November 2017

Theory and Practice of Supply Chain Synchronization

Michael Prokle

THEORY AND PRACTICE OF SUPPLY CHAIN SYNCHRONIZATION

A Dissertation Presented

by

MICHAEL PROKLE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2017

Mechanical & Industrial Engineering

THEORY AND PRACTICE OF SUPPLY CHAIN SYNCHRONIZATION

A Dissertation Presented

by

MICHAEL PROKLE

Approved as to style and content by:

Ana Muriel, Chair

Anna Nagurney, Member

Hari Balasubramanian, Member

Raj Subbu, Member

Sundar Krishnamurty, Department Chair
Mechanical & Industrial Engineering

DEDICATION

This dissertation is dedicated to my parents and sisters.

Thank you for your unconditional love and support.

I love you dearly.

ACKNOWLEDGMENTS

Many people have directly and indirectly contributed to this dissertation. Foremost, I would like to express my sincere gratitude to my academic advisor and mentor Professor Ana Muriel who guided me through my M.S. Theses and this PhD Dissertation. She taught me the necessary skills needed to approach and solve complex operations research and management science problems. I am thankful for her trust in me and the responsibility she gave to me during the course of many projects. I am particularly grateful for the opportunity to have experienced theoretical and applied research. I truly believe both are necessary to advance our society and create the world we all want. It is rewarding to see the significant practical impact that our research has for our industrial partners. I will always be indebted for her support and guidance.

Furthermore, I extend my warmest thanks to my dissertation committee for their insightful comments, time investment, and encouragement. John F. Smith Memorial Professor Anna Nagurney not only served in the dissertation committee, but also mentored me throughout my time at UMass and as part of her long time commitment as INFORMS student chapter faculty advisor. I want to express my sincere gratitude for her tremendous support and growth in knowledge she provided to me and my fellow students. Professor Hari Balasubramanian guided, taught, and mentored me since my very first day at UMass. His unique way of approaching problems and perspectives are truly inspiring and already impacted me as researcher and person. Dr. Raj Subbu provided the invaluable industry perspective, support, and mentoring. Having a practical impact is particularly important to me. My sincere appreciation for his expertise and encouragement.

I want to say thanks to Asli Ozen, Joanne Alvarez Oh, and Christian Huber. You are the greatest friends possible and were the best support imaginable going through this time. I met so many other friends and fellow students. I am very thankful for all your support.

I do want to highlight my gratitude to the fellowships I received. My time at UMass would not have been possible without the very generous support.

- Eugene M. Isenberg Scholar Award 2013 and 2014
- Cuneyt and Margo Oge fellowship
- Spyros and Yazmin Michail fellowship
- Mechanical and Industrial Engineering department fellowships

Multiple funding parties were involved in this research. Artaic along with Dr. Ted Acworth, Mike Trachtman, and Pratik Sheth supported the research presented in chapters 2 and 3. The work presented in chapters 4 and 5 resulted from a close collaboration with teams at Pratt & Whitney led by Dr. Raj Subbu and Dr. Robert Tomastik. I am thankful for the contributions, support, and comments. This dissertation was also funded in part by the National Science Foundation (NSF). This research would not have been possible otherwise. Thank you!

Furthermore, the large-scale industry projects included undergraduate and graduate students that supported parts of the project at different stages. In particular, I want to highlight Faried Beladi and Christopher Greene who both graduated with an M.S. from the University of Massachusetts Amherst, as well as Jessica Peterson, Kathryn Angevine, and James Heslin who participated during their undergraduate program. I am thankful for their contribution.

Chapter 5 includes collaborative work with the Case Western Reserve University, the University of Nebraska-Lincoln, and Pratt & Whitney. I want to thank Professor

Jennifer K. Ryan, Professor Robert X. Gao, and their PhD students for their support and insights. Dr. Robert Tomastik provided invaluable industry insights into the overall NSF study and the simulation framework presented in this chapter. Thank you for all the support!

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1301188, IIP-1439683, and industry members of the NSF I/UCRC Center for e-Design. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ABSTRACT

THEORY AND PRACTICE OF SUPPLY CHAIN SYNCHRONIZATION

SEPTEMBER 2017

MICHAEL PROKLE

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

DIPL.WI.-ING., KARLSRUHE INSTITUTE OF TECHNOLOGY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Ana Muriel

In this dissertation, we develop strategies to synchronize component procurement in assemble-to-order (ATO) production and overhaul operations. We focus on the high-tech and mass customization industries which are not only considered to be very important to create or keep U.S. manufacturing jobs, but also suffer most from component inventory burden.

In the second chapter, we address the deterministic joint replenishment inventory problem with batch size constraints (JRPB). We characterize system regeneration points, derive a closed-form expression of the average product inventory, and formulate the problem of finding the optimal joint reorder interval to minimize inventory and ordering costs per unit of time. Thereafter, we discuss exact solution approaches and the case of variable reorder intervals. Computational examples demonstrate the power of our methodology.

In the third chapter, we incorporate stochastic demand to the JRPB. We propose a joint part replenishment policy that balances inventory and ordering costs while providing a desired service level. A case study and guided computational experiments show the magnitudes of savings that are possible using our methodology.

In the fourth chapter, we show how lack of synchronization in assembly systems with long and highly variable component supply lead times can rapidly deteriorate system performance. We develop a full synchronization strategy through time buffering of component orders, which not only guarantees meeting planned production dates but also drastically reduces inventory holding costs. A case study has been carried out to prove the practical relevance, assess potential risks, and evaluate phased implementation policies.

The fifth chapter explores the use of condition information from a large number of distributed working units in the field to improve the management of the inventory of spare parts required to maintain those units. Synchronization is again paramount here since spare part inventory needs to adapt to the condition of the engine fleet. All needed parts must be available to complete the overhaul of a unit. We develop a complex simulation environment to assess the performance of different inventory policies and the value of health monitoring.

The sixth chapter concludes this dissertation and outlines future research plans as well as opportunities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND RESEARCH MOTIVATION

1.1 Supply Chain Management

Supply chain management (SCM) adopts fundamental manufacturing and logistics concepts and extends its scope aiming for an optimal integrated solution over various organizations and their individual characteristics. Since its beginnings in the early 1980s (see Oliver & Webber, 1982), SCM research became mainstream in the 1990s (see Mentzer et al., 2001), and continues to be a top priority for both industry and academia in our increasingly interconnected and fast paced world. Cooper, Lambert, and Pagh (1997) discuss SCM's early beginning and provide definitions and objectives distinguishing it from traditional logistics. The early years of SCM research brought up various definitions of the subject matter, each prioritizing certain SCM aspects differently. The study of Mentzer et al. (2001) provides a comprehensive discussion and review on the definition and aspects of supply chain (management) in the literature and concludes with the following definition of SCM:

> *"The systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole."*

Important concepts resulting from early SCM studies are now industry best practices, but are continuously revisited and adapted to evolving industry needs, new technological advancements, and newest research insights. Practical challenges remain, primarily caused by data management issues, supply chain network complexity, and the difficulty to manage the partnerships along the value and supply chain network (e.g., A.T. Kearney, 2008). Furthermore, recent years proved that the traditional supply chain (SC) design is changing. Decreasing costs has been the single objective for most SC partners in the past decades but has shifted to multiple objectives yielding to emerging external factors with additional requirements to the supply chain. For instance, it has been shown that it can be a competitive advantage to include the carbon footprint in SC decisions to meet customer sustainability expectations (e.g., Rao & Holt, 2005). Other aspects causing constant SCM adaptation and revisions are new legislator restrictions (e.g., traveling time restrictions) and the need for flexibility caused by global supply chain risks (e.g., disruptions), oil price volatility, and rising labor costs in emerging markets.

1.2 Research Motivation

Our study focuses on the manufacturing industry, which has traditionally been a vital part for the U.S. economy since the industrialization in the late 1800s. Recent numbers from 2013 underline that the manufacturing industry is indeed still important today. U.S. manufacturing supported 29.1 million jobs (directly and indirectly) and contributed with a gross output of $5.9 trillion or 35.4% to the GDP (Scott & Kimball, 2014). In the past decades, OEMs started to focus on their core competencies and began investing in production overseas as one promising way to lower costs and stay competitive in today's interconnected global market (e.g., Scott & Kimball, 2014; Gampenrieder, Damotte, Seel, Gates, & Mayor, 2015). Scott (2015) reports that the US economy lost about 6.6 million manufacturing jobs in the most recent 40 years (1973-2013). Thereof, almost half of the jobs (3.2 million) were lost in the past 12 years (2001-2013) (Scott & Kimball, 2014). This trend is in sharp contrast to recent studies which highlight the costs that companies experience when outsourcing overseas (e.g., supply risk, communication problems, loss of intellectual property, and reduction in innovation by separating R&D departments from manufacturing). These studies argue that shorter procurement lead times and the flexibility gained by producing locally may outweigh cost benefits overseas (e.g., De Treville & Trigeorgis, 2010; De Treville et al., 2014; Treville, Schürhoff, Trigeorgis, & Avanzi, 2014). Nevertheless, even companies that remain producing in the U.S. often have a significant portion of their suppliers overseas. This results in long extended supply chains that are vulnerable to disruption and, therefore, are variable in lead time. Main reasons for delays in long supply chains include weather conditions (see Boston Consulting Group, 2011), infrastructure and transportation modes (see Peck, 2005), congestion in foreign and domestic ports (see Boston Consulting Group, 2005), and politically imposed sanctions and export quotas (see Manuj & Mentzer, 2008). But there are other reasons why increased supply uncertainty can be observed. Today's single global market puts high financial pressure on companies to lower cost and strengthen their market competitiveness. For the supply chain, this means increasing efficiencies and automation, while reducing costly time and inventory buffers along the supply chain. The continuously tighter and stricter planning of processes and operations are more efficient but put additional stress on the system leaving little margins for unexpected and unplanned actions which then often result in process variations (e.g., lead times or quality) (Gampenrieder et al., 2015). Furthermore, this process automation in the past decades (e.g., SAP systems) has helped companies save money and better control their large-scale supply process. However, today's software platforms are still limited in capturing the entire dynamics of a system (e.g., dynamically changing lead times or inventory safety buffer) and, hence, fail under extreme situations that require continuous revision. The variability upstream the supply chain causes high inventory levels at the OEM when the majority of components of a complex product is waiting for a few very delayed components, and delays the product delivery to the customer significantly. In some instances, the OEM might be able to offer alternatives or alteration to the product to avoid a delay to the customers' agreed delivery date. Nevertheless, costs for OEMs are high when orders get lost or delayed and customers are displeased. Similar problems occur when demand downstream the supply chain is uncertain or highly variable which makes the planning process for the OEM extremely difficult. The most recent economic downturn of 2008/2009 meant for many industries a sudden decline in demand and demonstrated how volatile many companies are. Most of all, it uncovered companies' underlying hidden problems. This led many OEMs to reexamine and improve their existing processes, seek to find new innovative ways to stay more agile to change, and look for new methods to lower costs in order to stay competitive on the global market. Inventory has been shown to account for almost half of all logistics cost (Lancioni, 2000) and, hence, it is critical for companies (in particular with high cost components) to reduce inventory to a minimum. Most importantly, however, the essential foundation for value creation at the OEM (i.e., product manufacturing and assembly) is having the right amount of inventory available. This is the prerequisite for creating the subsequent financial stream from the customer at point of sale. Accounting for supply chain uncertainty in a cost effective manner is a key challenge. In particular, supply chain participants lack tools that let them easily incorporate historic and economic indicators to predict future system states and their associated risks. Furthermore, it is often unclear how to dynamically derive the optimal decisions (e.g., optimal ERP control values) resulting from these predictions. Traditionally, incoming and finished goods buffers are used to hedge against supply uncertainty upstream and final product demand variability downstream the SC. Further complexity is added when after-market sales downstream the SC constitute an additional important demand stream and source of uncertainty. The key challenge is to determine the right number and types of buffer that balance these three different sources of uncertainty. OEMs seek to find intelligent solutions to hedge against various uncertainties involved in the procurement and demand estimation process and, thereby, synchronize the inflow and outflow of goods. The objective is to create a flexible and agile supply chain that yields to company, industry, and economy specific (changing) conditions by effective prediction of the future and allows easy (pro-active) system adjustment (e.g., when facing an economic downturn). An optimal policy includes the cost-optimal supply order schedules, component inventory buffer levels, and finished good inventory levels that account for long transportation times, utilize economies of scale effects, and minimize system inventory while guaranteeing a desired customer demand service level. The optimal policy may be of dynamic nature and adapts to system conditions. This dissertation tries to address this objective in three very different settings, in industries that are inherently complex and challenging. First, we address supply uncertainty for assemblies with long and highly variable component lead times in the high-tech industry. Second, we spotlight the joint ordering of components under high fixed transportation costs in the mass customized manufacturing industry with an unwieldy product variety. Third, we present the spare part inventory management problem under advanced fleet sensor information in the aerospace industry under highly variable lead time and uncertain demand. This dissertation seeks to contribute to the field of supply chain management focusing on strategies that effectively synchronize OEMs' procurement and customer delivery in a stochastic system environment.

1.3 Dissertation Overview

This chapter introduces the reader to the general framework of OEM supply chain risk and provides motivation for our research on supply chain synchronization motivated by real-life problems. In Chapter 2, we study the deterministic joint replenishment problem under batch size restrictions. We illustrate the problem, review the literature, and present our modeling and solution approaches. In chapter 3 we extend the previous chapter by considering the case of stochastic demand. The objective is to find a joint part replenishment policy that balances inventory, and ordering costs while providing a desired service level. In a case study, our computational results show that a coordinated inventory ordering policy results in significantly lower costs by taking advantage of shipping economies of scale. In Chapter 4, we switch perspectives and introduce the component inventory management framework for assemblies in the high-tech industry focusing on the risk and uncertainty involved on the OEM supply side. Key challenges and industry specifics are highlighted. The framework is then illustrated and applied in a real-world case study of an aerospace assembly. In Chapter 5, we take on the challenge of spare part inventory management with advanced fleet condition information. We introduce the overall context, highlight key relevant literature, and present the framework building blocks needed to build a simulation that tests and optimizes condition-based inventory policies. In Chapter 6, we conclude this dissertation and highlight future research plans and directions.

CHAPTER 2

JOINT REPLENISHMENT PROBLEM WITH BATCH ORDERING: DETERMINISTIC CASE

2.1 Motivation

In many industry settings, supply orders for individual parts must be made in multiples of a batch size. This is required to drive efficiency in industries that produce in batches a high variety of small, relatively inexpensive products (e.g., screws, tile, or office supplies; see also the example of Spanish tile production in Bonavia and Marin (2006)). Product packaging may not easily adapt to variable order sizes and thus requires full container loads, palettes or boxes to be shipped (e.g., empty box space may result in quality problems). Likewise, it may be the incoming lot size that motivates companies to fully deplete and process the lot and thereby pass on batch restrictions down the supply chain (e.g., perishable items). In other cases, it might be resource allocation (e.g., full work shifts or process batches) that motivates the supplier to require customers to order in batches.

When considering one individual part under constant, deterministic demand, a simple EOQ solution rounded either up or down (whichever leads to lower cost) to a multiple of the batch size would provide the optimal inventory and ordering costs. A significant challenge arises, nonetheless, when multiple parts are jointly ordered to share a high common ordering cost from a supplier (e.g., overseas shipment in a container). In this situation, it is unclear in which time interval the joint orders should be placed, how this time interval should change over time, and which parts to include in each order to optimally balance supply ordering and inventory costs. This general problem is well-known as the deterministic joint replenishment problem (DJRP), and has been extensively studied in the literature, as we detail in the next section. The addition of batch restrictions, however, requires very different solution approaches, as the ZIO (Zero Inventory Ordering) property is no longer satisfied. Ordering points are not necessarily *regeneration points* where inventory is zero. As each part is ordered in batches of an exogenously given size, different quantities for each may be remaining at the time an order is placed. Furthermore, these quantities will change over time and require the number of batches ordered to change accordingly. This makes the formulation of the problem and computation of costs significantly harder. We must point out that we do not consider individual setup costs associated with the order of each part. The batch restrictions already force economies of scale in ordering and eliminate the need for unit-specific fixed costs.

Motivated by the joint replenishment problem that one of our industrial partners is facing, we research the DJRP with batch restrictions in this chapter. Although our examples reflect a particular industry, the work presented hereinafter is general and applicable to any industry setting with deterministic and constant demand, fixed joint setup costs, and batch restrictions. Porras and Dekker (2006) also consider supplier imposed order restrictions, as they study the DJRP under minimum order quantity (MOQ) constraints. Although their work provides excellent insights to the problem of batch restrictions, the results in unequal inventory and orders over ordering intervals, as mentioned above, require a different modeling approach. To the best of our knowledge, this problem has not been studied before and will extend existing literature.

The remainder of this chapter is organized as follows. We first review the available literature before we introduce the details of our models and analysis. In Section 2.3, we address the DJRP with constant demand and batch ordering. We characterize system *regeneration points*, derive a closed form expression of the average product inventory, and formulate the problem of finding the optimal constant joint reorder interval to minimize inventory and ordering costs per unit of time. We first consider demand to be in full units, and then generalize the analysis and formulation to the case of fractional demand. In Section 2.4, we model the DJRP problem with time-varying demand over a finite horizon. We formulate the problem as a mixed integer program (MIP) and explore new constraints to tighten the formulation. We show that the MIP can be applied to the case of constant demand over an infinite horizon, by considering a planning horizon equal to the *regeneration interval*. The resulting set of reorder intervals over a *regeneration interval* improves upon the constant reorder interval solution found in Section 3. We conclude the chapter with computational case study results and a discussion of future work.

2.2 Literature Review

The joint replenishment problem finds its application in the context of manufacturing and our context of procurement. In both settings, substantial setup costs can stimulate the consolidation of manufacturing operations or shipments of multiple parts to exploit economies of scale effects. In the procurement context, the JRP is practically observed when filling a full truck or container load with multiple parts. In manufacturing, furnace operations are one example that trigger joint production of different parts. Using auto glass as an example, the glass manufacturer jointly produces large batches of different types of parts of the same tint according to the OEM's fixed production schedule. When switching to a different tint, furnaces need to run empty for a significant amount of time first, in order to eliminate impurities. These time investments in switching production causes the high setup cost that motivates the JRP.

Extensive research has been done on the JRP in the past five decades and can generally be categorized into deterministic vs. stochastic demand models. Formulating the JRP necessitates the imposition of some structure to the replenishment schedule of the various parts. The strategies used have been classified into either direct or indirect part grouping (Van Eijs, Heuts, & Kleijnen, 1992). For the case of indirect grouping, the objective is to find (i) a fixed basic cycle interval T in which joint orders are placed, and (ii) the associated part-specific integer multipliers, k_j, indicating that the j^{th} part will be ordered every k_jT units of time. Hence, parts are indirectly grouped by k_j. In contrast, direct grouping divides parts into a predetermined number of groups, M, indexed by $j = 1, .., M$, where all parts in the group share a common reorder cycle T_j. Van Eijs et al. (1992) compare both strategies under various conditions and find indirect grouping strategies to slightly outperform direct grouping strategies.

Two JRP literature reviews are available. The early literature review of Goyal and Satir (1989) presents early studies from 1961 to 1988 whereas Khouja and Goyal (2008) follow up reviewing literature from 1989 to 2005. As shown in the reviews, early research tended to focus on finding quality solutions to the problem assuming deterministic demand (e.g., Silver, 1976; Federgruen & Zheng, 1992) whereas subsequent literature tackled the case of dynamic (e.g., Boctor, Laporte, & Renaud, 2004; Narayanan & Robinson, 2006; Robinson, Narayanan, & Gao, 2007; Kang, Lee, Wu, & Lee, 2016) or stochastic demand (e.g., Atkins & Iyogun, 1988; Viswanathan, 1997). Various authors adapted and extended the general JRP to fit special characteristics like quantity discounts (e.g., Cha & Moon, 2005; Duran & Perez, 2013), discrete time replenishment (e.g., Klein & Ventura, 1995), auto-correlated demand (e.g., Narayanan & Robinson, 2006), continuous unit cost change (e.g., Khouja, Park, & Saydam, 2005), storage and transportation capacities and budget constraints (e.g., Hoque, 2006), and pricing decisions with uncertain demand and yield (e.g., Li & Zheng, 2006).

A vast majority of previous research focused on determining (i) the optimal frequency of joint product orders, and (ii) the order frequency for each individual item, with the objective to minimize the total cost consisting of joint and individual replenishment cost as well as inventory holding cost. Since the DJRP is np-complete (see Joneja, 1990) developing faster algorithms as well as simple and effective heuristics has been an important aspect of the research in this field, in order to support quick and easy decision making in practice. For our deterministic demand case setting, a prominent example is the *power-of-two rule* that builds on the *economic order quantity* solution (see Jackson, Maxwell, & Muckstadt, 1985). The size of the reorder intervals is constrained to be a power of two, i.e., 2^bT for some integer b, of some basic period T. The optimal *power-of-two* solution can be easily calculated and is shown to yield holding and setup costs that are within six percent of those of the overall optimal solution.

In the setting that motivates this study, suppliers impose lot size restrictions that require the order to be in multiples of a batch size, rather than an individual product fixed cost that induces batch ordering. To our knowledge, this case has not been addressed in the literature.

The closest work to ours is Porras and Dekker (2006) who consider the case of minimum order quantity (MOQ) restrictions in the DJRP setting. For the MOQ case, customers face a similar dilemma as for batch ordering. Parts which have a demand that is smaller than the MOQ amount have to round their orders up accordingly. However, once the MOQ threshold is reached, an order for one part can be adjusted to (i) exactly reflect the demand until the next reorder interval so that the zero inventory ordering (ZIO) property is satisfied, and for the same reasoning also be (ii) synchronized with other parts for joint ordering. In the batch ordering case, however, only in rare circumstances will the inventory of different parts be fully depleted at the end of a replenishment cycle; thus, carrying inventory from one reorder interval to the next is inevitable. The ZIO property is not satisfied when batch constraints are present. As a result of positive ending inventories, the quantity ordered and inventory profile of a part varies over consecutive replenishment intervals. This greatly complicates the modeling of the problem. The following sections will further illustrate the challenges associated with synchronizing part orders under the batch restriction.

2.3 JRP with Constant Demand and Batch Ordering

Consider the joint replenishment of n products that share a joint fixed cost, A; that is, a cost A is incurred any time an order is placed regardless of the quantities ordered for each of the products. We measure time in periods, the smallest time unit over which ordering is feasible in a particular industry scenario; e.g., one day, in cases when multiple orders in a day would not be practical. Each product j, $j = 1, 2, , n$, has a constant demand rate of D_j units per period, and must be ordered in multiples of a batch (or box) size of B_j units. The batch requirement accounts for production, packaging, and handling economies of scale at the individual product level, and thus removes the need for additional fixed ordering costs associated with the individual products. Let the inventory holding cost for product j be h_j per unit per unit period. We seek to determine a constant, integer reorder interval T, so as to minimize the sum of long-run average ordering and inventory costs in the multi-product system over an infinite horizon. In the absence of individual fixed costs, all parts have the opportunity of being replenished at no additional cost every reorder interval, if needed. Each part, however, may be replenished in unequal frequencies over time, as dictated by the relative magnitude of its batch size versus demand. In our analysis, we will first assume that the demand per period is in full units and later consider the extension to the case of fractional demand per period.

2.3.1 Analysis

As a first step in developing a tractable formulation for the infinite horizon model, we identify the existence and timing of regeneration points. A *regeneration point* is a period where the ending inventory of all parts is 0, and thus the stationary system reverts back to the initial conditions. We denote the time interval between two consecutive *regeneration points* as a *regeneration interval.* The system behaves identically over each *regeneration interval.* Consequently, the average long-run ordering and inventory costs over the infinite horizon are equivalent to the average ordering and inventory costs over a *regeneration interval.*

In *Lemma 1*, we characterize the *regeneration points* of a single part. *Corollary 1* then extends that result to system *regeneration points*, when all parts regenerate.

Lemma 1. *Given a fixed system reorder interval T, the regeneration interval for a part with batch size B and demand D is $R = \frac{B}{g.c.d.(B,TD)}$ reorder intervals (that is, RT periods), where g.c.d. is the greatest common divisor.*

Proof. The system will regenerate when the inventory is 0 at the end of a reorder interval. For this to occur, the demand over the number of reorder intervals that make up the *regeneration interval* must be a multiple of the batch size B.

First observe that the system will naturally always regenerate after B reorder intervals. This is because the total number of boxes ordered and fully depleted over BT periods is an integer, TD, and thus no inventory will be left over at the end of BT periods.

However, the inventory will first reach 0 after $R = \frac{B}{g.c.d.(B,TD)}$ reorder intervals because demand over R reorder intervals is $RTD = \frac{BTD}{g.c.d.(B,TD)}$, a multiple of B.

No earlier *regeneration points* are possible. If an integer number x of reorder intervals is a *regeneration point* then $xTD = yB$, for some integer y. But the lowest x that makes $y = \frac{xTD}{B}$ integer is $R = \frac{B}{g.c.d.(B,TD)}$ by definition of the *g.c.d.*.

□

Corollary 1. *Given a fixed system reorder interval T, the regeneration interval for a multi-product system with individual batch sizes B_j and demand D_j for each product $j, j = 1, 2, .., n,$ is $R = l.c.m \left[\frac{B_j}{g.c.d.(B_j, TD_j)}, j = 1, .., n \right]$ reorder intervals, where l.c.m. is the least common multiple and g.c.d. is the greatest common divisor.*

2.3.2 Closed Form Expression of Average Product Inventory

Throughout this section, we consider a given reorder interval, T, and a single part. The ordering cost per period is simply $\frac{A}{T}$. The challenge lies in determining the long-run average inventory cost per period. For this purpose, we focus on characterizing the inventory over a *regeneration interval*.

Lemma 1 characterizes the *regeneration interval* for a part with batch size B and demand D as a number $R = \frac{B}{g.c.d.(B,TD)}$ of reorder intervals. *Theorem 1* uses this *regeneration interval* to derive a closed form expression for the average period inventory.

Theorem 1. *The average inventory in a system with batch size B and demand TD is:*

$$\frac{1}{2}(TD + B - g.c.d.(TD, B)).$$

Proof. The average inventory is calculated as the average of the inventory carried in the R identical reorder intervals between *regeneration points*. For each reorder interval, the average inventory is calculated as the sum of initial plus ending inventory divided by 2. To compute the overall average, we calculate and add the sum of initial inventory over the R reorder intervals to the sum of ending inventory over the R reorder intervals, and divide by $2R$. The four steps below highlight the details.

Step 1: Characterize the ending inventory dynamics over reorder intervals:

Let $N = \lceil \frac{TD}{B} \rceil$ be the number of boxes ordered, and $L = NB - TD$ denote the number of units leftover at the end of the first reorder interval. Observe that the inventory at the end of consecutive reorder intervals grows at a rate of L until the next reorder interval $x \geq 1$ such that $xL \geq TD - (N-1)B$. At the following reorder interval one fewer box (which would translate into an empty order when $B \geq TD$ and $xL \geq TD$) will be ordered and the inventory left at the end of that interval will be $(x+1)L - B$. The inventory at the end of the subsequent reorder intervals will then increase again at the rate of L, until the first interval y such that $(xL - B) + yL \geq TD - (N-1)B$ at which point again one fewer box is ordered. The process continues until the *regeneration point*. The inventory left over at the end of the i^{th} interval can be written as $0 < iL - z_i B < B$ for some unique integer z_i.

Step 2: Show that the sum of ending period inventory over the R reorder intervals within a regeneration interval can be written as:

$$\frac{1}{2}B(R-1)$$

Step 1 shows that inventory at the end of a period i, $i = 1, 2, .., B-1$ is $0 < iL - z_i B < B$ for some unique integer z_i. The inventory of period $R - i$ can thus be written as $0 < (R-i)L - z_{B-i}B < B$.

Adding up the two, we have $0 < RL - (z_i + z_{B-i}B) < 2B$. Since both terms in the subtractions are multiples of B, for the inequalities to hold we must have that $RL - (z_i + z_{B-i})B = B$.

That is, for any interval i, $1 \leq i < \frac{R}{2}$, we have that the sum of the ending inventory of interval i and its complement $(R - i)$ is equal to B. Observe that the ending inventory in interval $i = \frac{R}{2}$ given an even batch size quantity B must by the same argument be equal to $\frac{B}{2}$.

Thus, we can distinguish between two cases: If R is odd, the sum over all interval pairs is $\frac{B(R-1)}{2}$, since there are $\frac{(R-1)}{2}$ interval pairs. If R is even, the sum over all interval pairs is $\frac{B(R-2)}{2} + \frac{B}{2}$, since there are $\frac{(R-2)}{2}$ pairs with inventory B plus one interval of $\frac{B}{2}$.

Step 3: Show that the sum of initial inventory over the R reorder intervals within a regeneration interval can be written as:

$$RTD + \frac{1}{2}B(R-1).$$

Sum of initial inventory is calculated as the sum of all the orders RTD plus the sum of the ending inventory over all intervals $\left(\frac{1}{2}B(R-1)\right)$. This is true because each period starts from an inventory position equal to the previous interval's ending inventory plus the order received. The ending inventory in the last interval is 0, equal to the initial inventory position in the first interval, so the sum of initial inventory positions before orders are received in intervals 1 through R is equal to the sum of ending inventory positions in intervals 1 through R.

Step 4: Calculate average period inventory:

The sum of the average inventory over all reorder intervals within the *regeneration interval* is thus

$$\frac{1}{2}(RTD + B(R-1)).$$

Dividing by $R = \frac{B}{g.c.d.(B,TD)}$ yields the average inventory of

$$\frac{1}{2}\left(TD + B - \frac{B}{R}\right) = \frac{1}{2}(TD + B - g.c.d.(TD, B)).$$

$\square$

Definition 1. *Given a single-product system with reorder interval T, batch size B, and demand D, we define a new system with batch size $B' = R = \frac{b}{g.c.d.(B,TD)}$ and demand $TD' = \frac{TD}{g.c.d.(B,TD)}$, as its corresponding normalized system, where regeneration points occur exactly after the number of reorder intervals equals the batch size, and no earlier.*

Corollary 2. *The average inventory of a product with batch size B, demand D, and reorder interval T is equal to $g.c.d.(B,TD)$ times the average inventory of a normalized system with batch size $B' = \frac{B}{g.c.d.(B,TD)}$ and demand $TD' = \frac{TD}{g.c.d.(B,TD)}$.*

2.3.3 Illustrative Examples

For illustration, consider the simple example of box size of 9 units and demand over the reorder interval of 5 units, presented in Table 2.1.

