Modeling Supply Chain Dynamics

with Calibrated Simulation Using Data Fusion

by

Shanshan Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2010 by the
Graduate Supervisory Committee:

Teresa Wu, Co-Chair
John Fowler, Co-Chair
Michele Pfund
Jing Li
William Pavlicek

ARIZONA STATE UNIVERSITY

December 2010

ABSTRACT

In today's global market, companies are facing unprecedented levels of uncertainties in supply, demand and in the economic environment. A critical issue for companies to survive increasing competition is to monitor the changing business environment and manage disturbances and changes in real time. In this dissertation, an integrated framework is proposed using simulation and online calibration methods to enable the adaptive management of large-scale complex supply chain systems. The design, implementation and verification of the integrated approach are studied in this dissertation. The research contributions are two-fold. First, this work enriches symbiotic simulation methodology by proposing a framework of simulation and advanced data fusion methods to improve simulation accuracy. Data fusion techniques optimally calibrate the simulation state/parameters by considering errors in both the simulation models and in measurements of the real-world system. Data fusion methods - Kalman Filtering, Extended Kalman Filtering, and Ensemble Kalman Filtering - are examined and discussed under varied conditions of system chaotic levels, data quality and data availability. Second, the proposed framework is developed, validated and demonstrated in ‘proof-of-concept’ case studies on representative supply chain problems. In the case study of a simplified supply chain system, Kalman Filtering is applied to fuse simulation data and emulation data to effectively improve the accuracy of the detection of abnormalities. In the case study of the ‘beer game’ supply chain model, the system’s chaotic level is identified as a key factor to influence simulation performance and the choice of

data fusion method. Ensemble Kalman Filtering is found more robust than Extended Kalman Filtering in a highly chaotic system. With appropriate tuning, the improvement of simulation accuracy is up to 80% in a chaotic system, and 60% in a stable system. In the last study, the integrated framework is applied to adaptive inventory control of a multi-echelon supply chain with non-stationary demand. It is worth pointing out that the framework proposed in this dissertation is not only useful in supply chain management, but also suitable to model other complex dynamic systems, such as healthcare delivery systems and energy consumption networks.

DEDICATION

This dissertation is dedicated to my parents,

Mr. Shanwei Wang and Mrs. Qiaofen Chen

ACKNOWLEDGMENTS

I want to thank my advisors – Dr. Wu and Dr. Fowler for their continuing mentoring, guidance and support through the past five years. Not only being patient and inspiring, they always create great research opportunities for me. Their caring for students, dedictation to research, hard-working spirit and optimistic attitude will keep encouraging me in future career development.

Thanks go as well to Dr. Michele Pfund, Dr. Li Jing, Dr. Pan Rong, who gave me valuable advices during comprehensive exam and when I have research questions. I also want to thank Dr. Darsh, who previously worked on the IBM project and Dr. Shao-jen Weng, who co-authored the first paper with me.

I want to thank Dr. William Pavlicek for being an encouraging supervisor during my internship in Mayo and for being an exampler of excellent leader, lecturer and scientist.

Last, I would also like to thank my friends and lab mates, especially Ozgur, Liangjie Xue and Mengqi Hu, with whom that I spend good time and am always amazed with their talents and kindness.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Motivations

A supply chain is a system of business enterprises that are linked together to satisfy consumer demand (Riddalls 2000). It usually consists of multiple echelons, each can be considered as a generic production/distribution process to add value to the final product. The connection of neighboring echelons can be characterized by backward information flow (orders from downstream entity to upstream entity) and forward material flow (products from upstream entity to downstream entity). In simplified models, goods flow from one echelon to the next till they reach the end-customer. In more realistic models, a supply chain is a network with parallel flows between different levels, and relationships between entities are complicated and changing with time (Li et al. 2009).

Although Supply Chain Management (SCM) has been studied for a long time, new research questions continue to arise in response to a quickly changing business environment (global competition, shorter product life cycles, dynamic changes of demand pattern, product varieties and environmental standards, etc) and with the increased emphasis on customer-focused strategies (responsiveness, adaptation, etc). The failure to adapt to on-going changes potentially leads to the mismatch of supply and demand, which often causes dramatic financial loss. For example, Cisco Systems had to write-off $2 billion of excess inventory of its network infrastructure products when the actual market demands suddenly declined in 2001 (Hau 2004). Hendricks and Singhal (2005) studied 885

supply/demand mismatches reported by publicly traded firms and found on average 6.92% lower sales growth, 10.66% higher growth in cost, and 13.88% higher growth in inventories of those companies in the next year of the disruption. It is difficult to recover from such disturbance even after two years.

Haeckel (1999) proposed the concept of the Sense-and-Respond enterprise, in which the organizations are adaptive – they identify the changes in customer demands and the business environment "as they happen" and make an appropriate response to capture new opportunities. Proactively managing supply chain uncertainty has been recognized to be increasingly important since the uncertainties in demand, supply, and economic conditions have recently reach unprecedented levels (Johnson 2010). Empirical evidence advocates the good practices of a company include being responsive to environment (the market change, supplier conditions, economic trend, etc). Top-performing supply chain companies, such as Dell, Walmart and Amazon have the qualities of being Agile (quick responsive to short-term demand changes), Adaptive (monitor the structural changes and adapt) and Aligned (all portions of the supply chain share information, responsibility and work with the same goal) (Hau 2004). Companies that successfully build a responsive supply chain stay ahead of their rivals in competition and bring tremendous benefits to consumers and society. For example, Seven-Eleven Japan (SEJ), which has been recognized as 'one of the world's most profitable retailers' (Hau 2004), was able to deliver 64,000 rice balls to the assaulted city within six hours after the Kobe earthquake on January 17, 1995.

However, managing a supply chain system in real time and quickly responding is challenging. A supply chain is a large-scale complex system that is vulnerable to a variety of uncertainties. Individual companies are under risks in supply, demand, product and information management (Tang 2006). Since a supply chain is a complex system, the decisions and process changes made at one level impact the responsiveness-related costs in the upstream supply chain (Reichhart and Holweg 2007). Furthermore, each company is legally and economically independent and is seeking profits and the convergence of interests is not guaranteed; this leads to more dynamics. One well-known example of supply chain dynamics is the amplification of demand through echelons from end-consumers to upstream manufacturers, this is called the 'bullwhip effect'. Recent research points out that supply chains are Complex Adaptive Systems (CAS) (Pathak et al. 2007), characterized by deterministic chaos, randomness and evolutionary behaviors.

Fortunately, the current development of Information and Communication technologies (ICT) helps to build an information enriched environment (Gunasekaran and Ngai 2004). A great deal of data has been available along the supply chain. The implementation of an Enterprise Resource Planning (ERP) system integrates information at different levels and through different processes of a company (Cerpa and Verner 1998). The development of Electronic Data Interchange (EDI) (Webster 1995) and the Internet facilitates communication between buyers and suppliers and supports cost-efficient information sharing. Although information has the potential to help understand and monitor supply

chain dynamics, there are challenges as well. First, the information is not always as accurate, complete and precise as we hope (Li and Lin 2006). Even if information technology (e.g. EDI) is used to share information on end-customer demand and inventory levels, there is still often a discrepancy between the data on customer demand or inventory levels in information systems, and the real physical flow of products (Fleisch and Tellkamp 2005). Second, the information in the supply chain exists in different levels of details and is scattered in distributed autonomous nodes of a supply chain network. The reluctance to share information is prevalent since organizations traditionally perceive information disclosure as a loss of power (Mason-Jones and Towill 1997). We envision an integrated model that relates and consolidates distributed information and captures the complex behaviors of supply chain entities in real time will be a critical facilitator to manage supply chain adaptively and more effectively.

Extensive research has been done to analyze and model supply chain dynamics. Riddalls et al. (2000) divide the modeling methodologies into four categories - continuous time differential equation models (e.g. System Dynamic models), discrete time difference equation models (e.g. state-space models), discrete event simulation systems and operational research techniques (linear programming, queuing theory, Markov chains and dynamic programming). More recently, agent-based simulations (Swaminathan et al. 1998), petri-net based models (Blackhurst et al. 2008; Liu et al. 2007), traffic-like network models (Unver 2008; Schwartz 2008), artificial intelligent models (Lau et al. 2002;

Alfonso et al. 2007) and mathematical tools from the control literature (Sarimvels et al. 2008) have been introduced to the supply chain research community.

Among these methods, simulation-based methods are well known to have the flexibility to model dynamic behaviors and complicated structures of supply chain (Riddalls et al. 2000). Discrete-event simulations are generally used to capture operation-level details and to track the individual entities and their attributes (Paul et al. 2004). Continuous-time models are quick to build and run, which enable modelers to focus on abstracting the business/operation processes. Hybrid simulations are developed to take advantage of both techniques. Paul et al. (2004) show an example using discrete time simulation and system dynamic models to evaluate alternative decisions in 'Sense-and-Respond' systems. The development of the Parallel and Distributed Simulation (PDS) paradigm (Terzi and Cavalieri 2004) makes the simulation approach more attractive, since it is able to reduce simulation cycle time (Taylor et al. 2005) and to facilitate the expansion of scope of supply chain simulation (Mustafee et al. 2006).

While promising, conventional simulations usually lack the capability to incorporate real time data to keep up on the changes in real systems. Most simulation research is based on the assumption that simulation parameters and external input (e.g. demand) are known prior to execution, which limits the implementation of simulation in the real-time management of a supply chain when the level of uncertainty changes over time and unexpected events occur during the interaction between individual nodes. Additionally, a simulation is rarely a high-fidelity representation of a dynamic system and it is subject to errors

from estimated model parameters, neglected physics, sub-optimal solution techniques, etc. Thus, real-time calibration is beneficial, especially in a complex system, to compensate for simulation errors. In this dissertation, we focus to answer these questions: What is the best way to integrate inaccurate/imprecise data with simulation models to better understand supply chain dynamics? How can more accurate predictions be made in order to enable prompt response to be carried out to handle changes in a complex supply chain?

