs.

4.2.1 Project Proposal

It is important to begin this section by firstly defining "splits" as it is relevant to inventory allocation at Amazon. A split is defined as an order that gets broken down into more than one shipment. There are various root causes to a split including the size of ordered items, inventory availability and promised ship date constraints. For example, an order composed of a sortable (i.e. small) book and a non-sortable (i.e. large) microwave is split for practical purposes and shipped separately. These types of splits are outside of the control of Amazon. However, there are other types of splits that are within the control of Amazon that are labeled penalty splits. Penalty splits include those orders which could have been fulfilled by more than one FC but were not due to inventory availability or strategic inventory placement. Penalty splits increase both outbound transportation and labor fulfillment costs which are covered by Amazon as opposed to the customer.

Currently, the EU SC Team utilizes the network design model illustrated on Figure III-13 to determine both UK and DE inventory allocation (FAS) factors for each product line (PL). Through the minimization of transportation and labor fulfillment costs, the model provides Amazon with an optimal inventory allocation scheme for a given set of parameters, constraints and PL configuration. However, the current model makes two main assumptions:

1. Allocation factors do not have an effect on the likelihood of penalty splits
2. Current PL allocation across FCs is optimal

These two assumptions weaken the current model's validity. For example, the first assumption would neglect the likely event, and corresponding fulfillment costs, of penalty splits related to a scenario in which most books are allocated to one FC and most DVDs to another FC when in fact customers regularly place orders composed of both book and DVD items. The likelihood and related costs of having to source multi item orders across FCs should be considered. In any case, the intent of the network model is to minimize total fulfillment costs through optimal inventory allocation and split costs are an important driver of fulfillment costs. Similarly, it is not clear whether the current PL arrangement across FCs in any of the EU multi-node networks is optimal. That is, it is not clear whether a different PL arrangement might lead to more or less splits and/or total inventory, both of which are important drivers of fulfillment costs.

These assumptions were negligible in previous years given that the EU only had one multi-node network (i.e. the UK) and that the UK basically functioned as a single node with FC2 sharing a small fraction of the load. The situation in 2006 going into peak was very different though and these assumptions could no longer be neglected.

Phase I: A Short-Term Solution

At the onset of implementing the UK Q4 inventory allocation factors, the first assumption was questioned. Given peak demand forecasts and FC1's inventory capacity constraints, FC2 would now need to carry more than half the network inventory for books, music, video and DVDs (BMVD) to fully cover network demand (See Figure IV-13). As a result, the EU SC Team was concerned with whether the UK's Q4 strategic placement of inventory and product line configuration would inherently drive splits and thus fulfillment cost.

Figure IV-13: UK Inventory Allocation Factors[19]

Product Line	Non-Peak '06		Peak '06	
	FC1	FC2	FC1	FC2
Books	65%	35%	45%	55%
DVD	60%	40%	40%	60%
Music	60%	40%	45%	55%
Video	50%	50%	50%	50%
All Other	100%	0%	100%	0%

As a mostly uncontested business rule, FC1 carried all product lines whereas FC2 only carried high velocity (i.e. highly demanded) BMVD product lines. The shifted inventory to FC2 was thus composed of only high velocity BMVD inventory. The EU SC team questioned whether there would be enough single and multi high velocity BMVD-only orders during upcoming peak to justify having FC2 carry such a large percentage of all BMVD items while not carrying any other PLs. If this were not the case, then the inventory and PL arrangement would inherently drive strategic inventory related splits.

For example, consider an order composed of a high velocity book such as the latest Harry Potter book and a video game. In this case, the entire order would have to be fulfilled from FC1 to avoid a penalty split since FC2 does not carry video games. With the BMVD inventory shift, FC1 would now have a lesser percentage of Harry Potter books and as such the chances of having a stockout at FC1 would be greater. In the case of a stockout, the order would then have to be split between FC1 (ships video game) and FC2 (ships book).

Having peak around the corner, the UK's Q4 FAS factors had to be implemented as soon as possible. As a result, questioning the validity of the second assumption was held off and all

[19] Allocation factors are masked for confidentiality

efforts were focused on determining what the impact of the proposed factors coupled with the existing PL configuration would be on penalty splits.

In order to address this concern, an order profile analysis was carried out for last year's peak season to get a better understanding of whether the proposed factors were in-line with historical order behavior in the UK. A multi-tiered order analysis was carried forth for both peak and non-peak. Figure IV-14 is a boiled-down example of the process involved in obtaining order behavior.

Figure IV-14: Historical Order Behavior during 2005 Peak Season[20]

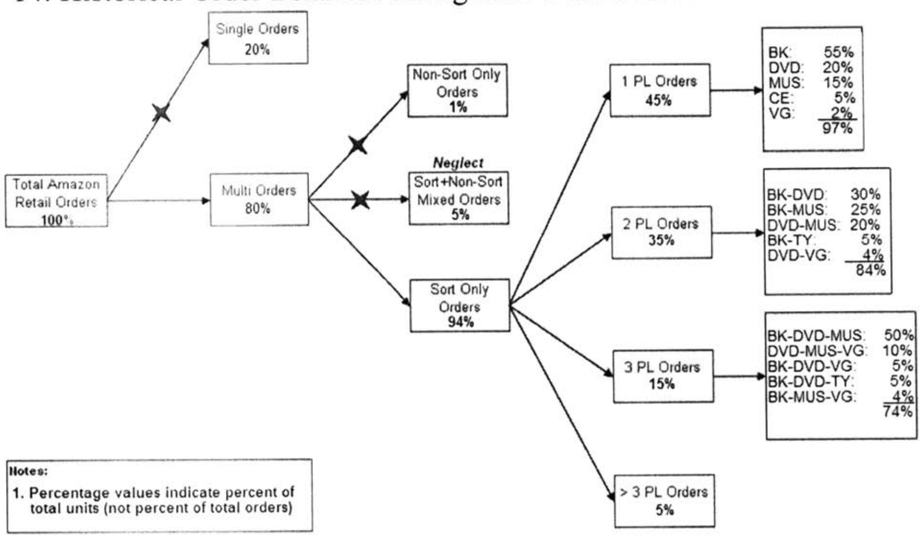


Figure IV-15 outlines useful decision-making information that was extracted from the entire order profile analysis to specifically gauge how the current PL network configuration inherently affected FC2 order fulfillment capacity.

Figure IV-15: Short-Term Guidance on UK Inventory Allocation[21]

Total Orders FC2 Could Fulfill	
Peak '06	**Non-Peak '07**
60%	70%

PL	FAS FC2 Healthy Upper Limits	
	Peak '06	**Non-Peak '07**
Books	55%	45%
DVD	85%	75%
Music	75%	65%
Video	65%	45%
Overall	**70%**	**55%**

The left set of numbers was obtained by determining *the percent of orders (single or multi) that were solely composed of items carried by FC2*, namely high velocity BMVD items.

The right set of numbers was derived in a similar manner. The "Healthy Upper Limit" is defined as *the percent of ordered items for a given PL that are both carried by FC2 and found within orders composed solely of items that FC2 carries.* For example, according to

[20] Percentage values are masked for confidentiality
[21] Guidance results are masked for confidentiality

the stylized values from the right table, during peak 55% of all ordered books are expected to be both high velocity books and found in orders (single or multi) which are solely composed of high velocity BMVD items.

The problem with setting FC2 FAS factors above the "healthy" limit is that in doing so, the number of multi orders that could potentially split would increase since the likelihood of having to fulfill a percentage of items of one multi order from both FC2 and FC1 increases. So in the example of books, if FC2 had an FAS factor of 60% then FC2 would be forced to fulfill 5% (=60%-55%) of ordered high velocity books which are coupled with items that FC2 does not carry. An example of such an order would be one composed of a Harry Potter book and a toy.

Since the proposed FC2 Q4 FAS factors for each PL were actually all below their "healthy" limit, the EU SC team was able to justify having FC2 carry greater than half of all BMVD items while not carrying any non-BMVD inventory in time for the required implementation of the FAS factors. In fact, even though there were other factors at play as well, it was still comforting to know that penalty splits did not increase after the new FAS factors were introduced on October.

Beyond this, the results offered the team useful insight on the differences in FC2 order fulfillment capabilities between peak and non-peak season. For example, given the difference in order behavior between seasons, FC2's current PL configuration enables it to "healthily" handle more volume during peak than non-peak. Finally, the results would also offer guidance in setting non-peak FAS factors for 2007.

Phase II: A Long-Term Solution

With Phase I of the project completed, the question still remained: Is the current PL configuration in the UK (and now also DE) optimal? As mentioned earlier, the current network model optimizes inventory allocation given a predetermined PL configuration. The EU's rapid growth and increasingly complex network demands a scalable network design model which takes into account the impact of PL configuration on network level fulfillment costs. Network complexity limits the value of studies such as that carried forth during Phase I. Furthermore, the study was only able to provide Amazon with limits as opposed to optimal allocation factors. To meet this growing need, an extension of the current network model was proposed through the addition of logic in an attempt to eliminate both of the existing model's aforementioned assumptions.