Reorder Interval	# Boxes Ordered	Beginning Inventory	Ending Inventory
0	-	-	0
1	1	9	4
2	1	13	8
3	0	8	3
4	1	12	7
5	0	7	2
6	1	11	6
7	0	6	1
8	1	10	5
9	0	5	0

Table 2.1. Example 1: $B = 9, D = 5$

The second column shows the number of boxes ordered in each interval. The third column describes the beginning inventory in the interval, which includes the units ordered in the period plus those available from the previous period. The last column states the inventory left over at the end of the reorder interval and carried over to the next. This example shows that ordering is not necessary in each "reorder" interval and inventory is carried over to the next interval. Only in interval 9, which is the last "reorder" interval of the *regeneration interval*, the demand of 5 exactly matches the inventory yielding to zero inventory at the end of the interval. Following our analysis on the previous pages, the number of reorder intervals in the *regeneration interval* matches the box size of 9 due to $g.c.d.(9, 5) = 1$.

Reorder Interval	# Boxes Ordered	Beginning Inventory	Ending Inventory
0	-	-	0
1	1	18	8
2	1	26	16
3	0	16	6
4	1	24	14
5	0	14	4
6	1	22	12
7	0	12	2
8	1	20	10
9	0	10	0

Table 2.2. Example 2: $B = 18, D = 10$

Table 2.2 illustrates the case of box size = 18, demand = 10 and $g.c.d.(18, 10) = 2$. This example demonstrates that we can see the exact same ordering pattern with doubled beginning and ending inventory and a *regeneration period* of only $\frac{18}{2} = 9$. This is also shown in Figure 2.1.

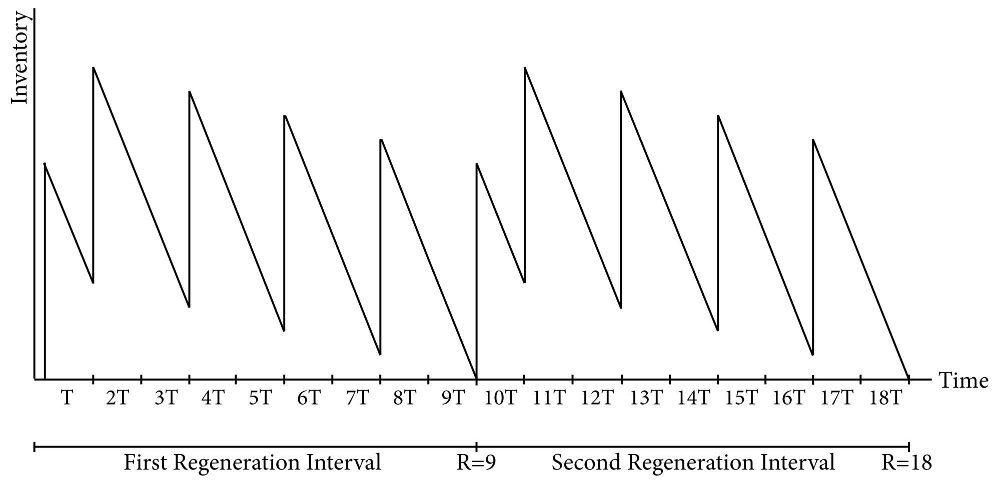

Figure 2.1. Graph for $B = 18, D = 10$

2.3.4 Problem Formulation

Find the constant reorder interval T that minimizes total ordering and inventory cost per unit of time:

$$Min_T \ \frac{A}{T} + \frac{h_j}{2} \sum_{j=1}^{n} \Big(TD_j + B_j - g.c.d.(TD_j, B_j) \Big)$$

We assume here that every reorder interval sees a positive number of boxes ordered. There are contrived cases, with low demand for all parts and high inventory costs relative to fixed costs, where some reorder intervals may have a zero order for all parts. Our approach will be overestimating the fixed costs over a *regeneration interval* then. Such cases, however, are rare if many different parts need to be jointly ordered and demand varies for each part type, as in the industry example that motivated this chapter.

2.3.5 Extension to Fractional Demand

In many cases the demand per period, D will be a fractional number. Let t be any integer such that tD is integer. Under any reorder interval of T periods, the system will always regenerate after tB reorder intervals, as demand over that time frame is an integer number of boxes.

Lemma 2. *Given a reorder interval length of T, the inventory system with batch size B and fractional demand D per period regenerates after $R = \frac{tB}{g.c.d.(tB, tTD)}$ reorder intervals, where t is any number of reorder intervals such that demand tTD is integer and g.c.d. is the greatest common divisor.*

Proof. The proof is analogous to that of *Lemma 1.*

□

Theorem 2. *The average inventory in a general system with reorder interval T, batch size B and fractional demand D is:*

$$\frac{1}{2}\left(TD + B - \frac{g.c.d.(tTD, tB)}{t}\right)$$

Proof. The proof is identical to that of *Theorem 1.* The only difference is that the number of reorder intervals in a *regeneration period* is $R = \frac{tB}{g.c.d.(tB, tTD)}$.

□

Once we have determined the average inventory of each part associated with any given reorder interval, we can formulate the objective function just as before. Observe that for each fractional part j we would need a multiplier t_j such that $t_j D$ is integer. In practice, if demands are given as fractions with up to x decimal points, then we can simply consider $t = 10^x$ for all parts, as this multiplier will make all demands integer.

We can thus write the problem as follows: Find the constant reorder interval T that minimizes total ordering and inventory cost per unit of time:

$$Min_T \ \frac{A}{T} + \frac{h}{2}\sum_{j=1}^{n}\left(TD_j + B_j - \frac{g.c.d.(t_j TD_j, t_j B_j)}{t_j}\right)$$

2.3.6 Solution Approach

We find the optimal constant integer reorder interval T through an exhaustive search over a bounded interval. This is a very fast algorithm since each iteration requires evaluating a very simple closed form expression. The interval bounds can be found in a similar fashion as in Porras and Dekker (2006), using the fact that the classical EOQ cost function is a lower bound on the actual cost curve under batch size restrictions. Let $C(T)$ be the actual cost per unit of time associated with reorder interval T, accounting for batch restrictions, and $C^{EOQ}(T)$ the classic EOQ cost function, without batch restrictions. Note that $C^{EOQ}(T) \leq C(T)$ for all T. The actual cost $C(T^{EOQ})$ associated with the EOQ optimal reorder interval T^{EOQ} (rounded to comply with the integrality requirement) is a feasible solution and thus an upper bound on the cost of the optimal solution. We can then calculate reorder intervals T^{LB} and T^{UB} such that the $C^{EOQ}(T^{LB}) = C(T^{EOQ}) = C^{EOQ}(T^{UB})$.

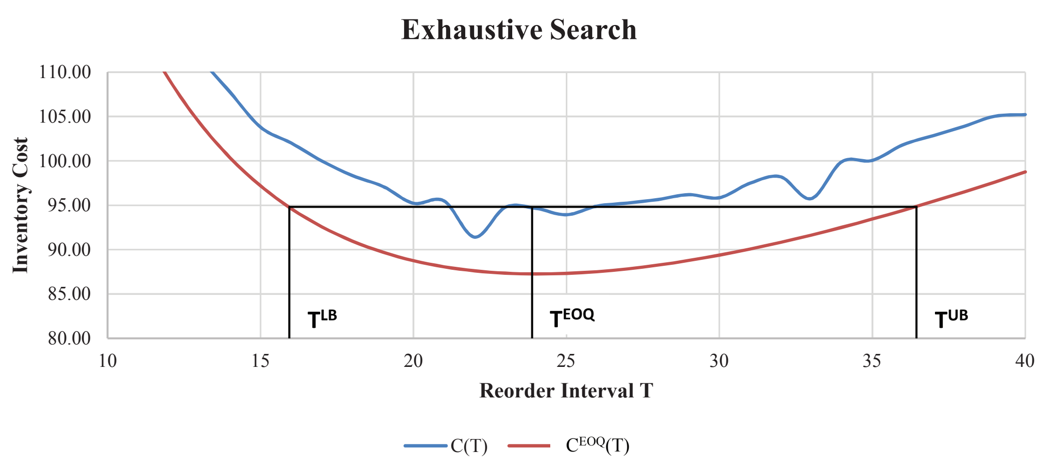

Figure 2.2. Exhaustive search bounded interval

2.4 JRP with Time Varying Demand and Batch Ordering

In this section, we consider that demand for each period is still known but may vary from period to period. The objective is to find the ordering periods to minimize joint ordering and inventory costs over a finite planning horizon. The problem can be formulated as a mixed integer program, as we show in the next section.

2.4.1 Mixed Integer Linear Programming Formulation

Notation	Definition
P	Number of periods in the planning horizon
Y_i	1, if an order is placed in period i, 0 otherwise, $i = 1, 2, ..., P$
D_{ij}	Demand for part j in period i
I_{ij}	Inventory of part j at the end of period i
N_{ij}	Number of boxes of part j ordered in the i^{th} period, an integer

Table 2.3. Mixed Integer Linear Programming notation

$$Min \sum_{i=1}^{P} \left(AY_i + \sum_{j=1}^{n} h_{ij} + I_{ij} \right)$$

subject to

$$\begin{aligned}
I_{0j} &= 0 && \vee j = 1, .., n \\
I_{ij} &= B_j N_{ij} + I_{(i-1)j} - D_{ij} && \vee i = 1, .., P,\ j = 1, .., n \\
N_{ij} &\leq M_{ij}^{UB} Y_i && \vee i = 1, .., P,\ j = 1, .., n \\
Y_i &\in \{0, 1\} && \vee i = 1, .., P \\
N_{ij}, I_{ij} &\geq 0 && \vee i = 1, .., P,\ j = 1, .., n \\
N_{ij} &\ integer && \vee i = 1, .., P,\ j = 1, .., n
\end{aligned}$$

Rather than using an arbitrarily large M value, we determine a tight upper bound on the number of boxes to cover demand over the remaining of the planning horizon; that is $M_{ij}^{UB} = \left\lceil \frac{D_j(i,P)}{B_j} \right\rceil$ where $D_j(i, P)$ is the demand over periods i through P.

2.4.2 Application to the Constant Demand Case

Section 2.3 focuses on finding the optimal fixed reorder interval, T, in the case of constant demand over an infinity horizon. Although the reorder interval and demand are constant, each ordering point over the *regeneration interval* may see different initial inventories, order quantities and ending inventories. There may well be ordering points where there is enough inventory of all parts to last for a few extra periods. Consequently, varying the size of the reorder interval over the *regeneration interval* may lower inventory costs while keeping ordering costs unchanged. Fortunately, the mixed integer program formulation in the previous section can be applied to find the optimal set of ordering intervals over the *regeneration interval.* The planning horizon in the MIP is the *regeneration interval*, i.e., $P = RT$, and demand each period is constant. Finally, observe that we could also use the MIP formulation to find the optimal constant reorder interval over a certain *regeneration period* or planning horizon by requiring $Y_i \leq Y_{ui}$ for all $u \leq \frac{R}{i}$. This expression assumes that the first period, where an order is always placed is $i = 0$. That way an order in period $i = 1$, given by $Y_i = 1$, means that the constant reorder interval is 1 and we need to order every period; an order in period i=2 means that we order every 2 periods, etc.

2.5 Computational Results

In the following, we present examples of the computational results. We calculate the optimal reorder interval T^* using (i) the traditional Economic Order Quantity Model, as well as (ii) the model presented in this chapter, and compare the resulting intervals and their performance.

We assume fixed costs of $A = \$1000$ and holding costs h of \$0.05 per lbs and week. We calculate the EOQ solution by rounding $T^{EOQ} = \sqrt{\frac{2AD}{Dh}}$ to the best-performing nearest integer.

Example 1

Here, we consider a problem with $B = 150$ and two parts with demand $D = 7$ and $D = 15$, respectively. The EOQ solution, rounded to best-performing nearest integer is $T^{EOQ} = 43$ with average period ordering and inventory costs of \$54.01. Using our iterative approach, we calculate ordering and inventory costs per period as T increases from $T = 1, .., T = 89$, to find $T^* = 50$ yielding average interval costs of \$50.00 (see Figure 2.3) and savings of 7.42%.

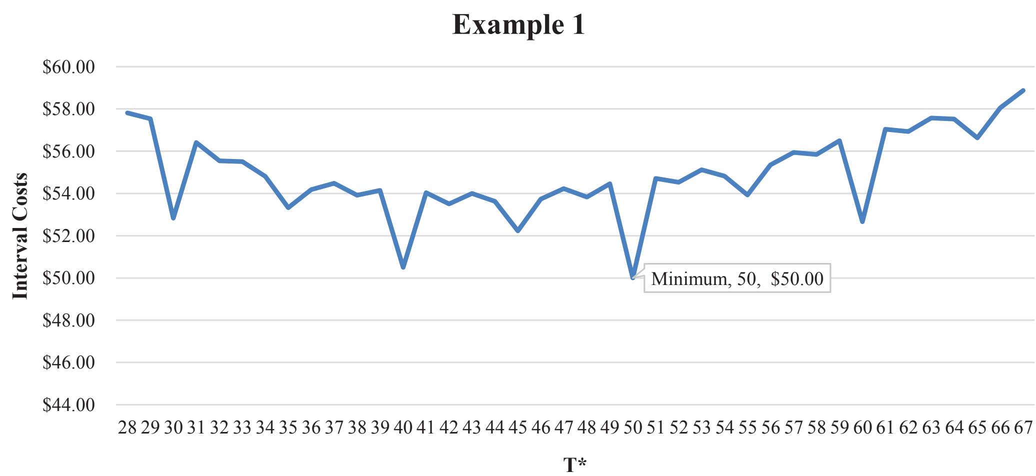

Figure 2.3. Example with D=7 and D=15

Example 2

Consider now $B = 19$ and a product consisting of 8 parts with demand $D = 7$, $D = 15$, $D = 3$, $D = 51$, $D = 18$, $D = 20$, $D = 13$, and $D = 100$, respectively. This yields $T^{EOQ} = T^* = 13$ for both the EOQ and our iterative approach with period costs of \$154.30 (see Figure 2.4).

In this example, we demonstrate the value of varying the length of the reorder intervals over the *regeneration interval*. We first use Excel's solver and the evolutionary solving method (Convergence: 0.001, Mutation Rate: 0.075, Population Size:

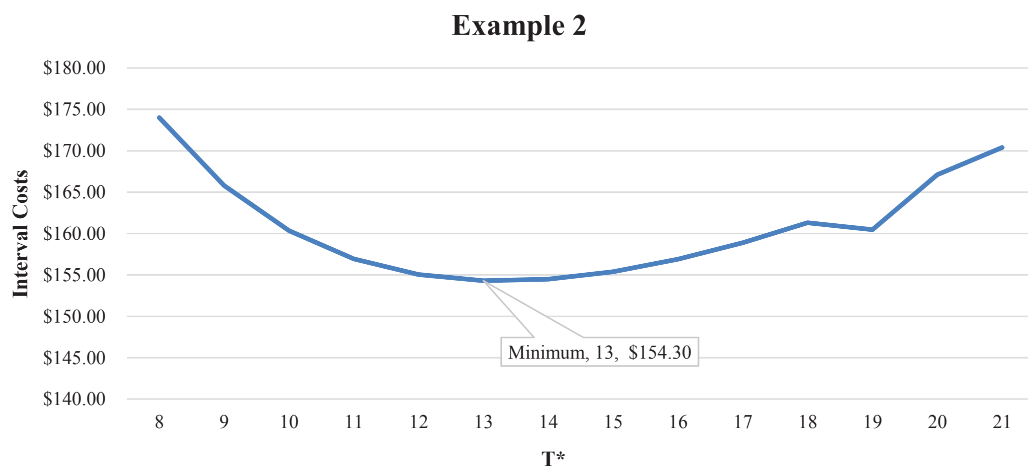

Figure 2.4. Example with 8 parts

100, and maximum time without improvement of 30s) for 10 trials and find a best improved solution yielding an interval cost of \$153.27 or 0.67%. The resulting interval length for each order interval T is presented in 2.4.

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	..	19
EOQ	13	13	13	13	13	13	13	13	13	13	13	13	13	13	..	13
Local	13	13	12	13	13	12	13	13	12	15	14	14	14	13	..	12

Table 2.4. Local shifting of order intervals

Solving the mixed integer linear programming formulation for this instance results in a similar (slightly improved) cost.

2.6 Conclusion

This chapter studies the deterministic joint replenishment problem under batch constraints. We characterize and proof the existence of *regeneration points*. This allows us to formulate the infinite horizon problem and derive a closed-form expression for the long-run average ordering and inventory costs under constant demand and a given reorder interval. A simple search algorithm can then be used to determine the optimal joint replenishment interval. Bounds on the search space can be derived from the EOQ solution to the problem ignoring batch restrictions.

The finite-horizon dynamic version of the problem is formulated as a mixed integer program. Using the MIP over a *regeneration interval* with constant demand, we show that a varying reorder interval attains better performance than the optimal constant reorder interval. A practical case study shows the savings associated with this practice.

A comprehensive computational study is needed to identify the settings in which the exact iterative approach and the varying intervals are most beneficial relative to a naïve EOQ solution. Further experiments are also necessary to tighten the MIP formulation and quantify scenarios where varying the reorder intervals yields highest savings. This can uncover structural properties of the optimal solution that can be used to refine the formulations. Finally, the effect of potential empty orders can be studied through computational experiments contrasting the solution to the MIP problem with that of the constant reorder interval search algorithm. Observe that the MIP will not generate any empty orders, while the constant reorder interval may in particular cases.

CHAPTER 3

JOINT REPLENISHMENT PROBLEM WITH BATCH ORDERING: STOCHASTIC CASE

3.1 Introduction

This chapter considers the joint replenishment problem with batch ordering described in the previous chapter, but incorporates an additional layer of complexity by considering demand to be stochastic. This reflects the industry setting that motivated our research and is the focus of this chapter. The challenge is to devise joint ordering policies to minimize inventory and ordering costs while maintaining a desired service level. We refer to this problem as the Stochastic Joint Replenishment Problem with Batch Ordering ($SJRP_B$). The practical setting involves the production of highly customized designer products that require a large number of low-cost parts sourced from overseas. The variety of colors, finishes, and materials customers can choose from makes for a high number of different parts. All parts of the same material are sourced from the same supplier and location. Consequently, joint ordering costs arise from the consolidation of orders into containers for ocean shipping. More specifically, parts are sourced from overseas via two channels: (1) air freight, with high variable costs and relatively quick lead times; and (2) ocean shipment, with steep fixed costs shared by all parts consolidated at the same port, and long lead times, but very low variable costs. The latter is the preferred shipping method given its low overall cost, whereas air freight offers an option of last resort to avoid stock-outs. Our objective is to develop a joint reordering strategy for ocean shipping with batch ordering requirements, using the additional cost associated with air transportation as a penalty cost for stockouts that allows us to calculate an appropriate service level to aim for. As in the previous chapter, the batch order restriction imposes economies of scale in production and transportation, and eliminates the need for part-specific fixed costs.

Although our examples and case study reflect this particular industry, the work presented hereinafter is general and applicable to any industry setting with fixed joint setup costs, batch restrictions, and variable demand. To the best of our knowledge, this problem has not been studied before and will extend existing literature.

3.2 Literature Review

The majority of the literature relevant for this study has been presented in the previous chapter. The two practical challenges observed under deterministic demand still hold. First, each part must be ordered in multiples of a batch size and, second, parts are jointly ordered to share a high common ordering cost from a supplier. When demand is highly variable and uncertain, additional inventory is required to maintain a desired service level and should result in a shorter optimal reorder interval, as in the stochastic JRP without batch ordering studied in Eynan and Kropp (1998). We refer to this problem as the Stochastic Joint Replenishment Problem with Batch Ordering ($SJRP_B$).

For a single part, the periodic stochastic inventory management problem with batch ordering has received significant attention. The seminal work of Arthur F. Veinott (1965) shows that an (R,Q) policy is optimal. A stream of recent literature extends it to multi-echelon serial and assembly systems (Chen, 2000; Chao & Zhou, 2009). Recent literature has focused on Q(s,S) policies, can-order policies, and correlated demands (Melchiors, 2002; Nielsen & Larsen, 2005; Larsen, 2009; Feng, Wu, Muthuraman, & Deshpande, 2015).

To our knowledge, the only previous work that considers the JRP with batch ordering (or JRP_B) is the work presented in chapter 2. There we derive a closed-form expression on the average inventory in the system and provide an algorithm to calculate the optimal reorder interval T^*. The work in this chapter builds on the results and insights derived there to account for random demand.

3.3 Model

As the previous chapter, we consider the joint replenishment of n parts that share a joint fixed cost, A; that is, a cost A is incurred any time an order is placed regardless of the mix of parts and quantities ordered for each of the parts. We measure time in periods, the smallest time unit over which ordering is feasible in a particular industry scenario; e.g., one day, in cases when multiple orders in a day would not be practical. Each part $j, j = 1, 2, , n$, has a random demand with a mean of D_j and a standard deviation of s_j units per period, independent and identically distributed over time. Orders arrive after a lead time of L periods. The quantity ordered for each part must be a multiple of a batch (or box) size of B_j units. The batch requirement accounts for production, packaging, and handling economies of scale at the individual part level, and thus removes the need for additional fixed ordering costs associated with the individual parts. Let the inventory holding cost for part j be h_j per unit per unit of time. We seek to determine a constant reorder interval T, so as to minimize the sum of long-run average ordering and inventory costs in the multipart system over an infinite horizon, while providing a desired cycle service level (probability of not stocking out in an ordering cycle).

In the absence of individual fixed costs, all parts have the opportunity of being replenished at no additional cost every reorder interval. Each part, however, may be replenished in unequal frequencies and unequal quantities over time as dictated by their batch sizes, even in the case of constant demand as we saw in the previous chapter. To formulate this complex problem, we approximate the inventory costs by the sum of the safety stock required to guarantee the desired service level, plus the cycle stock associated with the deterministic version of the JRP model with batch ordering. This approximation is common in the inventory management literature (Eynan & Kropp, 1998), and results in underestimation of inventory due to backorders being counted as negative inventory. The approximation is thus quite accurate when service levels are high. As in Eynan and Kropp (1998), we express the safety stock for part j required to achieve the desired service level as a multiple z_j of the standard deviation of demand forecast errors during $(T+L)$ periods, the interval of time before the next order arrives, during which the system is at risk of stockout.

The cycle stock under constant demand is characterized by the Theorem below, which was derived in the previous chapter. For simplicity, we use the result assuming demand per period to be in full units. The approach can be extended to the case of fractional demand, as demonstrated in that chapter.

Theorem 3. *(Adapted from previous chapter): Given a fixed reorder interval of T periods, the long-run average inventory of a part with constant demand rate of D and a batch size restriction of B is:*

$$\frac{1}{2}(TD + B - g.c.d(TD, B))$$

where g.c.d. is the greatest common divisor.

The problem of minimizing the average ordering cost plus cycle and safety inventory cost per period can thus be written as:

$$Min_T \ \frac{A}{T} + \sum_{j=1}^{N} h_j \left(\frac{1}{2}(TD_j + B_j - g.c.d.(TD_j, B_j)) + z_j\sqrt{T+L}s_j \right)$$

This objective function is not well behaved; see Figure 3.1 below for illustration. As a result, we solve the problem by performing an exhaustive search over the reorder interval T between a lower and upper bound. This procedure is very fast, since ordering can only be done in discrete periods and the objective function evaluation is extremely simple using the closed form expression given in *Theorem 3*. Observe that as the standard deviation of demand increases, the safety stock term will grow more quickly in T while all other terms remain the same. Consequently, the optimal reorder interval will decrease as the standard deviation of demand increases. An upper bound for T thus is that for the deterministic case ($s = 0$). We will use the upper bound T^{UB} derived from the EOQ solution for the deterministic case in Section 2.3 of the previous chapter, and perform an exhaustive search from 0 to T^{UB}.

3.3.1 Dynamic Ordering Quantity Calculation

Given a chosen reorder interval of T periods, the shared fixed cost is now a sunk cost and thus the ordering decision can be made independently for each part j. The number of batches of each part to order will depend on the current inventory position. Observe that this setting fits the newsvendor framework, with overage costs equal to Th_jB_j, i.e., the cost of carrying one batch over the reorder interval, and underage costs equal to the additional cost p_j associated with air shipping a batch. Let I_j denote the current inventory position of part j. Let X_j denote the demand for part j until the next order is received, that is, the demand over $T + L$ periods of time. It is optimal to order the $(n_j + 1)^{th}$ batch as long as the expected benefit in saved overage cost is greater than the expected carrying cost; that is, if:

$$p_j P[X_j > I + n_j B_j] > Th_j B_j P[X \leq I + n_j B_j]$$

Thus, the optimal number of boxes of part j to order for a single reorder interval with initial inventory I_j is $n_j + 1$, where n_j is the largest integer satisfying:

$$P[X_j \leq I_j + n_j B_j] < \frac{p_j}{p_j + Th_j B_j}$$

The implementation of this condition is simple. Calculate the base stock level S_j that corresponds to that critical fractile of $p_j/(p_j + Th_jB_j)$ and always order the minimum number of batches to bring the inventory up to or above that base stock level. In the case of normal demand distribution the order-up-to level and resulting batch ordering quantity are:

$$S_j = \mu_{x_j} + z_j\sigma_{x_j} \quad and \quad Q_j = \left\lceil \frac{(S_j - I_J)^+}{B_j} B_j \right\rceil$$

where

$$(x)^+ = max(x, 0) \quad and \quad \lceil x \rceil \quad is \quad the \quad ceiling \quad function$$

3.3.2 Service Level Determination

The first step in our approach was to calculate a reorder interval T, given a desired service level. This initial service level should be linked to the trade-off between holding and penalty costs discussed in the previous section. Observe that the inventory cost associated with overage depends on the length of the reorder interval. As an initial approximation, we consider the reorder interval given by the EOQ solution associated with the aggregate demand for all parts, which we denote by T^{EOQ}, and use for each part j a safety factor z_j, such that $P[Z < z_j] = p_j/(p_j + T^{EOQ}h_jB_j)$. Once the optimal T is calculated given these initial safety factors, the safety factors can be recalculated for that T, and the problem solved for these new factors. We can repeat this process iteratively until convergence is found.

3.4 Computational Study

In this section, we first carry out a guided computational study to demonstrate the savings associated with our proposed methodology under various parameter settings. In this study, we use an exemplary base case of [A=\$100; h=\$0.05; z=1.645, B=75; n=10; with demands independent identically distributed with μ=3 and $C.V.$=0.5], and test the sensitivity of the solution to various parameters. We determine the savings relative to the cost observed with the EOQ solution, $C(EOQ)$, and the solution presented by Eynan and Kropp (1998), $C(E\&K)$. Observe that in the absence of batch constraints and individual setup costs, the EOQ solution considering the aggregate demand of all parts is optimal.

We then apply the proposed joint replenishment policy under the demand and cost settings of our industrial partner to demonstrate savings in a real industrial context. In the scenario tested, 58 parts with means in the range [0, 3.54] per period and coefficients of variation in the range [0.09, 11.96] need to be jointly ordered from a supplier in boxes of 110lbs with a shared, fixed shipping cost of \$1050.00. Inventory cost per lb per week is h=0.03. The current policy is to air ship boxes individually, as needed, at a cost of \$2.85/lb.

3.4.1 Experimental Results

We first illustrate the complex shape of the total cost curve for our base case. As seen in Figure 3.1, the total cost follows a jagged curve as the reorder interval T increases, with T^*=25 and approximations of $T(EOQ)$=12 and $T(E\&K)$=11.

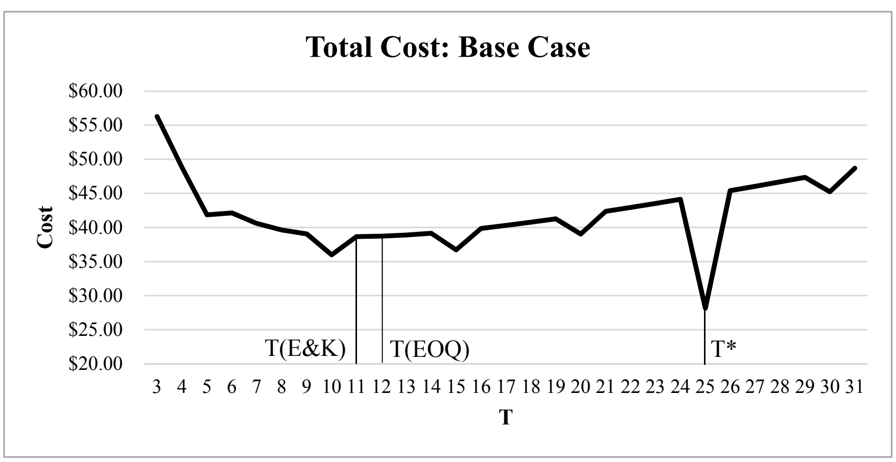

Figure 3.1. Total cost function for base case A=\$100; μ=3; $C.V.$=0.5; B=75; n=10

Table 3.1 shows the performance of our SJRP with batch ordering algorithm relative to the EOQ solution and the standard SJRP algorithm in Eynan and Kropp (1998), denoted by E&K. The first column presents the changes made to the base case (with the base case highlighted). The mean and standard deviation of all n parts is the same, except for the last set of cases where 5 parts have mean μ_1 and the other 5 parts μ_2. The results demonstrate the importance of accounting for the batch size when determining the joint ordering policy, as it can lead to up to 56% lower cost. The savings, however, vary wildly depending on the relative magnitude of the parameters. Further accounting for the safety stock, without considering batch restrictions, as in Eynan and Kropp (1998) reduces the reorder interval but tends to have little effect on costs for the cases tested where batch restrictions is the dominating factor.

3.4.2 Case Study

In our industrial case study, our proposed methodology promises savings for the different cases of $\frac{C(EOQ)}{C^*} = 1.013$ and $\frac{C(E\&K)}{C^*} = 1.010$ with $\frac{T(EOQ)}{T^*} = 1.05$ and $\frac{T(E\&K)}{T^*} = 0.69$.