1.2 Symbiotic Simulation

A simulation that interacts with the physical system in a mutually beneficial way is called 'symbiotic simulation' (or 'online simulation'). This term was coined by the parallel and distributed simulation working group at the Dagstuhl Seminar on Grand Challenges for Modeling and Simulation in 2002 (Fujimoto et al. 2002). In symbiotic simulation, the simulation model benefits from the continuous supply of the latest data and the automatic validation of its simulation outputs, whereas the physical system benefits from the improved performance obtained from the analysis of simulation experiments. A closely related research branch that also emphasizes the mutual relationship between simulation and physical world is Dynamic Data Driven Application System (DDDAS), sponsored by National Science Foundation (NSF) (http://www.nsf.gov/cise/cns/dddas/). Some applications of DDDAS are introduced by Darema (2005) and Ouyang et al. (2007).

We review some representative work in the fields of symbiotic simulation and DDDAS in order to summarize the existing usage of real-time data in

simulations. A majority of the research pulls out real-time data from business/ manufacturing information systems and feeds them into online simulations as initial states and inputs. Zeng et al. (2009) build an online simulation that automatically gets real demand information and solves inventory management and order scheduling problems for a chemistry supply chain. Given the real-time demand, order scheduling strategies are evaluated and the reorder point for each type of raw material is determined dynamically to meet the demand. Another example is the symbiotic simulation for semiconductor assembly and test operation developed by Low et al. (2005). Key performance indicators of the physical system, such as queuing length, are tracked. Online simulations are initiated to apply with simulation-based optimization when congestion is detected. Low et al. (2007) apply a similar approach in an aerospace spare parts logistics case study to carry out dynamic business process re-engineering. Besides, under the concept of DDDAS, Koyuncu et al. (2007) build an application in a semiconductor supply chain that monitors real-time machine states (temperature, pressure, vibration, status, sound), workshop environments (temperature, sound, air quality, humidity) and production data (type of process, material, operator skill level, product flow volume, time since last maintenance). A Bayesian Belief Network model is employed to analyze the potential root cause of abnormality and to decide the fidelity of online simulations. In sum, in these research efforts, real-time simulations are called on triggers and fed with real world data to analyze the most current situation.

Furthermore, real world data is used to validate simulation results. It is realized that simulations may deviate from real situations and need to be updated to model the evolving state of systems. Fujimoto et al. (2007) propose a multi-agent ad hoc simulation framework for transportation systems. Each agent represents a local simulation covering a small area of the transportation surface. The agents compare results from local simulations with measurements from the 'read world', which is represented by a larger-scale emulator. When the difference between simulation expectation and measurement is larger than a predefined threshold, the local simulation rolls back to a previous time point and re-start using new available data. The level of data aggregation is found critical (Hunter et al. 2006). Mitchell and Yilmaz (2008) propose an innovative multi-simulation approach for symbiotic simulation when 'real-world' data is learned by simulation agents. Rather than a single authorative model, an ensemble of plausible models collectively provides insights about the system state. Each agent incorporates real world data using reinforcement learning algorithm.

Through the literature review, we conclude that the existing data fusion methods in the field of symbiotic simulation usually ignore the errors in observations, that makes simulations less capable to accurately model a loosely-coupled and complex system like a supply chain. First, extracting detailed data from different supply chain nodes is difficult even for a major player in the supply chain. Second, the data in a supply chain environment often has noise in information release and in the transmission process, due to technologically motivated reasons, conflicting interests between supply chain partners, etc. A

methodology that can optimally incorporate noisy and incomplete data with simulation models is required.

In this dissertation, a framework of symbiotic simulation with a focus on studying online data fusion methods is proposed. The data fusion techniques, different from simple insertion methods, consider the errors in both the simulation and real world measurements. The estimation of a state is improved by extracting a maximum amount of information from both the measurements and the dynamical model, and this information is combined in an optimal way. Besides, the observations are integrated with the whole system state vector based on the covariance of variables, so that the states not directly measured are also estimated.

1.3 Data Fusion Method – Kalman Filtering

Data Fusion is a general category of techniques to combine information from multiple sources to generate better understanding of system states. Strictly defined, it is 'a dynamic process in the association, correlation and combination of information from multiple sources resulting in a fused product, which is more complete and accurate than any of the separate data elements' (Waltz and Linas 1990). The specific data fusion methods we study in the dissertation are Kalman Filtering (KF), Ensemble Kalman Filtering (EnKF) and Extended Kalman Filtering (EKF).

Kalman Filtering is a well-known tool in the control community for stochastic state and parameter estimation. It originates from the paper of R. E. Kalman back in 1960 (Kalman 1960). It is a recursive data processing technique that is optimal for a linear Gaussian system. 'Optimal' means that it is the linear

least square estimator (mean), the maximum likelihood estimator, and provides complete information on the probability distribution of state vector. When noise does not follow a Gaussian distribution, it is still the least square linear estimator (i.e., among all the estimators that are linear combinations of the measurements). An introduction of Kalman Filtering is provided in Section 2.3.

In a linear Gaussian system, Kalman Filtering could be considered as a special case of the Bayesian estimation method. It provides an efficient recursive computation of Bayesian method in a state-space model. A tutorial that relates Kalman Filtering to Bayesian estimation is provided in Barker et al. (1995).

In a nonlinear system, whose output is not directly proportional to its input, Extended Kalman Filtering and Ensemble Kalman Filtering are two approximate methods of Kalman Filtering (see Section 3.2 and Section 3.3). As an early attempt to adapt Kalman Filtering to nonlinear problems, the Extended Kalman Filtering is based on linearization of the nonlinear model system state transition function and observation function using the Jacobian matrix. It has been successfully applied and become a standard technique used in nonlinear estimation and machine learning applications (Wan and Van Der Merwe 2000). In comparison, Ensemble Kalman Filtering, introduced in 1990's, is a relative newly developed method. It is an integrated approach combining the Markov Chain Monte Carlo method with the Kalman Filtering updating scheme (Evensen 1994). It is especially successful in handling large-scale systems such as quasi-geostrophic ocean model (Evensen 1994) and wild land fire simulation (Mandel et al. 2008).

1.4 Research Overview

The objective of this dissertation is to develop an integrated framework of simulation and online data fusion to facilitate the study of adaptive management of large-scale complex supply chain systems. The study of the proposed framework is in three stages.

In the first stage, we study the use of Kalman Filtering (KF) to fuse simulation data and emulation data (representing the real world) in a linear supply chain system. A Discrete Event Simulation (DES) of a simplified semiconductor manufacturing process is built. The statistical control chart Wineglass (Wu 1988; Wu, Hosking et al. 1992) is introduced to signal the abnormality of order fulfillment. Experiments show that KF significantly improves the performance of estimation of system states, especially specificity (type I error), to filter out noise in observations. The tradeoff of specificity and sensitivity are also discussed with the use of control chart.

One limit of the first stage research is that Kalman Filtering assumes a linear state-space model. Considering the nonlinear and complex nature of supply chain systems, in the second stage, an integrated framework of online simulation coupled with nonlinear data fusion methods is proposed. Extended Kalman Filtering (EKF) and Ensemble Kalman Filtering (EnKF) are used in simulation calibration. The performance of data fusion is studied in supply chain systems with different levels of system chaos (measured by the Lyapunov Exponent (Larsen, Morecroft et al. 1999)). Experimental results indicate that under a median observation error (5% of end-customer demand) and daily information

update, the improvement is up to 80% in a highly chaotic system, and up to 60% in a stable system. The advantages and disadvantages of EKF and EnKF are analyzed according to varied system chaotic levels, the availability and the quality of the real data. EnKF is shown to outperform EKF in systems with high chaotic levels.

Finally, the online calibrated simulation framework is applied to an adaptive inventory control problem. Instead of using the result from a single simulation replication, an ensemble of simulation replications (called multi-simulations method) collectively represent the distribution of system states. Ensemble Kalman Filtering is adopted for data fusion. The calibrated simulation with adaptive order-up-to inventory policy is used in an N-echelon serial supply chain to deal with non-stationary demand. Information sharing on demands is assumed. In particular, the calibrated simulation consolidates noisy and distributed information from different information sources. A closed-form expression of Minimum Mean Square Error (MMSE) forecast is derived when the shared information does not have noise. Experiment shows that our Calibrated Simulation-based Forecast (CSF) generates comparable result to MMSE forecast. When the shared information is noisy and the derived MMSE model is not directly applicable, CSF is able to provide reliable forecast and facilitate cost saving. The impact of observation noise, the locations of information providing node in the supply chain, the number of nodes providing information are analyzed. Some managerial insights are gained from this study, which are related

to the value of information sharing and selection of information provider(s) in the supply chain.

To summarize, the contributions of the work include;

- A symbiotic simulation framework is proposed, developed and verified. The framework features the integration of simulations and the Kalman Filtering based data fusion method. The estimation of the real world is improved by extracting a maximum amount of information from both the measurements and the simulation model and the information in an optimal way.
- The strategies and configurations of data fusion methods are studied, including the choices of data fusion methods, the impact of data fusion frequency, data quality, sources of information, and single-data stream vs. multi-data stream.
- The integrated framework is demonstrated in 'proof-of-concept' case studies on representative supply chain problems, including early signaling of exceptional orders, the online monitoring of a complex supply chain system with noisy and partial observations and a system with adaptive inventory management. It is shown that calibrated simulation is able to better sense supply chain dynamics thus enables real-time proactive decision making.