The existing network model is composed of three main components:

1. Labor Costs & Constraints
2. Transportation Costs
3. Storage Constraints

Two additional components are proposed to extend the current model to be able to relax PL configuration constraints and account for those fulfillment costs brought about by a network's PL configuration and inventory allocation scheme:

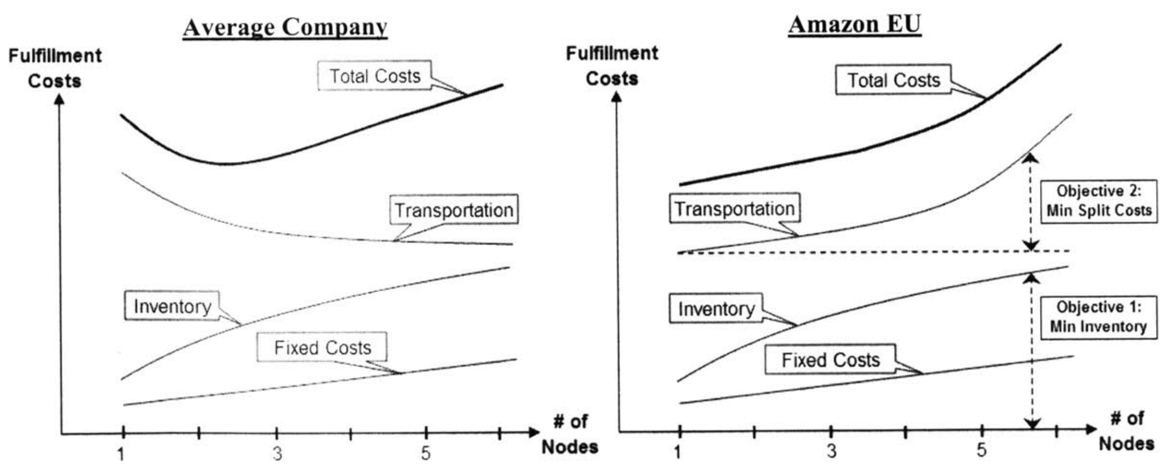
1. Safety Stock Holding Costs
2. Network Design Induced Split Costs

The purpose behind incorporating safety stock costs is to ensure that the model does not try to spread PL's (within storage constraints) to as many FC's as possible, a potential way of minimizing split costs. Network design induced split costs is meant to ensure that the model does not provide bipolar inventory stocking solutions. For example, having most of your books in one FC and most of your DVDs in another FC given that 10% (according to the stylized order profile figure shown above) of all multi orders are books with DVDs would most likely drive split costs for that specific order mix. In a sense, both of the proposed cost components check and balance each other out since minimizing the first would lead to PL centralization while the second would lead to PL and inventory spread.

Figure IV-16 further demonstrates the purpose behind these two additional components. It is important that note that even though in general, outbound transportation costs should decrease with an increasing number of distribution center nodes, this is not the case with Amazon EU. As previously mentioned, the main reason behind this is European mail carrier's rate structures which are not a function of distance. Thus, even though additional nodes would increase customer proximity, this would not lead to lower outbound transportation costs for Amazon EU. In fact, more nodes would most likely correspond to more splits and thus higher transportation costs. The objective behind including these two components is to influence PL and inventory allocation such that total fulfillment costs, including inventory and transportation cost contributions, is shifted downward.

Finally, it is important to note that Figure IV-16 in reality oversimplifies actual conditions since it excludes other costs and factors such as labor and real estate imposed capacity constraints, respectively. At a high level though, the plots are useful in demonstrating the need to incorporate both additional components.

Figure IV-16: Comparison in Fulfillment Cost Contribution



The formidable challenge for this project is in mathematically expressing and linking safety stock and network induced split costs to overall network design. This is especially the case considering that there is currently not much literature in this matter. Once acceptable and validated relationships are established, then the next step would be to incorporate these components into the current model's objective function:

$$\text{Minimize}\begin{cases}\text{Network Design Induced Split Costs}\\ \text{Safety Stock Holding Costs}\\ \text{Labor Costs}\\ \text{Transportation Costs}\end{cases}$$

With these two additional components in place, the current model's PL configuration constraints could then be relaxed to allow the model to optimally decide (1) where to allocate PLs and (2) what percent of inventory for each PL should be placed in each FC. The main output for the model would then be:

$FAS_{i,j,t}$ = Percent of PL i Inventory Allocated to FC j during Time Period t

The first step in improving upon the existing network model was in updating it given that the SPOT team had recently improved upon the accuracy, scope and flexibility of both the Labor and Storage sub-models. Having merged the revised sub-models to the existing model, the next step involved developing the mathematical expression for safety stock holding costs.

Safety Stock Component

Given that purchasing at Amazon EU is currently based on a Periodic Review (**Figure B-1**) policy[22], safety stock levels could theoretically be estimated as follows for any PL[23]:

$$SS = z * \sigma * \sqrt{r + VLT} \quad \textbf{(1)}$$

SS	Expected safety stock inventory for a PL
z	Multiplier that is a function of the chosen service level for a PL (Historical)
σ	Standard deviation of weekly demand for a PL (Computed from Equation 4)
VLT	Average vendor lead time in weeks from purchase order submission to receive (Historical)
r	Average review/planning period in weeks (Historical)

Thus, the safety stock holding costs could then be estimated as follows:

$$SSHoldingCost = h * SS * W \quad \textbf{(2)}$$

h	Weekly average per unit inventory holding cost (Historical)
W	Number of weeks during time period of interest (Given)

It is important to note that inventory holding cost (i.e. h) includes both the opportunity cost of tied capital and the physical cost of storing inventory (e.g. spoilage, obsolescence, pilferage, insurance, security, rent and lighting). As a result, each product line merits its own estimate for its related holding cost. More specifically to Amazon EU, the safety stock holding cost for any product line j could then be estimated as follows:

$$SSHoldingCost_j = h_j * z_j * W * \sum_i \left(\sigma_{i,j} * \sqrt{VLT_{i,j} + p_i}\right) \quad \textbf{(3)}$$

i	Fulfillment Center I
j	Product line j

[22] If a Continuous Review inventory policy is later adopted then the "r" must be simply removed from all equations

[23] Simchi-Levi, David et al. Designing & Managing the Supply Chain. 2nd Ed (2003) Chapter 3

Where the demand variation by FC and PL (i.e. $\sigma_{i,j}$) can be estimated for any inventory allocation ratio by using the relationship[24] shown below. The validity of this estimate is demonstrated later in this section.

$$\sigma_{i,j} = FAS_{i,j}^{\alpha_j} * \sigma_j \quad \textbf{(4)}$$

$\sigma_{i,j}$	Standard deviation of weekly demand for PL j in FC I
α_j	Regression-derived constant applicable to PL j (Historical)
σ_j	Standard deviation of weekly network demand for PL j (S&OP Forecasts)

The α constant for each PL during both peak and non-peak season were derived through log-log regressions while setting the intercept to 0. The regressed data sets consisted of n (=13 in peak and =39 in non-peak) weekly data points. The following relationship was used for the regression:

$$\sigma_{i,j,t} = FAS_{i,j,t}^{\alpha_j} * \sigma_{j,t} \quad \textbf{(5)}$$

$\sigma_{i,j,t}$	Standard deviation of daily demand for PL j in FC i during week t (Historical)
$FAS_{i,j,t}$	Percent of inventory for PL j in FC i during week t (Historical)
σ_j	Standard deviation of daily network demand for PL j during week t (Historical)

In simpler terms, Equation 5 is equivalent to:

$$A = B^{\alpha} * C$$
$$Log(A/C) = \alpha * Log(B)$$

Results from the regressions are shown in Figure IV-17. Furthermore, an example of one regression (books) is offered in **Figure B-2**. It is interesting to note that all alpha values are consistently greater than 0.7 yet less than 1.0. Further insight into the meaning behind these values and their relationship to demand and inventory are provided later in this section.

Figure IV-17: Alpha Regressed Coefficient Results by PL[25]

Product Lines	BK (Book)	DVD (DVD)	MUS (Music)	VHS (Video)	CE (Electronics)	VG (Video Game)	TY (Toys)	KI (Kitchen)	SW (Software)	HI (Home)	OL (Outdoor)
Regressed α	0.85	0.80	0.75	0.70	0.95	0.95	0.90	0.95	0.95	0.95	0.90

As noted in **Equation 3**, a major driver of safety stock inventory levels is a product line's standard deviation of demand (σ). Intuitively, the higher a product's sales volume and sales variability, the higher the safety stock should be to ensure against stockouts. A common form of inventory reduction in industry is by pooling inventory into as few locations as possible. In this manner, ensuring against stockouts need only occur in fewer locations.