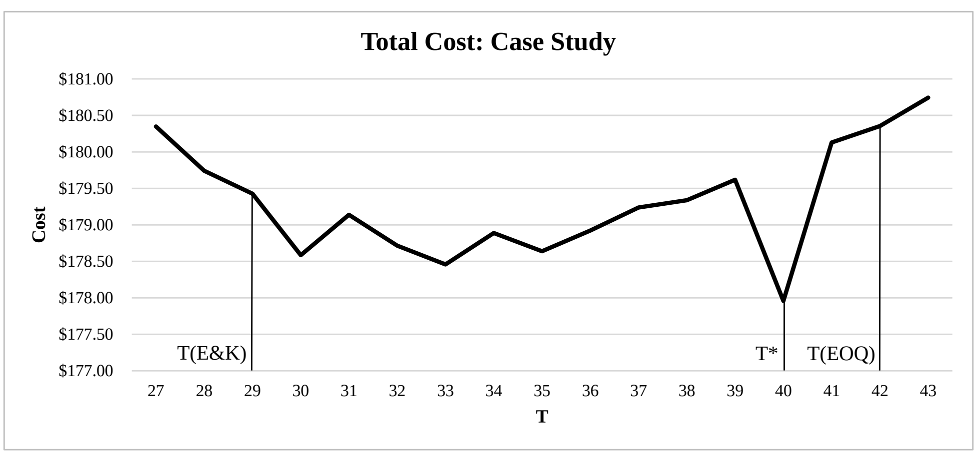

Figure 3.2. Total cost: Case study

While the safety stock under the highly variable demand observed in practice is a major factor and drives the $E\&L$ reorder interval to be less than 50% of that in the

n	T^*	$\frac{T(EOQ)}{T^*}$	$\frac{T(E\&K)}{T^*}$	C^*	$\frac{C(EOQ)}{T^*}$	$\frac{C(E\&K)}{T^*}$
1	50	0.76	0.72	$6.35	1.22	1.21
2	25	1.08	1	$8.83	1.38	1.00
10	25	0.48	0.44	$28.14	1.38	1.37
100	5	0.8	0.6	$238.63	1.10	1.10
B						
10	10	1.2	1.1	$22.14	1.08	1.09
25	10	1.2	1.1	$26.76	1.03	1.03
50	10	1.2	1.1	$31.37	1.06	1.06
75	25	0.48	0.44	$28.14	1.38	1.37
100	5	0.8	0.6	$238.63	1.03	1.02
C.V.						
0	25	0.48	0.48	$21.31	1.56	1.56
0.25	25	0.48	0.48	$24.72	1.46	1.46
0.5	25	0.48	0.44	$28.14	1.38	1.37
1	25	0.48	0.4	$34.97	1.26	1.18
2	25	0.48	0.36	$48.64	1.13	1.12
μ						
3	25	0.48	0.44	$28.14	1.38	1.37
10	5	1.40	1.20	$58.26	1.08	1.03
17	5	1	1	$81.58	1.00	1.00
79	2	1	1	$211.67	1.00	1.00
A						
100	25	0.48	0.44	$28.14	1.38	1.37
200	25	0.68	0.64	$32.14	1.44	1.43
500	25	1.08	1	$44.14	1.38	1.00
1000	50	0.76	0.72	$63.51	1.22	1.21
μ_1 μ_2						
3 **3**	25	0.48	0.44	$28.14	1.38	1.37
3 **12**	5	1.6	1.4	$53.89	1.03	1.02
3 **48**	5	0.80	0.80	$101.99	1.01	1.01
12 **48**	3	1.33	1.00	$113.33	1.00	1.00
12 **12**	5	1.2	1.2	$65.91	1.04	1.04
48 **48**	3	1	1	$151.36	1.00	1.00

Table 3.1. Comparison of optimal SJRP with batch ordering, EOQ and E&K policies

EOQ solution, the batch restrictions and the ensuing inventory accumulation when not synchronized, drive the optimal reorder interval up to be almost as high as the EOQ. Note that while the $E\&K$ reorder interval is much shorter, it leads to similar cost because the objective function is relatively flat around the EOQ reorder interval (see Figure 3.2). The true advantage of synchronizing supply through a joint ordering policy lies in the 38% cost savings achieved relative to the current company policy of air shipping all materials. Similar savings hold for two other real case studies we run.

3.5 Conclusion

In this chapter we study the stochastic joint replenishment problem with batch ordering, derive an approximate average cost function, and determine the corresponding optimal joint reorder interval. Once the reorder interval has been fixed, the batch order quantity of each part is calculated using a newsvendor approach. The inclusion of batch ordering and safety stock in the approximate model to calculate the reorder interval, rather than using a simple EOQ approximation, results in savings of over 1% in our case study and anywhere from 0-56%, depending on the parameters, in our guided computational experiments. The jagged shape of the total cost under batch ordering drives the savings. Further considering demand variability and the addition of safety stocks to the EOQ, as in the E&K model, resulted in very minor cost improvements in general, but a 38% gain in one of the cases tested. We thus conclude that jointly accounting for batch ordering and demand variability is necessary. A case study comparing the performance of this policy relative to the current industry practice of air shipping shows savings of 38%.

CHAPTER 4

COMPONENT INVENTORY MANAGEMENT FOR HIGH-TECH ASSEMBLY SYSTEMS

4.1 Introduction

Industries with high technology and innovative products, such as energy, transportation, defense, and aerospace, take on an important role in the U.S. economy (see aerospace case in Deloitte, 2012). In recent years, these industries have suffered from not meeting their production deadlines causing significant delays to their customers (e.g., Sanders & Cameron, 2011; Denning, 2013; Mann, 2016). These delays can in many cases be attributed to the industry's typically long and highly variable component lead times making delivery performance hard to predict for the OEM. There are several reasons why these industries are operating in such a difficult environment. First, these industries compete through innovative, high-tech product solutions (e.g., PricewaterhouseCoopers, 2011). Their products usually adopt and push the limits of the latest research in manufacturing, design, and materials. Tight design specifications are necessary to ensure performance, in particular for security relevant components used in defense and aerospace. The mix of product complexity, novel processes and materials, and tight design specifications induces a challenging production process along the supply chain. Second, more and more OEMs outsource and offshore a large portion of their production in order to focus on their core competencies (e.g., Bales, Maull, & Radnor, 2004). These have been promising methods in other industries to lower the OEM's costs for operation and labor, as well as to foster flexibility and agility, while being able to leverage external investments and expertise for components of the assembly. For complex high-tech products, the production process thus requires sub-assemblies produced in several stages involving multiple supplier tiers and geographical regions. Furthermore, these supply chains are being extended even further due to export control restrictions or proprietary in-house manufacturing steps, which may require components to travel from suppliers to OEM and back to suppliers for further processing. Third, OEMs in these industries face lower volumes and higher component costs than is common in other industries (e.g., automotive). In addition to the inherent manufacturing complexity, components are made out of expensive rare raw materials allowing only little or no inventory holding upstream in the supply chain. This results in small inventory buffers and, hence, longer response times along the supply chain. Fourth, replacement components may have very sparse demand and occur in batches, while requiring delivery within the component's underlying supply lead time. This introduces significant pressure in the supply chain, and may result in delays of the entire assembly if components routed for the assembly plant are funneled to cover spare component demand. Lastly, quality issues require extensive engineering analysis and tests, which may take months in some instances and, therefore, constitute another source of component lead time variability.

These industries tend to operate in a low volume environment but have the promise to grow and become increasingly important in the years ahead. Taking the aerospace sector as an example, Airbus predicts a growth of 4.7% per year within the next 20 years accounting for more than 29,000 new passenger aircrafts and freighters (Clearwater International, 2014). Hence, for a smooth production or ramp-up process, it is important to understand and evaluate how suppliers' stochastic delivery performance affects OEMs' inventory holding as well as the assembly process (see PricewaterhouseCoopers, 2012; Mann, 2016). For the OEM, the major problem is to determine appropriate component inventory buffer levels to hedge against the long and variable supplier lead times and mitigate their negative effects. The objective is to guarantee a high level of final assembly delivery performance while minimizing component inventory costs.

The best way to illustrate the impact of variable component deliveries is through a simple example. Consider a product assembly requiring eight different components. Demand is two assemblies per day. The OEM places component orders according to a certain Quoted Lead Time (QLT) that the supplier has agreed to. Delays, however, will occur because of the unpredictability in the lead times and will be normally distributed. Table 4.1 shows the delay distribution parameters, mean and standard deviation, and resulting component inventory buffers required to ensure a service level of 95% for each component.

Component	1	2	3	4	5	6	7	8
Mean	0	0	5	7	25	70	100	100
Standard Deviation	1	10	2	6	4	40	10	40
Inventory Buffer	4	33	7	20	14	132	99	132

Table 4.1. Inventory buffer example: Requirements

Despite significant inventory buffers being carried, this level of component availability is not at all sufficient for an assembly process of even just 8 components; it results in only a 66% probability of the final assembly being ready on time (day 0 in Table 4.2). Table 4.2 shows the service level of the final assembly at various points in time after its due date. Assembly service levels are calculated using the delay distributions and the multiplicative property of the service levels of independent components.

Table 4.2 reveals that the desired service level of 95% will only be reached after more than 10 days beyond the original planned production date; a high service level of 99% will only be achieved after 35 days. This simple example illustrates how fast the service level of an assembly degrades in the presence of multiple components with

	Probability of availability after given number of days										
Component	0	5	10	15	20	25	30	35	40	45	50
1	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.99	1.00	1.00	1.00
7	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.99	1.00	1.00	1.00
Final Assembly	0.66	0.89	0.93	0.96	0.97	0.98	0.98	0.99	0.99	0.99	1.00

Table 4.2. Inventory buffer example: Final assembly service level

supply lead time uncertainty. In practice, where assemblies typically involve several hundreds to thousands of components, on-time delivery performance is virtually impossible ($0.95^{100} = 0.006$) unless appropriate component buffers are maintained.

In this chapter, we address the problem of synchronizing component procurement in the assembly process of a product under long and highly variable supply lead times. Building on the previous literature, we use time buffers rather than physical safety stock, as they have been shown to be superior in this context. Our major contribution is providing simple, but effective tools for practitioners to (1) determine time buffer levels, (2) quantify the resulting inventory reduction and service level increase, (3) develop a phased implementation approach, and (4) assess the potential risks associated.

The rest of the chapter is organized as follows. This section provides the motivation for this study and discusses the relevant literature. The next section introduces the modeling framework including assumptions and analytical bounds that quickly identify the savings achieved through synchronization. We then follow up with stochastic optimization and simulation approaches to the problem, respectively. The computational results, based on a real industry case study, highlight the consistency of the output of the various approaches, and the drastic performance improvement associated with supply synchronization. Furthermore, they show the robust performance of time buffering strategies as supplier behavior evolves, and identify incremental implementation strategies to support the transition to a synchronized system in practice. Finally, we discuss limitations and future work and conclude with a summary of major insights.

4.2 Literature Review

Stochastic procurement lead times have been extensively studied for over 50 years (see for example the reviews in Bramson, 1962; Zipkin, 2000; Minner, 2003; Mula, Poler, Garcia-Sabater, & Lario, 2006; Tang, 2006; Tajbakhsh, Zolfaghari, & Lee, 2007; Colicchia & Strozzi, 2012; Aloulou, Dolgui, & Kovalyov, 2014). Dolgui and Prodhon (2007) survey the literature focusing on MRP systems under supply uncertainty, and highlight assembly systems with uncertain lead times as a promising and little studied research area highly relevant for both academics and practitioners. The interdependence of component inventories and the simultaneous consideration of uncertainties are identified as the main challenges. Dolgui, Ammar, Hnaien, and Louly (2013) review studies focusing on uncertain lead times under deterministic demand as well as studies in our context of assembly systems.

Stochastic procurement lead times have received increased attention in recent years driven by high competition, increased outsourcing, and the quest to further reduce operating costs. In our context of aerospace manufacturing and assembly, the industry underwent a general change from being a mostly vertically integrated supply chain to a product focused OEM with specialty suppliers (e.g., Bales et al., 2004; C. Rossetti & Choi, 2005). The outsourcing to suppliers located around the world, the increased complexity of products, and the continuously evolving technologies have led to significantly increased lead times and uncertainty which have challenged traditional procurement methods.

Procurement uncertainty has been categorized into three groups according to their outcome and underlying sources of variability. Supply uncertainty may lead to complete orders (i) individually not arriving over a longer period of time (disruption models), (ii) partially arriving at different points of time (random yield models), or (iii) arriving in full but at a random point of time (stochastic procurement models). All these research areas are related but account for the underlying circumstances originating from different root causes in each case. Our study falls into (iii) where the time between the placement of the order and its observed arrival varies significantly. Rather than being disrupted by a punctual external event, reasons for the delay may include (a) optimistic (competitive) quotes, (b) quality issues, (c) supplier congestion, or (d) spares cannibalization of incoming orders.

For a single-sourced component, there are generally three approaches to address uncertainty in lead times: (1) Safety stocks, (2) safety lead times, and (3) lot sizing (e.g., Dolgui, Louly, & Prodhon, 2005; Mula et al., 2006). In the safety-stock approach, uncertainty in procurement is addressed by physically stocking an additional quantity. In the case of safety lead times, components are ordered an extended period ahead of their planned usage. Lot-sizing rules combine both previous approaches and specify the order amount and timing. All approaches ultimately yield increased inventory. However, the underlying dynamic and timely distribution of accumulated inventory in each case is different. Whybark and Williams (1976) simulate the first two mitigation strategies for a MRP system which faces uncertainty in timing as well as in quantity (i.e., in both supply and demand). The authors conclude that a preference scheme exists allowing for a higher service level for the same average inventory. Namely, safety-lead time is preferred for uncertainty in timing whereas safety stock is the preferred method when the uncertainty relies in the quantity. The difference in service level is amplified when increasing the coefficient of variation. Intuitively, a static safety stock rule is designed for buffering sudden peaks in demand or supply disruption at a cost of an increased average inventory level. The dynamic safety lead time, however, is a buffer associated to a particular order and is designed to buffer for its arrival time. Hence, a safety lead time strategy is very limited in offsetting the quantity uncertainty, but reveals its advantages by only temporarily increasing inventory levels. This is echoed by Chang (1985) who further investigates the question of the interchangeability of safety stocks and safety lead times in a manufacturing planning setting. His study concludes that both buffering techniques are interchangeable, but only if planning flexibility is given, reflected by two distinct conditions. According to the author, the quantity uncertainty can be buffered by safety lead time when "(1) *The excessive demand is known before the actual production of the components in the lowest level*, (2) *The raw material at the lowest level is available. However, in most industry settings, these conditions are not realistic."* Melnyk and Piper (1981) demonstrate, through simulation, the value of adding safety lead times in an MRP implementation to ensure effective delivery performance in multi-product, multi-stage assembly systems. Molinder (1997) compares the three approaches to hedge against lead time and demand uncertainty in an MRP context. The study confirms that preferences should be given to safety lead times as lead time variability is high, demand variability is low, and stockout to inventory holding cost ratio is high.

Our study focuses on the assembly system of a single product consisting of hundreds or thousands of components with deterministic demand. There are a few studies that are most relevant to ours. M.-A. Louly and Dolgui (2011) and M.-A. Louly, Dolgui, and Al-Ahmari (2012) study a single assembly system consisting of multiple types of components. Lead times are stochastic, and a periodic order quantity rule is assumed in their modeled MRP environment that minimizes the sum of the average component holding costs, setup cost, and average backorder cost for the final product. Demand is constant and known. In M.-A. Louly and Dolgui (2011), the authors present a method that optimizes component-dependent planned lead times and a single periodicity parameter for all components. In M.-A. Louly et al. (2012), the authors extend their model to find the optimal MRP offsetting under service level constraints. Hnaien, Dolgui, and Wu (2016) also study an assembly for one product but consider stochastic demand to find both optimal component lead times and quantities. The authors develop a Branch and Bound algorithm (following results from M.-A. O. Louly and Dolgui (2009)) and compare it with five heuristics based on the newsvendor model. The computational results of up to 100 components favor the proposed branch and bound algorithm.

Jing-Sheng Song and collaborators have produced a significant stream of literature addressing variability in component lead times within assembly systems (e.g., Song, 1994; Song & Zipkin, 1996; Song & Yao, 2002; Song, Zhang, Hou, & Wang, 2010). Song, Yano, and Lerssrisuriya (2000) conclude that stochastic lead times may have a higher impact than stochastic demand, and that it is essential to consider stochastic lead times since even heuristics can improve performance significantly. Gallien and Wein (2001) study a single-item assembly system with Poisson demand, assuming uncapacitated suppliers with independent and non-identically distributed stochastic delivery lead times, instantaneous assembly, unsatisfied backordered demand, and the condition that sequential orders do not cross and mix. The authors focus on a finished good base stock policy with component postponement times. Using queuing theory and constrained mixed non-linear programming, the authors find an exact solution for the deterministic case and use an approximate decomposition method for the stochastic case.

In summary, we can conclude that uncertainty in supply lead times has been studied in many different contexts. We can note, however, that it has also been one of the least studied areas in supply chain management and production planning. The main focus has been on the challenging task of forecasting future demand as well as investigating the optimal lot sizing and inventory rules. The continuous offshoring and outsourcing, shorter reaction times, and the need for a more agile supply chain have made supplier performance in many industries more fragile and prone to delay or disruption. Many practitioners and academics have recognized the need and have taken on the challenge to close the gap (e.g., Dolgui & Prodhon, 2007). In our context of high-tech assembly systems, we focus on developing practical approaches to implement lead time buffers and evaluate their potential risks.

4.3 Modeling Framework

In this section, we introduce the assumptions and notations that we use as building blocks for the three modeling approaches we propose to study the supply synchronization problem.

4.3.1 Assumptions

We consider a specialized, high-cost, make-to-order environment in which the OEM typically receives orders months in advance, sets a production plan, and manages the supply chain according to the resulting (deterministic) demand for components driven by its MRP system. We focus on the final assembly of one product involving complex components and consisting of several hundreds to thousands of components. We assume that supplier lead times across components are independent and non-identical random variables with known distributions. We further assume that the lead time for each component is bounded by a finite 100^{th} percentile of the lead time distribution. Suppliers replenish the OEM's orders on a first-come-first-served basis, implying that consecutive orders do not cross in time. Single source and uncapacitated supply is assumed. In reality, capacity issues are a major challenge in this industry. The OEM, however, does not have visibility of the congestion state of the suppliers (who have other major sources of demand to satisfy as well), and thus simply observes variable lead times as a consequence of the capacity constraints. For simplicity, we assume the assembly time is negligible and all components need to be on-hand to start the assembly. A lengthy assembly sequence could be considered, but would require detailed accounting of the time phasing of each of the required components without changing the basic insights of the model. Lastly, we assume that unsatisfied customer demand will be backordered and satisfied on a first-come-first-served basis by the OEM.

4.3.2 Notation

Customer orders to the OEM are recorded, planned, and executed according to a MRP system. The component suppliers and the OEM contractually negotiate a lead time for each component j, which is referred to as the quoted lead time, q_j, and used by the OEM in placing supply orders.

Notation	Definition
J	set of components required for final product assembly, indexed by $j, j = 1, 2, .., J$
h_j	holding cost of component j
q_j	quoted lead time of component j
b_j	buffer time used to advance the ordering of component j beyond quoted lead time
X_j	lead time of component j, a random variable with mean X_j
$L_j := X_j - q_j$	delay (earliness/lateness) of component j, a random variable with mean $\bar{L}_j$
$\bar{L}_{max} := max\{\bar{L}_j\}$	maximum mean delay (earliness/lateness) over all $\bar{L}_j$
$L := max_j\{L_j - b_j\}$	delay (earliness/lateness) of final assembly, a random variable with mean $\bar{L}$
S	set of random scenarios considered, indexed by $s, s = 1, 2, .., S$
x_j^s	realized lead time of component j, under scenario s
$l_j^s := x_j^s - q_j$	delay (earliness/lateness) in days of component j, under scenario s
$l^s := max_j\{l_j^s - b_j\}$	final assembly delay under scenario s

Table 4.3. Model notation

Unfortunately, the complexity of the components and low demand volumes lead to significant variation in the actual delivery times and may cause delays. To buffer against this variability, we will consider the addition of buffer times b_j, which result in advancing the placement of orders by that additional time beyond their quoted lead times. That is, component j for a final assembly planned for delivery at time t will be ordered at time $t - q_j - b_j$. The timely delivery of the final assembly depends on the component arriving last. Hence, random variable L in Table 4.3 captures the lateness in days for the final assembly. Each component will be carried in inventory for the difference between the final assembly delay L and the individual component delay beyond the buffer time, i.e., $L_j - b_j$. Figure 4.1 depicts the lead time distribution of a particular component j, along with the other variables defined.

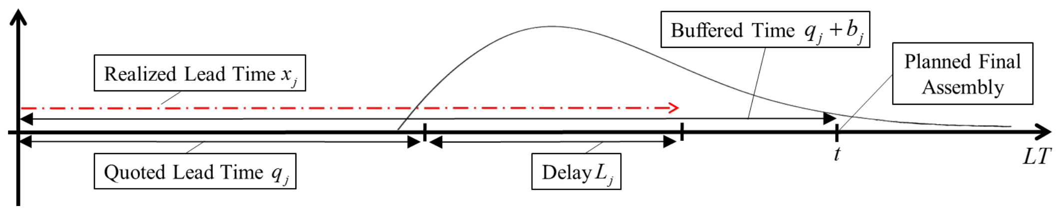

Figure 4.1. Example lead time distribution

4.3.3 Analytical Bounds

As a first step to assess the value of synchronizing the arrival of component supplies, we derive bounds on the performance of two extreme strategies:

1. *No-Buffer* Strategy: The firm orders components according to their given quoted lead time and holds no additional inventory.
2. *100%-Buffer* Strategy: The firm orders components in advance with a buffer time b_j^{100} equal to the maximum possible delay; that is, $b_j = P_j^{100}$.

A 100% time buffer is only possible if there is a known upper bound on the component supply lead time. In practice, we consider the worst delay seen in the past six months as the 100th percentile of the delay distribution. Practitioners almost universally expected the *100%-Buffer* to lead to perfect delivery performance at an unsustainable increase in component inventory in the system. The following lower bound on the expected inventory savings associated with moving to a *100%-Buffer* strategy shows otherwise. The lower bound on the expected difference is obtained by calculating a lower bound on the cost of a *No-Buffer* strategy and an upper bound on the cost of the *100%-Buffer* strategy as shown below.

4.3.4 Lower Bound on Cost of No-Buffer Strategy

The expected assembly delay

$$\bar{L} = E[L] = E\left[\max_j\{L_j\}\right]$$

has a rough lower bound at the maximum of the mean delays over all components j, i.e.,

$$E[L] \geq \max_j\{\bar{L}_j\} = \bar{L}_{max}$$

The expected system-wide holding cost incurred per final assembly is:

$$E\left[\sum_{j=1}^{J} h_j(L - L_j)\right] = \sum_{j=1}^{J} h_j(E[L] - E[L_j]) \geq \sum_{j=1}^{J} h_j(\bar{L}_{max} - \bar{L}_j)$$

This lower bound is admittedly rough as it simply represents the inventory resulting from the differences in the expected delays of the various components.

4.3.5 Upper Bound on Cost of 100%-Buffer Strategy

Clearly, the *100%-Buffer* strategy leads to no component shortage and thus no assembly delay. Components will no longer be late but early by an amount equal to the difference between the scheduled assembly time (synchronized for all components to be at time $t = q_j + P_j^{100}$ after their ordering time) and the observed lead time $X_j = q_j + L_j$. Therefore, each component needs to be carried on average for the time $P_j^{100} - \bar{L}_j$. If all components arrive early, assembly could possibly start at that earlier point with a corresponding decrease in inventory. Thus, the expected system-wide holding cost incurred per final assembly is at most:

$$\sum_{j=1}^{J} h_j(P_j^{100} - \bar{L}_j)$$

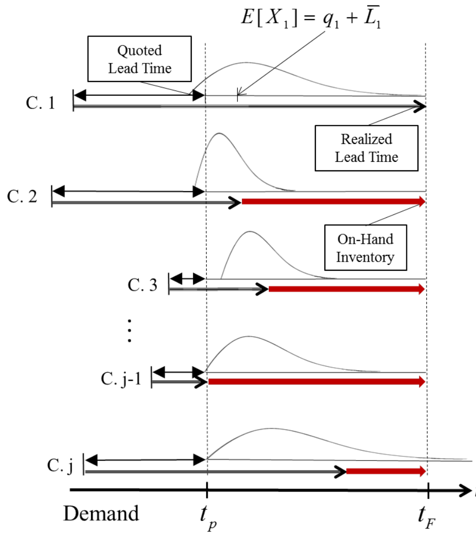

Figure 4.2. *No-Buffer* strategy

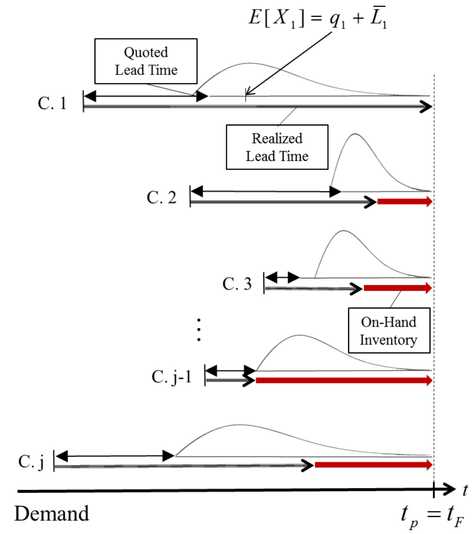

Figure 4.3. *100%-Buffer* strategy

Figure 4.2 and Figure 4.3 illustrate the supply ordering and delivery timeline under both strategies, the *No-Buffer* strategy on the left and the *100%-Buffer* strategy on the right, for an assembly with potentially hundreds of components, denoted as *C.1, C.2,.., C.3*. The time t_P represents the MRP planned assembly date and t_F the final assembly time, which requires the arrival of all components. Under the *No-Buffer* strategy orders are placed q_j days ahead of planned production, accounting for the agreed upon supply lead time. Components arrive randomly within the domain of their delay distributions, as marked on the figure using a black solid line with an arrow to the right, and are held in inventory until the final assembly time, t_F, as shown with a lighter arrow. Out of hundreds of components, the probability of one being at a high percentile of its right-skewed lead time distribution is large and as a result inventory will bloat as all other components wait for the arrival of the last few. Under the *100%-Buffer* strategy, on the other hand, the worst-case arrival scenarios of all components are synchronized to coincide with the planned assembly date, as shown in Figure 4.3. As a result, components are simply held in inventory from their actual arrival times until their planned worst-case (or 100^{th} percentile).

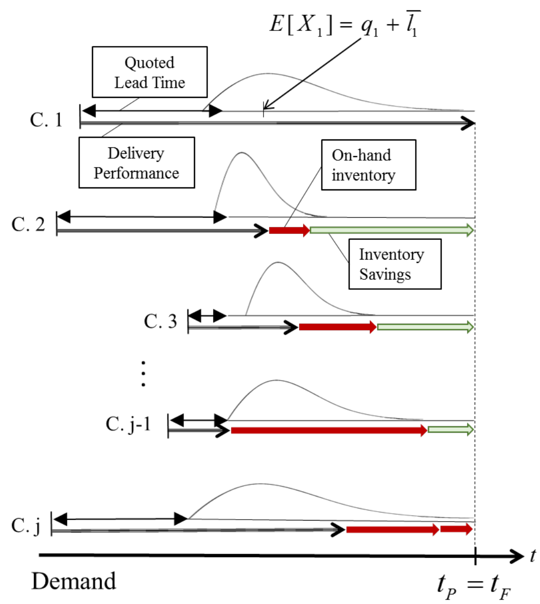

Figure 4.4. Illustrating inventory savings: *No-Buffer* versus *100%-Buffer* strategy

Figure 4.4 compares both strategies for the particular component lead time performance example shown in Figure 4.2 and Figure 4.3. For purpose of illustration, we depict the same figure as seen in the *No-Buffer* strategy but act as seen in the *100%-Buffer* strategy with $t_P = t_F$. This allows to visualize the observed savings using the *100%-Buffer* strategy in this example. In very rare cases, following the *100%-Buffer* strategy may lead to inventory surplus. This is, when (1) all components arrive close to their quoted lead time q_j yielding to no or very little inventory in the system, or (2) when the set of components exhibiting the worst delivery performance (i.e., long tail) arrive earlier than the maximum delay seen. Following the notion of our earlier numerical example, the probability for (1) to occur is close to zero. Our results show that (2) does occur but the savings significantly outweigh the inventory surplus observed (as seen in the example above).

Using the analytical bounds and the data provided by our industry partner, we show that perfect service could be provided at a significant reduction in inventory (almost 60%) by simply synchronizing the supply of the various components with the *100%-Buffer* strategy. The intuition behind it will be further discussed. The remainder of the chapter we explore three questions that arise from this realization:

1. Can we find a better balance between inventory and service than that provided by the extreme *100%-Buffer* strategy?

2. What is the right implementation process for the resulting synchronizing strategies?

3. What are the risks associated with this synchronization? What is the possible impact if our assumed worst cases (or 100th delay percentiles) turn out to be inaccurate?

4.3.6 Stochastic Optimization Model

The analysis in the previous section uncovers that significant savings can be achieved by synchronizing the worst-case arrival time of the different components using a blanket *100%-Buffer* strategy. Additional benefits should be possible by allowing the time buffers for the various components to be different; for example 98% (i.e., 98th percentile of the delay distribution) for a very expensive component with a highly right-skewed delay distribution. For that purpose, we formulate the problem of finding the time buffer b_j for each component j so as to minimize the expected component holding cost per final assembly subject to meeting a delivery time window, as a stochastic optimization model over a set of scenarios S.

$$Min \quad \frac{1}{S}\sum_{s=1}^{S}\sum_{j=1}^{J} h_j \Big[l^s - (l_j^s - b_j) \Big]$$

subject to

$$l_j^s = x_j^s - q_j \quad \forall\, j, s$$

$$l^s \geq l_j^s - b_j \quad \forall\, j, s$$

$$l^s \geq e \quad \forall\, s$$

$$l^s \leq d \quad \forall\, s$$

$$b_j \geq f \quad \forall\, j$$

For each scenario s, the delay of the entire assembly l^s is required to be at least as long as the largest delay of all components beyond the buffered time (constraint #2) and not allowed to be earlier than a certain amount e (constraint #3) or later than a certain delay d (constraint #4). In our numerical examples, we allow assemblies to occur up to 7 days earlier than planned, and a potential delay of up to 28 days, as this was of interest to our industrial partner. Early assembly will need to be accommodated in the assembly plant. A potential delay, or grace period, must have been negotiated with customers. No delay, $d = 0$, would necessitate our *100%-Buffer* strategy. How much can we lower component inventory by allowing a grace period? Our decision variable b_j describes the optimal component specific buffer time in days and must be of positive nature (constraint #5). In practice, we randomly generate 10,000 of these scenarios using empirical data.

4.3.7 Simulation Model

To test the performance of the proposed strategies in a dynamic setting where the lead times of subsequent assemblies will necessarily be correlated (since they never cross in a practical setting with first-in-first-out allocation of the delivered components), we use discrete event simulation. Besides the bill of materials and their cost, the main input parameter is the historic delivery performance for each component j to construct lead time distributions. Demand for the final product is set to be a constant (one final product per week).

The granularity chosen for the simulation is weekly buckets in which orders are placed and earlier orders are received. Supply delay distributions are constructed based on the time from order to delivery observed over the most recent six months as compared to the quoted lead time. Six months was found to be the right timing given the trade-off between having enough data points to construct the distribution and providing an accurate picture of current rather than past supplier performance.