The proposed integrated approach can be applied not only in supply chain management, but also for a wide range of large-scale dynamic systems, such as healthcare delivery management and disease control.

1.5 Dissertation Organization

The remainder of dissertation (Chapter 2, 3 and 4) describe the research efforts of each stage. The dissertation is summarized and future research is discussed in Chapter 5.

Chapter 2

AN INTEGRATED APPROACH TO MONITORING SUPPLY NETWORK DYNAMICS USING KALMAN FILTERING AND CONTROL CHARTS

In today's global market, a critical issue for companies to survive increasing competition is how to manage perturbations in their supply network. Literature indicates that most existing decision tools and models have focused on decision strategies that attempt to mitigate the negative consequences from actual and potential perturbations. Comparing to these "post" activities, we envision proactively sensing the states of the supply network operation is critical to understand the network dynamics and make necessary responses in a timely manner. In this chapter, we explore the application of Kalman Filtering to monitor the dynamic states of the supply network. Further, we integrate the Kalman Filtering with a control chart to enhance the sensitivity of detection. An industry supply network case is presented to illustrate how the integrated approach can be used to improve monitoring the supply network dynamics. We show that Kalman filtering is superior to the non-Kalman Filtering approach in specificity (the proportion of negatives which are correctly identified). Also we demonstrate how to set the time-varying statistical control limits of the re-estimation of system states at each milestone. By choosing the appropriate confidence levels, managers could tradeoff sensitivity and specificity to achieve a favorable detection performance.

2.1 Introduction

A supply network consists of a group of business entities including suppliers, manufacturers, distribution centers and customers. The globalization of the supply network makes it vulnerable to various types of risk associated with the uncertainty or unpredicted events affecting one or more parties in the network. Managing supply chain risk can be very challenging due to the complexity and dynamics of the system. Risks in such systems exist in demand requirements, capacity, delivery time, manufacturing time and cost (Taylor and Brunt 2001), which are usually categorized as internal risk factors. Whereas terrorist attacks, natural disaster, increased competition from opponents are often referred to external risk factors. These risks, if not properly accounted for, will have enormous negative impact on supply network performance. For example, Intel reported that in the fourth quarter of 2006 profit fell 39% from the year-earlier period, to $1.5 billion (Intel New Release, 2007). The profit was hurt by lower average selling prices since Intel was forced to cut prices after AMD gained share in the market for chips used to power PCs and servers. An explosion at Apache's Varanus Island gas plant in Western Australia on June 3, 2008 cut off 30% of the Australia's domestic gas supply and this crisis was estimated to cause $6.7 billion in losses (Gav 2008), assuming energy suppliers are fully recovered by December. These external risks are characteristics of high-impact/low probability situations which makes it difficult to collect historical data for analysis. However, the effect of relatively lower impact internal risks cannot be neglected. Hendircks and Singhal (2003, 2005) reported that firms affected by supply network

disruptions can expect a reduction in their operating income by 107%, return on sales by 114% and return on assets by 92%. A recent study indicated quality, cost, on-time delivery, continuity of supply and engineering/production are the five key factors associated with supply network performance (Wu et al. 2006). For instance, General Motors lost $900 million and Boeing lost $2.6 billion when suppliers failed to delivers components on time. Thus, for supply network managers to successfully assume the responsibility of contingency planning and execution, they must, first and foremost, be able to accurately "sense" subtle changes in network state for making proactive "response".

Recently, a notable growing interest in the area of supply network management, specifically, associated with disruption, is on the solutions that "sense" what is going on in a business, alert users to demand signals, plant failures, supply outages, and similar occurrences. These tools, named Supply Network Event Management have concentrated on the design of the information infrastructure to obtain and update real-time information on imminent events (Buckley 2006). However, the collected data tends to be incomplete despite the huge volume and the data tends to be stored distributed. Moreover, inconsistent data exists due to the fact that noise and uncertainty is often introduced during data collection. The provision of accurate estimates based on such incomplete data or error-prone data is therefore questionable. In this chapter, we first explore the application of Kalman Filtering to monitoring the dynamics of the system under uncertainty. Kalman Filtering is then integrated with a control chart to enhance the sensitivity of detection. A high-end server fulfillment supply network

is studied to demonstrate the effectiveness of the integrated approach. The remainder of this section is organized as follows: Section 2.2 reviews the literature of "sense and response" with focus on sensing in the supply network applications. Section 2.3 presents an introduction of Kalman Filtering with illustrative examples. The proposed integrated approach is detailed in Section 2.4 followed by a case study in Section 2.5. Finally, the managerial insights and conclusions are drawn in Section 2.6.

2.2 Literature Review

In Supply Network Disruption Management, a number of research tasks from addressing the uncertainty in location and transportation problem for supply network design, multi-echelon inventory decisions, shop floor controls, to the linkage between stages such as vendor buyer partnerships have all attracted great attention from both academia and industry. Techniques such as mathematical programming, simulation, network based modeling have proved to be useful for the investigation. For example, Blackhurst et al. (2004) present a network-based approach to model uncertainty in a supply network. The approach allows for the inclusion of stochastic variables so that uncertainty in the operation of a supply network can be modeled. Lu and Li (2006) use Fuzzy sets and possibility theory to model supply networks in an uncertain environment. Lu indicates that decision makers often consider risk by determining the trade-offs between customer service levels and inventory investment in the supply network. Li et al. (2002) model the entities in a supply network from three aspects - functionality, process flow and co-ordination mechanism. Their model aims to capture the complexity

of supply networks under the different scenarios. To quantitatively analyze the impacts of disruptions, some studies focus on the cost analysis of the supply-chain when the disruption takes place. Wu and Blackhurst (2005) present modeling methodologies to manage the synthesis of a supply network by linking hierarchical levels of the systems and to model and analyze disruptions in the integrated supply network. The work is further extended in Wu et al. (2007) to determine how changes or disruptions propagate due to supply network disruptions. Qi et al. (2004) analyze the cost in one supplier-one retailer supply network under a disruption in demand. In addition, Xiao et al. (2005) introduce an indirect evolutionary game model with two-vertical integrated channels to study evolutionarily stable strategies of retailers in a quantity-setting duopoly situation with homogeneous goods and analyze the effects of the demand and raw material supply disruptions on the retailers' strategies.

To our knowledge, most existing research focuses on offline impact analysis and develops mitigation strategies after the perturbations take place. While this approach is useful, we believe that the cornerstone of improved supply network performance is online control of both material and information flow. In the last two decades, there are significant advances in data capturing and analysis systems that offer great potential. For instance, the application of bar codes and RFID has provided an opportunity to create useful databases that can drive better decisions. Also, the developments within telecommunications have complemented other technologies such as microprocessors, EDI, global positioning systems, etc. by providing organizations with access to operational

real-time information never before available. However, the volume of the available data, the noise inherent in the data and the distributed nature of collecting data from the supply network elements, together raise a challenging question, that is: what is an efficient and effective approach to assessing supply network performance given the large volume of data with noise embedded? In 2004, Wu and O'Grady (2004) embed nonlinear Kalman Filtering into a Petri Net model to estimate supply network states. In 2006, Parmar et al. (2006) make the first attempt to use Kalman Filtering for exception detection in the production of a server fulfillment supply network Later, Chen et al. (2006) extend Parmar's work to improve the estimates of job completion times in a whole server fulfillment supply network. While promising, Kalman Filtering in the studies has been used mainly as a technique for predicting delivery due-dates. Even though Upper Control Limits (UCLs) and Lower Control Limits (LCLs) are used to trigger alerts for troubling operations, the control limits are derived from business operation rules and are fixed (e.g., the control limit of the server is defined as 48 hours). Considering the dynamics of a supply network, the control limits should vary as the performance of different stages (tiers) at the supply network is assessed. For example, the threshold for triggering the alert should be larger in the early stages of the supply network and become tighter as the operation proceeds. Such ideas have been well studied in various control chart methods, one of which is termed WineGlass (Wu 1988; Wu et al. 1992). WineGlass has been successfully applied to monitoring the dynamics of financial systems. It statistically tests the consistency between the on-going financial performance

(e.g., sales, expenses) and historical performance, initiates the "off-track" alert when the performance is beyond the control limit. One evident merit in WineGlass is that the UCL and LCL are sensitive to the stages being measured. As it approaches the end, the UCL and LCL get tighter. This motivates the research to integrate Kalman Filtering and time-varying control chart like WineGlass to more accurately and sensitively estimate the supply network states.

2.3 Kalman Filtering Basics

2.3.1 Background

Kalman Filtering is a well-known and often-used tool for stochastic state estimation from noisy measurements. Under certain assumptions, Kalman Filtering is an optimal, recursive data processing or filtering algorithm. It makes an overall best estimate of a state based on all information, i.e., the values of the variables of interest. Such an estimate is achieved by incorporating knowledge about the system dynamics, statistical description of system noise, measurement noise, uncertainty in the dynamic model, and any available information about initial conditions of the variables of interest. However, Kalman Filtering is recursive, that is, not all data needs to be kept in storage and re-processed every time. For example, when a new measurement arrives, information gained in successive steps is all incorporated into the latest result.

In Kalman Filtering, the state $x \in R^n$ of a discrete-time controlled process is estimated through the transition equation and measurement equation.