The following example demonstrates the advantages of inventory pooling. Assume a firm sells only one product to two regions. The standard deviation for each region and the correlation in demand is equal to σ_1 & σ_2 and $\rho_{1,2}$ respectively. The firm has two options:

Decentralized Inventory		Centralized Inventory	
Scenario	Two small warehouse	Scenario	One large warehouse
Total Std Dev	$\sigma_1 + \sigma_2$	Total Std Dev	$[\sigma_1^2 + \sigma_2^2 + 2*\rho_{1,2}*\sigma_1*\sigma_2]^{1/2}$

[24] Rosenfield, Donald et al. The Logistics Handbook: Section IV, Chapter 14: Demand Forecasting. (1994)
[25] Alpha values are masked for confidentiality

Safety stock as shown by **Equation 1** is proportional to the total standard deviation of demand. All else equal, a decentralized inventory will typically have more safety stock. The percentage increase in inventory is defined as:

$$SSIncrease = \frac{\sigma_1 + \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + 2 * \rho_{1,2} * \sigma_1 * \sigma_2}} \quad \textbf{(6)}$$

Assuming the standard deviation of demand from each region is equal, then the inventory increase would actually equal to:

$$SSIncrease = \sqrt{\frac{2}{1+\rho}} \quad \textbf{(7)}$$

This phenomenon is referred to as the "Square Root Law"[26]. For simplicity, industry assumes independent demand between regions (which is typically incorrect) and thus a correlation of 0. This leads to an inventory increase of √2 (i.e. 41.4%). This is a rough approximation however that rarely holds.

In the more realistic scenario where demand is correlated and standard deviations are not equal, safety stock increase as per **Equation 6** is therefore a function of two main variables: correlation of demand (ρ) and the variance ratio ($r_{1,2} = \sigma_1/\sigma_2$).

$$SSIncrease = \frac{1 + r_{1,2}}{\sqrt{r_{1,2}^2 + 2 * \rho_{1,2} * r_{1,2} + 1}} \quad \textbf{(8)}$$

The goal of **Equation 4**, which is empirically-based, is to emulate the theoretical formulation shown in **Equation 6** while relating to the network model's inventory allocation decision variable (i.e. FAS). Validation for the empirical formulation can be demonstrated by using the same illustrative example of a firm serving two regions.

$$\sigma_1 = \sigma_T * FAS_1^{\alpha} \quad \text{and} \quad \sigma_2 = \sigma_T * FAS_2^{\alpha}$$

$$SSIncrease = \frac{\sigma_1 + \sigma_2}{\sigma_T} = FAS_1^{\alpha} + FAS_2^{\alpha} \quad \textbf{(9)}$$

Figure IV-18 fully demonstrates the effectiveness of the empirical model in accurately emulating theory. The comparison shown in Figure IV-18 also clearly demonstrates that:

$$FAS_1^{\alpha} + FAS_2^{\alpha} \sim \frac{1 + r_{1,2}}{\sqrt{r_{1,2}^2 + 2 * \rho_{1,2} * r_{1,2} + 1}}$$

Moreover, the comparison study also reveals that the variance ratio ($r_{1,2}$) is proportional to inventory allocation (FAS). Similarly, the correlation of demand (ρ) is proportional to the regressed alpha coefficient (α). For example, an alpha of 1.0 is equivalent to a demand correlation of 100% whereas an alpha of 0.5 is equivalent to statistically independent

[26] Simchi-Levi, David et al. Designing & Managing the Supply Chain. 2nd Ed (2003) Chapter 7

demand. The highest safety stock increase occurs when both demand is independent and inventory is spread out equally among FCs (FAS = 50% for a bi-nodal network).

Conversely, there should be no safety stock increase with perfectly correlated demand. This insight was valuable to Amazon since demand at an FC-PL level is highly correlated[27]. For example, all FC book's outbound volume increases and decreases with peak and non-peak, respectively. It is for this reason that the regressed alphas shown on Figure IV-17 all approximate 1.0. Furthermore, FCs are unlikely to have equal variance in demand given that variance is proportional to FAS factors and inventory is rarely spread evenly across FCs for any given PL in the EU. Thus, Amazon EU's safety stock levels would increase only a fraction of industry's common "square root of 2 law" when spreading PLs across FCs.

Figure IV-18: Empirical vs. Theoretical Safety Stock Formulation Comparison

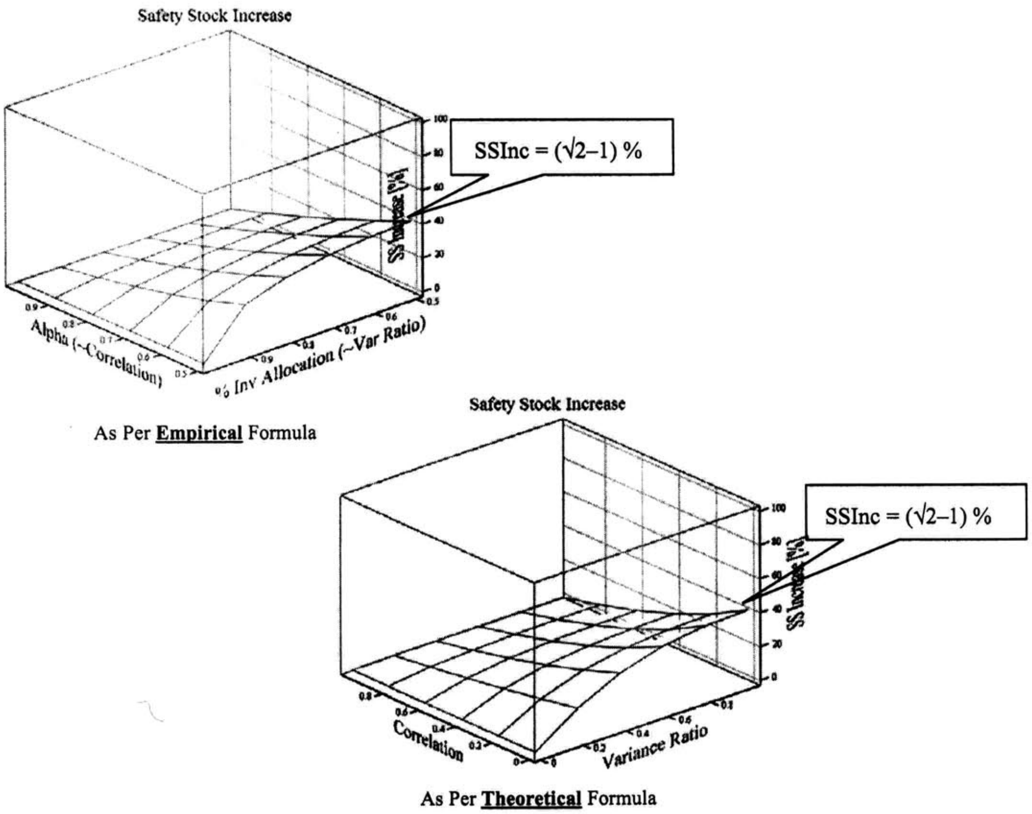


Figure IV-19 and Figure IV-20[28] better demonstrate the magnitude in which the regressed alphas affect safety stock levels. For illustrative purposes, both plots assume "all else equal". In other words, lead time, review period and service levels for each PL and FC are equal and constant. Figure IV-19 plots **Equation 9** as a function of inventory allocation (FAS_1) for FC1 utilizing the regressed alphas for books, music and DVD. Figure IV-20 plots safety

[27] This is partly the case due to the lack of geographical market segments within EU networks
[28] Alpha values are masked for confidentiality

stock increase as a function of the number of FC nodes assuming inventory is uniformly spread out across all FCs. For this figure, the following formula was used:

$$SSIncrease = Nodes * \left(\frac{1}{Nodes}\right)^{\alpha} \quad \textbf{(10)}$$

Figure IV-19: Increase in Safety Stock as a Function of % of Inventory Allocation (FAS)

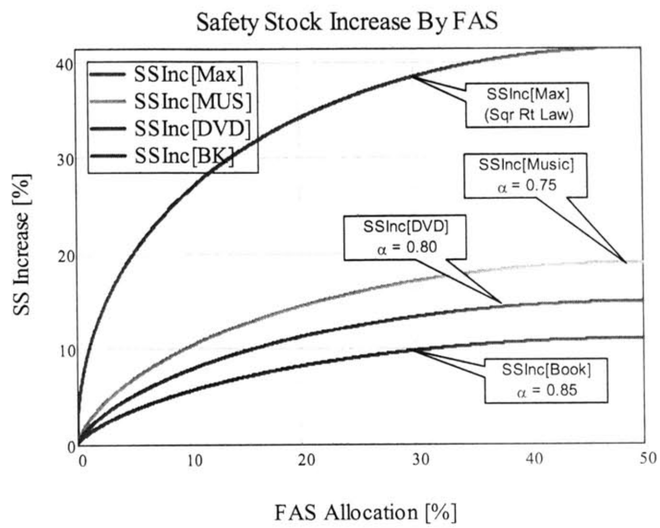
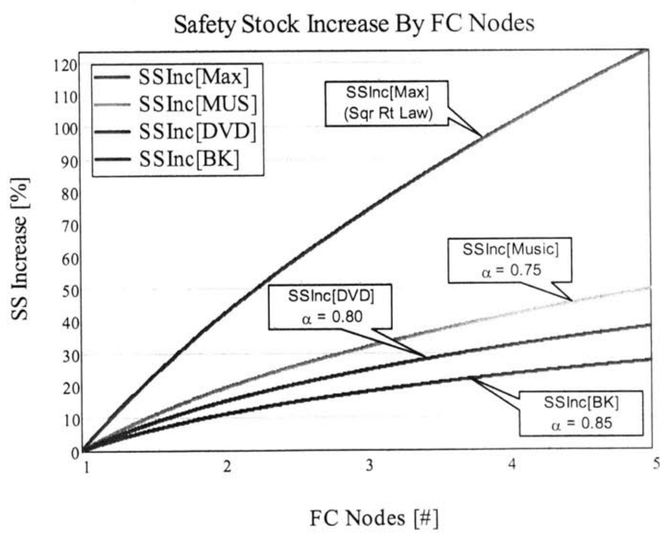


Figure IV-20: Increase in Safety Stock as a Function of Number of FC Nodes



Having laid the down the empirical framework for estimating safety stock levels, the next and final step involved introducing this component into the network model such that it flowed with the rest of the model's format and layout. The following equations are equivalent to **Equations 3 and 4** yet modified to flow with the model.