The major challenge is the generation of a series of weekly lead times $q_j + l_j(w)$ for orders of component j in week w to closely replicate those observed in practice. The simulated lead times must reflect the same discrete distribution observed in practice but cannot be generated as independent draws from it because this would lead to order crossing, i.e., later orders arriving earlier than previous ones. Given that delivered materials will be used in a first-come-first-served basis, order crossing would lead to shorter than desired realized lead times in the simulation. To avoid order crossings, we must ensure that $l_j(w+1) \geq l_j(w) - 1$. This can be achieved by appropriately defining a new modified delay distribution L'_j to randomly draw values $l'_j(w)$, for each week w, and generating non-crossing delays, $l_j(w+1) := max(l'_j(w+s), l_j(w)-1)$, that match the original distribution observed in practice. The modified delay distribution L'_j can be found using the theory of Markov processes; please refer to the appendix.

Our simulation tool can adapt to both physical and time buffering of inventories. In the case of time buffering, the final product demand for a particular week t triggers an order for component j placed at time $t - q_j - b_j$. A delay l_j will be randomly drawn using the modified delay distribution and the order will thus be delivered at time $t - b_j + l_j$. The simulation allows us to collect the following performance metrics:

1. Long-run average of on-hand inventory in the system.

2. Final assembly delivery time distribution.

3. Fill rate over different "grace periods" providing insights into meeting customer demands within a certain time window.

4.4 Case Study

4.4.1 Analytical and Computational Results

In this section, we present the results of applying the three modeling approaches described above to the aerospace industry data for one product consisting of over 1500 components. Other products within this industry were analyzed in a similar fashion and led to comparable conclusions. The data for each individual component includes the units per assembly, the quoted lead time, cost, and past supply order and delivery data over multiple years. Supply delay distributions are constructed based on the time from order to delivery observed over the most recent six months as compared to the QLT.

In the following sections, we use these data to (1) test and compare the performance of our three modeling approaches - analytical, simulation, and optimization; (2) assess the value of synchronization, both under an encompassing *100%-Buffer* and under an optimal time buffer mix; (3) identify an effective phased implementation approach where time buffers (advanced ordering) are sequentially applied to more and more components over time; and (4) evaluate potential risks associated with synchronization due to the unpredictable evolution of supplier delays.

4.4.2 Evaluation of Synchronization Strategies

Table 4.4 below provides a comparison of the inventory cost projected using each of the modeling approaches: (1) analytical bounds, (2) stochastic optimization model, and (3) simulation. Four different inventory management strategies are considered:

1. *No-Buffer* or base case following current practice.

2. Time buffer to the 98th percentile of the delay distribution.

3. *100%-Buffer* or time buffer to the 100th percentile of the delay distribution.

4. Optimal or time buffer to the optimal percentiles suggested by the stochastic optimization approach.

The analytical bounds are only available for strategies one and two. The reported "optimization" costs associated with non-optimal strategies are simply the values of the objective function of the optimization model under the respective time buffers.

We use the *No-Buffer* or base simulation case as reference with a normalized value of 1, and normalize the inventory cost of all other strategies by dividing by the reference cost. The insights, however, are kept intact since the ratios stay the same:

					Savings	
Comparison	No-Buffer	98th	100%-Buffer	Optimal	No-Buffer Vs 100%-Buffer	*100%-Buffer* Vs Optimal
Simulation	1	0.43	0.42	0.41	58.47%	1.94%
Optimization	1.06	0.46	0.37	0.36	65.40%	1.58%
Analytical Bounds	0.90	-	0.37	-	59.27%	-

Table 4.4. Comparison of normalized inventory costs using three approaches

The three approaches consistently estimate the benefits of full synchronization in providing on-time delivery while reducing inventory by roughly 60%! The analytical bounds provide a fairly accurate estimate of the inventory costs and ultimate savings associated with synchronization. Simulation and optimization calculations are not a perfect match, as the simulation captures the dynamics of the system and how the delays evolve over time.

In what follows, we provide further detail on the results obtained through the optimization and simulation approaches.

Stochastic Optimization Approach

The *100%-Buffer* strategy provides perfect delivery and striking inventory savings. Nonetheless, it may result in overstocking of very expensive components to cover the end tail of a very skewed delay distribution. The stochastic optimization approach allows us to determine individual time buffers for each component to minimize inventory cost while achieving the desired delivery performance. Based on the needs of our industry partner, we allow a 4-week grace period over which delivery of the final product is acceptable beyond its due date. We use 10,000 randomly generated component lead time scenarios. This is found to provide sufficient accuracy of the expected inventory levels, as it results in a normalized 95% confidence interval for the mean inventory cost of [0.9985 1.0015].

The stochastic optimization approach finds that the optimal solution relaxes the *100%-Buffer* requirement for fifteen components to levels between the 96^{th} and 99^{th} percentiles, yielding inventory cost savings of 1.58%. The cost savings further increase to 1.62% when allowing assembly to occur as early as one week before the planned production date if all components are on-hand. The worst-case, or maximum amount of inventory that the firm may be saddled with out of all possible scenarios, is also of interest. Given the buffer times of each component, we can readily calculate the worst case across scenarios (see Table 4.5). We can also slightly modify the stochastic program to find the buffer times that lead to lowest worst-case inventory holding across scenarios (see last column in Table 4.5).

Optimization	100%-Buffer $l_s = 0$	$0 \leq l_s \leq 7$	$0 \leq l_s \leq 14$	$0 \leq l_s \leq 21$	$0 \leq l_s \leq 28$ Base Case	$-7 \leq l_s \leq 28$	Optimal Worst Case
Average Cost	1.016	1.005	1.002	1.001	1	0.999	1.122
Worst Case	1.169	1.244	1.358	1.491	1.592	1.727	-
Average l_s	0	0.545	0.865	1	1.128	-5.749	3.467

Table 4.5. Average cost and final product delay for different scenarios

Using the performance of the case with a 4-week grace period ($0 \leq l_s \leq 28$) as reference, Table 4.5 shows that narrowing the assembly window requirement gradually increases cost while slightly improving delivery performance. Furthermore, it highlights that the extreme *100%-Buffer* strategy is in fact very attractive; it achieves close to optimal average inventory costs, most robust cost across scenarios (lowest worst case), and best delivery performance with a simple, blanket policy. Figure 4.5 graphically depicts the cost savings as the grace period grows for two cases: (1) allowing early assembly as far as 7 days before MRP date and (2) requiring the assembly to occur no earlier than the MRP date.

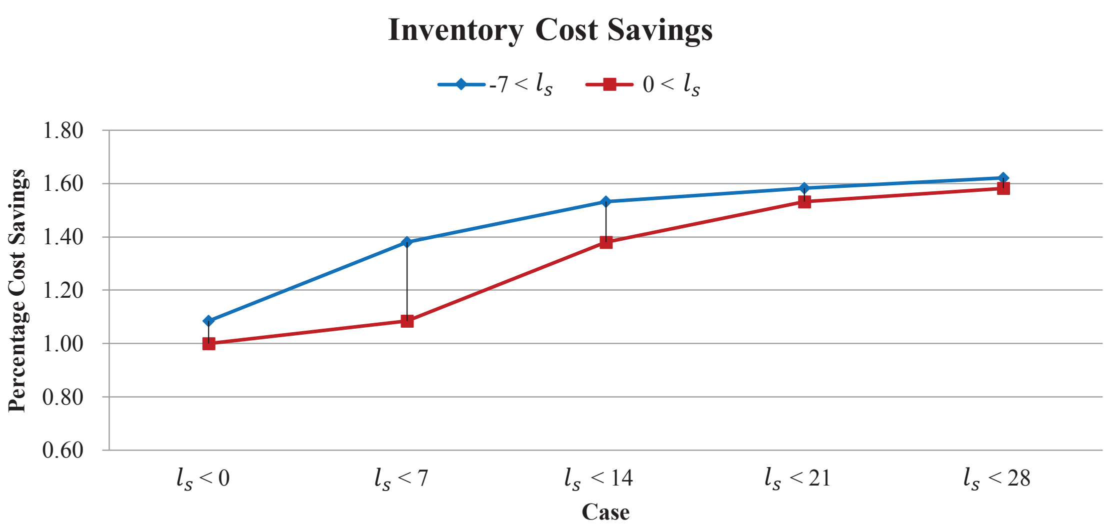

Figure 4.5. Inventory cost savings depending on final assembly allowance

Interestingly, allowing for an assembly window of four weeks still results in the majority of components to be assembled at the original MRP date. As shown in Figure 4.6, the earliest possible assembly time is reached with a probability of over 80%. An assembly within the first week is over 94% likely. An assembly at the latest day (day 28) is highly unlikely with a probability of 0.52%.

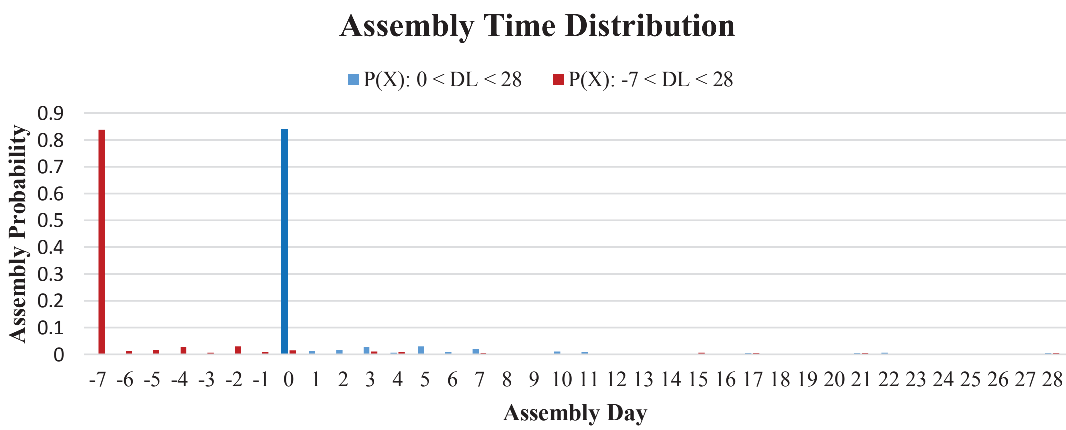

Figure 4.6. Distribution of assembly delays in 10,000 scenarios for two scenarios

Simulation

In order to capture the dynamics in the system, we create a discrete-event simulation using Matlab R2014a. We are simulating a 52-week period for 250 iterations. This was found to provide sufficient accuracy; a 95% confidence interval for the average normalized annual inventory cost was calculated to be [0.9981 1.0018]. We include a warm-up period in the beginning for each component simulated over which we assume deliveries to be on time. We require each component to have gone through at least one full ordering cycle during the warm-up period. Table 4.6 shows the resulting performance of the four inventory management strategies considered, including fill rate over different delivery windows. To comply with confidentiality restrictions, the results shown for average lateness, inventory cost, and inventory turns are scaled to one (*No-Buffer* case).

With no buffer implemented, the fill rate at the MRP assembly date is zero. Even four weeks after, the high variability and long lead times across hundreds of components still make assembly impossible. Only 100% time buffering over the lead time distribution, i.e., perfect synchronization, guarantees on-time delivery according to MRP. Buffering to a lower percentile (98th) of the lead time distribution may

Simulation	Fill Rate: MRP	Fill Rate: 2 Week	Fill Rate: 4 Week	Avg. Lateness	Inventory Cost	Inventory Turns
No-Buffer	0.00%	0.00%	0.00%	1.00	1.00	1.00
Buffer 98th	0.02%	97.70%	100.00%	0.05	0.43	2.08
100%-Buffer	100.00%	100.00%	100.00%	0.00	0.42	2.14
Optimal	90.10%	99.62%	100.00%	0.05	0.41	2.17

Table 4.6. Overview of simulation results using the different strategies

still yield an acceptable fill rate over a grace period of 2 or 4 weeks; however, it is dominated by the *100%-Buffer*, as it results in higher inventory. As seen in Table 4.6, the average lateness, inventory cost as well as inventory turns improve significantly when implementing the *100%-Buffer*. Simulation of the optimal strategy suggested by the stochastic optimization model shows that inventory costs can be further reduced and inventory turns increased while keeping a high service level and low average lateness.

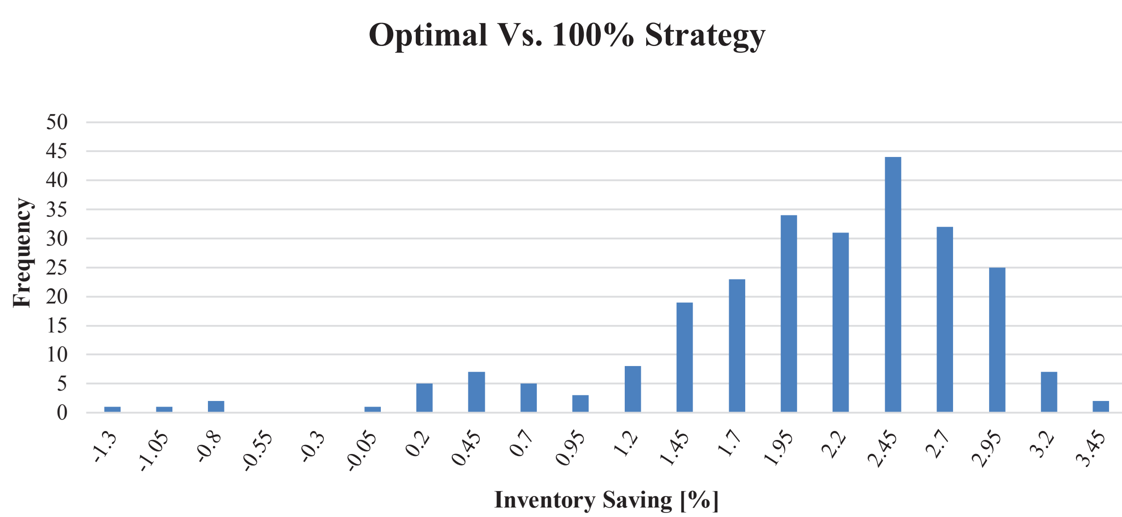

Figure 4.7. Additional relative inventory savings when using the optimal strategy

On average, the optimal buffer strategy provides further inventory savings. Is this consistent over the years? Could there be cases (plausible annual scenarios) where the optimal solution would significantly underperform? For each of the 250 years simulated (each year is one replication of the simulation), we compare the yearly inventory cost for the *100%-Buffer* strategy and the inventory cost for the optimal buffer strategy. The histogram in Figure 4.7 shows that there are only 4 of our 250 randomly generated 52-week simulations for which a *100%-Buffer* strategy would have been cost advantageous. In fact, all other iterations show significant cost savings, which accumulate over the 250 iterations to 1.94%.

In Figure 4.8, we compare the observed inventory savings when using the optimal strategy relative to a *No-Buffer* strategy. In the simulated 250 iterations we can consistently observe high inventory savings between 55.5% and 63%.

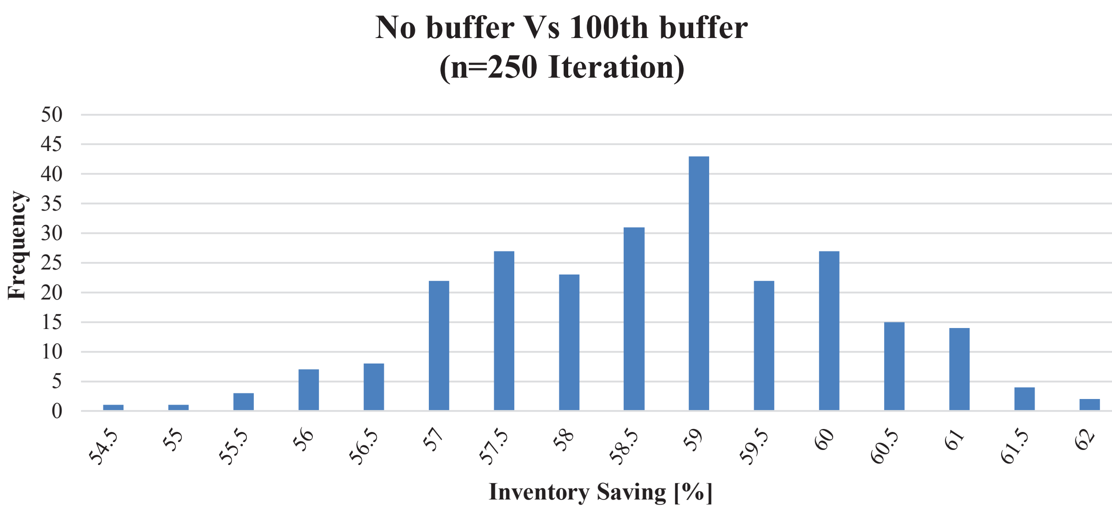

Figure 4.8. Relative inventory savings occurred using the optimal strategy

4.4.3 Phased Implementation

While the benefits of appropriate inventory buffers across all components required for final assembly are striking, firms may not have the initial capital and human resources necessary to transition all components at once to the new buffering strategy. It is important to understand that the implementation of time buffers will not only require immediate cash outflow, but also significant personnel effort. The procurement department will need to manage the expectations and behavior of suppliers who, at a whim, see all their requirements pulled up by a substantial amount of time. Under such constraints, what would be the right components to transition first? Does component prioritization have a significant impact on performance?

To answer these questions, we simulate the performance of the assembly system as more components are transitioned into the buffering strategy under three plausible component prioritization schemes. These schemes focus on identifying the components that have the largest impact on assembly delays and overall system instability. The prioritization schemes define the order in which components will be transitioned into the buffering strategy as follows:

1. *Average Days Late*: Components are ranked from highest to lowest average days late.

2. *Standard Deviation of Days Late*: Components are ranked from highest to lowest standard deviation of days late.

3. *Maximum Days Late*: Components are ranked from highest to lowest maximum days late (or, equivalently, 100th percentile of the delay distribution).

4. *No Sorting*: This is a base case for comparison.

Figure 4.9 shows the importance of carefully selecting the components to transition into the buffering strategy over time. Utilizing the wrong sequence would result in bloated inventories followed by lack of trust and probably abandonment of the buffering strategy altogether. Buffers are deployed sequentially for subsets of components. Deploying buffers for components with highest maximum delays first is by far the most effective strategy to reduce system inventory. Furthermore, as we have demonstrated earlier, low inventory and high delivery performance go hand in hand

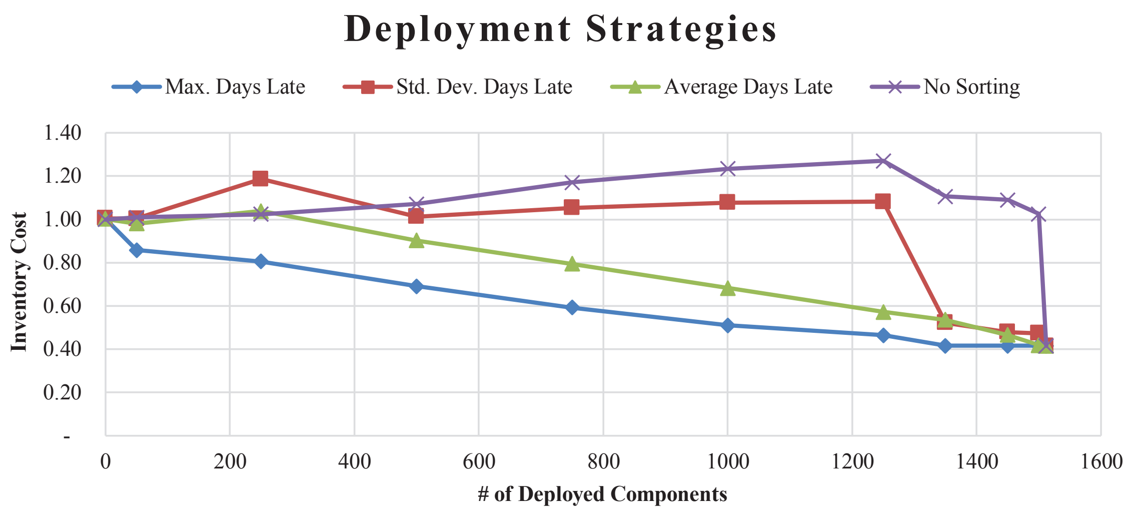

Figure 4.9. Inventory performance under different component prioritization strategies

in the assembly process. Thus, not only is work in process reduced but also on-time delivery fill rates are improved under this prioritization strategy. Components with highest maximum delays impact the performance of the system most severely.

# Deployed Components	Fill Rate: MRP	Fill Rate: 2 Week	Fill Rate: 4 Weeks	Normalized Inventory
0	0%	0%	0%	1.00
50	0%	0%	0%	0.86
250	0%	0%	0%	0.81
500	0%	0%	0%	0.69
750	0%	0%	100%	0.59
1000	0%	100%	100%	0.51
1250	0%	100%	100%	0.46
1350	100%	100%	100%	0.42
1450	100%	100%	100%	0.42
1500	100%	100%	100%	0.42
1512	100%	100%	100%	0.42

Table 4.7. Fill-rates observed for the deployment by maximum days late strategy

Table 4.7 shows the observed fill rates for the best deployment strategy (deploy by maximum days late). 100% fill rates are obtained before all components have transitioned to having time buffers since the remaining components experience good supply delivery behavior to start with.

4.4.4 Risk Analysis

Our recommendations and optimal buffer strategy are based on supplier delivery performance over the most recent 6-month period, as a perfect predictor of future performance. How will unexpected shifts in supply delivery performance affect overall system inventory? Will inventory bloat to the point that the system is worse off under synchronization than under the *No-Buffer* strategy?

In this section, we analyze how the system behaves when the underlying supplier performance changes and the delay shifts. For that purpose, we define three different scenarios:

1. *Scenario 1*: A single component delivery is delayed by 4, 8, or 12 weeks beyond its previously believed 100^{th} percentile of the delay distribution;

2. *Scenario 2*: Deliveries for a single component over 4, 8, or 12 consecutive weeks are delayed by 4 weeks beyond the 100^{th} percentile of the delay distribution;

3. *Scenario 3*: A percentage of components (1%, 3%, 5%, 10%, 15%, and 20%) experience random deliveries that are 4 weeks late beyond their 100^{th} percentile of the delay distribution.

We observe that in order for a delivery to be late by 4 extra weeks, the deliveries associated with orders over the following three weeks will be delayed by a corresponding 3, 2, and 1 week(s), respectively, since otherwise orders would cross and the 4-week delay would not materialize as such. The unexpected delay may lead to very different performance depending on the characteristics of the particular component experiencing the delay. To understand this effect, we rank all components based on their maximum delivery delay (100^{th} percentile) and classify them into *bad*, *medium*, and *good* delivery performance based on their ranking on the first, second, and last thirds, respectively. In testing Scenario 1 and Scenario 2, we select a single component from each of those categories. In Scenario 3, the delayed components are randomly selected.

4.4.5 Scenario 1

The lower line in Figure 4.10 shows how inventory costs increase when the *100%-Buffer* strategy is implemented as the delay of the single component grows. Costs are identical regardless of the type of delayed component (*good*, *medium*, or *bad*). In the unsynchronized *No-Buffer* case, on the other hand, a delayed *good* or *medium* component has very little effect on the observed inventory.

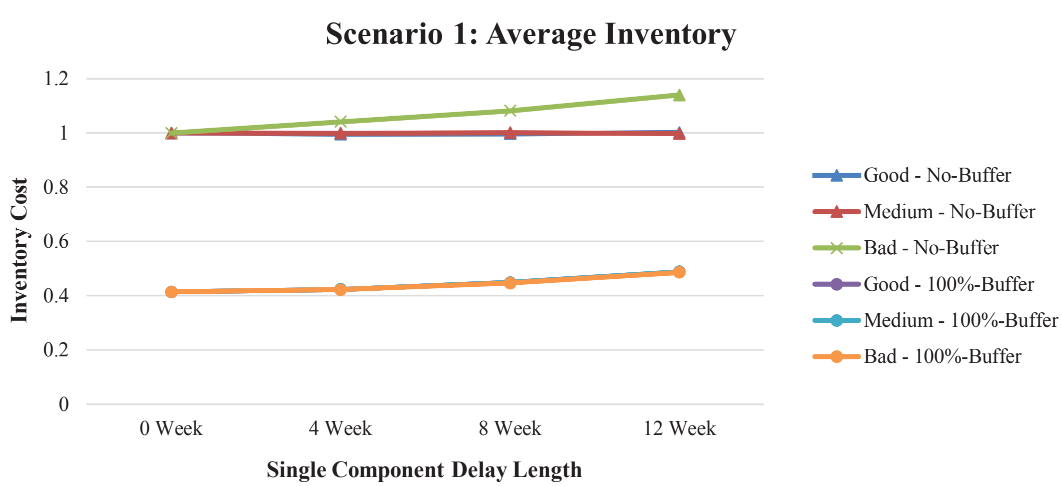

Figure 4.10. Inventory cost increase as a single component delay grows

This is because the assembly point marked by the last arriving component tends to happen much after the arrival of the delayed component. In fact, in this case, it is even possible to save inventory costs by having a *good* or *medium* component be late and thus closer to the actual assembly time. This is not the case for a *bad* component. The further delay of a *bad* component will cause it to be the ”pacing” component of the assembly and, hence, also delay the entire assembly further in the unsynchronized case.

4.4.6 Scenario 2

In Scenario 2, we assume that a 4-week delay will continue for consecutive orders over 4, 8, or 12 weeks. We observe the same effects as in Scenario 1 but an increased magnitude of inventory cost as the length of the interval with recurrent delays increases.

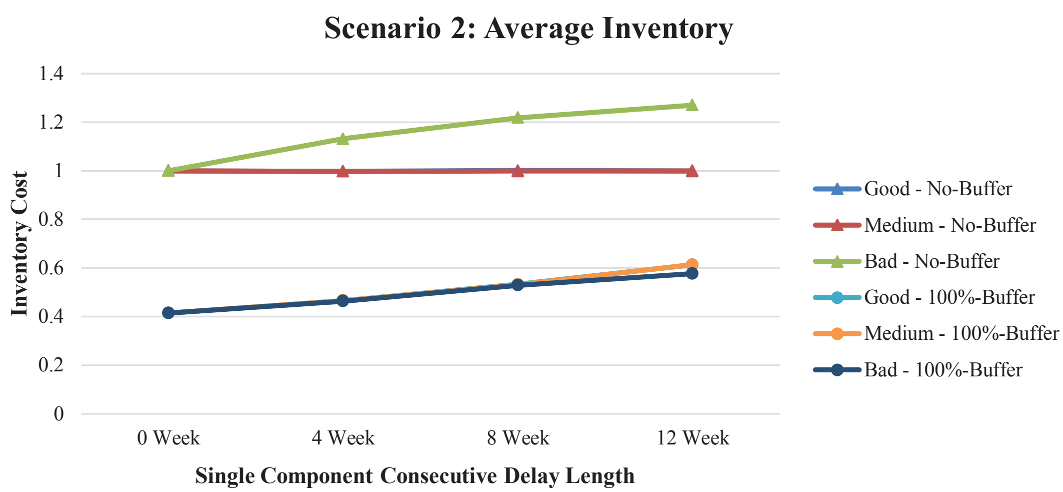

Figure 4.11. Inventory cost increase as single consecutive delay length grows

4.4.7 Scenario 3

Scenario 3 tests the performance of the system as a higher percentage of random components experiences a significant delay of 4-weeks beyond its worst case at random points in time over the year. Here again, in the unsynchronized case of *No-Buffer*, we can see that delayed components have little effect on the overall inventory costs. As shown in the previous scenarios, the results might be different depending on the characteristics of the delayed components. Delays of *bad* components will likely affect the final time of assembly (and therefore make all other components wait longer) while delays of *good* components will result in lower inventory as they arrive closer to the actual time of assembly. For the synchronized case of the *100%-Buffer*, any component delay (as seen in the other scenarios) has a significant effect on the time of assembly and therefore causes inventory to rise. Inventory levels stabilize after a significant percentage of the components observe a 4-week delay at a random time, since then delays are commonplace and having more than one component delayed at a point in time has little effect on inventory (in fact, it will reduce it).

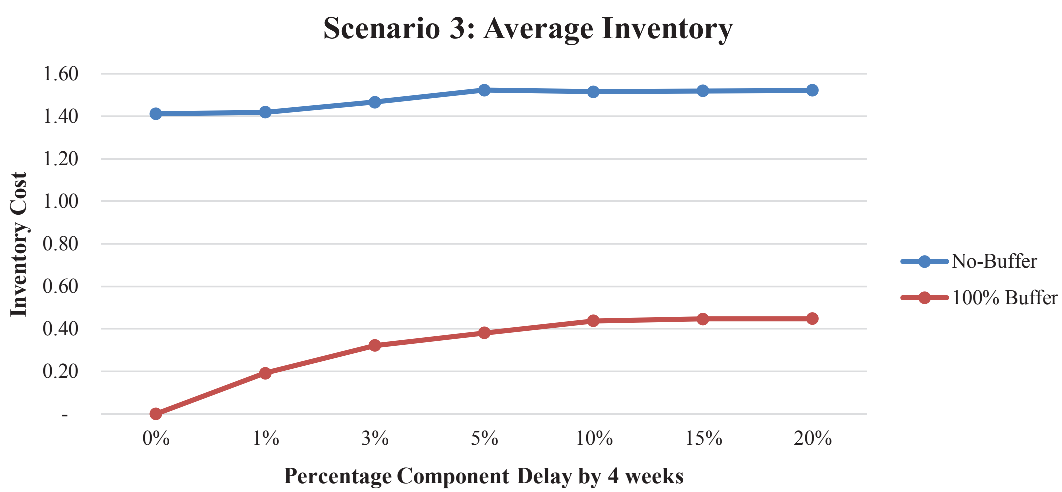

Figure 4.12. Inventory increase as a percentage of components getting delayed

The MRP fill rate in Scenario 3 depicts clearly how fast the system degrades when a synchronized system observes multiple components being delayed.