Transition Equation:

$$x_k = Gx_{k-1} + Iu_{k-1} + \varepsilon_{k-1}, x \in R^n, k = 1...K \quad (2.1)$$

Measurement Equation:

$$z_k = Hx_k + \eta_k, z \in R^m, k = 1...K \quad (2.2)$$

where G is the system state matrix that relates the state at the previous time step k-1(x_{k-1}) to the present step k(x_k). The matrix I relates the optional control input at previous time (u_{k-1}) to the current state x_k. The matrix H relates the system state (x_k) to the measurement (z_k). The random variables ε_k and η_k represent the process and measurement noise respectively. They are assumed to be white noise with independent normal distributions: $p(\varepsilon_k) \sim N(0, Q_k)$ and $p(\eta_k) \sim N(0, R_k)$. Kalman Filtering applies time update equations and measurement update equations iteratively to obtain estimates of the future state of the system. For details of Kalman Filtering, please refer to Kalman (1960). We use a simple example to demonstrate the application of Kalman Filtering to detect the dynamics of a supply network.

2.3.2 Illustrative Example – A simple supply network

Let us consider a supply network consisting of a number of 2nd tier suppliers (T_1), 1st tier suppliers (T_2), transporters (T_3), manufacturers (T_4) and distribution centers (T_5). For one specific order, assume there is a critical path of tasks that will decide the order completion time, T_{order}. The tasks could be the procurement of materials, ordering the shorted components from suppliers, critical manufacturing steps, waiting in queue for available machines, shipment of parts from one site to another, etc.

Without loss of generality, we assume that the time when the order starts the first task is 0, thus, the order completion time $T_{order} = T_1 + T_2 + \cdots + T_5$. When an order is received, two pieces of information are provided to the customers: an expected delivery time $E(T_{order}) = E(T_1) + E(T_2) + \cdots + E(T_5)$ and maximal possible delay represented as a control limit $r \geq 0$. The end of each task is considered as a "time" milestone when we estimate the system state and detect potential tardy orders. For this simplified example, the system state x_k refers to the estimated completion time of task T_k for the order (k=1,..5). Using the updated x_k, the order completion time is re-estimated and a control chart is drawn with UCL being fixed, that is, UCL = $E(T_{order}) + r$. Note UCL is of interest in this study as we assume the tardiness of order delivery is critical in the supply network. The application of Kalman Filtering to the problem is outlined as following, where $E(T_k)$ and $V(T_k)$ are the mean and variance of T_k:

Step 1: The initialization at the first task (T1). $\hat{x}_k^-$ is called the "prior estimation of order completion time", and P_k^- is the prior estimation error variance:

$$\hat{x}_k^- = E(T_1) \tag{2.3}$$

$$P_1^- = V(T_1) \tag{2.4}$$

Step 2: For task k (k >1), prior estimation of task completion time and its variance at the beginning of each task is calculated based on prior estimation of order completion time at previous stage ($\hat{x}_k^+$):

$$\hat{x}_k^- = \hat{x}_{k-1}^+ + E(T_k) \tag{2.5}$$

$$P_k^- = P_{k-1}^+ + V(T_k) \tag{2.6}$$

When the order starts task k, the system state increases by the expected time of T_k , $E(T_k)$, and the variance of the system state increases by the variance of T_k , $V(T_k)$.

Step 3: Posterior estimation of task completion time and its variance after observing the real completion time of the task, denoted by $\hat{x}_k^+$ and P_k^+ , are calculated as follows:

$$K_k = P_k^- / (P_k^- + R_k) \tag{2.7}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \tag{2.8}$$

$$P_k^+ = (I - K_k)P_k^- \tag{2.9}$$

After the task k is finished, we use the observed completion time point of task z_k, to update the order completion time and its expected variance by (2.7)–(2.9). Kalman Filtering tolerates the measurement error in the observation z_k and R_k denotes the variance of error. When $R_k = 0$, that is, the observation is accurate, the updated system state $\hat{x}_k^+$ equals to the observation z_k. When R_k increases, $\hat{x}_k^+$ tends to get closer to $\hat{x}_k^-$.

Step 4: Re-estimate the order completion time and determine if the order is on track.

$$E(T\_order|x_k) = \hat{x}_k + \sum_{l=k+1,\ldots,} E(T_l) \tag{2.10}$$

With the "posterior estimation of task completion time" $\hat{x}_k$, we re-estimate the order completion time by (2.10) and plot it in a control chart (Figure 1). As shown in Figure 1, assuming the expected delivery time is 60 days and r is 2 days, the upper limit is 62 days. At the end of T1, we estimate the completion time is 58

days. It increases to 63 days at the end of T4. Since 63 days is out of UCL (62 days), a signal will be triggered at the end of T4 to draw the manager's attention.

Step 5: k=k+1, return to Step 2 until k=5, then stop.

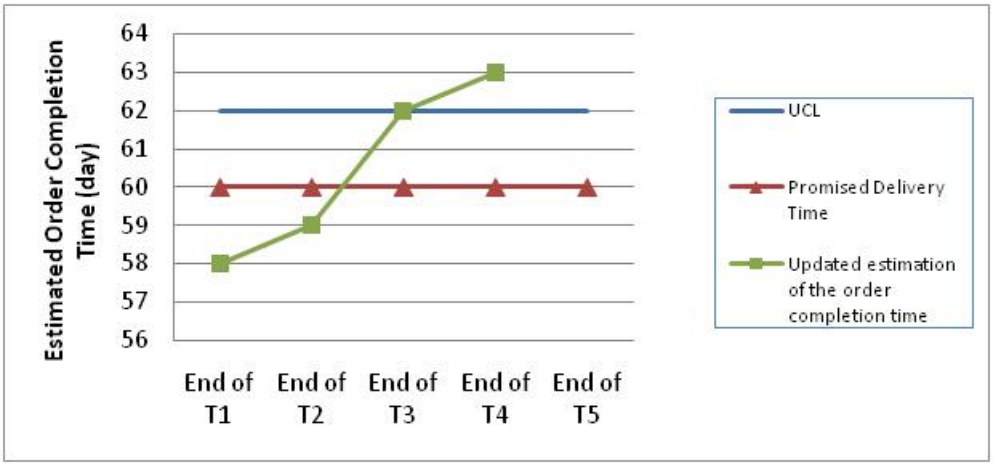


Figure 1. Example of a control chart of Kalman Filtering approach

Clearly, Kalman Filtering has focused on fusing data to improve the estimations of the system states from both the system transitions (expected supply network performance) and measurements (observed supply network performance) which inherently have errors. To apply Kalman Filtering to detecting abnormalities in a supply network, control decisions on whether the current state implies an exception should be made. Earlier research explores the use of a fixed control limit for control decisions (Chen et al. 2006; Parmar et al. 2006). The challenge lies in identifying the control limit. If the control limit is small, too many false alarms will likely be triggered. On the other hand, fewer true positive alarms might be triggered as the order approaches the end if the control limit is large. Thus, the control limit needs to be sensitive to the current stage of the operation. Proposed herein is an integration of Kalman Filtering with the

WineGlass, a control chart that focuses on the determination of a time-dependent UCL with respect to a given confidence level.

2.4 Proposed Approach – Integrating Kalman Filtering with WineGlass

2.4.1 Wineglass Background

WineGlass is a statistical process control method, used to track whether the stage-by-stage state is consistent with the overall plan (Wu 1988). When monitoring a process, the overall time is divided into a number of small time periods. At the end of each time period, we set a "time" milestone k (k = 1, … K) and measure the performance. Assume the performance target for the time period between the milestone k-1 and k is denoted by U_k. Let u denote the actual performance in the period. The actual-to-track ratio (g_k) at the milestone k is calculated as:

$$g_k = \sum_{i=1..k} u_i \,/\, \sum_{i=1..k} U_i, \text{ k=1..K} \qquad (2.11)$$

WineGlass assumes that u_k follows a normal distribution $N(gU_k, w^2 g^2 U U_k)$, where $U = \sum_{i=1..K} U_i$, $g = \sum_{i=1..K} u_i \,/U$, and w^2 is a constant coefficient obtained from historical data. The structure of the variance $w^2 g^2 U U_k$ is based on two assumptions: 1) variance of u_k is proportional to U_k and 2) w^2 is a dimensionless coefficient representing the dynamics of the process. More details about these assumptions are discussed in (Wu 1988). Later, Wu et al. (1992) determine the conditional distribution of g_k is normal with the mean and variance in (2.12) and (2.13). Since g_k is the actual-to-ratio rate of state at the end of the process, $g_k = 1$ means that the final state is exactly the same as planned. The WineGlass technique then generates the upper limit ($UCLg_k(q)$) and lower limit

($LCLg_k(q)$) for the "time" milestone k (k=1,…K) given a confidence level, q, as follows:

For k=1,2,…K,

$$E(g_k \mid g_K = 1) = 1 \tag{2.12}$$

$$\text{var}(g_k \mid g_K = 1) = w^2 (U_{k+1} + ... + U_K) / (U_1 + ... + U_k) \tag{2.13}$$

$$UCLg_k(q) = E(g_k \mid g_K = 1) + z_{q/2} \sqrt{\text{var}(g_k \mid g_K = 1)} \tag{2.14}$$

$$LCLg_k(q) = E(g_k \mid g_K = 1) - z_{q/2} \sqrt{\text{var}(g_k \mid g_K = 1)} \tag{2.15}$$

where $z^{q/2}$ is the z-score for type I error rate (1- q/2). At the "time" milestone k, if $g_k \notin (LCLg_k(q), UCLg_k(q))$, WineGlass concludes g_k has abnormally deviated from expected states. As shown in (2.14) and (2.15), the control limits are associated with the time period k. As k increases, the variance decreases, the $LCLg_k(q)$ and $UCLg_k(q)$ are getting closer to the expected states (as shown in Figure 2).