Figure IV-21: Safety Stock Holding Cost Component[29]

Objective Function Addition
(Equation 3)

$$\text{SSCost} = \sum_p \left[\text{InvHolding}_p \cdot \text{ServLvlMult}_p \cdot (\text{End} - \text{Start} + 1) \cdot \sum_f \left(\text{FC\_StdDev}_{f,p} \cdot \sqrt{\text{VLT}_{f,p} + \text{PlanPeriod}_p} \right) \right]$$

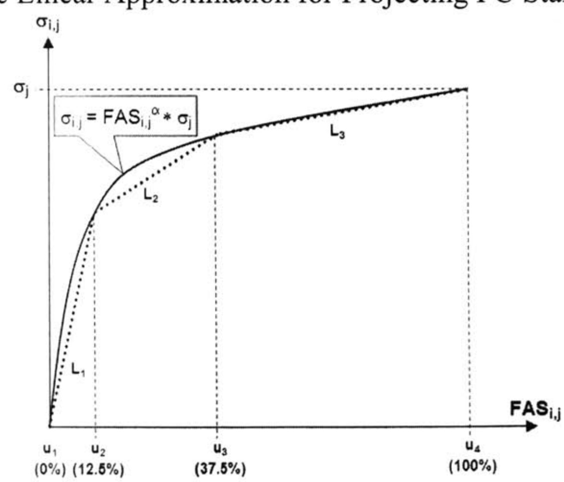
Constraint Addition
(Equation 4)

$$\text{FC\_StdDev}_{f,p} = \left(\sum_s \sum_v \sum_t \frac{\text{AllocInv}_{f,p,s,v,t}}{\text{Start} - \text{End} + 1} \right)^{\text{SSConstant}_p} \cdot \text{SaOP\_StdDev}_p$$

Note the similarities between these equations and **Equations 3 and 4**. The main difference is the naming of variables and parameters. For example, SaOP_StdDev_p stands for the standard deviation of S&OP's network demand level forecasts for every PL (i.e. σ_j). A main difference is in the formulation of the main decision variable $FAS_{i,j}$. In the network model, the inventory allocation factor is determined for every possible product entity (PE) throughout each week. For practicality though, an additional constraint is added to ensure that factors do not change across time. A product entity is a PL-sortability-velocity (i.e. p-s-v) combination. An example of a PE could be a sortable, fast book. This specific PE would include books like the latest Harry Potter.

A disadvantage of the formulation is that it is non-linear. The non-linearity is introduced through the raising the decision variable to the power of alpha. The entire network model is otherwise linear. A linear program (LP) is highly preferred over a non-linear program (NLP) for various reasons but mostly for tractability. The LP's low processing time and convenient analysis is highly valued by the SPOT team considering that the network model is sometimes required to churn out nearly 5,000 decision variables. It was therefore decided to create a piecewise-linear mathematical expression for the safety stock constraint (i.e. **Equation 4**) as shown in Figure IV-22. In this way, the constraint would no longer be considered non-linear.

Figure IV-22: Piecewise Linear Approximation for Projecting FC Standard Deviation



[29] Indices p, f, s, v and t stand for PL, FC, sortability, velocity and time, respectively

The expression was broken down into 3 piecewise-linear approximations. The first segment ranged from an FAS of 0% (u_1) to 12.5% (u_2), the second from 12.5% (u_2) to 37.5% (u_3) and the third from 37.5% (u_3) to 100% (u_4). The first pieces are shorter to better accommodate the shape and form of a typical function where the non-negative variable is less than one and is raised to the power of [0.5 to 1.0]. The number of pieces was limited to 3 for simplicity while sufficiently enough to capture non-linearity. The ranges can be modified by the user.

$$\sigma_{i,j} = FAS_{i,j}^{\alpha_j} * \sigma_j \quad \textbf{(4)}$$

Piecewise-linear formulation:

$$\sigma_{i,j} = \sigma_{i,j}(u_1) + \sum_{k=1}^{3} \left[\frac{\sigma_{i,j}(u_{k+1}) - \sigma_{i,j}(u_k)}{u_{k+1} - u_k} \right] * FAS_{i,j,k} \quad \textbf{(11)}$$

Additional constraints:

$$\begin{aligned} y_2 * (u_2 - u_1) \le FAS_{i,j,1} \le y_1 * (u_2 - u_1) \\ y_3 * (u_3 - u_2) \le FAS_{i,j,2} \le y_2 * (u_3 - u_2) \\ 0 \le FAS_{i,j,3} \le y_3 * (u_4 - u_3) \end{aligned} \quad \textbf{(12)}$$

Additional parameters:

$$[u_1, u_2, u_3, u_4] = [0\%, 12.5\%, 37.5\%, 100\%]$$

The additional constraints are implemented in order to ensure that each segment of the linear approximation reaches its upper bound prior to jumping to the next segment. As a result, three new binary decision variables were incorporated (y_1, y_2 and y_3). Making sense of the segmented decision variable (FAS) is then as simple as summing up the pieces.[30]

$$FAS_{i,j} = \sum_k FAS_{i,j,k} \quad \textbf{(13)}$$

The constraint equation shown in Figure IV-21 was modified accordingly to "linearize" it as per **Equation 11**. In addition, the three extra constraints and binary decision variables shown in **Equation 12** were also added to the network model. The main decision variable now also carried an extra index (i.e. k) and was revised accordingly throughout the model. With this component now fully developed and incorporated, attention was focused in developing the Network Induced Split Costs component.

Network Induced Splits Component

A way for approaching this problem was through a part probabilistic-part heuristic model. In this case, the formidable challenge was in accurately estimating the probability of a split occurrence. This can be mathematically expressed as follows:

$$P(S) = \sum_g [P(S \mid OM_g) * P(OM_g)] \quad \textbf{(14)}$$

$P(S)$	Probability that the "next" sort-only multi item order gets split due to network design
OM_g	Order mix g (Historical)
$P(OM_g)$	Probability that the "next" sort-only multi item order is equal to OM g (Historical)
$P(S \mid OM_g)$	Probability that the "next" order of mix g gets split due to network design (Heuristic)

[30] Rardin, Ronald. Optimization in Operations Research. (1998) Chapter 14

Using **Equation 14**, fulfillment costs associated to network design induced splits could be estimated as follows:

$$NetworkInducedSplitCost = P(S) * \frac{SOI * Multi}{\mu_m} * (\mu_s - 1) * UC \quad \textbf{(15)}$$

SOI	Total number of sort-only items to be ordered across all PLs (S&OP forecasts)
Multi	Percentage of ordered items that are part of multi item orders (Historical)
μ_m	Avg number of items per sort-only multi order (Historical)
μ_s	Avg number of boxes shipped per every sort-only multi order that is split (Historical)
UC	Additional fulfillment costs from shipping an extra box (Historical)

An objective behind the improved model is to avoid solutions where PLs are positioned and their inventory is allocated in such a manner that they are not aligned with order behavior, in other words, common order mixes. Proper order and inventory alignment ultimately leads to the minimization in likelihood of having to source orders from multiple FCs. Thus, the expression for P(S) needs to incorporate typical order behavior and the network model's main decision variable: FAS. The following simplified case example is used to illustrate a possible heuristic to achieve this objective.

Simplified Case Example:

Assuming Amazon only had 2 FCs and 2 PLs with the following inventory allocation and expected inventory availability.