4.5 Discussion

4.5.1 Data Limitation

The most recent 6 months of supplier performance data are used to generate the component delay distributions. Nevertheless, the delays captured in this data set are

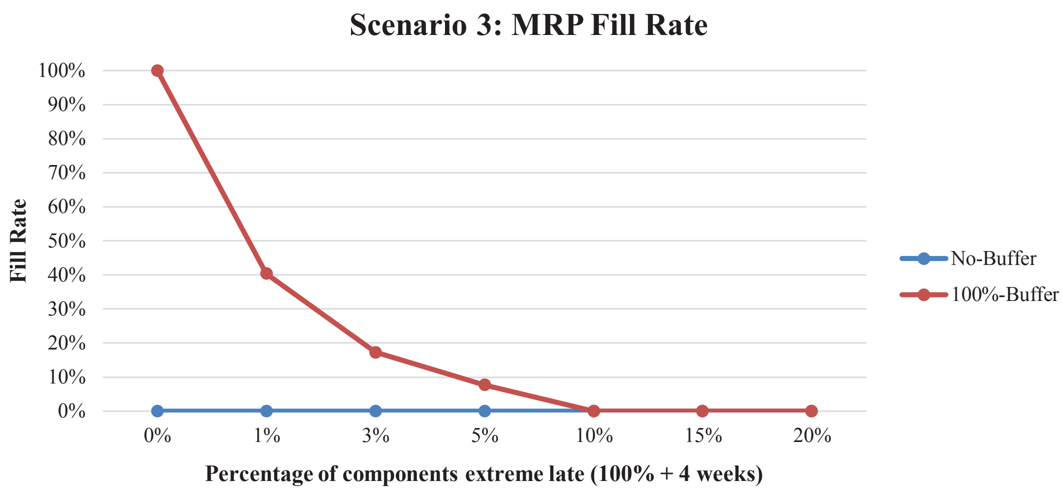

Figure 4.13. Fill rate decrease as a percentage of components getting delayed

naturally influenced by situation-dependent human intervention driven by the state of the system at any point in time. This factor may induce biases and result in an inaccurate picture of supplier lead time performance. The component delay for the next order is sequentially measured as the difference between the time delivery of the last unit occurs and the MRP date associated with that order. We present two examples of situations that unduly influence the delay distributions. First, the purchasing department may dynamically change supplier expectations over time. Advanced knowledge about a critical component delay may lead the buyer to allow other suppliers to delay their delivery resulting in a recorded delivery delay unreflective of the supplier's actual performance; the delay simply led to savings in inventory holding. Second, changes to the MRP demand are allowed within the quoted lead time making the order instantly late or early, again to no fault of the supplier. This is common for customer orders of spare components. For the above reasons we pruned data points that could be clearly identified as outliers (e.g., $L_j \geq 200$ days).

4.5.2 Behavioral Limitation

The presented approach shows that a smart ordering approach can result in tremendous savings. However, "cheating" the system by inflating lead times may also lead to a reverse effect. Plossl and Wight (1971), talk about it in the context of a firm's own production system:

> *"Putting in safety time really doesn't tell the system the truth...Priorities are distorted and by such cushions, work-in-process inventories are inflated and operating people soon learn that they have more time to get parts than the due dates indicate. The resulting 'credibility gap' can easily offset the benefits of having safety allowances."*

Suppliers may adapt to the inflated lead times and, hence, interpret delivery due dates differently and prioritize other customers. It is thus key for the OEM to manage supplier expectations proactively and firmly measure their performance against the advanced delivery times required under synchronization.

4.6 Conclusion

High-tech, low-volume industries struggle with the optimal management of component inventories. High inventory costs coupled with sub-optimal delivery performance are common, due to the large number of specialized, expensive components with long and variable lead times that constitute their products.

We have demonstrated that excess inventory accumulations occur when thousands of components that arrive wait for a highly delayed few in order to proceed to product assembly. A fully synchronized system achieved through time buffering provides both desired observed delivery performance and reduced inventory levels - an initially unexpected win-win situation. Furthermore, our time buffer optimization model showed that modest (1.58%) additional inventory savings could be achieved by lowering the time buffers for a handful of components.

Skepticism about the true savings, and an abundance of caution led us to carry out a comprehensive simulation study to characterize the dynamics of the inventory levels as component-level time buffer coverage was systematically increased over time, as well as when unforeseen delays occur. The three-pronged approach: Analytical derivations, simulation-based concept validations, and optimization, showed the robustness of the time buffering strategy in providing optimal delivery performance and inventory levels. To the best of our knowledge no existing article has shown the striking value of full supply synchronization under long and highly variable procurement lead times, the importance of choosing the right sequence of components when following a sequential implementation of time buffers, and the robustness of this strategy to changes in supplier behavior. We believe this work will be relevant for practitioners as well as future studies in this field. Future work will further extend the model to account for (1) stochastic demand of the end product, (2) the case of a spare part demand stream with aggregate service level constraint, and (3) the interaction of both production and spare part demand streams.

CHAPTER 5

SPARE PART INVENTORY MANAGEMENT WITH ADVANCED FLEET CONDITION INFORMATION

5.1 Introduction

The cost efficient management of spare parts is inherently difficult. The stochastic part life-time deterioration makes the prediction of needed maintenance timing and scope extremely challenging. This is particularly true in our chosen research area of jet engines. These engines consist of many expensive high-tech parts of low and intermittent demand volume, which make the holding of safety stock costly and risk-prone. They also incur high opportunity costs associated with the engine being held-up on ground unutilized during overhaul. Following the economic downturn of 2008/2009 which led to a build-up of tremendous amounts of inventory, the industry is seeking improved methods to cost efficiently manage and lower the associated risk of spare parts inventory.

Distributed sensors in jet engines in the field promise to have a significant positive impact in forecasting the uncertain future fleet demand of spare parts. Condition monitoring involves collecting real-time sensor information from a functioning device to make predictions regarding the health condition and lifetime of the unit. By aggregating over the condition of an entire engine fleet, this information not only promises improved maintenance scheduling but also better management of the resources needed - in particular spare parts. This is reflected in Peng, Dong, and Zuo (2010) who define condition-based maintenance (CBM) as:

> "Condition-based maintenance is a decision-making strategy to enable real-time diagnosis of impending failures and prognosis of future equip-

ment health, where the decision to perform maintenance is reached by observing the "condition" of the system and its components."

Our work is part of a multipronged and interdisciplinary study which seeks to research the methodologies necessary to utilize sensor readings from a large number of distributed working units as a reliable forecast parameters in spare part inventory policies for maintaining those units. The research consists of four key milestones (as outlined in the NSF Abstract #1301188):

1. *"Advancing sensing methods and the interpretation of signals to diagnose equipment condition"*.

2. *"Developing procedures for transforming these data into predictions of time-to-overhaul and resource-requirements"*.

3. *"Building part forecasting methods and inventory policies that aggregate this information across equipment, under consideration of field usage and economic conditions"*.

4. *"Creating a simulation tool for the monitoring and maintenance of a large fleet to validate the methodology"*.

This chapter focuses on the essential last milestone of this study and builds the simulation environment that validates, compares, and further optimizes the study's proposed methodologies and inventory policies. The ultimate objective is to highlight the economic value of advanced sensing techniques. Section 2 discusses the need and potential impact of the overall study. Section 3 broadly describes the aerospace industry, its supply chain, and complex maintenance operations. Section 4 reviews the most relevant literature on fleet management simulation and spare part inventory. Section 5 provides an overview of the simulation framework developed. Section 6 describes the many modules that comprise the simulation of such a holistic fleet management simulation, and highlights the challenges faced within each. Section 7 illustrates the output of the simulation through a case study. We end with a conclusion and a discussion of impact and limitations in Section 8.

5.2 Motivation

The aerospace market has undergone tremendous changes over the past thirty years. Flying emerged as a mass mode of transportation affordable for the majority of people in the western world and, hence, travel demand increased by almost 400% between 1981 and 2012 (Deloitte, 2014). This trend is persistent as globalization and migration are rising and flying is becoming increasingly affordable. Emerging markets like India, China, or Brazil, in particular, drive new demand for the market (Clearwater International, 2014). Hence, the industry is anticipating significant growth in the next twenty years. The aircraft fleet is expected to double, resulting in over 32,600 new passenger and freight aircrafts (single-aisle, twin-aisle, and very large aircrafts) (e.g., Clearwater International, 2014; Deloitte, 2014; Airbus, 2014). The beneficiaries are the actors in the aerospace supply chain which consists of (1) OEMs (e.g., the two biggest players in the market - Boeing and Airbus) taking on the design, manufacture, and assembly functions, (2) tier one suppliers (e.g., United Technologies, General Electric, or Rolls Royce) providing essential aircraft components like engines, flight control systems, or fuel systems, and (3) tier three suppliers manufacturing parts required by tier two suppliers. Historically, OEMs sell engines at cost or even below but have high markups on after-sale parts generating most of their earnings. The low margins on the product sale became even more prevalent in recent years when airlines pushed for lower product prices in favor of reduced product warranty. Hence, jet engine OEMs rely even more on after-market revenues but have faced the risk of disintermediation and, hence, losing revenue to suppliers that offer spare parts directly to their customers (C. L. Rossetti & Choi, 2008).

A major share of airplane MRO revenues (43%) is associated with the maintenance of the aircraft engine as engine components represent around 27% of the value of the aircraft (Clearwater International, 2011). Consequently, it is not surprising that MROs of aerospace jet engines is a key cost driver for commercial airlines and military fleets. In fact, for military fleets, it has been reported as taking up to 70% of all aircraft related costs accounting up to 10% of the total defense budget (McKinsey & Company, 2010).

The associated spare part inventory management for MRO services is critical to the availability of fleets but is inherently challenging for the following reasons: (1) A high fraction of parts are very capital intensive since jet engines not only push the limits on what is technically possible in terms of operational performance, but also provide a safety critical service that use the best material and design to ensure reliability. (2) Tier one suppliers need to account for a large number of distinct spare parts for which safety stock and inventory easily adds up to a large operating capital in that matter. For example, Mabert, Soni, and Campbell (2006) report that Pratt & Whitney stock more than 22,000 distinct parts. (3) While part costs are often very high, the demand rates are very low and often lumpy and intermittent which can lead to long and costly inventory holding. High demand service levels, however, are critical to keep the customers' fleet highly utilized. (4) The demand forecasting process is very challenging and distinguishes itself from most other products. Demand for these spare parts may result from actual part failure, operators' economic decision-making to procure and stock these parts, or decisions to perform maintenance out of schedule to accommodate lease agreements, optimize overall fleet operations, or take advantage of maintenance contract terms. (5) Fleet maintenance and the engine spare parts used are subject to macroeconomic and company specific business conditions. This causes expensive overhauls to be postponed creating a highly variable and unpredictable spare part demand stream at the tier one level. (6) Tier one suppliers experience the industry specific long and highly variable part procurement lead-times (compare to Chapter 4 of our study). In order to counteract the uncertainty on the demand and supply side, the implementation of highly expensive safety stocks is the common practice to prevent stock-outs.

These challenges can be further illustrated in the light of the most recent 2008/2009 economic downturn. As depicted on Figure 5.1, the available seat kilometers followed closely the world real GDP growth. Aircraft usage declined significantly and airlines were facing a significant loss of revenue. Accounting for the high cost of ownership of an airplane, deferring MROs helped airlines to reduce fleet costs. However, it led to significant inventory build-up of replacement parts impacting tier one and tier two suppliers significantly.

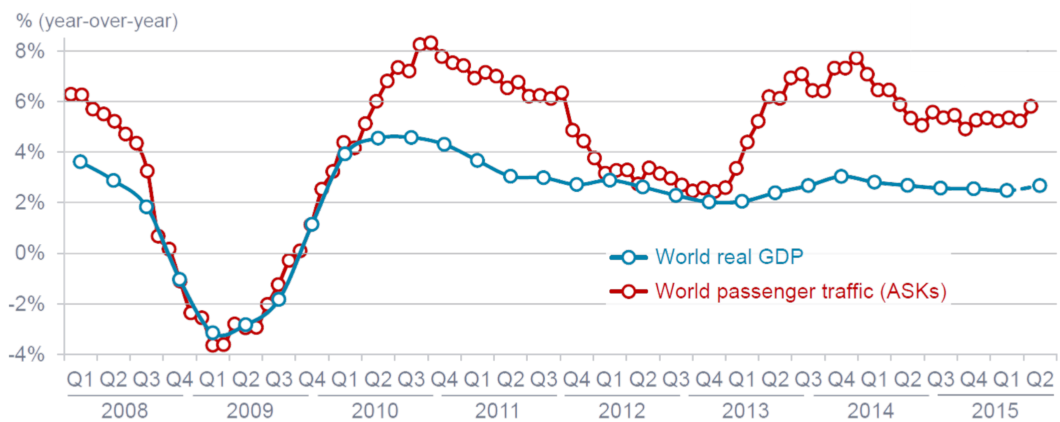

Figure 5.1. Available seat kilometer from 2008 to 2011 (Source: Airbus (2014))

In fact, the value of inventory of parts stocked at tier one, tier two, and small part suppliers has been estimated at $40 billion in 2010 (Clearwater International, 2011), a number that corresponds to the entire MRO market size in 2010 (Reals, 2010). In an industry that is already inherently difficult, this triggered companies to re-examine their operations and spare part inventory management. In this capital intensive environment, small improvements have significant financial impact.

Advanced aircraft engine condition monitoring and diagnostic technology is capable of transmitting real-time condition data to the ground during the flight and providing essential information on observed changes in temperature and pressure. This condition information can be used to identify degraded parts, physical faults, and the engines' damage propagation. The latter has been shown to be a complex task, but engine gas temperatures (EGT) readings have been used as the basis to model the degradation of engine condition (e.g., Saxena, Goebel, Simon, & Eklund, 2008). Linking all information from the fleet allows to not only build the analytics framework to create easy fleet condition reporting but, most of all, make better predictions, on overhaul schedules, engine specific predicted workscopes, and the spare parts needed at different points of time, possible. The need for this kind of integrated framework has been recognized in the industry. In fact, Capgemini (2009) lists predictive maintenance as one key challenge that companies need to overcome.

The results of this study are also of special interest for jet engine manufacturers or other service providers offering engine service contracts like 'Power-by-the-Hour'. These services are designed to guarantee asset availability to customers under a predefined fixed cost model helping airlines to build stable financial plans by lowering the risk of unexpected operational expenses. For this purpose, advanced sensor information helps service providers maximize profit and assure delivery of the promised available hours in a cost efficient way (see Nowicki, Kumar, Steudel, & Verma, 2008; Justin & Mavris, 2015).

In summary, the aerospace industry is expected to keep its strong economic position and is expected to grow substantially. Even with new jet engines becoming more efficient, optimized MRO operations are essential for airline operators to keep their expensive fleet highly utilized and free of disruptions. This particularly holds true for the defense industry where available aircrafts are indispensable for military operations and critical for the safety of its personnel. Engine manufactures rely on important aftermarket sales. Their challenge is to predict spare part demand to guarantee very high customer service levels while minimizing on-hand inventory. The 2008/2009 economic downturn has illustrated the importance of effective spare part inventory management even more. Our study seeks to provide an integrated solution that transforms sensor data and diagnostic output from sets of aircrafts into reliable forecasts and inventory policies for the parts required to repair those units. Therefore, sensor data must be translated into decisions regarding when to service given units and predict the corresponding spare parts needed for the generation of maintenance schedules, joint replacement policies, customer usage plans, and economic indicators.

5.3 Background: Aerospace Industry

In order to understand the underlying dynamics of the problem and the research that has been done in the past, we briefly introduce the reader to the aerospace supply chain and its context of spare parts and maintenance, repair, and overhaul.

5.3.1 Aerospace Supply Chain

The aerospace supply chain underwent significant changes over the past decades. Traditionally, the OEM was vertically integrated and acted as a centralized hub controlling and directing the majority of the supply chain (i.e., information and material flow). Hereby, the OEM led most of the transformational manufacturing and management of raw material procurement and inventory while suppliers acted in a supporting function to production (Bales et al., 2004). Competition increased with amplified globalization and led OEMs to outsource much of their in-house part production in favor of cheaper production overseas and focus on their core competence. The outsourcing peaked in the 1990s (C. Rossetti & Choi, 2005) and fostered increased risk sharing with all supply chain participants, imposing the OEMs to give up control towards a decentralized supply chain. Suppliers became more specialized and knowledgeable, hence, taking an active role (i.e., partnering) in developing complex parts and slowly moving into an expert role integrating design into their previously purely mandated manufacturing role (e.g., Lorell, 2000). Figure 5.2 illustrates the shift in manufacturing and design.

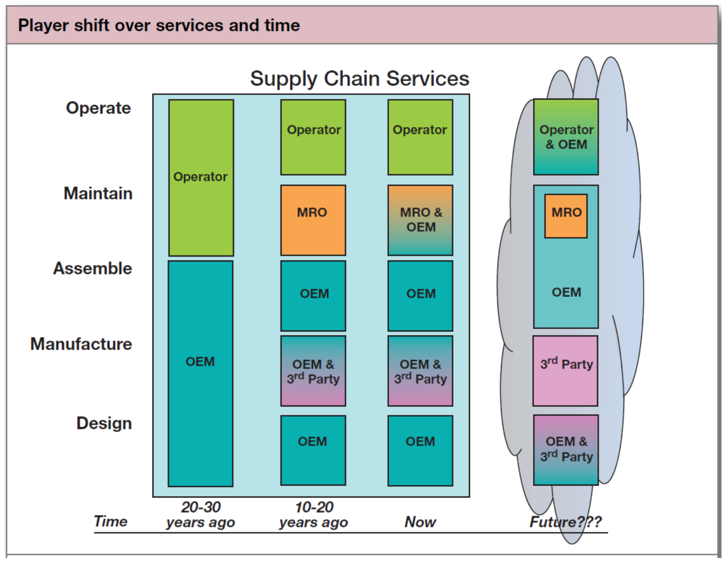

Figure 5.2. Player shift over service and time (Source: Capgemini (2009))

Most recently, the aerospace industry has seen examples where OEMs shift back in time and vertically integrate suppliers in-house in response to past problems of long supply chains. This helps them to tackle the expected growth in the industry requiring reliable, high-quality sourcing to successfully compete with other OEMs (Linebaugh, 2013).

5.3.2 Maintenance, Repair, and Overhaul

Maintenance, repair, and overhaul is a key cost driver for airline operators. Nevertheless, it is in the airlines' best interest to perform these MROs in the best possible manner in order to:

1. Keep engines in operational and reliable condition.

2. Retain their current and future value by physical deterioration.

3. Fulfill the regulatory requirements which specify maintenance and inspection standards.

New technological innovation makes new airplanes and engines become less failure prone and increasingly more efficient regarding time, frequency, and complexity of MROs (e.g., GE Aviation, 2010). However, with typical lifetimes of over 30+ years (GE Aviation, 2010), new airplanes and engines will only slowly replace the old. The growth in airplane fleets, with a mix of old and new aircraft, may very well outrun the newly gained efficiencies, ensuring increased demand for the MRO service industry.

In the past decades, responsibility for MRO services has shifted (see Figure 5.2). Historically, airline operators were responsible for most of the maintenance or their assets. Some developed an expertise in MRO service operations, optimized their business, and act today as service providers to other airlines (e.g., Delta TechOps). The *Airline Deregulation Act of 1978* fostered more competition between U.S. airlines. Some airlines started to focus on their core business by outsourcing and increasingly favoring outside MRO service (e.g., Lorell, 2000; Garg & Deshmukh, 2006; Wilkinson et al., 2009). As experts of their product, OEMs (e.g., Airbus or Boeing) and jet engine manufacturers started shifting towards integrating MRO services into their product portfolio helping airlines to manage and maintain their fleet under various business and service models (e.g., Power-by-the-Hour, Airbus Flight Hour Service, Boeing Edge). This not only serves as an additional revenue source, but also helps OEMs move closer to their customer, collect fleet information in the field (and thus enable optimization of MRO operations), underline their value offering, and protect their intellectual property rights (Lorell, 2000).

Figure 5.3 shows how MRO service providers change over the product life cycle. For new products, the OEMs still provide the warranty for the product and have the competitive advantage in expertise and spare parts sourcing.

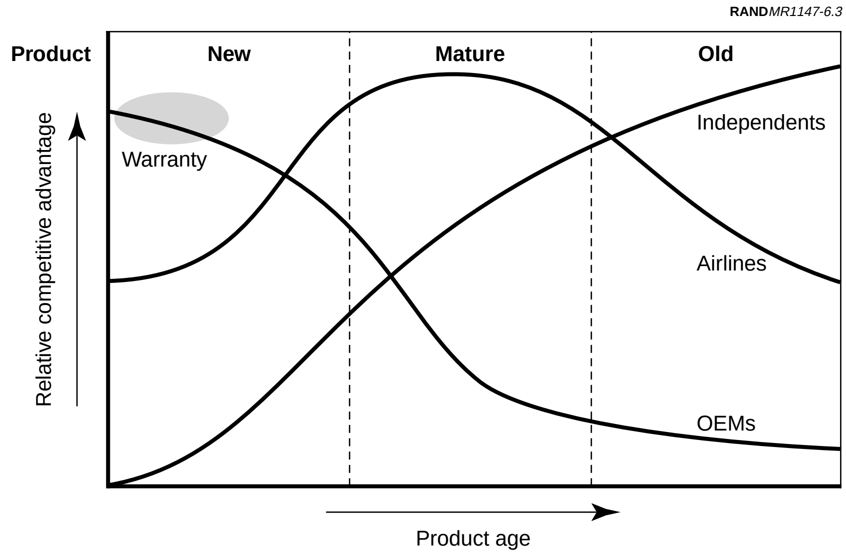

Figure 5.3. MRO service provider over product age (Source: Lorell (2000); Canaan Group (1996))

As the product matures, airlines may take over to use their own infrastructure for better availability, or make use of independent service providers. Independent specialized firms enter the market to take over MRO services as the product is phased out of production (Lorell, 2000). Today, only approximately 20% of aircraft maintenance is performed by U.S. carriers in-house versus 80% in the 1970-1980s (Clearwater International, 2011, 2014).

5.4 Literature Review

Our study requires a synthesis of the state of the art knowledge from multiple traditional management science research streams like demand forecasting, inventory management, and maintenance scheduling combined with insights from aerospace part degradation modeling and the interpretation of on-board condition monitoring sensor signals. There is a tremendous number of studies that take on the complex individual tasks comprising this study (or joint combination of some individual aspects). However, previous research that takes on the holistic approach of translating condition-based sensor readings into maintenance scheduling decisions and spare part inventory holding decision for fleets of aircraft engines is very sparse and this collaborative study is one of the first ones attempting it. General research in this area has been driven primarily by real industrial problems and practical incentives to solve them. Therefore, it is not surprising that many available studies are rooted in industry or defense environments. Due to confidentiality reasons, the details of such studies are often restricted for publication in regular scientific journals. Many studies we found, however, have been presented at conferences and were published in conference proceedings, often a preferred outlet for practitioners. Many authors implement their results in software packages due to their immediate practical relevance (e.g., Stranjak, Dutta, Ebden, Rogers, & Vytelingum, 2008). These case studies provide excellent insights but are often limited in revealing the full scope of technical details of the simulation and modeling portions of the research. Many spare part inventory management and sensor condition information aspects have often not been disclosed in detail in these studies. Nevertheless, there are a tremendous number of academic studies that tackle the individual important aspects of the problem.

In what follows, we first discuss a number of review papers from the past 20 years that provide a high-level perspective of the state of the art on the various problem areas within fleet MRO operations. These review papers provide the reader a reference and introduction to the complex areas that this study touches upon as well as more background information on existing challenges. We then focus on research papers in the area that is most related to our study: Simulation of fleet maintenance operations.

Pham and Wang (1996) provide an early review paper on imperfect maintenance and define the different types of repair that are most commonly used. The authors see imperfect maintenance models as most useful in practical applications and classify them into eight categories. Various concepts for imperfect maintenance have been developed by different studies helping to answer questions like (1) how does preventive repair affect the failure rate of the working unit? (2) How does the failure rate change over time and engine age? (3) How should one model individual part lifetimes? The authors review and summarize relevant studies from 1978 to 1996 and discuss existing types of reliability measures to model imperfect repair and find optimal maintenance policies. The review paper of Dekker (1996) focuses on maintenance optimization, which originated in the early sixties and has seen many studies and reviews since then. We highlight this review since it takes on an application-focused view by seeking to evaluate the value of maintenance optimization for management and its general relevance in practice. In doing so, the author focuses on studies that (I) provide tools to support maintenance optimization or (II) describe actual model application. Even though this study explicitly excludes spare parts, it provides insights into the practical importance of maintenance, its optimization, and implications to spare parts. The author found 112 studies (between 1969 and 1996) that include maintenance optimization with a relevant practical focus. Most of these studies were written in close academic cooperation with only a few originating from pure industry settings. The author finds that the early models of 'block and age replacement' have been investigated most often, with 'equipment overhaul' being the most popular application reaching over one quarter of the studies found. The author notes, however, that many studies use a tailor-made solution indicating the need to account for the application specific environment. The study highlights several existing challenges for maintenance optimization which include the (I) complexity of available solutions, (II) diversity of existing maintenance problems and application areas which make the generic application of current models difficult, (III) need to formulate problems, models, and decision support systems in a manner that leaves no space for misinterpretation (e.g., regarding assumptions), and (IV) need for available deterioration and failure data which must be collected under strict rules to reflect the true system. Furthermore, the author discusses the gap between theory and practice highlighting

(1) the complexity of studies, which lack in offering a general model and are difficult to understand, (2) existing papers that are *"written for math purposes only"* and do not offer solutions to real problems, (3) companies not being interested in publishing making it more difficult for academics to be exposed to industry problems, (4) optimization which might not always be necessary since benefits may not outweigh its cost (i.e., application complexity), and (5) existing optimization models that not always focus on the most effective maintenance type. Even though significant problems exist and the practical impact of maintenance optimization has been limited, the author describes the future prospects of maintenance optimization as optimistic based on the technical evolution that will solve some of the challenges. The continuous deployment of technology will make efficient maintenance in the future even more important.

H. Wang (2002) reviews and surveys maintenance policies of deteriorating systems to provide a classification scheme of existing maintenance models. Popular policies include *"age replacement, random age replacement, block replacement, periodic preventive maintenance, failure limit, sequential preventive maintenance, repair cost limit, repair time limit, repair number counting, reference time policy, mixed age policy, preparedness maintenance policy, group maintenance policy, and opportunistic maintenance policy"*. Generally, the study distinguishes between single and multi-unit systems but focuses on the former. Garg and Deshmukh (2006) review 142 papers classifying maintenance literature into the six areas of (I) maintenance optimization models, (II) techniques, (III) scheduling, (IV) performance measurement, (V) information systems, and (VI) policies. Thereby, the authors subcategorize and map the six areas to relevant methodologies and subtopics before discussing relevant studies in each subcategory. Jardine, Lin, and Banjevic (2006) review condition-based maintenance and discuss the three modeling steps necessary to utilize condition monitoring information with the objective to reduce unnecessary preventive maintenance operations. These are (I) data acquisition, (II) data processing, and (III) maintenance decision support (e.g., remaining useful life and monitoring interval). The authors also discuss studies that consider multiple sensors for which data needs to be consolidated (i.e., data fusion) to predict the overall health of the system. Finally, the study concludes with a discussion of the research needed to enhance the effectiveness of condition monitoring, in which the *"establishment of efficient validation approaches"* is explicitly mentioned.

Other literature reviews focus on the aspect of multi-part systems and its implications for maintenance. Multi-part systems may have dependencies that have previously been categorized to be of economic, structural, or stochastic in nature. For the case of aircrafts, economic dependencies refer to the economies of scale effects that occur when multiple components (e.g., landing gear and engine) are maintained at the same time, aiming to minimize costs associated with the time that an aircraft is on ground. For an aircraft engine, this refers to the simultaneous maintenance of not only defective parts but also those that are considered to be still in reliable condition (for joint replenishment optimization refer to the work of Sun, Zhao, Luh, and Tomastik (2004, 2008); Tu, Luh, Zhao, and Tomastik (2004)). Once a specific section of the engine is opened-up, simultaneous maintenance saves future downtime and set-up costs. Similarly, there are parts for which maintenance of one will automatically lead to the maintenance of others due to structural dependencies. Lastly, stochastic dependence in our context refers to either (I) parts whose health state influences the remaining useful life (RUL) distribution of other parts or (II) external influence that correlate the RUL of parts (e.g., weather conditions). The consideration and modeling of these dependencies are generally very complex, especially when considered together. We refer to the literature reviews of Van der Duyn Schouten (1996), Dekker, Wildeman, and Van der Duyn Schouten (1997), and Nicolai and Dekker (2008) for a closer look. Peng et al. (2010) provide the most recent literature review on machine prognostic and condition based maintenance for different application domains. The authors classify studies into (I) physical model-based, (II) knowledge-based (e.g., expert opinion or fuzzy logic), and (III) data-driven methodologies (e.g., multivariate statistics or neural networks) as well as (IV) models that utilize multiple approaches. Sharma, Yadava, and Deshmukh (2011) focus their review on optimization models in maintenance operations providing a good classification of existing literature based on used optimization criteria. They also give an overview of case studies that are driven by real data. The authors conclude that there is a need for studies that are able to evaluate and optimize costs for different combinations of maintenance modeling approaches. The use of simulation models to optimize maintenance operations is mentioned as an emerging trend. This is echoed by Alrabghi and Tiwari (2013) who review simulation-based optimization in maintenance operations. The authors' results indicate that discrete event simulation combined with genetic algorithms for optimization is the dominant technique. The study highlights that the majority of maintenance studies consider the manufacturing industry (i.e., machines) and only few consider operational products (e.g., jet engines) in the field. Furthermore, the authors conclude that mathematical models are limited in capturing the problem complexity and, therefore, simulation is the preferred methodology. They discuss the lack of a framework to evaluate different maintenance policies combined with advanced optimization methods as well as the shortage of real life case studies verifying existing models.

Kennedy, Patterson, and Fredendall (2002) review spare part inventory management, highlighting the unique aspects of spare parts relative to regular products, and the resulting management challenges. The authors first discuss these challenges before discussing specific research areas. One of these research streams assumes fixed age-based replacement policies and investigate the joint optimization of age replacement and spare part ordering decision. Another research stream focuses on multi-echelon contexts in different network configurations studying the optimal placement of spare parts in the network, their appropriate inventory levels, or the location of maintenance facilities that service the items. The authors also discuss studies considering repairable items, which significantly impacts the spare parts orders needed over time to guarantee high spare availability. Smith and Babai (2011) review spare parts forecasting specifically for bootstrapping. The authors introduce the historic background of bootstrapping before discussing methods for spare part forecasting. In their conclusions for further research, the authors indicate that there is an opportunity to review commercially available best practice software and compare academic models to their results. Commercial software has been shown to be very effective. For example, the company MCA Solutions (today merged with PTC, Inc.) reported that their software helped to *"decrease spare parts inventory by up to 66%, increase first-time fill rates by up to 26%, and drive service levels up to 98%"* (BusinessWire, 2011).

The remainder of the literature review takes the perspective of a centralized actor that stocks spare part inventory and possibly manages MRO operations for large fleets of aircrafts engines aiming to utilize real-time condition monitoring to set the optimal spare part inventory stocking levels. We emphasize studies that simulate the complex management of fleet maintenance and overhaul operations while highlighting the connection to condition monitoring and spare part inventory management. Traditionally, these two research streams pertained to two different entities (airline and OEM, respectively) and, therefore, have most often been either considered separately or sequentially (Sarker & Haque, 2000; Elwany & Gebraeel, 2008; W. Wang & Syntetos, 2011). However, maintenance operations and spare parts inventory management are closely correlated (i.e., once maintenance, and its scope is scheduled, the spare part demand can be estimated) and therefore need to be considered simultaneously. In light of today's dual offering of life-cycle maintenance contracts and spare part by aircraft engine OEMs, the integration of both streams has become more relevant than ever.