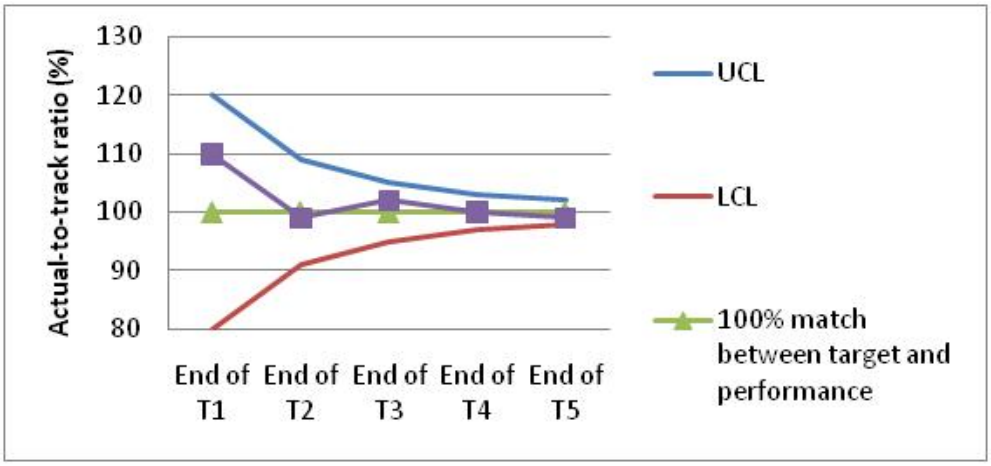


Figure 2. Example of a Wineglass chart

2.4.2 Kalman Filtering and WineGlass Integration for Supply Network State Monitoring

The integrated approach extends Kalman Filtering by relying on the WineGlass to make time dependent control decisions. Since we are only interested in tardy orders, we set one-side control limits in our study. First, we will estimate parameters for WineGlass which can be used to generate the dynamic $UCLg_k(q)$ with respect to actual-to-track ratio g_k. Secondly, the $UCLg_k(q)$ is converted to UCL_k' for the completion time of operation task T_k. Adding UCL_k' with the estimated time of uncompleted tasks, we could get the stage-dependent UCL_k for the order completion time. Finally, Kalman Filtering is used to estimate the supply network states which are judged by UCL_k.

In WineGlass, w^2, a constant coefficient representing the dynamics of the system, is estimated from historical data. The larger variance of the process is, the larger w^2 is. Let us assume the historical information of n orders with each order going through m tasks is available. Assume $T_{j,k}$ (j=1,…n, k =1,…m) denotes the expected completion time for task k of the j^{th} order, and $Y_{j,k}$ (j=1,…n, k =1,…m) denotes the actual time taken by task k of the j^{th} order. The best estimation of w^2 is calculated as $\tilde{w}^2$:

$$\tilde{g}_j = \frac{\sum_{k=1,...m} Y_{j,k}}{\sum_{k=1..m.} T_{j,k}}, j = 1,...n \tag{2.16}$$

$$\tilde{w}_j{}^2 = \frac{1}{(m-1)} \frac{\sum_{k=1..m} (Y_{j,i} - \tilde{g}_j T_{j,k})^2}{\tilde{g}_j{}^2 \sum_{k=1..m} T_{j,k}}, j = 1,...n \tag{2.17}$$

$$\tilde{w}^2 = \frac{\sum_{j=1..n} \tilde{w}_j^{\,2}}{n} \tag{2.18}$$

Next, we need to determine U_k, the targeted completion time of task k. Let us assume $E(\mathrm{T_k})$ represents the expected completion time of task k, the expected order completion time is $E(T_{order})$, the order delay control limit r is allowed delay of a normal order at the last stage, given order tardiness is of interest in this study, we define:

$$U_k = E(T_k) \times (1 + \frac{r}{E(T_{order})}) \tag{2.19}$$

From (2.12) - (2.14), the UCLgk with the confidence level q is calculated as:

$$UCLg_k(q) = 1 + z_q \tilde{w} \sqrt{\frac{U_{k+1} + ...U_K}{U_1 + ... + U_k}}, \; k = 1, 2.... \tag{2.20}$$

where z_q is the z-score for type I error rate =1- q

The confidence level q controls the sensitivity of the WineGlass chart. When q is 0, there is no tolerance for the difference between target and actual performance. When q becomes larger, the control chart becomes less strict thus fewer orders will be estimated to be tardy. Given confidence level q, $\mathrm{UCLg_k(q)}$ represents the control limit on the ratio, that is, the probability current performance is on track. Since Kalman Filtering is applied to estimate the supply network states in terms of order completion time, we convert $\mathrm{UCL_k(q)}$ to the control limit on the task completion time by (2.21), and obtain the control limit on the order completion time by (2.22):

$$\mathrm{UCL'_k(q) = UCLg_k(q) \times \sum_{i=1..k} U_i} \tag{2.21}$$

$$\mathrm{UCL}_k(\mathrm{q}) = \mathrm{UCL}'_k(\mathrm{q}) + \sum_{l=k+1,\ldots,} E(T_l) \quad (2.22)$$

In the integrated monitoring approach, Kalman Filtering updates the estimation of task completion time following the procedure as described in Section 2.3.2, so that the order completion time is re-estimated at the end of each task. The control limits derived from WineGlass are used to make the control decisions. The next section provides the examples and experimental results for Kalman Filtering and the integrated approach in a server fulfillment supply network. We focus on the manufacturing stages in a server assembly factory to validate our approaches. However, the approaches are general enough to apply to the whole supply network.

2.5 Industry Case Study

To validate the proposed approach, a high end server fulfillment supply network of a multinational computer technology and IT consulting corporation is studied. The production of computer server is facing a big challenge to meet quotas at the end of each quarter (Parmar et al. 2006). This server fulfillment supply network is complicated, including customers, server fulfillment production, a peripheral supply of Integrated Circuits (ICs), and warehouses. From the end-customer point of view, good service means on-time receipt of servers ordered and the required quantities (Fordyne 2001). However, on-time delivery is not always assured. A delay in any intermediate supply chain entity not only influences the fulfillment of specific orders, but also impacts other orders in the manufacturing line. For example, one of main issues that affect the performance of server fulfillment supply network is the quality of Integrated

Circuits (ICs) from the supplier. In some cases, when ICs are installed into the server, defective ICs are not detected until the whole server assembly passes through testing. Once defective servers are identified, they reenter the production line which requires additional production capacity and materials (chips), thus, the fulfillment is delayed and it potentially causes congestion of the production line.

2.5. 1 Problem Specification

A simplified representation of the general process in the server fulfillment center is shown in Figure 3. The ICs provided by outside suppliers along with other peripherals are put on the boards in Panel Assembly. After panel assembly, the assembly is tested and is put together with addition peripherals to form a basic untested server system. The basic untested server system then goes for system test. After the test, the server system is disassembled (Dekit) and the resulting tested components are put into storage to fulfill a future customer order. Once a customer order is issued by the server demand generator, servers are configured depending on the actual requirements. The configured customer servers are then tested and sent to packing and shipping.

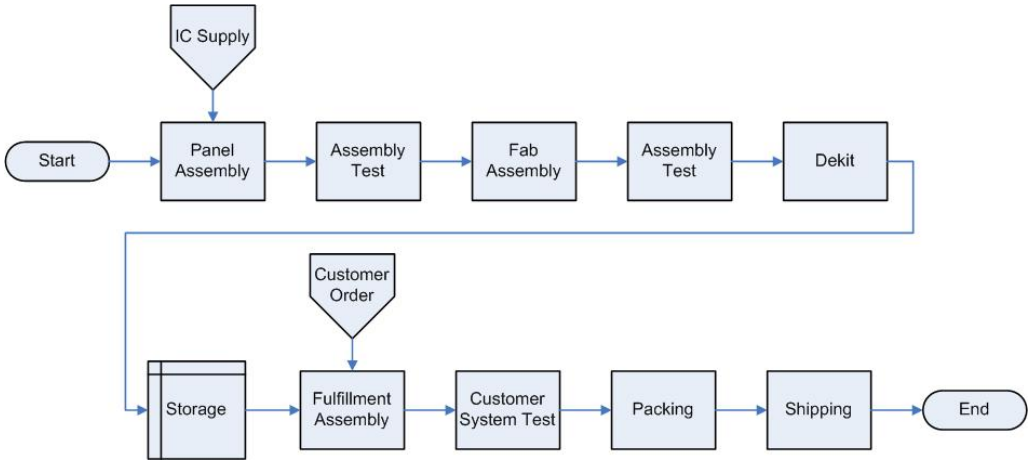


Figure 3. General manufacturing process in server fulfillment center

2.5.2 Simulation and Performance Metric

In this study, we develop a simulation model to collect data for validation. Specifically, the simulation models 6 out of 10 server manufacturing stages, a six step serial stage that is identified as show in Figure 4 (P_A to P_F). The simulation is implemented using SIMUL8®. Based on industrial data, the processing time distributions for all six processes are assumed triangular with the parameters shown in Table 1. The time between arrivals of entities (orders) is exponentially distributed with an average of 14.9 hours.