FAS Factors		
PL	**FC1**	**FC2**
BK	60%	40%
MUS	30%	70%

Availability		
PL	**FC1**	**FC2**
BK	60%	65%
MUS	70%	75%

In estimating splits induced by network design, the term $P(S \mid OM_g)$ from **Equation 14** should be a function of the network design configuration. In addition, we know from internal company reports[31] that penalty splits are in fact statistically linked to availability. On this basis, the following part probabilistic-part heuristic method was developed and applied to this simplified case example for explanatory purposes. Consider an order composed of book and music items. Then,

$$P(S \mid Bk + Mus) = P(F_{1,2}) + P(F_{2,1}) \quad \textbf{(16)}$$

$F_{1,2}$	Event where all book items are shipped from FC1 & all music items are shipped from FC2
$F_{2,1}$	Event where all book items are shipped from FC2 & all music items are shipped from FC1

More generally,

$$P(S \mid OM_g) = \sum_{i1}\sum_{i2} P(F_{i1,i2}) \quad \textbf{(17)}$$

$F_{i1,i2}$	Event where all book items are shipped from FC i1 & all music items are shipped from FC i2 where i1 is not equal to i2

[31] Ng, Nicholas. Six Sigma Green Belt project (Aug '06) demonstrates a correlation between splits and availability

The right side of **Equation 17** can be calculated using Bayes' rule:

$$P(F_{i1,i2}) = \sum_{m} P(F_{i1,i2} \mid A_m) * P(A_m) \quad \textbf{(18)}$$

Where the term $P(A_m)$ can be calculated as follows:

$$P(A_m) = \prod_{i} \prod_{j} P(A_{i,j})^{A_{m,i,j}} * [1 - P(A_{i,j})]^{(1-A_{m,i,j})} \quad \textbf{(19)}$$

$P(A_{i,j})$	Probability that PL j is available in FC i (Historical)
$A_{m,i,j}$	Availability of PL j in FC i under Availability Scenario m (1 if available & 0 if not)
A_m	Availability scenario corresponding to each set of $A_{i,j}$ combination $m = 1, 2, \ldots, 2^{i*j}$
$P(F_{i1,i2} \mid A_m)$	Probability that PL1 and PL2 are shipped from FC i1 and i2, respectively, given the availability scenario A_m. This value is a function of FAS factors.

In a scenario where there is availability in both FCs for both PLs, all else equal, one would expect an order to get assigned to the FC with the largest amount of inventory for both PLs. This and all other scenarios are reflected in the term $P(F_{i1,i2} \mid A_m)$. Values for all $P(F_{i1,i2} \mid A_m)$ combinations are enumerated and illustrated in **Figure B-3**. However, for further illustration, two basic scenarios are calculated in detail in this section beginning with the aforementioned availability scenario (i.e. $A_{m=AllBoth}$) where both FCs have both PLs available. In this case, a split would not be likely and thus

$$P(F_{1,2} \mid A_{AllBoth}) = P(F_{2,1} \mid A_{AllBoth}) = 0\%$$

However, in the unlikely case that at the time of the order neither FC had availability for neither PL (i.e. $A_{m=NoneBoth}$) then the likelihood of a split is estimated as follows

$$P(F_{1,2} \mid A_{NoneBoth}) = 42\% \; (= FAS_{FC1,Book} * FAS_{FC2,Music} = 0.6*0.7)$$

$$P(F_{2,1} \mid A_{NoneBoth}) = 12\% \; (= FAS_{FC2,Book} * FAS_{FC1,Music} = 0.4*0.3)$$

The chance of this last availability scenario is low however. The term $P(A_m)$ incorporates the probability of each availability scenario into the estimation of the overall probability of a split in a Bayesian format as shown in **Equation 18**. Values for all $P(A_m)$ are enumerated and illustrated in **Figure B-3**. However, for further illustration, the two basic scenarios shown above are once again calculated in detail in this section beginning with the probability of the scenario $A_{m=AllBoth}$ where both FCs have both PLs available

$$P(A_{AllBoth}) = P(A_{1,1} \cap A_{1,2} \cap A_{2,1} \cap A_{2,2})$$

$$= P(A_{1,1})*P(A_{1,2})*P(A_{2,1})*P(A_{2,2}) = 0.60*0.70*0.65*0.75 = 20.5\%$$

On the other hand, the probability of having the scenario $A_{m=NoneBoth}$ where neither FC has neither PL available

$$P(A_{NoneBoth}) = P(\overline{A}_{1,1} \cap \overline{A}_{1,2} \cap \overline{A}_{2,1} \cap \overline{A}_{2,2})$$

$$= [1-P(A_{1,1})]*[1-P(A_{1,2})]*[1-P(A_{2,1})]*[1-P(A_{2,2})] = 0.40*0.30*0.35*0.25 = 1.1\%$$

Equation 19 essentially does the calculations shown above for all possible scenarios. The binary exponent $A_{m,i,j}$ simply mathematically defines for each scenario m whether there should be availability for PL j in FC i, in other words $P(A_{i,j})$, or whether there should be no availability for PL j in FC i, in other words $[1-P(A_{i,j})]$. Thus, the user would only have to provide as additional input the parameter $A_{m,i,j}$. Manual enumeration such as that shown in

Figure B-3 would therefore not be necessary since the mathematical program would automatically run through all scenarios using **Equation 19**.

Completely following the outlined procedure for the simplified case example would then lead to a P(S | Bk+Mus) equal to 14.4% (See **Figure B-3**). Finally, one would input this value along with the historical profile results for $P(OM_g = Bk+Mus)$ and the other historically-based constants into **Equations 14 & 15** to obtain the expected additional fulfillment costs resulting from network design inefficiencies. This simplified example can be further extended by adding more nodes and PLs.

4.2.2 Next Steps & Challenges

This project was broken down into two phases: a short- and long-term solution. The former involved the development of UK inventory allocation guidance limits for both peak 2006 and non-peak 2007. The latter involved the proposal of incremental improvements to the current inventory allocation optimization model to extend its capabilities beyond offering optimal inventory allocation guidance so that it may also flexibly offer optimal product line placement guidance across FCs in the UK as well as any other EU multi-node network. The objective was to have the product and inventory allocation optimization model not only attempt at minimizing transportation and labor costs, but also at minimizing safety stock and the likelihood of sourcing orders from multiple FCs.

Given the duration of the internship, Phase II of this project was not fully completed since validation and implementation did not take place. However, all findings and progress were submitted to the SPOT Team for further inspection and consideration. Next steps for this phase involve translating the mathematical expression for network induced split costs into the network model and further validating both components. In addition, the extended network model should be run side-by-side to the existing network model to gauge the impact of incorporating both components on inventory allocation guidance.

The last phase of this project was highly challenging from various standpoints. First of all, there is currently no literature related to the split costs component. Secondly, heuristic proposals require significant time dedicated to validation studies. Also, the network model demands a large scope. As a result, various system dynamics are inevitably missed in the simplification process. The size of the model demands tractability which limits the use of non-linear optimization in order to remain convenient. This imposes restrictions in the level of accuracy and flexibility behind the mathematical formulation of system effects. This was made apparent in the formulation of both the safety stock and split cost components.

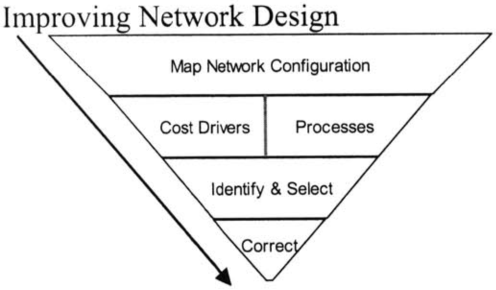
[THIS PAGE IS INTENTIONALLY LEFT BLANK]

PART V: FRAMEWORK FOR IMPROVING NETWORK DESIGN

5.1 Procedural Steps

This part of the thesis draws from the experience of both projects and provides the reader with a generalized framework for identifying, selecting and improving upon a firm's logistics network. The process is broken down into five steps. While the first three steps seem overly basic, they are nonetheless highly instrumental in leading one to discovering those opportunities with most potential and promise for returns. Following these steps can be just as valuable to a veteran employee as to an outsider. This is the case since those on the "inside" can many times loose perspective on the basics amidst the day-to-day operational challenges. Moreover, networks, processes and cost drivers are likely to change with time and these changes along with their indirect consequences can easily be overlooked and go unperceived. This is especially the case with fast growing companies. As such these first three elemental steps serve as the foundation for identifying and as a filter for selecting improvement opportunities.

Figure V-1: Framework for Improving Network Design



It is important to note that this framework is only applicable to improvements upon logistics networks at a tactical and operational level. Frameworks are also available for approaching strategic level decision-making (e.g. number, size and location of facilities) particularly those related to the first component mentioned in Part II. For a practical approach at developing a facilities strategy, the reader is referred to Beckman and Rosenfield (2007) Chapter 5: Facilities Strategy and Globalization.

5.1.1 Map Existing Network Configuration

A comprehensive understanding of the entire supply chain is required in order to even conceive of any improvement effort. The flow of inventory from source to user and how the firm fits into the value-add process has a fundamental influence on a firm's short- and long-term decisions. Naturally then, the relevance of any change effort is influenced by its effect on the position of the firm within the supply chain and its relationship with external stakeholders. For example, any change efforts affecting proximity to customers for a firm within an industry requiring short order fulfillment lead times is of high relevance.

Closer to the control of a company is its value chain. The method in which the firm procures raw materials, processes them and distributes its finish goods to fulfill customer demand is defined by all four components of its logistics network configuration:

1. The number, size and location of facilities
2. The chosen transportation mode
3. Inventory flow
4. Order flow

Identifying and mapping the first and last two components, respectively, provides a holistic view of the company's value chain. This, along with an understanding of a firm's position in the supply chain is a fundamental step in identifying areas for opportunity.