Several studies simulated fleets for better decision making in maintenance scheduling and operations. In collaboration with Bombardier, Gharbi, Girard, Pellerin, and Villeneuve (1997) seek to optimize a yearlong maintenance program for the Canadian fighter CF-18. The program consists of a maintenance schedule for a fleet of aircraft consisting of major repair and overhaul projects defined by a planned start and end date. Each project consists of different work scopes represented as a network of work steps that consume some of the limited overall resources. Traditional maintenance scheduling is used to optimize resource utilization. In their case study, historically, this resulted in fill rates of only 0-40% of all project end dates. The authors attribute this fact to the variability caused by changes in work scope authorizations, deterministically estimated work-step durations, and unplanned maintenance activities which delayed and shifted resources between different projects (amongst other reasons). The study's proposed simulation model uses the initial production plan (status quo) as input to their model and delivers an improved final production schedule as output. The authors incorporate stochastic elements like resource constraints, unpredicted failures, or variability in work durations, and allow for specific user input to create an updated feasible production plan. The authors successfully validate their model using historical data, and find the model to not only produce superior production plans, but also provide high value in performing 'what-if' analysis, helping to negotiate maintenance aspects with customers. The authors only provide very limited discussion of the technical details of their simulation. Gatland, Yang, and Buxton (1997) approach a similar capacity and facility loading problem for a fleet of engines and provide insights from Delta Airlines who perform maintenance not only for their own fleet, but also for the fleet of other (smaller) airlines. This practice known as insourcing creates a new problem in having multiple customers with different priorities competing for the same resources. To understand the capacity of the facility, the authors built an ARENA simulation model which analyzes the impact of varying

(1) engine removal times, (2) engine disassembly start times, (3) disassembly work schedules, and (4) engine workscope mix. KPIs considered in the study include engine turntime, engine throughput, engine service level, machine utilization, personal utilization, part turntime, part throughput, average throughput, and work-in-progress. The industry funded study of Stranjak et al. (2008) uses a multi-agent approach to the problem of overhaul prediction and scheduling to navigate the competing objectives of minimizing operational maintenance costs and decreasing waiting times. In addition, the model also accounts for unforeseen events and strategic decision-making (i.e., investments in resources like overhaul capacity and spare engines). The authors' software platform tackles the four application areas of *"(1) multi-agent negotiation for scheduling and adaptive re-scheduling, (2) modeling whole engine reliability, (3) response to unforeseen events, and (4) post-analysis of stored performance data."* The agents that are being modeled are (I) fleet manager, (II) fleet planner, and (III) different overhaul bases. The reliability of a whole engine is being approximated by the Weibull function by aggregating the part specific probability distributions. Using the function's scale and shape parameters, the authors distinguish between different life stages of the engine and its types of disruptions (i.e., infantile, random, and wear-out) and schedule overhauls as close as possible to their predicted optimal overhaul date under capacity restrictions. An algorithm minimizes the distance to the optimal overhaul date for overlapping schedules. Engines are being swapped for spares at maintenance events and sent to overhaul base locations. Besides utilization, turnaround times, and aircraft-on-ground occurrences, the authors also capture the impact of the number of spare engines available. The simulation tool is built using Java combined with the JADE agent platform. Painter, Erraguntla, Hogg Jr, and Beachkofski (2006) use Arena to simulate fleets of engines using military mission profiles. The authors are specifically interested in estimating the long-term cost effects (i.e., life-cycle costs) of maintenance policy decisions influencing key performance in-

dicators like expected time-on-wing, cost-per-engine flight hour, and the operational fleet availability. The authors argue that historic data is a bad estimator for future costs. Instead, they develop a simulation coupled with data mining techniques to, first, generate a data set of maintenance history and cost statistics, and, then, build a life-cycle cost model that uses appropriate static and non-static cost estimation parameters (e.g., mission profiles or operating environments). The authors fit field data of aircraft engine maintenance history and reliability characteristics to Weibull distributions to model failure modes and resulting maintenance requirements for different modules. Statistical sampling determines scope, timing, and location of failure. Data-mining, regression, classification, and clustering techniques are used to identify key life-cycle limit cost drivers. Mattila, Virtanen, and Raivio (2008) also simulate flight missions and model the maintenance of fighter aircrafts for the Finish Air Force under normal and conflict situation conditions with the objective to improve decision-making in fleet maintenance operations. In their Arena model, the author considers three types of maintenance needs: (I) periodic maintenance (model criteria are cumulative flight hours and predetermined service intervals), (II) failure repair (modeled as time between failures), and (III) battle damage (type of damage modeled as pass-fail probabilities). The configuration of fleet and maintenance operations constitutes the simulation input, and aircraft availability, maintenance, and flight performance statistics are generated as output. Maintenance network locations are considered and incorporated into their model. Due to confidentiality reasons of material handling data, spare parts are not considered in this study. However, historical statistical data and subject matter experts were available to define the probability density function for (I) the time between failures, (II) duration for each type of repair and maintenance, (III) time between flight missions, (IV) duration of a mission, (V) probability for each failure type, and (VI) maintenance requirements for each type of maintenance or repair. The author highlights the challenge of the scarcity of data, its confidentiality, and the important insights of subject matter experts to overcome some of the unknown factors of the model.

In summary, we can see that the research areas this study incorporates are expansive and multifaceted. At the same time, little work has been done to integrate all research areas, model the entire process from condition monitoring to spare part inventory management and evaluate the value of condition monitoring sensors. This chapter's objective is to build on the work of our collaborators and contribute to fill that gap. We found several studies that simulate fleets of aircrafts or engines providing insights to our study. Nevertheless, many of these studies have been adapted to a specific industry setting or are limited in the details that are provided.

5.5 Simulation Framework

This chapter addresses the fourth and last milestone of the collaborative study described in the introduction, whose ultimate objective is to provide a methodology to use the stream of condition information collected by sensors from units in the field to improve spare part inventory control and, consequently, the management of fleet maintenance operations. In particular, this chapter focuses on the (1) development of the theoretical simulation framework, (2) practical implementation of the simulation environment, and (3) identification of future research on basis of the simulation developed herein. We are using an agile and iterative design approach and continuously refine, add features, and test for simulation performance based on the feedback of our industrial partner. The latter is an essential requirement to ensure that assumptions and input data reflect reality. Furthermore, the integration of multiple research areas, and the many practical decisions made in the field, requires simplification and focus on the most influential modules relevant to our objective. Only then computationally feasible and meaningful results can be achieved that are not tampered by the complexity of the system.

Generally, sensor data must be translated into probabilistic information regarding:

1. When the engine will be overhauled.
2. What modules will be included in the workscope.
3. What parts will need replacement.

In practice, the final decisions are very hard since they involve various decision makers and must be made in accordance with (1) spare engine locations and availability, (2) maintenance schedules and associated capacity constraints, (3) joint replacement policies, (4) economic indicators, (5) FAA regulations, (6) airline route schedules, (7) airline corporate strategies (e.g., cash flow), and (8) maintenance service contracts. Many of these aspects have been approached as individual research studies and simulation frameworks (see Literature Review).

Previous studies have used multi-agent approaches for fleet maintenance operations in specific industry settings. Our approach allows for the addition of the most relevant agents in the system but focuses on evaluating and optimizing the system performance on a macro level. Based on the growing popularity of new contracts types (e.g., Power-by-the-Hour) that shift MRO responsibility to the engine OEM, the role of individual agents has sharply decreased in recent years, which makes a centralized decision-making approach a good approximation.

The following takes a high-level view on the simulation framework. The modules of the discrete-event simulation environment that seeks to assess the value of incorporating fleet sensor information into spare part inventory management will be described in detail in the next section.

We consider main input parameters in four major categories: (1) economic indicators, (2) engine profile, (3) cost parameters, and (4) sensor information. Figure 5.4 depicts these four categories and their dependence. The figure highlights specific parameters we consider most relevant in this study. In the engine profile, we use flight cycles as the key parameter affecting the engine condition since airplane starts put most stress on the engine and are the major driver of degradation. We assign a status to each engine reflecting its activity in the fleet (e.g., spare engine or currently in overhaul).

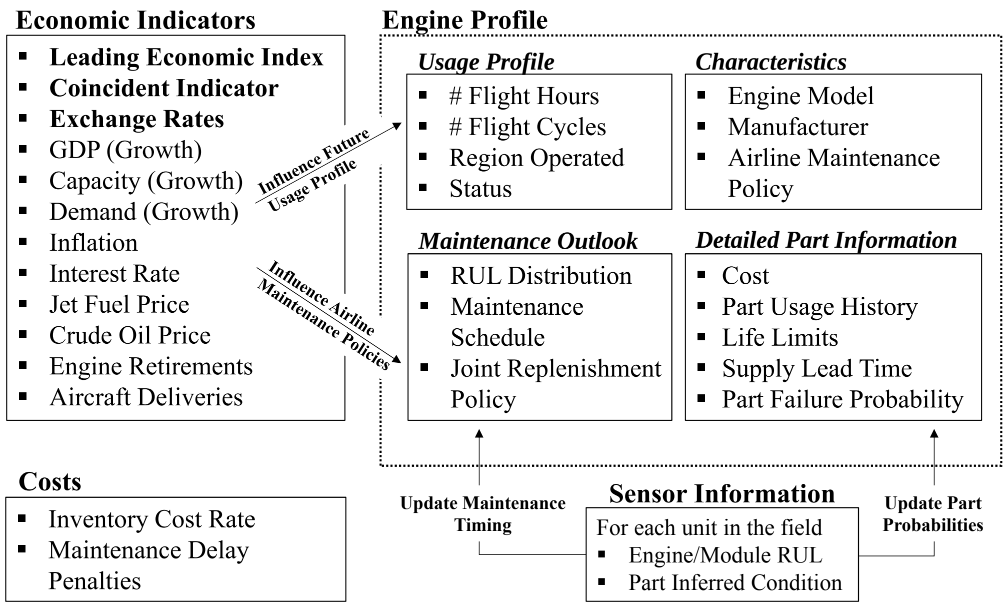

Figure 5.4. Simulation input

Parameter values of the engine profile and sensor information are updated in a weekly interval. Economic indicators change slowly over time and have a longer-term influence on flight cycles flown and airline behavior. Hence, economic indicators follow a slower quarterly time interval.

Figure 5.5 shows the macro outline of our simulation framework that seeks to compare the traditional approach to our proposed condition-based inventory control and fleet management process.

After an initialization step, we simulate the fleet operation for each time-step (i.e., each week) and generate RUL distributions for individual engines, and potentially their modules and parts, based on condition information readings and the economic forecasting model. Sensor readings are transformed into probability distributions of

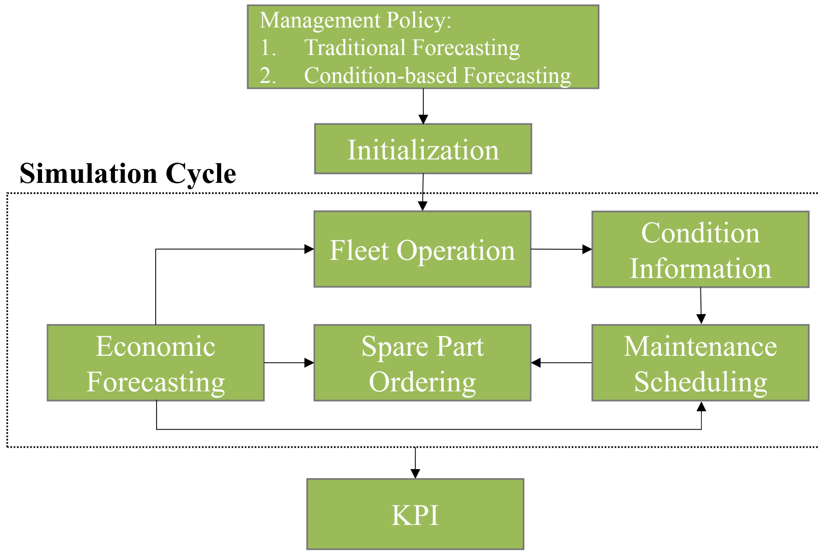

Figure 5.5. Simulation framework

the number of cycles (i.e., flights for commercial engines) until overhaul; economic conditions are used to predict engine usage, the number of cycles per week. The distribution of spare part demand over its lead time is required to compute the weekly order quantity. This is done by aggregating the probability distribution of demand during the part lead time over all the engines in the fleet. Comparing the performance of traditional versus condition-based policies allows to quantify the value of incorporating the fleet condition into the spare part forecast.

As illustrated in Figure 5.6, we update the engine profile, part inventory, and, quarterly, economic indicators in each simulation period (week). In the figure, the black color indicates externally defined and given parameters. The production schedule is set by the jet engine manufacturer and the engine usage profile is driven by airline policies and economic condition. In our simulation, we generate random engine cycle usage profiles for each engine that corresponds to the economic profile and overall fleet flight cycles observed. Individual airline policies, contract types, and maintenance regulations influence how maintenance operations are scheduled. Engine availability is determined by maintenance events and part failure in every period. The dark green color indicates key modules for which models are needed. These modules are described in Section 5.6.

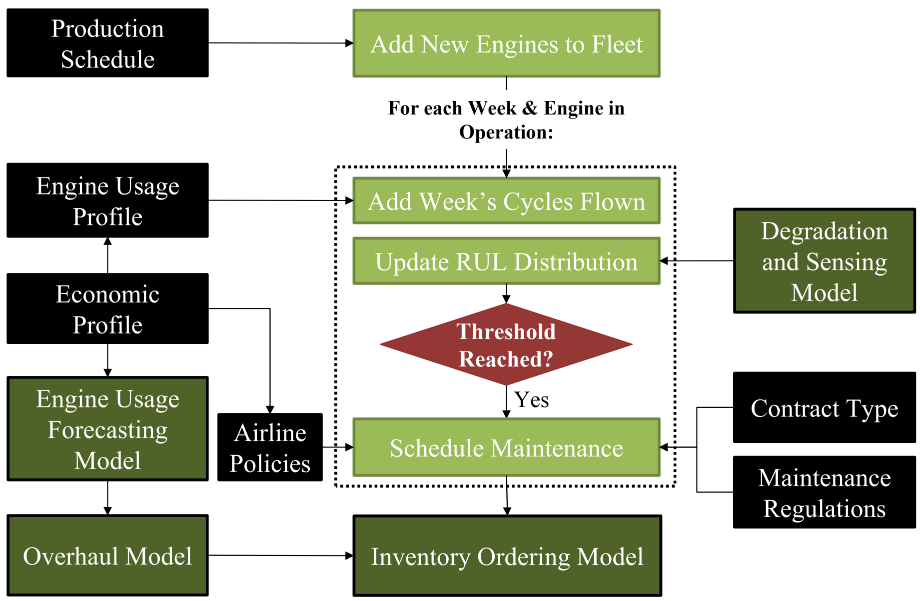

Figure 5.6. Simulation cycle

The hybrid sensing class will estimate engines' RUL based on sensor readings simulated following actual degradation data. Maintenance regulations, contract type, and economic outlook influence the maintenance schedule. The spare part ordering class will order according to predefined policies under available condition information.

To schedule MRO operations, we define workscopes on the engine and module level. Each workscope has an associated spare part demand (distribution). Maintenance overhaul operations increase engines' health index. An engine will then follow a different degradation path. Figure 5.7 provides insight into the simulation's modeling of a maintenance event. Each event may have limited resources (e.g., spare parts or mechanics) and may infer delay. We focus on maintenance events that require engine removal and an available spare engine as replacement. Swapped engines and its defective parts are then being repaired and returned into inventory for reuse, or scrapped.

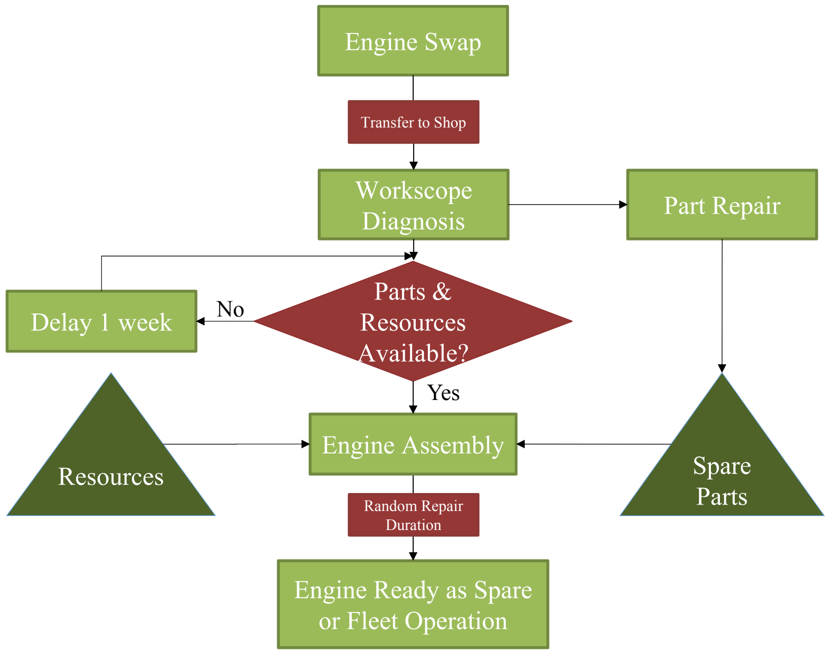

Figure 5.7. Simulation maintenance event

Figure 5.8 illustrate an example of a engine life cycle. The following provides a description of the major stages our simulation is currently considering:

1. The engine is being produced and entered into service. We add the engine to the existing fleet of engines and assign random weekly flight cycle usage. The engine stays in service until an overhaul is scheduled due to abrupt fault, specified part life limit, or regular wear and tear. The duration of the engine in service is highlighted in the graph as a solid line.

2. The engine has reached the overhaul criteria. This is a prespecified threshold defined as either a fixed number of flight cycles remaining with a specified probability or an individual part life limit. This triggers a spare engine into service and sends the engine into overhaul.

3. The engine has reached the end of overhaul. The maintenance duration is randomly assigned and might further be increased in weekly increments when parts or resources are not available. The different overhaul duration is highlighted in our figure as dotted line. The engine is added to the pool of spare engines.

4. The engine switches its status from spare to service and replaces an incoming engine to be overhauled. The duration in the pool of spare engines is highlighted as straight dotted line.

5-8. These steps are identical to steps 1-4, and the cycle will be repeated until the engine is retired. However, duration and timing might change significantly based on evolving system conditions and the stochasticity in each simulated period.

9. The engine reached a specified threshold in age measured as number of lifetime flight cycles and is retired.

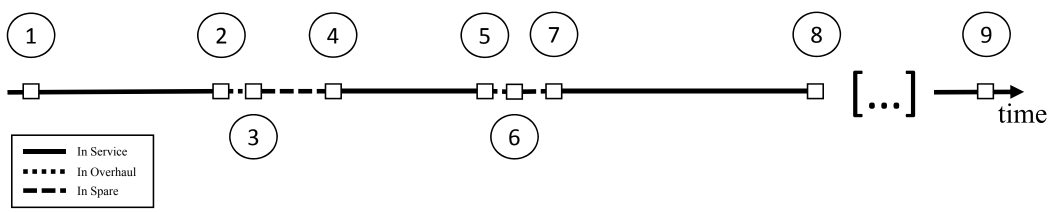

Figure 5.8. Engine life cycle example

5.6 Simulation Modules

The simulation is developed in Matlab 2017a environment to seamlessly interface with the previous work of our collaborators (Milestone 1-3 in introduction). The following describes the key modules that are initially considered in our simulation.

5.6.1 Engine Usage Forecasting Model

As Figure 5.1 illustrates, the total number of flight cycles flown in a given time frame is strongly correlated with the economic conditions that the plane is operated in. In economic tough times, airlines have incentives to delay any costly MRO spending, operate planes on different routes or keep them on the ground. An overall decline in flight activity and the postponement of overhauls has direct impact on maintenance operations and spare part inventory needed in a given period. Hence, it is crucial to include the economic environment in the simulation model.

Figure 5.4 lists the most relevant economic indexes as well as key indicators for the airline industry available to us. The economic forecasting model seeks to forecast flight usage measured in flight cycles per month by using the available economic data as predictor variables in a regression model. Therefore, we need to select the economic indexes which predict the future total flight cycles flown best without overfitting the data.

Feature Selection

Using the pool of economic and airline data, we seek to predict the total number of flight cycles for a given quarter. We update these predictions only on a quarterly basis since economic conditions represent a high level view and change slowly over time. For our simplified forecasting model, we select the most relevant subset of the economic indicators (features) on-hand. We use Matlab's R2017a Statistics and Machines Learning Toolbox and its sequential feature selection function *sequentialfs* which builds pools of subsets and sequentially adds and test features. Random partitioning of training and test sets as well as tenfold cross-validations assure statistical validity (please refer to Mathwork's online documentation for further details).

Regression Model with ARIMA Time Series Errors

Once the most reliable subset of features has been selected, a regression model is built using Matlab's Econometrics Toolbox. More specifically, we create a regression model with ARIMA time series errors (*regARIMA* class). This class allows to estimate regression coefficients, forecast future flight cycles, and automatically calculate confidence intervals. Furthermore, this class allows to account for the flight cycle typical seasonality while also testing time lags of the features used for the forecast.

Figure 5.9 shows preliminary results on predicting four quarters ahead. In the study, we apply the above described feature selection algorithm resulting to the following subset of variables for forecasting the number of flight cycles:

- Worldwide GDP
- Number of Installed Engines
- Worldwide Rate of Inflation
- OECD Composite Leading Indicator (MEI)

The ARIMA model applied to the selected features results in a Mean Absolute Percentage Error (MAPE) of 2.79%.

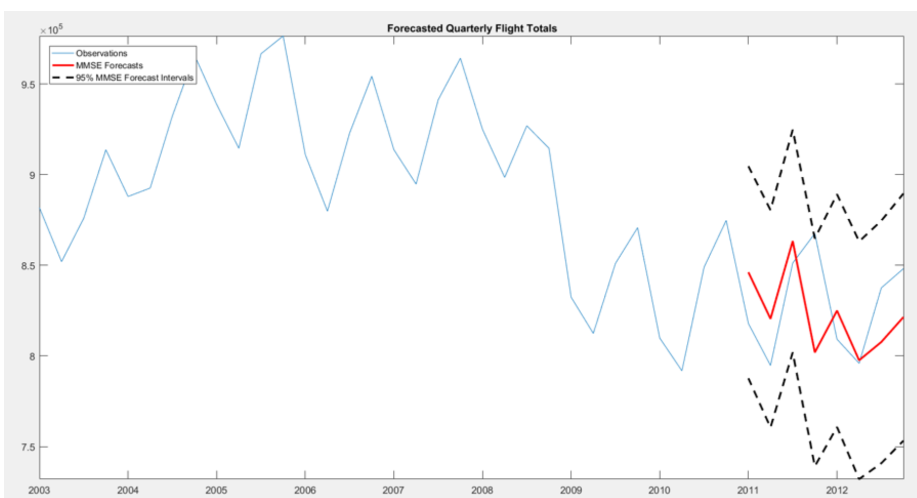

Figure 5.9. ARIMA regression model

The engine usage forecasting model is used to predict the number of flight cycles flown by each engine over a particular time period. In the simulation, we use this information to calculate the number of flight cycles flown during the part lead time for each engine and spare part modeled. The projected flight cycles flown value is an input to the inventory ordering model and used to calculate the probability of the engine failing during the part lead time. Hence, it influences the number of spare parts ordered in each period.

5.6.2 Degradation and Sensing Model

As a starting point, our current study uses sensor information to determine RUL distributions at engine level. Future development in degradation modeling of the various modules and major parts will be incorporated later to generate RUL distributions at the module and part levels. Modeling engine modules and parts is a long-term effort and a key driver to future progress in our research context since it has the potential to refine the individual forecast of workscopes and associated spare parts needed during an overhaul processes. Figure 5.10 illustrates typical modules considered and current sensors available in jet engines today (adapted from Gao and Wang (2015)). Condition information from additional sensors may be needed to capture the higher granularity in RUL distribution on part level. Our simulation framework can be used to assess the value of additional condition information through new sensors which is one of the long term research objectives.

Various methodologies and techniques for health and condition monitoring are currently investigated by the research community as prerequisite for reliable RUL predictions. Most recent work of our collaborators includes using deep convolutional neural networks for health monitoring and fault classification (P. Wang, Yan, & Gao, 2017), automated performance tracking (P. Wang & Gao, 2017), as well as Bayesian approaches and particle filtering techniques for wear predictions and lifetime estima-

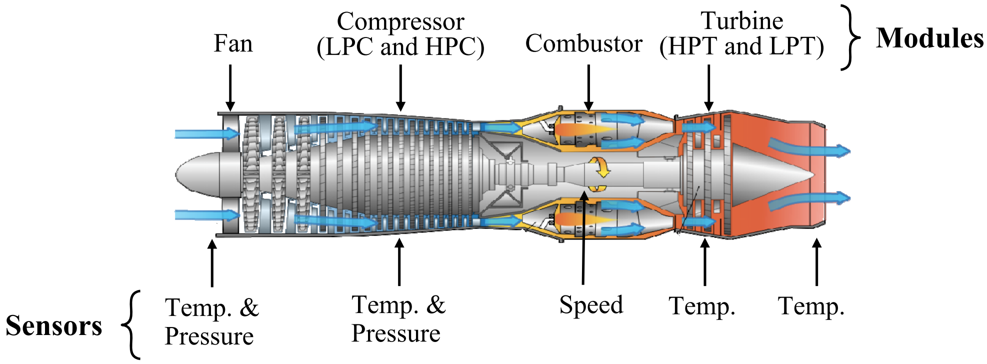

Figure 5.10. Engine modules and sensors (adapted from Gao and Wang (2015))

tion (P. Wang & Gao, 2016, 2015; J. Wang, Wang, & Gao, 2015). Applying these techniques to real jet engine sensor data allows to estimate overhaul scopes, associated parts needed, and the time of the overhaul. A detailed description of the work of our collaborators on deriving RUL distributions is beyond the scope of this chapter. However, we want to provide a high level description to provide an understanding for RUL distribution and its derivation.

Generally, the gas path analysis aims to detect physical faults in a part (e.g., fan, compressor, or turbine) which caused changes in performance (i.e., efficiency or flow capacity) producing changes in measurable parameters (i.e., pressures, temperatures, or speeds). The analysis can estimate the state (i.e., efficiency) of a given part based on the sensed parameters. The posterior distribution for the state (i.e., efficiency) is estimated using particle filter Bayesian approaches on the weekly updated observable parameters.

The particle filter requires a model describing the evolution of the state (i.e., efficiency), x_k , over time:

$$x_k = f_k(x_{k-1}, \theta_k, v_k)$$

for which θ_k are model parameters to be estimated and v_k the process noise.

Furthermore, the particle filter requires a measurement model relating the observable measurements (i.e., pressures, temperatures or speeds), z_k, to state (efficiency) x_k:

$$z_k = h_k(x_k, w_k)$$

for which w_k is the measurement error.

Particle filtering uses a set of random samples (i.e., particles) with associated weights to construct a posterior distribution for model parameters and system state. In order to predict the RUL at a time t_k, samples with estimated model parameters θ_k and x_k are necessary. Each sample (i.e., particle) is propagated using the state evolution model $x_k = f_k(x_{k-1}, \theta_k, v_k)$ with no further updating of the model parameters. This lets us determine when each particle's state reaches an exogenously specified failure threshold. These values can then be used to compute the probability distribution for the remaining useful life of parts or the overall engine:

- $RUL_{ie}(t)$: Remaining useful life for part i on engine e as estimated at time t.

Hence, in every simulation cycle, we can use the engine's specific RUL distribution and calculate the probability $P(RUL_{ie}(t) \leq \tau)$ that part i will be required for engine e at time of next maintenance event, τ, as estimated at time $t \leq \tau$.

Figure 5.11 illustrates how sensor data is continuously measured and used to estimate the RUL of the engine. The left figure illustrates the measurement of sensor data over 100 flight cycles. Although measurements naturally are very noisy (as seen in the Figure), a trend can be observed. Using the previous measurement, a future path can be predicted. The noise and uncertainty in the system provides multiple paths results in a range of possible remaining flight cycles. This notion is further illustrated on the right side where different paths generate a RUL distribution with an expected number of remaining flight cycles.

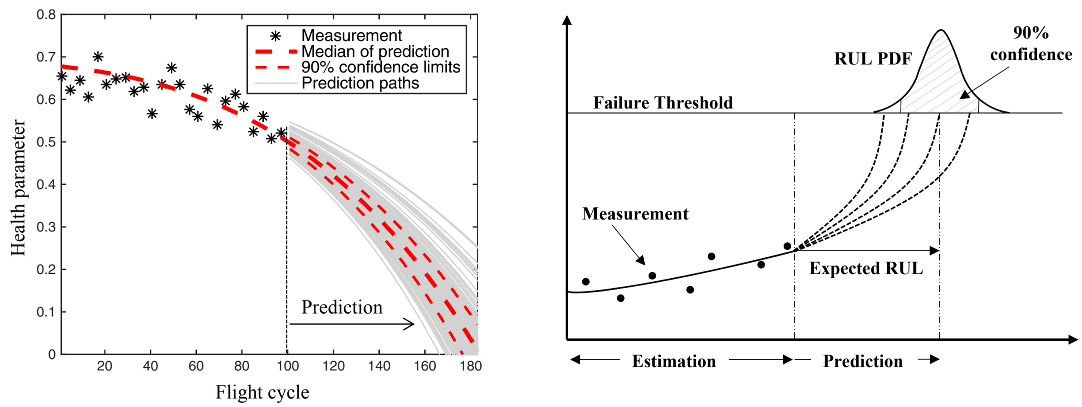

Figure 5.11. Remaining useful life estimation

Between MRO operations, each engine follows a specific EGT degradation curve. Some engines may experience abrupt flight disruptions (e.g., bird strikes), which are captured by a sharp step decline in the degradation curve. RUL is defined as the number of cycles until the EGT reading reaches a certain threshold. Sensor readings and the subsequent RUL predictions are updated weekly according to the number of flight cycles flown in the corresponding week. In the general case, the first module or part that reaches its threshold or life limit defines the next overhaul.

In the simulation environment, every engine is assigned a certain degradation profile. Based on that profile, sensor information is available after each flight (or cycle) and the particle filtering method produces a set of 500 particle predictions of RUL, which are then used to build the RUL distribution at that point in the engine's life. In each simulation cycle, we check the health status of the engine and schedule the engine to overhaul when (i) the probability of a remaining useful life of the engine is below 150 flight cycles with a probability of $P \geq 0.95$ or (ii) the life-limit of a part has been reached.