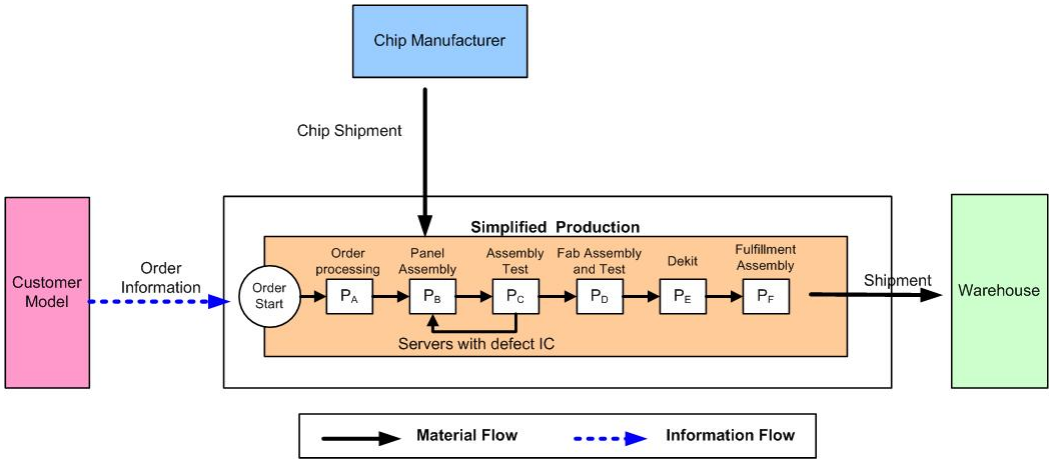


Figure 4. Six-stage serial process of simplified production (P_A to P_F)

Table 1. Assumed parameters of triangular distribution for processing time and number of machines at each process (Parmar et al. 2006)

Process	Low (in hours)	Mod (in hours)	High (in hours)	Resources
A	6	8	17	1
B	30	60	75	5
C	18	35	92	5
D	30	60	80	6
E	8	13	17	2
F	25	45	60	4

In each simulation run, 1000 orders are released and the first 500 orders are used as a warm-up period and the second 500 orders are used for the study. We first collect data from 30 simulation runs as historical data to obtain the estimates of the mean and variance of the completion time of the orders. Then, another run is conducted as an emulation, which mimics the actual processing of incoming orders. The 30 simulations are used to characterize the transition process, and the emulation run represents the measurement in the context of Kalman Filtering.

Assume there is a rework route between process B and process C because of the defect issue of one type of IC required in process B. After installing the chip into the products (XYZ-servers in this study), the testing process of the XYZ-server will determine whether this chip fails or passes. If this chip fails, the current XYZ-server which has a failed chip needs to be reworked and a new chip is requested for process B. Since this study aims to study how to sensitively initiate the triggers for out of control, we purposely differentiate emulation from

simulation, that is, we set the chip defect rate to be zero for the 30 simulation runs, and 1% defect rate for the emulation. The study on the impact of different defect rates will be reported separately. Meanwhile, we assume the promised order time is the average order completion time from the 30 simulation runs. To define a true tardy order, we apply control limit r (in hours) on the emulation results, i.e. if the completion time of an emulated order is later than the promised order completion time by r hours, it is considered to be a true tardy order. Both Kalman Filtering and the integrated approach monitor each order at each stage. They will decide whether to send out a signal to mark the order as an abnormal (potential tardy) order stage by stage. Thus, at each stage, let:

- CS: True tardy order correctly signaled as being tardy
- FS: Non-tardy order wrongly signaled as being tardy (Type I error)
- FN: True tardy order wrongly non-signaled (Type II error)
- CN: Non-tardy order correctly non-signaled

We introduce two metrics: sensitivity (2.23) measures the proportion of actual tardy orders that are correctly identified and specificity (2.24) measures the proportion of the non-tardy orders that are correctly identified.

$$\text{Sensitivity} = \frac{\#\text{ of CS}}{\#\text{ of (CS + FN)}} \tag{2.23}$$

$$\text{Specificity} = \frac{\#\text{ of CN}}{\#\text{ of (CN + FS)}} \tag{2.24}$$

A Receiver Operating Characteristic (ROC) chart (Hanley and McNeil 1982) is used as a graphical plot to demonstrate the performance of Kalman Filter, the integrated approach in terms of sensitivity and specificity. The horizontal axis

of ROC chart is 1-specificity, and the vertical axis of an ROC chart is sensitivity. A good monitoring performance means both high sensitivity and high specificity, which is denoted by a point close to (0,1) in a ROC chart. The larger the area under ROC curve is, the better the monitoring performance is.

2.5.3 Experimental Results

We collect the statistics of mean and variance of the time that an order spends in each manufacturing stage from the 30 simulation runs (Table 2). Also, we use the simulation data to estimate the WineGlass coefficient w^2 following (2.16) - (2.18), which is 0.01437.

In the emulation run, we assume the order delay tolerance r is 24 hours. In this case, we have 109 true tardy orders among the total 500 orders.

Table 2. Mean and variance of processing time in each stage (in hours)

	Process A	Process B	Process C	Process D	Process E	Process F	T_order
E(T)	21.44	59.82	49.53	57.69	13.54	47.89	249.91
Q(T)	30.01	125.89	263.63	117.86	8.82	98.15	644.36

Experiment I – Comparison Study on Approaches with and without Kalman Filtering

To apply Kalman Filtering for supply network dynamics detection, important parameters include mean and variance of processing time (Table 2) and white noise from measurements (represented as the emulation measurement errors, R_k). This set of experiments is conducted to (1) demonstrate the advantage of applying Kalman Filtering to monitor system dynamics, and (2) study the performance of

Kalman Filtering based detection, under varied measurement noise R_k over [0, 16, 36, 64, 100, 144, 196]. To signal potential tardy orders, we use the fixed UCL, which equals to the historical order completion time adding a predefined tolerance, that is, $E(T_{order}) + r$ = 249.01 + 24 = 273.91 hours at each stage. Table 3 and Table 4 summarize the sensitivity (Se) and specificity (Sp) of detections at the end of process A through E given the emulation measurement errors. The results of the detection based on the direct use of noisy emulation observations are shown in Table 3, and the results of detection using the fused simulation and emulation data based on Kalman Filtering are shown in Table 5. As seen from (2.7) - (2.9), the updated state estimations equal to the emulation observations when R_k =0, that is why the results from Table 3 and Table 4 are the same for R_k =0.

Table 3. Experiment I - Non-Kalman Filtering approach (r=24 hours)

R_k	Process A		Process B		Process C		Process D		Process E	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0	0.21	0.98	0.43	0.97	0.70	0.96	0.78	0.96	0.77	0.96
16	0.18	0.98	0.45	0.96	0.69	0.96	0.75	0.96	0.73	0.96
36	0.22	0.97	0.48	0.96	0.72	0.96	0.75	0.94	0.72	0.94
64	0.21	0.96	0.42	0.96	0.65	0.94	0.73	0.93	0.75	0.94
100	0.22	0.95	0.43	0.94	0.64	0.94	0.72	0.94	0.74	0.94
144	0.17	0.95	0.46	0.93	0.67	0.93	0.74	0.94	0.72	0.92
196	0.28	0.92	0.49	0.92	0.60	0.91	0.73	0.91	0.70	0.91

Table 4. Experiment I – Kalman Filtering approach (r=24 hours)

R_k	Process A		Process B		Process C		Process D		Process E	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0	0.21	0.98	0.43	0.97	0.70	0.96	0.78	0.96	0.77	0.96
16	0.10	0.99	0.40	0.97	0.66	0.96	0.76	0.96	0.75	0.96
36	0.01	1.00	0.37	0.98	0.69	0.97	0.73	0.96	0.70	0.97
64	0.00	1.00	0.27	0.98	0.62	0.97	0.72	0.96	0.76	0.97
100	0.00	1.00	0.27	0.99	0.57	0.97	0.75	0.96	0.73	0.95
144	0.00	1.00	0.12	1.00	0.57	0.97	0.67	0.97	0.74	0.96
196	0.00	1.00	0.14	0.99	0.50	0.97	0.65	0.96	0.70	0.96

Figure 5 shows the ROC charts obtained with and without Kalman Filtering for R_k = 16, 36, 64, 100, 144, 196. We conclude that the overall performances of both methods improve as processes move from A through E. This implies that the identifications of the tardy orders are more accurate as the order approaches the last process. Secondly, the Kalman Filtering approach consistently outperforms the approach without Kalman Filtering under various observation errors (R_k). Considering the sensitivity (vertical axis), both approaches achieve comparable results, however, the specificity (horizontal axis) of Kalman Filtering is always better. This indicates the approach without Kalman Filtering tends to misclassify non-tardy orders more thus generate more false alarms. We also observe from Figure 5 that with the increase of R_k , the performance differences between the two approaches increase. We conclude that in the cases there exist larger observation noise the approach without Kalman Filtering deteriorates greatly. While promising, both approaches have lower

sensitivity (less than 80%). Therefore, in the second experiment, we explore the integration of Kalman Filtering with WineGlass for improved performance.

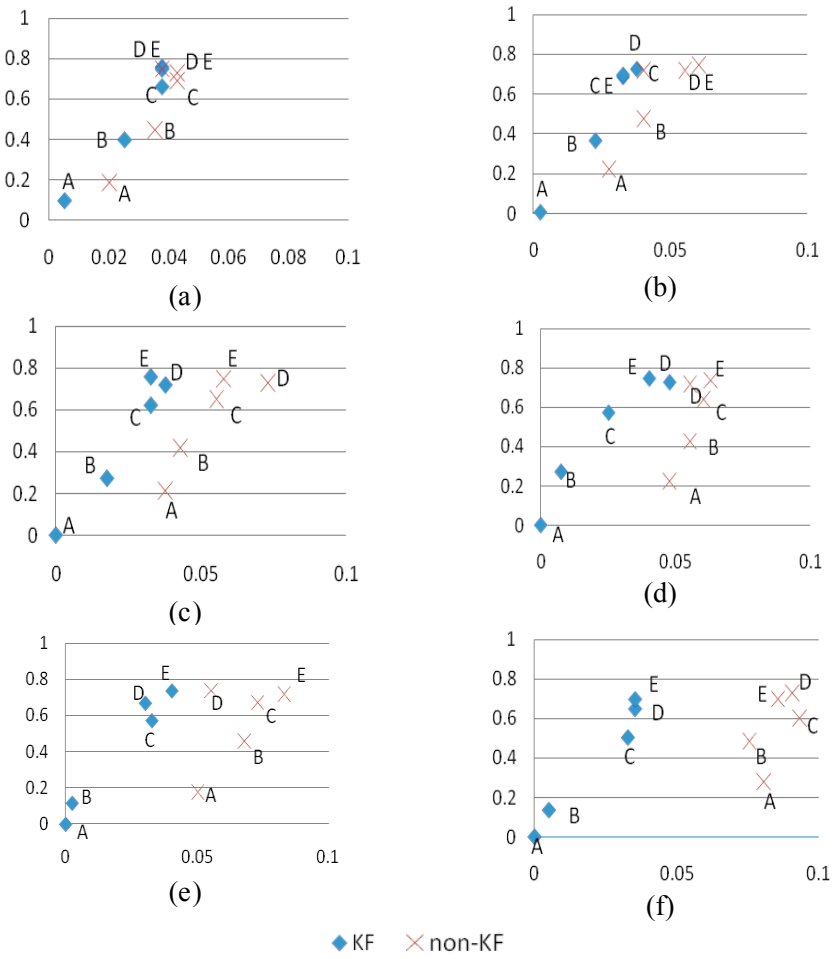


Figure 5. The ROC charts of Kalman Filtering vs. Non-Kalman Filtering approach (R_k = 16, 36, 64, 100, 144, 196)