5.1.2 Identify Main Operational Cost Drivers

The next step involves identifying the firm's main operational cost drivers. Each firm has its unique cost structure and respective cost drivers. The relative proportion of fixed to variable costs is a strategic decision that does not necessarily remain constant throughout an industry. Examining a firm's cost structure, understanding its cost drivers and comparing this to other firms within the industry offers a glimpse into the firm's competitive strategy. More importantly, it offers a sense of what activity-based cost pools the company should focus on minimizing and where a marginal percent reduction equates to large gains. Several examples of cost factors include: Procurement, labor, inventory, transportation and property, plant and equipment costs.

For example, firms with high fixed costs generally focus on improving economies of scale through centralization efforts to use this as a pricing advantage. Firms with high variable costs on the other hand typically focus on lower volume, customized products. Firms on the latter end of the spectrum have a vested interest in improving upon processes that drive incremental costs. This is the case with Amazon, and understanding this quickly led me to understand the reasoning behind its process-driven culture. Furthermore, identifying the main drivers of fulfillment costs served as a sounding bound in filtering and finally selecting both projects.

5.1.3 Link Existing Processes to Cost Drivers

After identifying a firm's main operational cost drivers, it is important to determine how activities and processes relate to each cost driver. This provides direction as to which processes to further investigate with the aim of identifying areas for opportunity. Improvements upon those processes that directly affect main operational cost drivers would garner greater internal support and yield greater savings on a marginal basis. The objective of this step would then be to identify a filtered set of processes for the purpose of further detailed investigation.

5.1.4 Identification & Selection of Improvement Opportunities

The next step involves a detailed investigation of those processes previously selected. The best place to start is with the process owners themselves. They naturally have an intimate understanding of the inner workings. The intent is not to become an expert as well but to instead better understand how the process fits within the firm's value chain. In this sense, a valuable contribution one can provide is a critical outlook on how the inner workings are aligned with and supportive of other processes and systems. Communication and knowledge sharing between functional silos is vital considering the interdependent nature of logistics networks. For example, those associated with purchasing might have no idea how purchased inventory is assigned; similarly those associated with order assignment might not be aware of how inventory is assigned. A holistic perspective can lead to interesting findings.

Interviewing process owners and relevant stakeholders are the best source for leads into discovering improvement opportunities. In the end, they are the ones that deal with problem-solving on a day-to-day basis. Equally valuable from this process is in noting any inconsistencies in perspectives and recognizing internalized assumptions. Investigating the source of inconsistencies as well as testing the validity of assumptions can lead to interest findings. Section 5.2 offers further direction in identifying areas for improvement.

Once several potential opportunities have been identified, it is important to select a project(s) on the basis of several factors:

1. Supervisor and stakeholder feedback
2. Cost driver relation and ranking
3. Available resources
4. Organizational "pull"
5. Time constraint

Organizational "pull" is often underestimated. The culture of a company plays a big factor in this matter. Some firms such as Amazon are highly data-oriented. These firms typically place a lot of weight on validation results of a change initiative as opposed to the rank or position of the employee driving change. However, no matter how data-oriented a firm is, project alignment with stakeholder incentives is always a critical determinant in the viability of any project.

5.1.5 Corrective Action

The next steps following project selection involves finding and implementing a solution. This process might very well require revisiting previous steps. Whether this is required or not depends on the nature of the problem and solution. Heifetz (1994) categorizes those projects whose problem statement or solution is not clear as requiring adaptive work[32]. For example, the first project of this internship is considered a Type II since the problem was clear yet the solution required further learning. The second project on the other hand is considered a Type III since both the problem and solution required further learning. Recognizing whether a project requires adaptive work is helpful in setting realistic milestones and stakeholder expectations.

Throughout this step, it is important to keep stakeholders involved. To do so, it is critical to begin by identifying and relating project stakeholders. Recognizing expectations, support levels and conflicts of interest between stakeholders offers insight on the level of difficulty of developing and implementing a solution. This stakeholder mapping exercise is also helpful in developing strategies to gain stakeholder support.

In a similar manner, the firm's culture should also be kept in mind when developing and attempting to implement a solution. The strategy for obtaining buy-in for a change proposal should match a firm's culture. For example, a data-driven culture will require hard data as proof of benefit, whereas a hierarchical-driven culture will require support from a high-level executive(s).

---

[32] Heifetz, Ronald. Leadership Without Easy Answers (1998)

Whatever the solution may be, whether it is a more automated and streamlined process or an improved decision-support system, general stakeholder approval and effective ownership transfer is important for the long term viability of the project's solution. To ensure the latter, Standard Operating Procedures (SOP) should be developed and updated to ensure consistency in operation and enable future ownership transfers. The SOP should offer insight into the need and purpose behind the solution along with directions for operation.

5.2 Areas Rich in Opportunity

This section is intended to provide further guidance in the search for areas of improvement in a firm's logistics network.

5.2.1 Risk Management

All firms must execute decisions in the face of variability and forecast error. Incorporating variability into decision-making is often considered a tedious process whose benefits are often underestimated. Deterministic analysis and decision-making is the general rule in industry since it is a common misconception that useful prescriptive results can be obtained only if stochastic variation is ignored[33]. This misconception is not without good cause since it is only until recently that simulation techniques are widely accessible through commercially available software.

The lack of risk management techniques is therefore not surprisingly common in industry. As a result, opportunities to improve upon analysis and decision-making models through the inclusion of risk management techniques should not be uncommon.

5.2.2 Internal/External Change

In the day-to-day workload, the effect of internal or external change upon the validity of existing processes and systems can often go unperceived. The purpose and validity of assumptions behind existing processes, systems and decision-making tools are affected by both the internal and external environment. Any relevant internal (e.g. growth, restructuring, mergers) or external (e.g. regulation, technology, competition, currency fluctuation) change can invalidate the method in which effective business is ran.

The lack of response to change can lead to sub-optimal conditions. With any major change, a re-evaluation of the different components of the logistics network and its respective processes can lead to improvement opportunities.

5.2.3 IT & Decision-Support Systems

The value of sophisticated IT systems can often be underestimated by firms. The main reason behind this is the difficulty in placing a savings figure associated to the value of having real-time, accurate company-wide data accessible to the average employee. These systems require time and money to setup and maintain. However, these systems give employees at all levels the opportunity to understand the status of operations, track the effect of decisions and make data-driven decisions. Examples of such IT systems include ERPs, Wiki sites, company-wide accessible databases and internal websites. Similarly, there may

[33] Rardin, Ronald. Optimization in Operations Research. (1998) Chapter 1

also exist opportunities to automate non-value added activities and upgrade or simplify existing IT systems.

On a similar note, decision-support systems are another way of enabling management with data-driven semi- or fully automated guidance in everyday decision-making. These systems, such as the one described in the first project, can be built to collect data from sophisticated IT systems and analyze and interpret results for the purpose of guiding decision-making.

The lack of IT tools facilitating extensive data collection and the absence of practical decision-support systems enabling data-driven analysis and guidance can lead to sub-optimal decision-making and present an opportunity for improvement.

5.2.4 Modeling Tradeoff

Overestimating the added complexity to running a model by increasing its accuracy, scope or flexibility discourages any improvement efforts. Before discarding such opportunities, it is important to re-evaluate the modeling tradeoff considering the recent emergence of commercially available decision-support systems and Excel-integrated platforms that enable tailored in-house decision-support systems. Moreover, the increase in power and speed of today's computers allow the practical use of non-linear optimization techniques.

The lack of sufficient accuracy, scope or flexibility of existing models can lead to inadequate guidance. Improving upon any of the three aforementioned dimensions without jeopardizing practicality should be pursued as long as the marginal returns outweigh the costs of doing so.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

PART VI: CONCLUSION

This paper documents my efforts as an LFM 6-month intern to improve upon Amazon's EU network design to enable total fulfillment cost reduction. The actual steps followed in identifying and selecting projects is outlined; moreover, the paper also provides the reader with a generalization of these steps in order for it to be applicable to any other firm.

The majority of the paper however focuses on those two main projects that were ultimately selected. Mainly due to the length of the internship and the initial intent to carry projects all the way through implementation, these projects only focus on tactical and operational mid- to short-term horizon decisions involved in shaping a logistics network, particularly inventory and order flow. Thus, strategic and longer-horizon network defining aspects such as those involving the determination of the optimal number, size and location of warehouses for a network are not covered. Given the tactical and operational nature of this paper, both projects focus on tactical and operational cost and service-related fulfillment aspects including labor processing, transportation, inventory holding costs and service level requirements (i.e. "splits" reduction).

The intent for the majority of the paper is to offer the reader with a case approach at two types of projects. The first which was aimed at minimizing UK outbound transportation costs is intended to be representative of a project that involves improving upon internal processes and systems that are built around an existing and fixed network configuration. The second which was targeted at minimizing total fulfillment costs through the extension in scope and flexibility of the existing EU inventory allocation optimization model is intended to be representative of a project that involves the significant alteration of a main component of a network's design (i.e. inventory flow) through the adjustment of product and inventory distribution.