5.6.3 Overhaul Model

In practice, overhauls are highly complex and uncertain. Even when engine information is available, it might not reach all actors involved in the planning and execution of the maintenance operation. Different maintenance workscopes are defined and assigned to engines to predict the resources and parts needed before more information is obtained when the engine reaches the repair shop and is opened. In our simulation we model different part types. Parts are modeled with the following properties: (1) Part Type, (2) Part ID, (3) Life Limit, (4) Cost, (5) Lead Time, (6) Part Failure Probability, and (7) Flight Cycles Flown. The latter allows to model defective parts to be repaired and introduced back to the fleet while keeping track of the part life limit. For engines, we model (1) Engine ID, (2) Time of Entry Into Service, (3) Flight Cycles Flown per Week, (4) Total Life Time Flight Cycles Flown, (5) Total Flight Cycles since Last Overhaul, (6) Overhaul History, (7) Part List, (8) RUL Distribution, and (9) Engine Status. For the latter we distinguish between 1 = in service, 2 = in shop, 3 = spare, and 4 = retired (see engine logic flow diagram in Figure 5.12).

For overhaul operations, we initially start with two policies that determine which parts are exchanged during an overhaul:

1. *Deterministic Policy:* Exchanges a predefined set of parts in every overhaul, and

2. *Random Policy:* Uses a random number generator and part specific probabilities of failing to determine which parts are being exchanged.

For each engine, we record all overhauls and parts exchanges for KPI calculation.

5.6.4 Inventory Ordering Model

The spare parts inventory ordering model uses the condition of the individual units in the field to provide an aggregate view of the distribution of part demand over

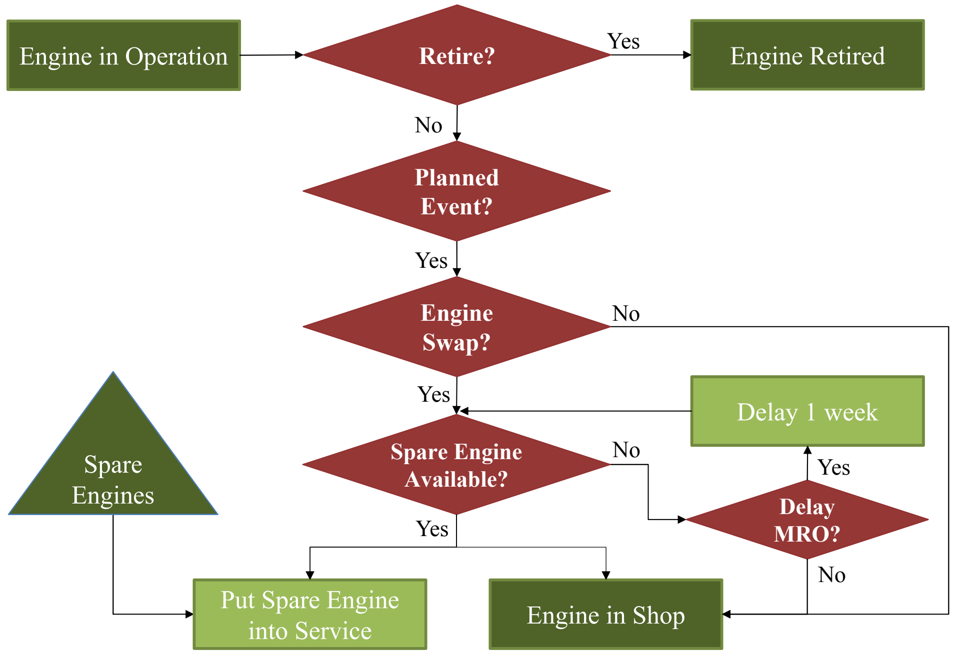

Figure 5.12. Engine logic flow diagram

the uncertainty period (lead time plus reorder interval) and generate a base-stock level as follows:

1. For each particular engine e and time t:

   (a) The sensing module provides a distribution of the number of cycles remaining useful life of the engine.

   (b) This distribution is transformed into a distribution of remaining useful life $RUL_e(t)$ in calendar time, using the predictions on cycles flown per week for that particular engine.

   (c) For a part i with lead time L_i, the probability of engine e requiring part i over the part lead time plus the reorder interval (1 week) can then be determined as $P_{ei}(t) = P[RUL_e(t) < L_i + 1] * p_{ei}$, where p_{ei} is the probability of part i being required for the overhaul.

2. The aggregate demand for part i over the lead time plus reorder interval, that is over the relevant uncertainty period $[t, t + L_i + 1]$ can then be approximated by a normal distribution with mean $\mu = \sum_e P_{ei}(t)$ and standard deviation $\sigma = \sum_e P_{ei}(t) * (1 - P_{ei}(t))$.

3. The base-stock of order-up-to level for part i at time t required to achieve a desired service level α is given by

$$S_i(t) = \sum_e P_{ei}(t) + z_\alpha * \sum_e P_{ei}(t) * (1 - P_{ei}(t)),$$

where z_α is the standard normal safety factor.

This provides a basic inventory model built upon the aggregation of the condition information of units in the field. A major thrust in the future work is to improve upon this model. In particular, the current inventory ordering model considers each part in isolation and thus ignores the assemble-to-order nature of overhaul operations, where all required spare parts in that specific overhaul need to be there to proceed with the re-assembly of the product. Therefore, we could formulate our inventory problem as a large-scale, multi-product assemble-to-order problem subject to order fill rate constraints with non-stationary demands. This notion is further illustrated in Figure 5.13.

In ATO systems, orders are received for final products that require multiple parts. The number of units of a part i required to assemble one unit of product p may be a random variable and orders cannot be filled until all required parts are available. In our setting, an order corresponds to an overhaul where each overhaul follows a prespecified overhaul policy. The latter specifies which parts will be inspected and the probabilities that inspected parts may need replacement. The overhaul cannot be completed until all identified parts are available.

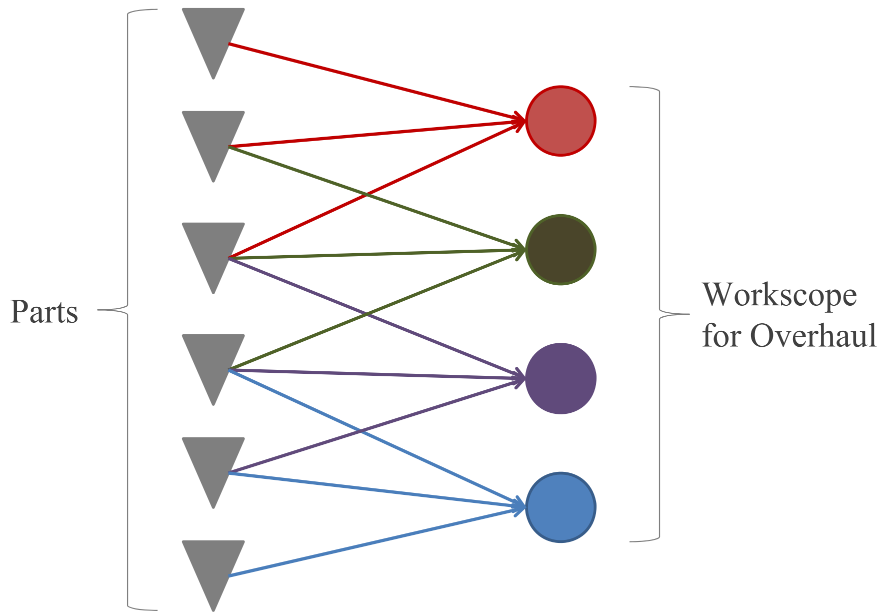

Figure 5.13. ATO principle applied to parts and workscopes for overhaul

The ATO model would use a base-stock policy for each part set to meet a desired aggregate order fill-rate. The order fill-rate only considers the fraction of orders for which all needed parts can be provided within the specified time window without delay. In the ATO literature, the demand process is generally assumed to be stationary. Our setting, however, involves a demand process that is continuously changing based on the condition of the fleet of engines as they age and accumulate flying hours.

5.6.5 Key Performance Indicators

To assess the performance of the proposed inventory policy, we use the following key performance indicators:

System Performance

- *Inventory Cost:* The overall cost incurred for parts held in inventory over the simulation period.

- *Part Fill Rate:* Percent demand for individual spare parts, as needed in maintenance operations, satisfied directly from stock.

- *Part Induced Delay:* The average delay in engine overhaul that an individual part caused during maintenance operations.

- *Fill Rate of Spare Engines:* Percent demand for spare engines satisfied directly from stock, to replace engines in the field grounded for maintenance operations.

- *Maintenance Fill Rate:* Percent of engine overhauls for which all spare parts are directly available from stock.

- *Average Engine Maintenance Delay:* Average engine delay during maintenance operations.

- *Average Delay of Delayed Engines:* Average engine delay, considering only engines delayed during maintenance operations.

Policy Comparison and Optimization

Using the information on system performance, we can characterize the value of condition monitoring by comparing (i) demand-based stock levels and (ii) condition-based stock levels. Furthermore, we can iteratively refine and test both inventory policies as we better understand their impact on overall fleet management performance. Finally, we can assess and characterize the value of placing additional sensors and virtual sensing methods.

Contract Comparison

Two contract types are of interest in this context. First, fleet-hour agreements (FHA) for which airlines buy fixed-price service agreements from the engine OEM. For this contract type, the OEM is responsible for service operations and guarantees fixed engine availability. Incorporating fleet condition information allows for better overhaul and spare part ordering decisions. Second, these newer FHA agreements can be compared to the traditional time and material (T&M) agreements for which airlines dictate the overhaul schedule while the OEM is responsible for providing the needed parts. For both contracts types, the simulation makes it possible to assess and compare the value of spare part inventory management under fleet condition information.

5.7 Case Study and Results

In this section, we describe a simple case study carried out to highlight the power and capabilities of the current version of the simulation. Increased detail of the complex industry environment and improved decision-making models are still in the works but are beyond the scope of this dissertation. As an example, we simulate four parts and a limited number of engines, and then address scalability issues to incorporate a higher number of critical parts. We use the following Lead Time (LT in weeks), Cost (C), and Start Inventory (SI) characteristics:

- *Part 1:* 5 (LT) — 2000 (C) — 10 (SI)
- *Part 2:* 13 (LT) — 2000 (C) — 15 (SI)
- *Part 3:* 26 (LT) — 1000 (C) — 20 (SI)
- *Part 4:* 36 (LT) — 1000 (C) — 30 (SI)

The simulation runs over a period of 20 years (1040 weeks) and simulates weekly engine production of two engines for the first ten years, i.e., from week one to week 520. In addition, spare engines are introduced to the fleet on a continuous basis to fulfill a 10% requirement of spare engines in the fleet. Each engine is assigned to a fixed degradation profile, and associated noisy sensor readings (see Figure 5.14 on the left). 10% of the engines are assigned a fixed degradation profile that includes an abrupt fault (see Figure 5.14 on the right). The number of flight cycles flown are randomly assigned to each engine and assumed to be constant in the short-term, and only newly assigned after either (i) new economic conditions are incorporated each quarter of a year, or (ii) the engine finishes maintenance operations and is installed in a new aircraft. The assigned flight cycles reflect the U.S. flight cycle numbers for passenger airplanes in the years 2003 to the end of 2012 available from the United States Department of Transportation. We repeat the 10-year flight cycle two times to reflect the simulation length of 20 years. Engine retirement age is set to 40,000 flight cycles.

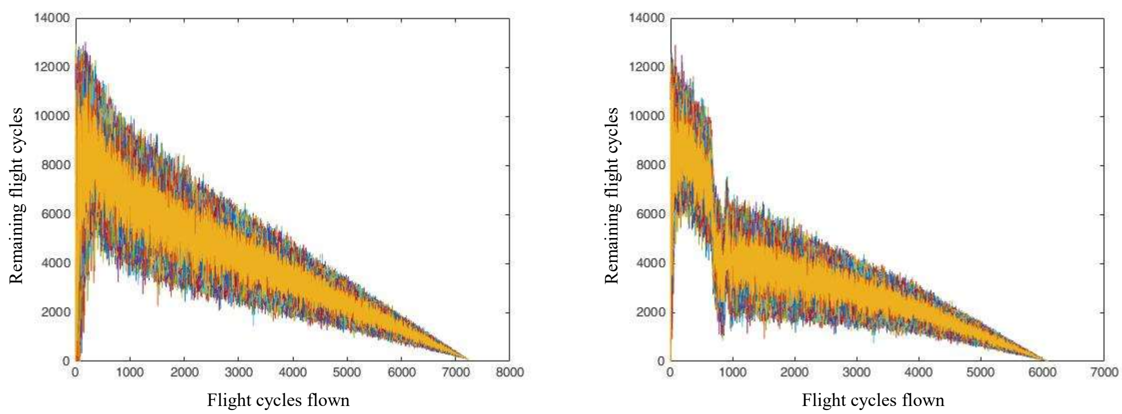

Figure 5.14. Case study RUL distribution

Durations of maintenance operations are randomly assigned according to industry expert knowledge. Parts are required to be in physical inventory three weeks before the scheduled end of the random maintenance durations. This reflects the time that is needed to reassemble the engine after the required replacement parts are available. Initially, we consider a policy that requires all parts to be exchanged at each maintenance point of time. A missing part delays the maintenance by one additional week until all parts are available. Spare engines are installed in the fleet once an engine requires an overhaul. In this case study, we apply the inventory ordering model described in Section 5.6, but use the actual flight cycles flown during the part lead time, rather than forecasting the flight cycles using the economic conditions observed at the inventory ordering point. This is due to the limitations to our current economic data input. With additional data on-hand, this assumption can be easily be removed to include the uncertainty in flight cycles flown and their prediction. Please see Appendices B, C, D for data sources, assumptions, system architecture, and pseudocode of the simulation.

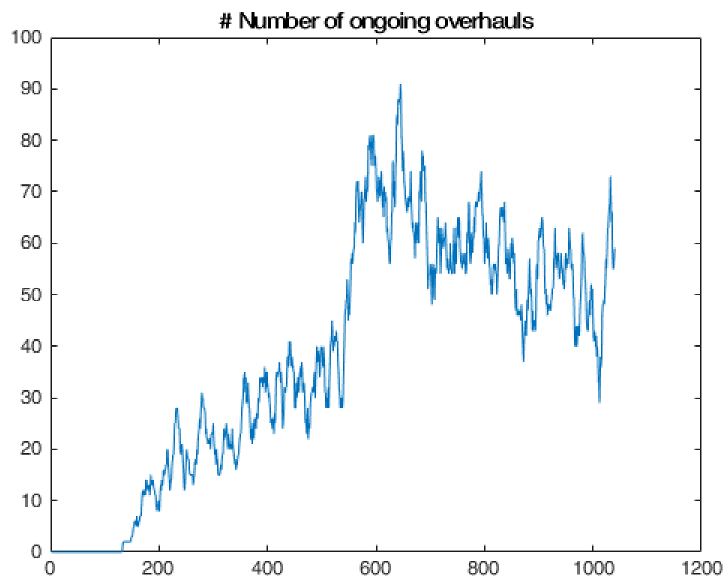

Figure 5.15. Number of ongoing overhauls (y-axis) over simulation length (x-axis)

Figure 5.15 shows the total number of overhauls in each period over the simulation length. New engines are introduced to the fleet at a rate of two starting period 1 (until period 520). The figure shows how overhaul operations only start around period 150. Different RUL distributions and flight cycles flown influence the random time of the overhaul. A sudden spike of overhauls can be seen at period 580. Older engines requiring their second major overhaul start to overlap with newer engines requiring their first overhaul. Observe that we are not modeling the end of the engine programs life cycle, since most engines do not reach their life-span limits within the 20 years simulated. As a result, the number of overhauls is still close to its peak in the final weeks of the simulation.

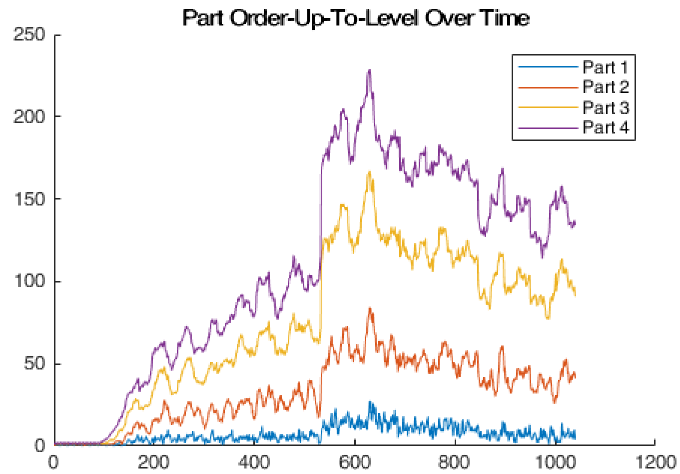

Figure 5.16. Part order-up-to level (y-axis) over simulation length (x-axis)

Following the number of overhauls at any point in time, we can see in Figure 5.16 that our condition-based inventory order-up-to methodology supports the trend in number of overhauls correctly. The order-up-to level at a given point in time also reflects the lead time of the part. Parts 3 and 4 with the highest lead time have a higher inventory buffer than Parts 1 and 2 with shorter lead time. In this example, we specified a part service level of 80% for the calculation of the base-stock level with the normal distribution approximation.

The physical on-hand inventory shows high fluctuations as seen in Figure 5.17. Orders arrive after their deterministic lead time and inventory is depleted according to the number of engines that require maintenance operation at a given point in time. We start with an arbitrary initial inventory. The figure shows how inventory is depleted when the number of maintenance operations start in period 150. Maintenance duration is randomly distributed between 6 and 27 weeks with the highest probability associated with the interval between 9 and 11 weeks. Inventory only reaches zero twice in this simulation run in approximately periods 190 and 650.

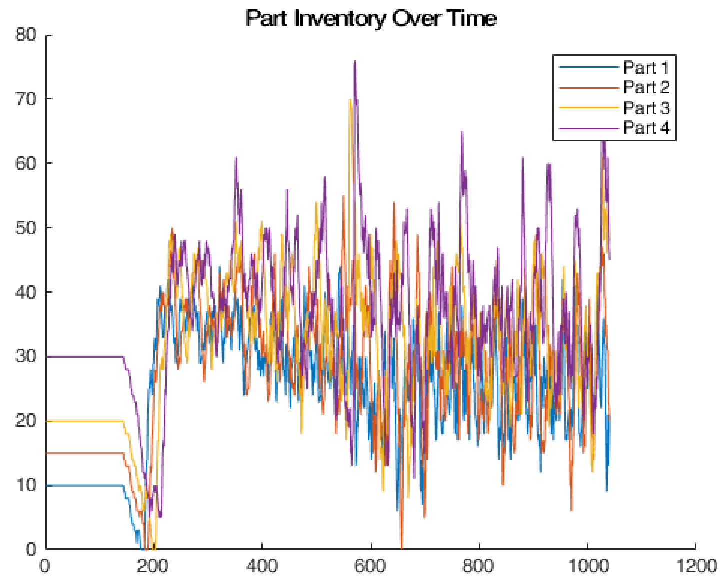

Figure 5.17. Part inventory (y-axis) over simulation length (x-axis)

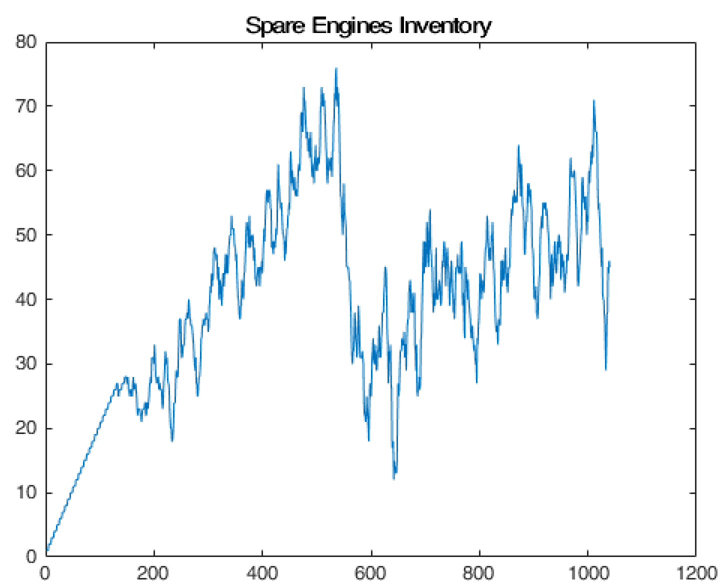

Figure 5.18. Spare engine inventory (y-axis) over simulation length (x-axis)

Lastly, we picture the inventory of spare engines over the run time in Figure 5.18. While no maintenance operations are performed in the early weeks, the number of spare engines is growing according to the specified 10% spare engine requirement. The figure clearly shows that the 10% spare engine production requirement overestimates the actual need for spare engines. The sudden spike seen in Figure 5.15 is reflected here in a sudden decline, since every incoming engine requiring maintenance is replaced by a spare engine. Engines finished with maintenance operations are added to the spare engine pool, explaining the growth in spare engines even after engine production is stopped in period 520.

The uncertainty in the timing of maintenance operations and the ensuing difficulty in planning for spare parts, maintenance resources, and spare engines underlines the value of incorporating condition information in engine fleet management.

Other key performance measures include:

- Number of Overhauls: 2999
- Spare Engine Fill Rate: 100%
- Average Engine Maintenance Delay: 0.014 weeks
- Average Delay of Delayed Engines: 2.625 weeks
- Maintenance Fill Rate: 99.60%
- Average Inventory of Spare Engines: 39
- Part Fill Rate: 99.00%

The fill rate reported is the average over the simulation period. Observe, this includes the initial 150 weeks where little to no engine failure occurs and spare parts are set to a relatively high starting inventory position. This explains the high part fill rate observed.

5.8 Conclusion

This is one of the first studies that integrate multiple research streams (sensing, degradation modeling, RUL predictions, economic conditions, part forecasting, and inventory management) into one framework. Various assumptions are essential to narrow down the most important building blocks for the study and reduce the complexity. Similarly, many extensions are possible in continuously refining the assumptions to better capture reality and understand the impact of various parameters and policy decisions. This will require continued close collaboration with our industry partners and much data gathering and analysis.

We see the broad impact of our work in the following. First, this study will help to demonstrate the economic value of condition monitoring for improved spare part demand forecasting. Second, our condition-based inventory management approach should contribute to the reduction of inventory costs as well as increase fleet availability for commercial and military aircrafts. Third, the simulation framework can be used to evaluate the strategic implications for future development of sensor technologies by identifying the operational value of adding different sensors to the engine. Fourth, the study supports better decision-making for MRO service contracts. The engine manufacturer can better assess the risk for engine failures and is, therefore, able to better assess their service offering (e.g., guaranteed availability of an engine) and set the pricing of the service contracts accordingly. Fifth, further research on condition sensing, data gathering, and analytics evaluating engines' health status contributes to keeping airplanes safe and reliable.

CHAPTER 6

CONCLUSION AND FUTURE RESEARCH DIRECTION

6.1 Conclusion

Our study highlights the value of supply chain synchronization in three very relevant but different practical industry contexts. Our results clearly show the value of (1) order coordination in the presence of joint setup costs and batch restrictions, (2) time buffering when faced with uncertain lead times and (3) condition information in maintenance and overhaul operations, to achieve synchronized flow in the supply chain. In Chapter 2, we consider the deterministic joint replenishment problem with batch restrictions and high setup costs. Our analysis shows that a ZIO policy is not feasible because the batches of the various products will get depleted at different times. However, regeneration points can be determined and used to formulate the infinite horizon problem and calculate and exact expression for the average inventory in the system. The optimal constant reorder interval can then be found using a simple search algorithm. In addition, we show that, despite demand and all parameters being constant, a constant reorder interval is not optimal under batch ordering restrictions. A mixed integer program, which calculates the optimal ordering periods over a finite horizon with (potentially) time-varying demands, is proposed to solve for the optimal varying reorder intervals within a regenerations period. In chapter 3 we extend the analysis to incorporate stochastic demand, derive an approximate average cost function and determine the corresponding optimal joint reorder interval. The inclusion of batch ordering and safety stock in the approximate model to calculate the reorder interval, rather than using a simple EOQ approximation, results in savings of over 1% in our case study and anywhere from 0-56%, depending on the parameters, in our guided computational experiments. In Chapter 4, we show how to synchronize an ATO system consisting of possibly hundreds of components by advancing supply orders. Our methodology not only promises on-time delivery but also achieves significantly lower expected inventory cost. Extensive simulation of various settings shows that the advanced ordering policing is robust to changes in supplier delivery performance, and identifies the sequence of parts whose orders to advance first in a phased implementation of the time buffering strategy. In Chapter 5, we seek to use condition information from a large number of distributed working units in the field to improve the management of the inventory of spare parts required to maintain those units. We develop a general fleet management simulation framework that will evaluate the overall impact of using advanced condition information over the life cycle of an engine program.

6.2 Future Research Directions

The research presented in this dissertation has been conducted as part of large-scale, long-term industry projects and collaborations. The complexity of each project immediately allows for the formulation of extensions and future research directions to extend the work presented here within. The following sections outline work currently in progress and future research directions.

6.2.1 Research Opportunities for the Joint Replenishment Problem with Batch Ordering

For the Joint Replenishment Problem with Batch Ordering, we distinguish between the two cases we presented:

Deterministic Case

Future work will further build on the work on this chapter and show its practical value through a comprehensive computational study under a wide-range of parameter settings. Comparing the resulting constant reorder interval and its cost to the EOQ solution will allow us to identify the settings in which the exact iterative approach is most beneficial. This may allow us to further characterize properties of the optimal solution. The computational study will also determine the additional benefit associated with varying the reorder interval length over the regeneration interval. Finally, further work is also needed in capturing the effect of potential empty orders. This can be done using the mixed integer program proposed in the chapter to find the periods when orders will be placed at optimality.

Stochastic Case

The more complex practical case of highly stochastic demands at our industrial partner required us to develop heuristic approaches that are beyond the scope of this dissertation. These heuristic algorithms are currently being used by our industrial partner and have significantly lowered their operating costs. The next step will explore incorporating advanced demand information obtained from a job opportunity pipeline that captures, for each potential order, the current status of the client customization process necessary to land their order. However, the realization of the clients order, exact part composition of the job, and timing of the job are uncertain given the status. It therefore provides only limited information.

A simulation has been developed and will be used to evaluate practical heuristics and optimization approaches for the practical context. Furthermore, the practical case requires an integration into a enterprise resource planning system with other resource restrictions. This further influences the demand planning and the inventory needed for production.

6.2.2 Research Opportunities for Component Inventory Management for High-Tech Assembly Systems

Future work could further extend the model to account for (1) stochastic demand of the end product, (2) the case of a spare part demand stream with aggregate service level constraint, and (3) the interaction of both production and spare part demand streams.

Other extensions to consider involve capturing more details of the supply and assembly processes: (1) The current model assigns a certain supply delivery performance to each component based on its recent history. When a component is sourced from two or more different suppliers, however, modeling each separately may be necessary. (2) Supplier responsiveness is often tightly linked to capacity limitations; in this case, congestion would need to be modeled requiring a queuing approach. (3) A more detailed model of the assembly process would involve time-phasing the need for components at each assembly stage and the addition of assembly capacity constraints. The particular stages where delays occur could affect the amount of inventory carried. We conjecture that the time-phasing effect is not significant, since our current simulation replicates current performance metrics reasonably well.

Understanding the impact of advance ordering through detailed accounting will reveal the true overall costs/savings (e.g., cost of inventory, idle assembly capacity induced by delays, penalty payments from supplier and to customer). The financial flows may very well be asynchronous to the physical flows and, hence, will deliver further insights into the actual benefits of the time buffering strategies.

A full computational study that aims to analyze the *100%-Buffer* strategy under various product and delay parameters (e.g., number of components considered, component delay distributions) is needed to understand the value of time buffering in different settings and identify when it is most critical.

6.2.3 Research Opportunities for Spare Part Inventory Management with Advanced Fleet Condition Information

The simulation framework developed allows for consideration of different policies and models within each of its modules. In the current simulation model, we have chosen a simple engine introduction and retirement schedule, a particular method of forecasting cycles flown given current economic conditions, a basic method for the inventory control of spare engines, a simple spare part inventory management policy considering the condition information, etc. Further research is needed for the careful selection of other models and policies to use within each module. Extensive experiments need to be run to evaluate the performance of different inventory policies and understand the impact of the many parameters at our disposal.

There are various relevant research directions and extensions beyond this proposal and the proposed work for this dissertation.

First, our research should have significant impact on MRO service contracts. A higher confidence in spare part forecast will have significant effects on promised contract conditions (e.g., service levels) and pricing of the service (see Nowicki et al., 2008; Justin & Mavris, 2015). This is an extension that could be further evaluated.

Second, there are multiple streams of demand for a part, such as military spares, commercial spares, military production assembly, and commercial production assembly. Each demand stream has different requirements such as demand lead time or service level. Modeling the interaction of all streams appropriately is complex but is of practical relevance (see Koçağa & Şen, 2007).

Third, there are multiple agents that are involved or influence MRO decisions which include (I) Fleet Planners, (II) Fleet Managers, (III) Inventory & Supply Chain Planners, (IV) Financial Planners, (V) Business Developers, and (VI) Strategic Planners. A future study could further model agents' impact and evaluate their individual competing objectives.

Fourth, the fleet condition methodology could be extended to capture the optimal number of spare engines in the system. This needs to be integrated with the engine production schedule, customer contracts, and the flexibility for delivering engines.

APPENDIX A

CHAPTER 4: LEAD TIME DISTRIBUTION CHALLENGE*

The major challenge faced in the development of the simulation model is the generation of non-crossing lead times that match those observed in practice. The distribution of simulated lead times must match the empirical distribution observed in practice. We cannot, however, simply draw samples from the observed lead time distribution because order crossing would naturally occur. To overcome this challenge, we construct a discrete random variable X, with probability distribution $P(X = i) = r_i$ for $i = 0, 1, 2, ..$ that will yield the observed true delay distribution once it is being independently sampled from over time under non-crossing requirements. Let $X_t, t = 1, 2, ..$ be the random process defined by independently sampling from distribution X at each time t. Imposing the non-crossing requirement, we define $Y_t = max X_t, Y_{t-1} - 1$. This random process, Y_t, must match the true empirically observed delay distribution. Observe that Y_t is a Markovian random process, where the probability of reaching a future state only depends on the current state. In this scenario, the states represent the array of possible lead times.