Experiment II – Comparison Study on Kalman Filtering and Integration Kalman Filtering with WineGlass

From Table 2, the average variance of each process is 107. To reduce the bias on observation, we decide to use R_k =100 to test the performance of the

integrated approach. In the integrated approach, the UCL'$_k$ is determined jointly by the confidence level q and control limit r. To keep consistency with Experiment I, assume r=24 hours. First, based on the mean processing time of each stage (shown in Table 2), we calculate the targeted completion time U_k using (19). Note $E(T_{Order})$ = 249.91 is determined by assuming the mean processing time of the six tasks. We obtain U_A=23.50, U_B=65.57, U_C=54.28, U_D=63.23, U_E=14.84, U_F=52.49. Next, we calculate the UCLg$_k$, UCL'$_k$ and UCL$_k$ following (2.20) - (2.22). For example, assuming q=0.9, the corresponding z-score=1.28, for process A, we get:

$$UCL - g_A = 1 + z_q \tilde{w} \sqrt{\frac{U_B + ...U_F}{U_A}} = 1 + 1.28 * \sqrt{0.01437 * \frac{65.57 + ...52.49}{23.5}} = 1.5$$

$$UCL'_A = UCL_A \times \sum_{i=A..F} U_i = 1.5 * 23.5 = 35.25$$

$$UCL_A = UCL'_A + U_B + ...U_F = 285.7$$

The UCL of each process with r=24 and q=0.9 is summarized as follow.

Table 5. UCLs for order completion time at the end of Processes A-F (r =24, q=0.9)

	Process A	Process B	Process C	Process D	Process E	Process F
UCL	285.70	293.63	294.93	292.03	290.48	273.91

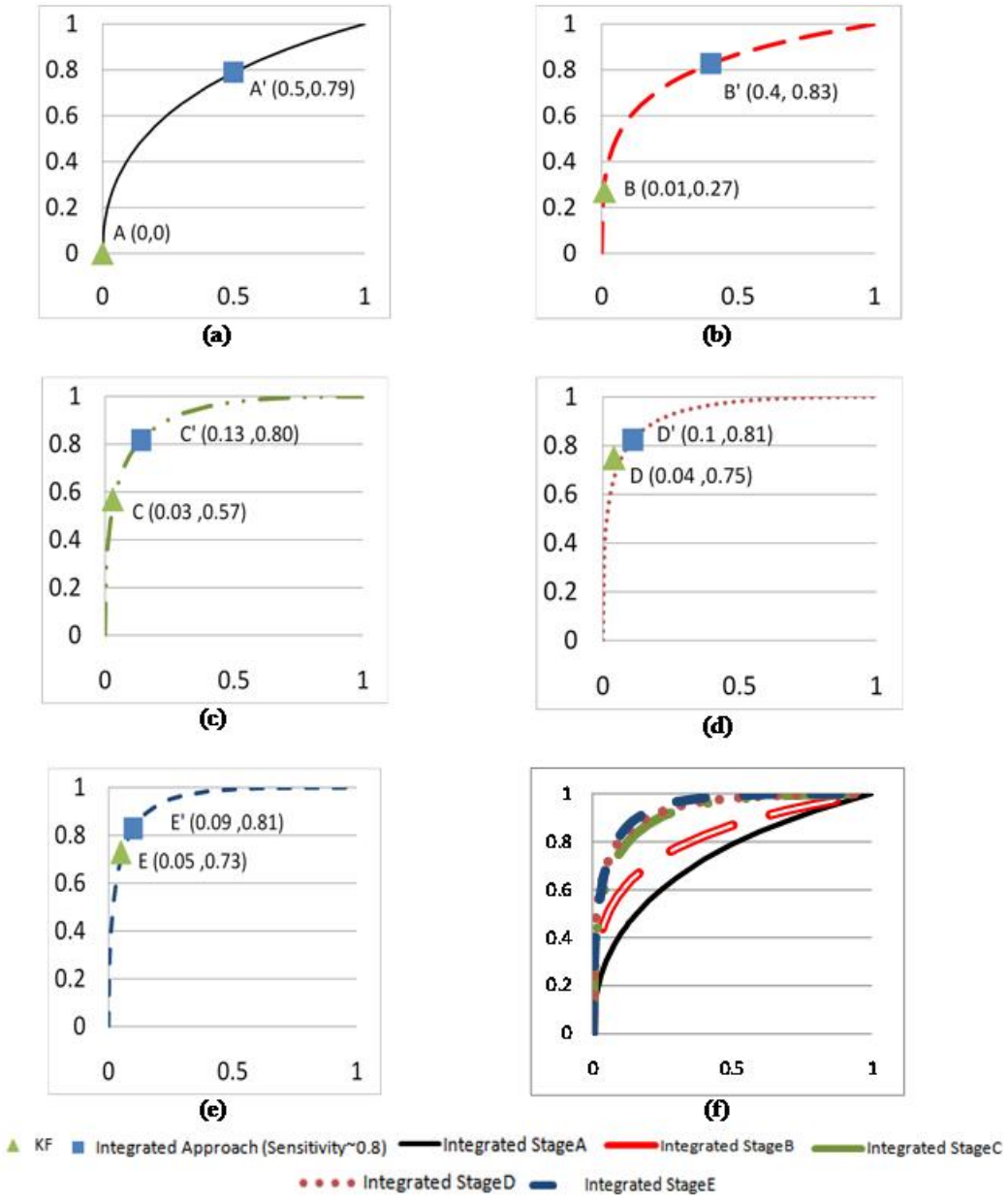
Table 6. Experiment II – the integrated approach (R_k=100, r=24 hours)

	Process A		Process B		Process C		Process D		Process E	
q	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0.01	1.00	0.00	0.99	0.06	1.00	0.27	1.00	0.47	1.00	0.53
0.05	1.00	0.00	0.92	0.31	0.98	0.48	0.95	0.65	0.97	0.71
0.1	1.00	0.02	0.92	0.43	0.97	0.58	0.93	0.72	0.93	0.77
0.2	0.91	0.18	0.83	0.58	0.94	0.71	0.92	0.81	0.90	0.85
0.3	0.83	0.50	0.80	0.74	0.89	0.80	0.87	0.87	0.88	0.89
0.4	0.63	0.75	0.68	0.83	0.83	0.85	0.83	0.91	0.83	0.91
0.5	0.46	0.87	0.57	0.89	0.76	0.89	0.79	0.94	0.81	0.93
0.6	0.28	0.95	0.49	0.93	0.68	0.93	0.75	0.96	0.73	0.95
0.7	0.15	0.99	0.39	0.96	0.58	0.96	0.64	0.97	0.67	0.97
0.8	0.05	1.00	0.27	0.99	0.49	0.99	0.53	0.99	0.56	0.99
0.9	0.00	1.00	0.15	1.00	0.30	0.99	0.37	1.00	0.40	1.00
0.95	0.00	1.00	0.15	1.00	0.28	1.00	0.29	1.00	0.37	1.00
0.99	0.00	1.00	0.01	1.00	0.03	1.00	0.11	1.00	0.13	1.00
KF	0.00	1.00	0.27	0.99	0.57	0.97	0.75	0.96	0.73	0.95
Non-KF	0.22	0.95	0.43	0.94	0.64	0.94	0.72	0.94	0.74	0.94

Varying q from 0.01 to 0.99, we can observe the tradeoff between true position rate (sensitivity) and false positive rate (1-specificity) of the integrated approach (Table 6). Note the last two rows in Table 6 are the results from Kalman Filtering and non-Kalman Filtering approaches. We use ROCKIT to fit bi-normal ROC curves. Figure 6 shows the resulting ROCs of detections by the integrated approach at process A to E. In each figure, the performance of the Kalman Filtering based approach with fixed control bound = 273.91 is highlighted with ▲. We also mark the point on the integrated approach by ■ where sensitivity is

close to 0.8. It is observed that the approach with fixed control bound suffers low sensitivity in the early process (Figure 6a). As the sensitivity improves from process A through E, the specificity decreases (Figure 6b-e). In the integrated approach, q can be used to balance sensitivity and specificity. Since high-end server costs $1M and even higher, the supply network manager will focus more on satisfying customers' order and reducing any delays. Thus, sensitivity may be more important than specificity which drives the need for the integrated approach, especially, at the earlier processes of the supply network. The points marked by ■ in Figure 6(a)–(e) are the examples when monitoring the processes with sensitivity ≈ 0.8. Figure 6(f) combines the ROC chart and Kalman Filtering solution of each process. The results confirm that as the operation progresses from A through F, the overall performance improves. Figure 6(f) compares the ROCs at all five processes. Area Under Curve (AUC) is usually used as an indicator to assess the detection performance (Hanley and McNeil 1982). The results indicate AUC increases from 0.73, 0.82, 0.92, 0.95 to 0.96 for process A through E.