The main takeaway from this paper should however be that there is significant hidden potential and value in organizations to apply analytical techniques to improve upon their network designs. While these techniques may seem at first impractical, that is no longer the case with the availability of many useful decision-support systems that provide practical access to optimization and simulation techniques such as Crystal Ball, LogicNet, ProModel and PowerChain Inventory to name a few. For highly tailored and complex projects, there are tools such as Excel's VBA and Excel integrative software such as Crystal Ball that can be used by firms in order for them to be able to develop and implement their own tailored decision-support tools.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

PART VII: APPENDICES

7.1 Appendix A: UK Transportation Cost Minimization

Figure A-1: Standard Operating Procedure (SOP) for the Stochastic Model

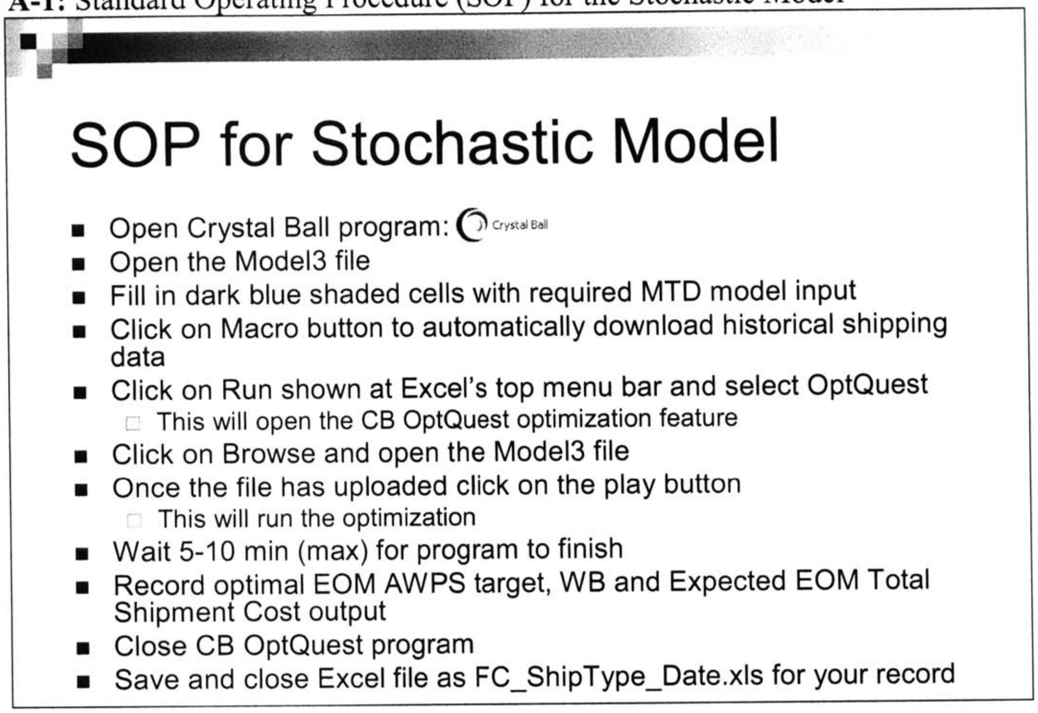


Figure A-2: Screenshot - Stochastic Model's Excel User Interface

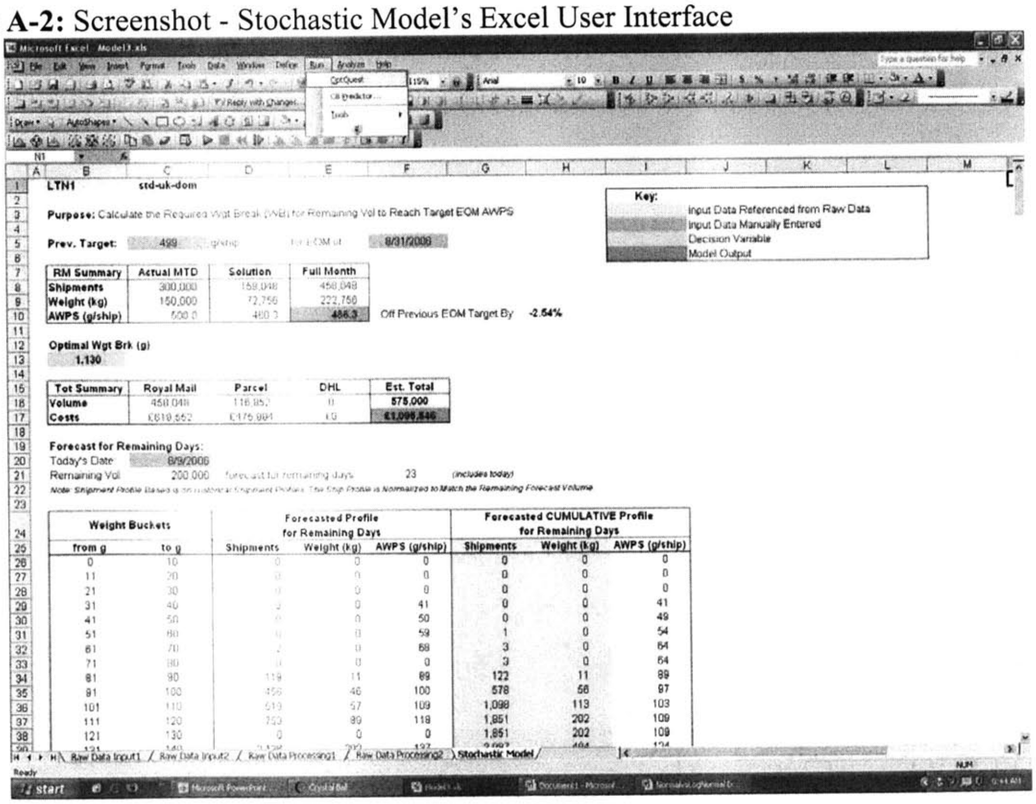


Figure A-3: Screenshot - Stochastic Model Crystal Ball Monte Carlo Simulation

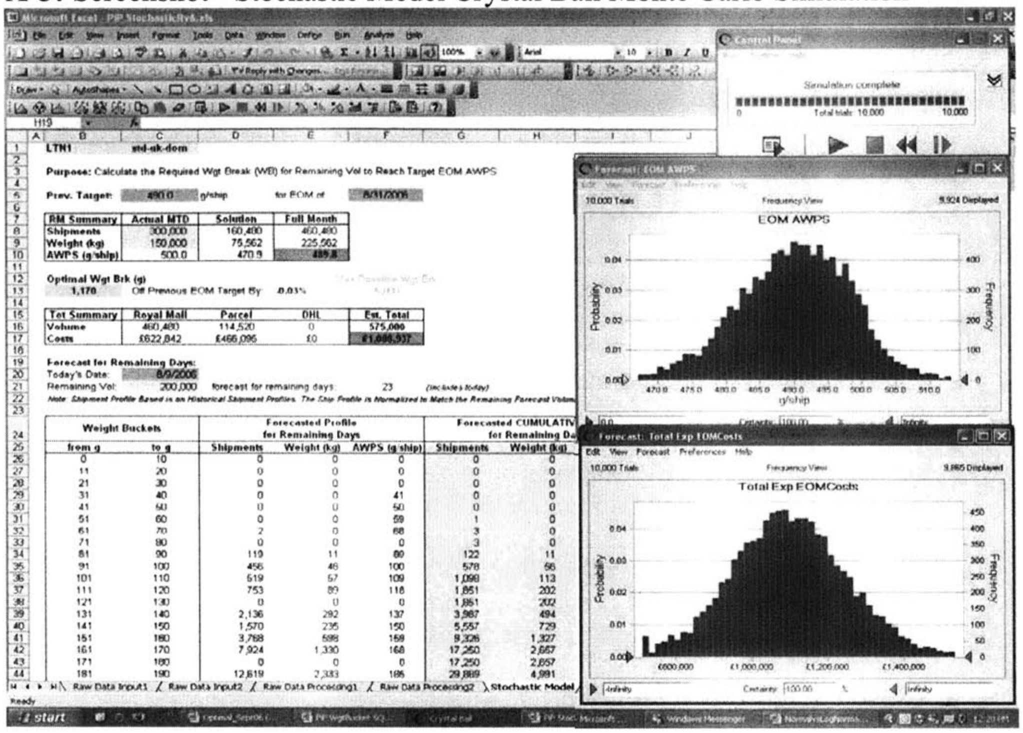
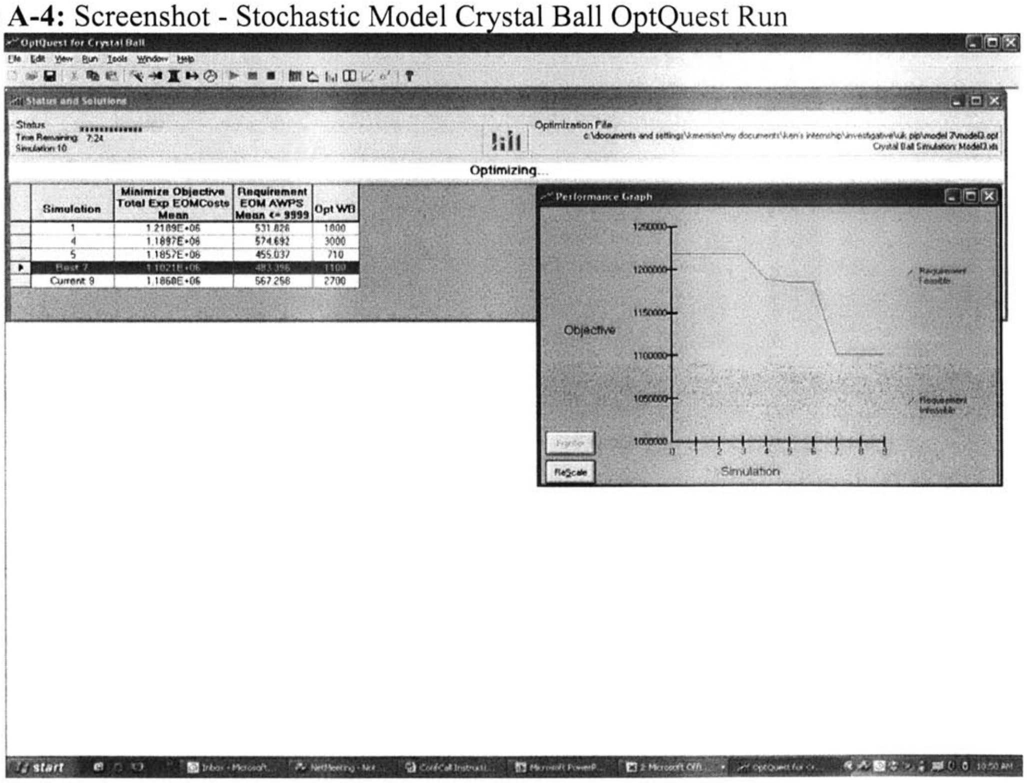