Figure A.1 depicts an example of the Markov process Y_t. The transition probability from a state i into any future state j can be written as:

$$p_{ij} = \begin{cases} r_j & \text{if } j \geq i \\ \sum_{k=0}^{i-1} r_k & \text{if } j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

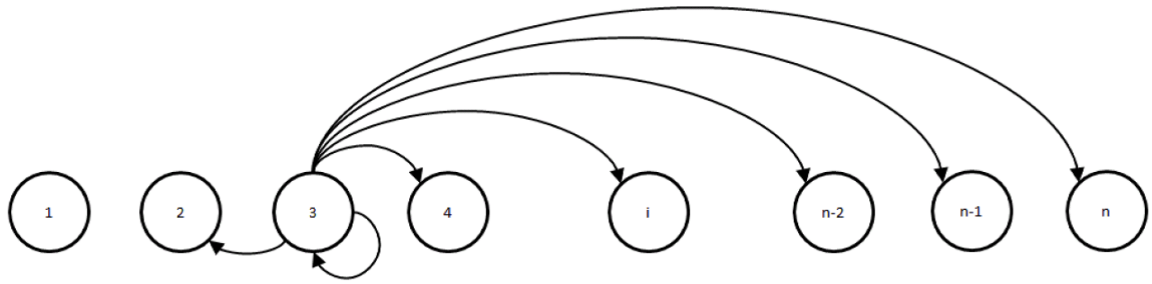

Figure A.1. Markov random process Y_t, limited by non-crossing requirement

The steady-state probability π_j of the random process Y_t being in state j is calculated as $\pi_j = \sum_{i=1}^{n} p_{ij}\pi_i$.

Imposing that those values match the observed delay distribution, the probability distribution of X can be calculated recursively as follows:

$$r_j = \begin{cases} \frac{\pi_1}{\pi_1+\pi_2} & \text{if } j = 1 \\ \frac{\pi_j - \pi_{j+1}R}{\Pi} & \text{if } j = 1..n-1 \\ \pi_j & \text{if } j = n \end{cases}$$

where $R = \sum_{i=1}^{j-1} r_i$ and $\Pi = \sum_{i=1}^{j+1} \pi_i$.

*Adapted from Beladi'S M.S. Thesis

APPENDIX B

CHAPTER 5: CASE STUDY ASSUMPTIONS AND DATA SOURCES

List of data sources and assumptions:

1. **Turbofan engine degradation simulation data set:** Based on the C-MAPSS tool and available at the NASA Prognostics Repository.
   (a) URL1: https://c3.nasa.gov/dashlink/resources/139/.
   (b) URL2: http://ti.arc.nasa.gov/c/6/
2. **Overhaul duration:** We currently rely on qualitative information based on expert opinion. We assume a minimum of 7 weeks and a maximum of 27 weeks for overhauls with an expected duration of around 11 weeks.
3. **Economic data:** Most data can be found in publicly accessible data sources (e.g., stats.oecd.org). A history of the number of flight cycles flown can be found on the Bureau of Transportation Statistics website (https://www.transtats.bts.gov).
4. **Weekly flight cycles:** We currently assume average daily flight cycles of 4.5 and assign a probability P=0.3 to 4 & 5 daily flight cycles as well as P=0.2 to 3 & 6 daily flight cycles.
5. **Part failure probabilities:** We currently assume a policy which replaces all parts. This can later be extended to part failure probability rates empirically observed in practice.

6. **Life-limited parts:** As a proof of concept, we currently assume life-limited parts to not operate longer than 25,000 flight cycles and will initiate an overhaul when that limit is exceeded. More information can be found at the Federal Aviation Administration webpage: www.faa.gov

7. **Spare engine pool size:** We currently assume a fixed 10% of spare engines in the fleet. The optimal number of spare engines is an open research question. It depends on the condition of the fleet and may change significantly over time, especially as fleet size is changing, engines enter and leave the fleet, and shop visit volume increases dramatically at the point where some engines require subsequent major overhauls while others are on their first shop visit.

APPENDIX C

CHAPTER 5: SYSTEM ARCHITECTURE AND VARIABLES

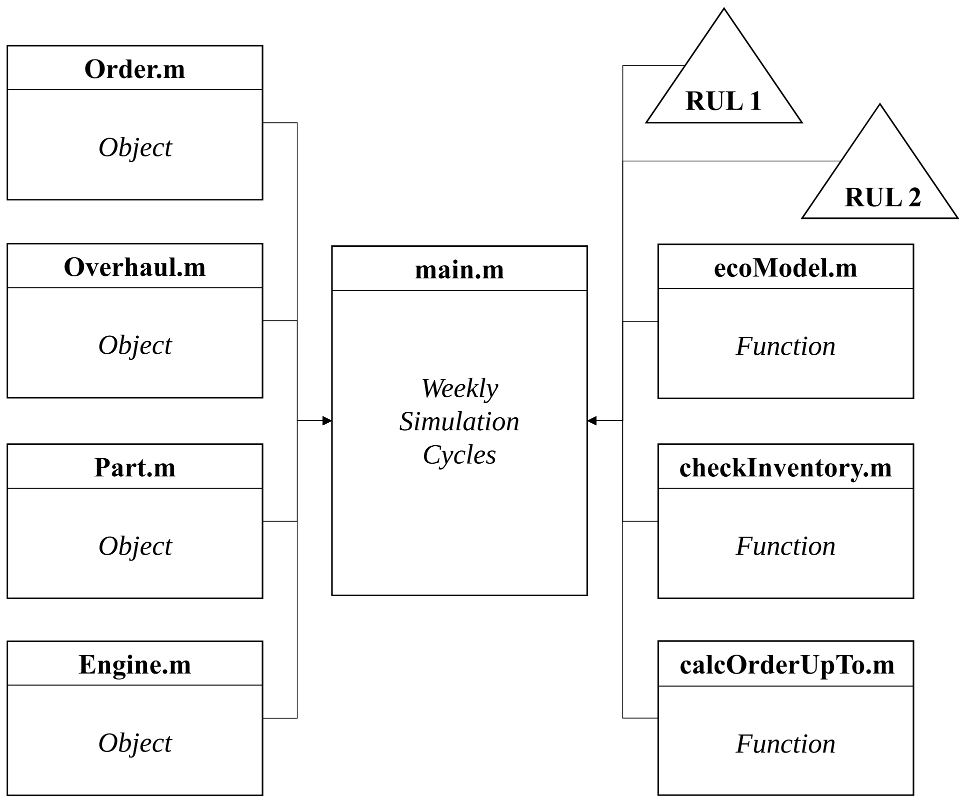

Figure C.1. Simulation system architecture

1. **Main.m:** Main simulation class containing all parameters.

2. **Order.m:** *Order* object containing all variables to track an outgoing order to a part supplier. Variables tracked include part type, time of order, time order is expected to arrive, amount ordered, and time of actual delivery. When parts

are ordered, we track *Order* instances in an array of all outstanding orders for the specific part. Once an order arrives, the *Order* instance is saved in a list that contains all previous orders. This way, we can calculate all KPIs that are associated with part ordering.

3. **Overhaul.m:** *Overhaul* object containing all variables to track the overhaul of an engine. Variables include ID of engine being overhauled, start of the overhaul, scheduled end date (random distribution), actual end date, marker if overhaul was successful, and the overhaul policy being used. All *Overhaul* objects are stored in an array to calculate KPIs at the end of the simulation.

4. **Part.m:** *Part* object containing all variables that define *Part* instances. Variables describe the part type, specific part instance ID, life limit, cost, lead time, and part failure probability. We use the *Part* object also to track inventory for each part modeled in a master list. Hence, we also include variables that describe the inventory, inventory position, inventory history, inventory position history, and order-up-to level.

5. **Engine.m:** *Engine* object containing all variables of a modeled engine. This includes an unique ID, time of entry into service, number of flight cycles assigned per week, effective number of flight cycles flown per week (influenced by the economy), total life time flight cycles flown and total flight cycles flown since last overhaul. Furthermore, an *Engine* instance saves all *Overhaul* and *Part* instances for the particular engine. Lastly, *Engine* instances include variables for the RUL distribution and the engine status.

6. **calcOrderUpTo.m:** Function that includes the core inventory model to be tested in the simulation. The function returns the part specific order-up-to level in each week reflecting the engine fleet condition. We place orders in every period according to the dynamically changing part order-up-to level.

7. **checkInventory.m:** Function that is used during the overhaul process to assure that all parts are available. The function uses a list of parts as input (parts needed for the overhaul) and returns a binary number (1=inventory available, 0=no inventory available).

8. **usageForecastingModel.m:** Function that is called to forecast the future total number of flights of the fleet at any given point of time. The functionality follows the description in Section 5.6.

9. **rulEstimation.m:** Engine degradation paths along with their sensing estimation uncertainties are loaded into the simulation and assigned to engines when (i) a new engine enters the fleet or (ii) after an overhaul is finished. These degradation paths are given by the research results from our collaborators and follow sensing module logic described in Section 5.6. They are given as a matrix where for each number of flight cycles flown there are 500 readings of the RUL which were generated by the particle filtering method. From these 500 readings, the simulation will create a discrete distribution of weeks to overhaul given the current number of cycles flown by the engine under consideration.

APPENDIX D

CHAPTER 5: SIMULATION PSEUDOCODE

1. Initialize all parameters (e.g., production start/end date, economic data or list of parts used), probability distributions (e.g., overhaul durations), and policies (e.g., overhaul policy).

2. Initialize all variables to track KPIs.

3. Start weekly simulation cycle

   (a) Introduce engines to the fleet by initializing *Engine* instances. We assign a random usage profile reflecting the economy in the particular period. We also assign a random RUL distribution to the engine reflecting a regular or abrupt fault degradation profile.

   (b) Check if the minimum spare engine pool is fulfilled. Otherwise, introduce a new engine to the spare engine pool.

   (c) Assign the weekly flight cycles flown for each engine according to economic conditions. We experiment with different policies, but currently keep flight cycles flown for each engine constant over a quarter of a year reflecting that an airplane is operated on the same route for multiple weeks at a time.

   (d) Calculate the week's order-up-to levels for each part simulated according to the inventory model tested.

   (e) Iterate through all outstanding part orders to supplier and update inventory numbers if order arrives in current period.

(f) Place new part orders to suppliers if the inventory position is below the period's order-up-to level.

(g) Save period's inventory (position) numbers for KPI calculations.

(h) Decide which engines need overhaul in the current period based on the probability to fail within a specified flight cycle range ahead. Create an *Overhaul* instance for each of those engines and add each of them to the list of engines with 'in shop' status.

(i) Check if spare engines are available and introduce spare engines to the fleet to replace engines going into overhaul.

(j) Change status for affected engines and keep track of all KPI measures.

(k) Iterate through all engines with status 'in shop' and check for part availability in physical inventory. We distinguish between different overhaul policies and check which parts need to be replaced. We initially assume parts are needed 3 weeks before planned (random) overhaul end.

(l) Delay maintenance by a week for those engines in need of parts not yet available.

(m) Check which engines reached their overhaul end and set status to 'spare'. Introduce engines to spare engine pool with a newly assigned degradation path.

(n) Update all simulation cycle data in all *Engine* and *Part* instances.

(o) Save cycle KPI data.

4. Display simulation progress.

5. Calculate and display all KPIs

REFERENCES

Aloulou, M. A., Dolgui, A., & Kovalyov, M. Y. (2014). A bibliography of non-deterministic lot-sizing models. *International Journal of Production Research*, *52*(8), 2293–2310.

Alrabghi, A., & Tiwari, A. (2013). A review of simulation-based optimisation in maintenance operations. In *15th international conference on computer modelling and simulation (uksim)* (pp. 353–358).

Arthur F. Veinott, J. (1965). The optimal inventory policy for batch ordering. *Operations Research*, *13*(3), 424–432.

Atkins, D. R., & Iyogun, P. O. (1988). Periodic versus can-order policies for coordinated multi-item inventory systems. *Management Science*, *34*(6), 791–796.

Bales, R., Maull, R., & Radnor, Z. (2004). The development of supply chain management within the aerospace manufacturing sector. *Supply Chain Management: An International Journal*, *9*(3), 250–255.

Airbus. (2014). *Global market forecast 2015-2034.* Retrieved 11/06/2016, from `https://goo.gl/4owmfE`

A.T. Kearney. (2008). *Integrated value chains in aerospace and defense.* Retrieved 12/31/2015, from `https://goo.gl/XVpx0o`

Boston Consulting Group. (2005). *Avoiding supply chain shipwrecks: Navigating outsourcing's rocky shoals.* Retrieved 10/31/2016, from `https://goo.gl/LlQppN`

Boston Consulting Group. (2011). *Rethinking the supply chain: After japan's earthquake.* Retrieved 10/31/2016, from `https://goo.gl/QaUd9H`

BusinessWire. (2011). *Mca achieves record result in 2010.* Retrieved 11/06/2016, from `https://goo.gl/QHzt1q`

Canaan Group. (1996). The changing character of aviation maintenance: From technical imperative to business enterprise. *Commentary*.

Capgemini. (2009). *Maintenance, repair and overhaul (mro).* Retrieved 11/06/2016, from `https://goo.gl/7HnDYv`

Clearwater International. (2011). *Aerospace global report.* Retrieved 11/06/2016, from https://goo.gl/5vJacu

Clearwater International. (2014). *Aerospace global report.* Retrieved 11/06/2016, from https://goo.gl/3Xpt5k

Deloitte. (2012). *The aerospace and defense industry in the u.s. - a financial and economic impact study.* Retrieved 10/31/2016, from https://goo.gl/qxNY7c

Deloitte. (2014). *Global aerospace and defense industry outlook.* Retrieved 11/06/2016, from https://goo.gl/pswDXQ

GE Aviation. (2010). *Jet engine life span.* Retrieved 11/06/2016, from https://goo.gl/CTpl4x

McKinsey & Company. (2010). *Mckinsey on government.* Retrieved 11/06/2016, from https://goo.gl/B8ZPm5

PricewaterhouseCoopers. (2011). *A&d insights: Gaining technological advantage.* Retrieved 10/31/2016, from https://goo.gl/HgfgWn

PricewaterhouseCoopers. (2012). *Soaring or stalling - can aircraft manufacturers prevent rate ramp-up problems?* Retrieved 10/31/2016, from https://goo.gl/lgYzW0

Boctor, F., Laporte, G., & Renaud, J. (2004). Models and algorithms for the dynamic-demand joint replenishment problem. *International Journal of Production Research*, *42*(13), 2667–2678.

Bonavia, T., & Marin, J. A. (2006). An empirical study of lean production in the ceramic tile industry in spain. *International Journal of Operations & Production Management*, *26*(5), 505–531.

Bramson, M. (1962). The variable lead-time problem in inventory control: A survey of the literature–part 1. *OR*, *13*(1), 41–53.

Cha, B., & Moon, I. (2005). The joint replenishment problem with quantity discounts under constant demand. *OR Spectrum*, *27*(4), 569–581.

Chang, C. A. (1985). The interchangeability of safety stocks and safety lead time. *Journal of Operations Management*, *6*(1), 35–42.

Chao, X., & Zhou, S. X. (2009). Optimal policy for a multiechelon inventory system with batch ordering and fixed replenishment intervals. *Operations Research*, *57*(2), 377–390.

Chen, F. (2000). Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research*, *48*(3), 376–389.

Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: A new methodology for a systematic literature review. *Supply Chain Management: An International Journal*, *17*(4), 403–418.

Cooper, M. C., Lambert, D. M., & Pagh, J. D. (1997). Supply chain management: More than a new name for logistics. *The International Journal of Logistics Management*, *8*(1), 1–14.

Dekker, R. (1996). Applications of maintenance optimization models: A review and analysis. *Reliability Engineering & System Safety*, *51*(3), 229–240.

Dekker, R., Wildeman, R. E., & Van der Duyn Schouten, F. A. (1997). A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research*, *45*(3), 411–435.

Denning, S. (2013). *What went wrong at boing?* Retrieved 10/31/2016, from `https://goo.gl/2l1viU`

De Treville, S., Bicer, I., Chavez-Demoulin, V., Hagspiel, V., Schürhoff, N., Tasserit, C., & Wager, S. (2014). Valuing lead time. *Journal of Operations Management*, *32*(6), 337–346.

De Treville, S., & Trigeorgis, L. (2010). It may be cheaper to manufacture at home. *Harvard Business Review*, *88*(10).

Dolgui, A., Ammar, O. B., Hnaien, F., & Louly, M.-A. (2013). A state of the art on supply planning and inventory control under lead time uncertainty. *Studies in Informatics and Control*, *22*(3), 255–268.

Dolgui, A., Louly, A. O., & Prodhon, C. (2005). A survey on supply planning under uncertainties in mrp environments. *IFAC Proceedings Volumes*, *38*(1), 1–12.

Dolgui, A., & Prodhon, C. (2007). Supply planning under uncertainties in mrp environments: A state of the art. *Annual Reviews in Control*, *31*(2), 269–279.

Duran, O., & Perez, L. (2013). Solution of the spare parts joint replenishment problem with quantity discounts using a discrete particle swarm optimization technique. *Studies in Informatics and Control*, *22*(4), 319–328.

Elwany, A. H., & Gebraeel, N. Z. (2008). Sensor-driven prognostic models for equipment replacement and spare parts inventory. *IIE Transactions*, *40*(7), 629–639.

Eynan, A., & Kropp, D. H. (1998). Periodic review and joint replenishment in stochastic demand environments. *IIE transactions*, *30*(11), 1025–1033.

Federgruen, A., & Zheng, Y.-S. (1992). The joint replenishment problem with general joint cost structures. *Operations Research*, *40*(2), 384–403.

Feng, H., Wu, Q., Muthuraman, K., & Deshpande, V. (2015). Replenishment policies for multi-product stochastic inventory systems with correlated demand and joint-replenishment costs. *Production and Operations Management*, *24*(4), 647–664.

Gallien, J., & Wein, L. M. (2001). A simple and effective component procurement policy for stochastic assembly systems. *Queueing Systems*, *38*(2), 221–248.

Gampenrieder, E., Damotte, E., Seel, K., Gates, D., & Mayor, T. (2015). *Global manufacturing outlook 2015.* Retrieved 10/31/2016, from `https://goo.gl/WYOXlX`

Gao, R. X., & Wang, P. (2015). Stochastic modeling for engine remaining life prediction. *Pratt & Whitney Seminar Series*.

Garg, A., & Deshmukh, S. (2006). Maintenance management: Literature review and directions. *Journal of Quality in Maintenance Engineering*, *12*(3), 205–238.

Gatland, R., Yang, E., & Buxton, K. (1997). Solving engine maintenance capacity problems with simulation. In *Proceedings of the 29th conference on winter simulation* (pp. 892–899).

Gharbi, A., Girard, J., Pellerin, R., & Villeneuve, L. (1997). Bombardier turned to simulation to validate the cf-18 maintenance program. *Interfaces*, *27*(6), 22–34.

Goyal, S. K., & Satir, A. T. (1989). Joint replenishment inventory control: Deterministic and stochastic models. *European Journal of Operational Research*, *38*(1), 2–13.

Hnaien, F., Dolgui, A., & Wu, D. D. (2016). Single-period inventory model for one-level assembly system with stochastic lead times and demand. *International Journal of Production Research*, *54*(1), 186–203.

Hoque, M. (2006). An optimal solution technique for the joint replenishment problem with storage and transport capacities and budget constraints. *European Journal of Operational Research*, *175*(2), 1033–1042.

Jackson, P., Maxwell, W., & Muckstadt, J. (1985). The joint replenishment problem with a powers-of-two restriction. *IIE Transactions*, *17*(1), 25–32.

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510.

Joneja, D. (1990). The joint replenishment problem: New heuristics and worst case performance bounds. *Operations Research*, *38*(4), 711–723.

Justin, C. Y., & Mavris, D. N. (2015). Aircraft and engine economic evaluation for fleet renewal decision-making and maintenance contract valuation. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 1–14.

Kang, H.-Y., Lee, A. H., Wu, C.-W., & Lee, C.-H. (2016). An efficient method for dynamic-demand joint replenishment problem with multiple suppliers and multiple vehicles. *International Journal of Production Research*, *55*(4), 1–20.

Kennedy, W., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics*, *76*(2), 201–215.

Khouja, M., & Goyal, S. (2008). A review of the joint replenishment problem literature: 1989–2005. *European Journal of Operational Research*, *186*(1), 1–16.

Khouja, M., Park, S., & Saydam, C. (2005). Joint replenishment problem under continuous unit cost change. *International Journal of Production Research*, *43*(2), 311–326.

Klein, C., & Ventura, J. (1995). An optimal method for a deterministic joint replenishment inventory policy in discrete time. *Journal of the Operational Research Society*, *46*(5), 649–657.

Koçağa, Y. L., & Şen, A. (2007). Spare parts inventory management with demand lead times and rationing. *IIE Transactions*, *39*(9), 879–898.

Lancioni, R. A. (2000). New developments in supply chain management for the millennium. *Industrial Marketing Management*, *29*(1), 1–6.

Larsen, C. (2009). The q (s, s) control policy for the joint replenishment problem extended to the case of correlation among item-demands. *International Journal of Production Economics*, *118*(1), 292–297.

Li, Q., & Zheng, S. (2006). Joint inventory replenishment and pricing control for systems with uncertain yield and demand. *Operations Research*, *54*(4), 696–705.

Linebaugh, K. (2013). *Ge brings engine work back.* Retrieved 11/06/2016, from https://goo.gl/YmrDBV

Lorell, M. (2000). *Cheaper, faster, better? commercial approaches to weapons acquisition.* Santa Monica, CA: RAND Corporation.

Louly, M.-A., & Dolgui, A. (2011). Optimal time phasing and periodicity for mrp with poq policy. *International Journal of Production Economics*, *131*(1), 76–86.

Louly, M.-A., Dolgui, A., & Al-Ahmari, A. (2012). Optimal mrp offsetting for assembly systems with stochastic lead times: Poq policy and service level constraint. *Journal of Intelligent Manufacturing*, *23*(6), 2485–2495.

Louly, M.-A. O., & Dolgui, A. (2009). Calculating safety stocks for assembly systems with random component procurement lead times: A branch and bound algorithm. *European Journal of Operational Research*, *199*(3), 723–731.

Mabert, V. A., Soni, A., & Campbell, G. (2006). A case study of pratt and whitney aircraft's commercial spares planning. *Communications of the IIMA*, *6*(3), 1–13.

Mann, T. (2016). *United technologies pratt struggles with supply chain for new jet engine.* Retrieved 10/31/2016, from http://on.wsj.com/1UhbVTN

Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management. *Journal of Business Logistics*, *29*(1), 133–155.

Mattila, V., Virtanen, K., & Raivio, T. (2008). Improving maintenance decision making in the finnish air force through simulation. *Interfaces*, *38*(3), 187–201.

Melchiors, P. (2002). Calculating can-order policies for the joint replenishment problem by the compensation approach. *European Journal of Operational Research*, *141*(3), 587–595.

Melnyk, S. A., & Piper, C. J. (1981). Implementation of material requirements planning: Safety lead times. *International Journal of Operations & Production Management*, *2*(1), 52–61.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, *22*(2), 1–25.

Minner, S. (2003). Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, *81*, 265–279.

Molinder, A. (1997). Joint optimization of lot-sizes, safety stocks and safety lead times in an mrp system. *International Journal of Production Research*, *35*(4), 983–994.

Mula, J., Poler, R., Garcia-Sabater, J., & Lario, F. C. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics*, *103*(1), 271–285.

Narayanan, A., & Robinson, E. (2006). More on models and algorithms for the dynamic-demand joint replenishment problem. *International Journal of Production Research*, *44*(2), 383–397.

Nicolai, R. P., & Dekker, R. (2008). Optimal maintenance of multi-component systems: A review. In *Complex system maintenance handbook* (pp. 263–286). Springer.

Nielsen, C., & Larsen, C. (2005). An analytical study of the q (s, s) policy applied to the joint replenishment problem. *European Journal of Operational Research*, *163*(3), 721–732.

Nowicki, D., Kumar, D. U., Steudel, H., & Verma, D. (2008). Spares provisioning under performance-based logistics contract: Profit-centric approach. *The Journal of the Operational Research Society*, *50*(3), 342–352.

Oliver, R. K., & Webber, M. D. (1982). Supply-chain management: Logistics catches up with strategy. *Outlook*, *5*(1), 42–47.

Painter, M. K., Erraguntla, M., Hogg Jr, G. L., & Beachkofski, B. (2006). Using simulation, data mining, and knowledge discovery techniques for optimized aircraft engine fleet management. In *Proceedings of the 38th conference on winter simulation* (pp. 1253–1260).

Peck, H. (2005). Drivers of supply chain vulnerability: An integrated framework. *International Journal of Physical Distribution & Logistics Management*, *35*(4), 210–232.

Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, *50*(1-4), 297–313.

Pham, H., & Wang, H. (1996). Imperfect maintenance. *European Journal of Operational Research*, *94*(3), 425–438.

Plossl, G. W., & Wight, O. W. (1971). *Material requirements planning by computer*. Washington: American Production and Inventory Control Society.

Porras, E., & Dekker, R. (2006). An efficient optimal solution method for the joint replenishment problem with minimum order quantities. *European Journal of Operational Research*, *174*(3), 1595–1615.

Rao, P., & Holt, D. (2005). Do green supply chains lead to competitiveness and economic performance? *International Journal of Operations & Production Management*, *25*(9), 898–916.

Reals, K. (2010). *Forecasts detail 10-year outlook for global mro market.* Retrieved 11/06/2016, from `https://goo.gl/OqM8Kh`

Robinson, E., Narayanan, A., & Gao, L. (2007). Effective heuristics for the dynamic demand joint replenishment problem. *Journal of the Operational Research Society*, *58*(6), 808–815.

Rossetti, C., & Choi, T. Y. (2005). On the dark side of strategic sourcing: Experiences from the aerospace industry. *The Academy of Management Executive*, *19*(1), 46–60.

Rossetti, C. L., & Choi, T. Y. (2008). Supply management under high goal incongruence: An empirical examination of disintermediation in the aerospace supply chain. *Decision Sciences*, *39*(3), 507–540.

Sanders, P., & Cameron, D. (2011). *Boeing sees long-delayed 787 deliveries starting in third quarter.* Retrieved 10/31/2016, from `https://goo.gl/GCzefH`

Sarker, R., & Haque, A. (2000). Optimization of maintenance and spare provisioning policy using simulation. *Applied Mathematical Modelling*, *24*(10), 751–760.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *International conference on prognostics and health management* (pp. 1–9).

Scott, R. E. (2015). The manufacturing footprint and the importance of u.s. manufacturing jobs. *EPI Briefing Paper, January*. Retrieved 10/31/2016, from `https://goo.gl/hl2MwH`

Scott, R. E., & Kimball, W. (2014). China trade, outsourcing and jobs. *EPI Briefing Paper, December*. Retrieved 10/31/2016, from `https://goo.gl/13l28Z`

Sharma, A., Yadava, G., & Deshmukh, S. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, *17*(1), 5–25.

Silver, E. A. (1976). A simple method of determining order quantities in joint replenishments under deterministic demand. *Management Science*, *22*(12), 1351–1361.

Smith, M., & Babai, M. Z. (2011). A review of bootstrapping for spare parts forecasting. In *Service parts management* (pp. 125–141). Springer.

Song, J.-S. (1994). The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science*, *40*(5), 603–613.

Song, J.-S., Yano, C. A., & Lerssrisuriya, P. (2000). Contract assembly: Dealing with combined supply lead time and demand quantity uncertainty. *Manufacturing & Service Operations Management*, *2*(3), 287–296.

Song, J.-S., & Yao, D. D. (2002). Performance analysis and optimization of assemble-to-order systems with random lead times. *Operations Research*, *50*(5), 889–903.

Song, J.-S., Zhang, H., Hou, Y., & Wang, M. (2010). The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research*, *58*(1), 68–80.

Song, J.-S., & Zipkin, P. H. (1996). The joint effect of leadtime variance and lot size in a parallel processing environment. *Management Science*, *42*(9), 1352–1363.

Stranjak, A., Dutta, P. S., Ebden, M., Rogers, A., & Vytelingum, P. (2008). A multi-agent simulation system for prediction and scheduling of aero engine overhaul. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems: Industrial track* (pp. 81–88).

Sun, T., Zhao, Q., Luh, P. B., & Tomastik, R. N. (2004). Joint replacement optimization for multi-part maintenance problems. In *Ieee/rsj international conference on intelligent robots and systems* (Vol. 2, pp. 1232–1237).

Sun, T., Zhao, Q., Luh, P. B., & Tomastik, R. N. (2008). Optimization of joint replacement policies for multipart systems by a rollout framework. *IEEE Transactions on Automation Science and Engineering*, *5*(4), 609–619.

Tajbakhsh, M. M., Zolfaghari, S., & Lee, C.-G. (2007). Supply uncertainty and diversification: A review. In *Trends in supply chain design and management* (pp. 345–368). Springer.

Tang, C. S. (2006). Perspectives in supply chain risk management. *International Journal of Production Economics*, *103*(2), 451–488.

Treville, S., Schürhoff, N., Trigeorgis, L., & Avanzi, B. (2014). Optimal sourcing and lead-time reduction under evolutionary demand risk. *Production and Operations*

Management, *23*(12), 2103–2117.

Tu, G., Luh, P. B., Zhao, Q., & Tomastik, R. N. (2004). An optimization method for joint replacement decisions in maintenance. In *43rd ieee conference on decision and control* (Vol. 4, pp. 3674–3679).

Van der Duyn Schouten, F. (1996). Maintenance policies for multicomponent systems: An overview. In *Reliability and maintenance of complex systems* (pp. 117–136). Springer.

Van Eijs, M., Heuts, R. M. J., & Kleijnen, J. P. C. (1992). Analysis and comparison of two strategies for multi-item inventory systems with joint replenishment costs. *European Journal of Operational Research*, *59*(3), 405–412.

Viswanathan, S. (1997). Note. periodic review (s, s) policies for joint replenishment inventory systems. *Management Science*, *43*(10), 1447–1454.

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, *139*(3), 469–489.

Wang, J., Wang, P., & Gao, R. X. (2015). Enhanced particle filter for tool wear prediction. *Journal of Manufacturing Systems*, *36*, 35–45.

Wang, P., & Gao, R. X. (2015). Adaptive resampling-based particle filtering for tool life prediction. *Journal of Manufacturing Systems*, *37*(2), 528–534.

Wang, P., & Gao, R. X. (2016). Markov nonlinear system estimation for engine performance tracking. *Journal of Engineering for Gas Turbines and Power*, *138*(9), 091201.

Wang, P., & Gao, R. X. (2017). Automated performance tracking for heat exchangers in hvac. *IEEE Transactions on Automation Science and Engineering*, *14*(2), 634–645.

Wang, P., Yan, R., & Gao, R. X. (2017). Virtualization and deep recognition for system fault classification. *Journal of Manufacturing Systems*, *In press*.

Wang, W., & Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E: Logistics and Transportation Review*, *47*(6), 1194–1209.

Whybark, D. C., & Williams, J. G. (1976). Material requirements planning under uncertainty. *Decision Sciences*, *7*(4), 595–606.

Wilkinson, A., Dainty, A., Neely, A., Johnstone, S., Dainty, A., & Wilkinson, A.

(2009). Integrating products and services through life: an aerospace experience. *International Journal of Operations & Production Management*, *29*(5), 520–538.

Zipkin, P. H. (2000). *Foundations of inventory management.* Boston, MA: McGraw-Hill.