Figure A-4: Screenshot - Stochastic Model Crystal Ball OptQuest Run



7.2 Appendix B: EU Product & Inventory Allocation Optimization

Figure B-1: Periodic Review Inventory Policy (As applied to any line item)

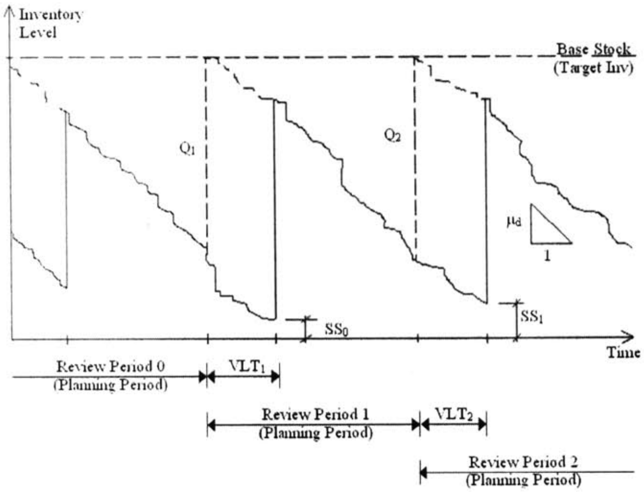


Figure B-2: Case Example for Safety Stock Regressed Alpha Coefficient

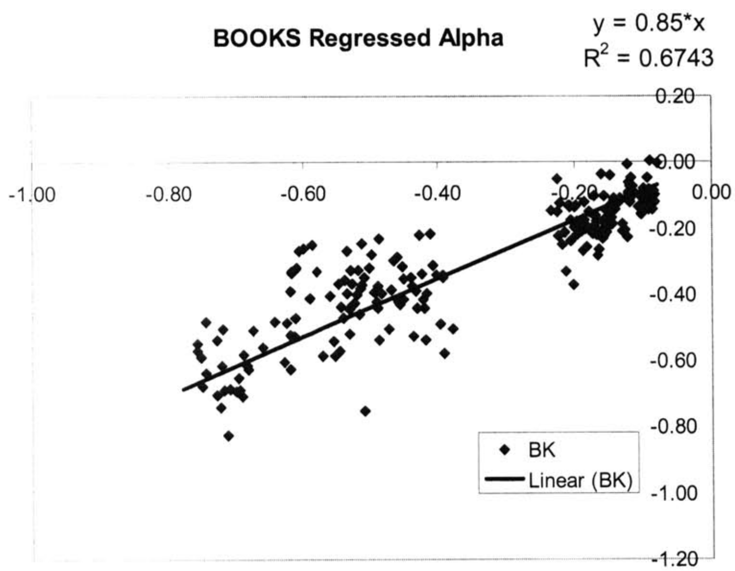


Figure B-3: Case Example for Application of Network Induced Splits Cost Heuristic

PL	FC1	FC2	Availability	P(Avail Event)	P(F1 | Ai)	P(F2 | Ai)	P(F12 | Ai)	P(F21 | Ai)
BK MUS	A11 A12		Only FC1	3.7%	100.0%	0.0%	0.0%	0.0%
BK MUS	A11 A12	A21 	Only FC1	6.8%	100.0%	0.0%	0.0%	0.0%
BK MUS	A11 A12	 A22	Only FC1	11.0%	100.0%	0.0%	0.0%	0.0%
BK MUS	A11 	A21 A22	Only FC2	8.8%	0.0%	100.0%	0.0%	0.0%
BK MUS	 A12	A21 A22	Only FC2	13.7%	0.0%	100.0%	0.0%	0.0%
BK MUS		A21 A22	Only FC2	5.9%	0.0%	100.0%	0.0%	0.0%
BK MUS	A11 A12	A21 A22	All Both	20.5%	39.1%	60.9%	0.0%	0.0%
BK MUS	A11 	A21 	Part Both	2.9%	30.0%	70.0%	0.0%	0.0%
BK MUS	 A12	 A22	Part Both	7.4%	60.0%	40.0%	0.0%	0.0%
BK MUS			None Both	1.1%	18.0%	28.0%	42.0%	12.0%
BK MUS	A11 		1/2 @ FC1	1.6%	30.0%	0.0%	70.0%	0.0%
BK MUS	 A12		1/2 @ FC1	2.5%	60.0%	0.0%	0.0%	40.0%
BK MUS		A21 	1/2 @ FC2	2.0%	0.0%	70.0%	0.0%	30.0%
BK MUS		 A22	1/2 @ FC2	3.2%	0.0%	40.0%	60.0%	0.0%
BK MUS	A11 	 A22	Part Both	4.7%	0.0%	0.0%	100.0%	0.0%
BK MUS	 A12	A21 	Part Both	4.6%	0.0%	0.0%	0.0%	100.0%

P(F1)	P(F2)	P(F12)	P(F21)
37.0%	48.6%	8.2%	6.2%

P(No Split | BK+Mus)	P(Split | BK+Mus)
85.6%	**14.4%**

PART VIII: REFERENCES

Allgor, Russell J, Maomao Chen. *Company Report: "Planning Labor in Amazon's Fulfillment Network"*. Strategic Planning and Optimization Team, Amazon.com. June 12, 2006.

Anupindi, Ravi, Sunil Chopra, Sudhakar D. Deshmukh, Jan A. Van Mieghem, and Eitan Zemel. Managing Business Process Flows. Second ed. Saddle River, NJ: Pearson Education, Inc., 2006.

Beckman, Sara and Donald B. Rosenfield. Chapter 5: Facilities Strategy & Globalization in Operations Leadership: Competing in the 21st Century. McGraw-Hill/Irwin, 2007.

Campbell, Bill et al. *Company Report: "2006 Inventory Imbalance-UK"*. Amazon.com. October 18, 2006.

Diebold, Francis X. Elements of Forecasting. Second ed. Cincinnati, Ohio: South-Western Thomson Learning, 2001.

Devore, Jay L. Probability and Statistics for Engineering and the Sciences. Fourth ed. Belmont, CA: Wadsworth Inc., 1995.

Fourer, Robert, David M. Gay, and Brian W. Kernighan. AMPL: A Modeling Language for Mathematical Programming. Second ed. Pacific Groove, CA: Thomson Brooks/Cole, 2003.

Heifetz, Ronald. Leadership without Easy Answers. Belknap Press (Harvard University Press), 1998.

Hum, Marcus. *Company Report: "UK Network Model Overview"*. Strategic Planning and Optimization Team, Amazon.com. July 28, 2005.

Kassman, Dean, and Russell Allgor. *Company Report: "Supply Chain Design, Management, and Optimization"*. Amazon.com. 16th European Symposium on Computer Aided Processes Engineering & 9th Symposium on Process Systems Engineering, Volume 21, 2006.

Klein, Janice A. True Change: How Outsiders on the Inside Get Things Done in Organizations. San Francisco, CA: Jossey-Bass, 2004.

Mendelson, Haim. Amazon.Com: Marching to Profitability (Case Study EC-25). Graduate School of Business Stanford University, 2006.

Rardin, Ronald L. Optimization in Operations Research. Upper Saddle River, NJ: Prentice Hall, Inc., 1998.

Rosenfield, Donald B. The Logistics Handbook: Section IV, Chapter 14: Demand Forecasting. New York, NY: The Free Press, 1994.

Shea, Peter et. al. *Company Report: "UK Domestic Ship Method Modeling".* Amazon EU Transportation Team, Amazon.com. May 4, 2006.

Simchi-Levi, David, Philip Kaminsky, and Edith Simchi-Levi. Designing & Managing the Supply Chain Concepts, Strategies, & Case Studies. Second ed. New York, NY: McGraw-Hill/Irwin, 2003.