In summary, our first experiment was conducted to compare the approach with Kalman Filtering and without Kalman Filtering. We conclude Kalman Filtering achieve comparable sensitivity but outperforms greatly in specificity. Due to the fact that sensitivity may be more important for high-end server supply network, we explore the integration of WineGlass to Kalman Filtering to balance sensitivity and specificity. Experimental results indicate that sensitivity of the earlier stages of the supply network can be greatly enhanced by the use of WineGlass.



Horizontal -axis: 1- specificity. Vertical -axis: sensitivity.

Figure 6. ROC Charts for Process A, B, C, D, E and all five processes

2.6 Conclusions and Future Work

Most existing research in responsive supply network management has focused on mitigation strategies during the onset of disturbances/perturbations. This research attempts to proactively sense the system status, which enriches the practices in the field of "sense and response". We study Kalman Filtering, a control theory based approach in applying for system status sensing. Noting the undesirable property of a fixed control bound used in the Kalman Filtering approach, we integrate Kalman Filtering with a control chart method called WineGlass for better detection of abnormalities. A simulation model representing a server fulfillment supply network is developed to validate the proposed approach. In our experiments, 30 simulation replicates are made to collect processing statistics and a single emulation run is conducted to represent the real-time measurement. Thus, fusing data from simulation and emulation based on Kalman Filtering and an approach integrating Kalman Filtering and WineGlass are studied. We conclude that Kalman Filtering greatly improves the specificity and its sensitivity can be then enhanced by employing the WineGlass method. In the future, we will conduct more experiments including different defect rates (in this study, we set 1% for emulation). We will further explore other time sensitive control chart methods. In addition, we will apply the approach to other service industries such as retailer businesses, and/or transportation systems.

Chapter 3

MONITORING AND STUDYING SUPPLY CHAIN DYNAMICS: A DATA FUSION APPROACH

Effective supply chain management is becoming increasingly important for companies to be competitive in today's global market. It is a challenging task to manage a supply chain due to the complexity and dynamics inherent in the system. Though a great amount of research efforts have spent on modeling supply chains, the accuracy of models has always been of concern. First, the estimated parameters for the model may not be accurate. Secondly, the assumptions for the model formulation and model initialization may not truly reflect the dynamics in a real supply network operation. Another level of difficulty added is that the dynamics and disturbance may not be well accounted in developing and initializing the models. To address these issues, the application of data fusion techniques to real-time calibrate the running models is proposed. Specifically, a simulation based on discrete time difference equation model for a well studied beer game supply chain is implemented. Two data fusion methods including Extended Kalman Filtering and Ensemble Kalman Filtering are explored to monitor the supply chain dynamics. The advantages and disadvantages of each calibration method are analyzed under varied system dynamics, different levels of information availability and quality of the real time data. Experimental results indicate that dynamic data calibrated simulation model is significantly improved in terms of prediction accuracy. With appropriate tuning, the accuracy

improvement up to 80% can be achieved in chaotic system; and up to 60% in stable systems.

3.1 Introduction

A supply chain is a system of business enterprises that are linked together to satisfy consumer demand (Riddalls 2000). It usually consists of multiple echelons, each can be considered as a generic production/distribution process to add value to the final product. In simplified models, goods flow from one echelon to the next till they reach the end-customer. In more realistic models, a supply chain is a network with parallel flows (both information and material) between different echelons, the inter-relationships between entities are more complicated and will evolve over time (Li et al. 2009).

Quickly changing business environment including global competition, shorter product life cycles, dynamic changes of demand pattern, product varieties and environmental standards along with the increased emphasis on customer-focused strategies calls for an effective supply chain management. Yet the challenges facing the decision makers are the prevalence of uncertainties stemmed from the suppliers, the customers and/or the manufacturers. Research has explored ways to model the uncertainties in order to alleviate the negative impacts. As a result, supply chain risk management has gained great attention lately. Tang (Tang 2006) provides a comprehensive overview and categorizes the supply chain risk research based on supply, demand, product, information management as well as risk mitigation strategies. Most research is based on the assumptions that the risk and uncertainties are known as a prior. This limits the

implementation of adaptive supply chain which is designed to make real time adjustment in responding to the changing environment. Instead of assuming the uncertainties are known, the key of adaptive supply chain is to "sense/know" the operating environment so that timely responses can be made. One example is "sense and response" enterprise proposed by Haeckel (1999) in which the system can identify the changes in customer demands and business environment "as they happen" and make appropriate response to capture new opportunities. The failure to adapt to the on-going changes potentially leads to the mismatching of supply and demand, which often causes dramatic financial loss (Hendricks and Singhal 2005). Fortunately, the advancement in information and communication technologies (Gunasekaran and Ngai 2004) nowadays makes large amount of data readily available for the participants in the supply chain to detect the supply chain dynamics. The challenges from this data rich environment are lying in the large quantity of data and data quality (e.g., incomplete data, noisy data and conflicting data). Thus, an adaptive supply chain requires a robust sense method to ensure the effectiveness of responses.

In this chapter, two data fusion techniques based on Kalman Filtering are studied for better sensing the supply chain dynamics. Kalman Filtering is known as a technique for model calibration by recursively fusing data from real measurement into model. Parmar et al. (2006) make the first attempt to study the application of Kalman Filtering to detect product quality in a server fulfillment supply network with the focus on the production line. In Chapter 2, we propose the use of Kalman Filtering to monitor a simplified server supply network

performance (e.g. delivery date) with various uncertainties, such as supplier quality, demand variation. While promising, Kalman Filtering is designed for use in linear state system which may limit its applications to a complex nonlinear supply network. Therefore, Extended Kalman Filtering and Ensemble Kalman Filtering (Evensen 2006) for nonlinear system are of particular interests in this study. Two sets of comparison experiments are conducted to gain the insights on the performance of each in supply chain model calibration. The supply chain model studied is a four-echelon beer game model consisting of a retailer, a wholesaler, a distributor and a manufacturer. We are particularly interested in this model because it can present different levels of chaos and dynamics depending on the parameter setting (Hwarng and Xie 2008; Larsen et al. 1999) which will make the comparison more meaningful.

The chapter is organized as follows. Section 3.1 summarizes the literature in supply chain dynamics and data fusion. It is followed by the introduction of Extended Kalman Filtering and Ensemble Kalman Filtering. The detailed experiment studies and analysis are explained in Section 3.4.The conclusions with future directions are provided in Section 3.5.

3.2 Supply Chain Dynamics

The dynamics and complexity of a supply chain increase uncertainties which potentially lead to chaos within the network (Wilding 1998). There exist notable research efforts to capture and analyze supply chain dynamics. Riddalls et al. (Riddalls 2000) categorize the modeling methodologies into four: discrete event simulation systems, operation research models, continuous time differential

equation models and discrete time difference equation models. Operation research models such as linear programming, nonlinear programming, queuing theory, stochastic programming suffer from their limits in scalability. That is, the large-scale supply network introduces a large number of variables which prohibits the applications of these operation research methods. Secondly, the convergences of nonlinear programming model have always been an issue. Thirdly, most optimization techniques provide less insight on how system parameters affect the solution. In comparison, discrete event simulation is well positioned to take account of stochastic factors. However, it requires extensive domain expertise and time to develop discrete event simulation models. In addition, though discrete event simulation models are known to help understand the system dynamics, whether the model is sufficiently accurate and precise to represent the real-world scenarios has always been a difficult question to tackle. When there are divergences between simulation results from real system observations, no simple strategy exists to reverse the result deviation by adjusting the initial setting and simulation parameters to "rebuild and restart" the model. To monitor and control the system in real-time, models based on differential equations are more appealing. This is because many influential characteristics can be explicitly presented in the differential equations. Tools and methodologies can then be implemented to gain insights into the system dynamics. Consider that the dynamics of any supply chain are fundamentally discrete in nature. That leaves discrete time difference equation model, a hybrid approach which enables classical controls within a discrete framework, a suitable method to model supply

chain dynamics. Such model could be further enhanced through the calibration using real time data as shown in this chapter.

However, fusing real time data is not without challenges because (1) redundant or even conflicting data may be collected; (2) data may be inaccurate or missing as a result of manual entries; (3) there exists unavailable information due to the reluctance to share private information among the supply chain participants. There is a need for a robust data fusion model that takes into account both the noise, incompleteness from the real time data, and the uncertainty, inaccuracy from the model. Data fusion is 'a dynamic process in the association, correlation and combination from multiple sources resulting in a fused product, which is more complete and accurate than any of the separate data elements' (Waltz and Linas 1990). Known as one of optimal (minimum error variance) unbiased linear filters, the recursive process of Kalman Filtering by fusing measurement data (e.g., from sensor) into system data (from model) makes it an attractive tool for linear system calibration. For a complex nonlinear system such as supply chain, extensions of Kalman Filtering, such as Extended Kalman Filtering and Ensemble Kalman Filtering should be considered.

3.3 Methodology

3.3.1 Kalman Filtering Basics

Kalman Filtering (KF) is the optimal state estimation method in a linear state-space model (Welch and Bishop 2001). Let x_k be an n-dimensional vector process, called the state of the system. Let z_k be an m-dimension vector process,

representing the measurement of the system. The discrete-time state-space system (without control input) can be mathematically described by two equations:

State Transition Equation: $x_k = G_k x_{k-1} + \varepsilon_k$ (3.1)

Measurement Equation: $z_k = H_k x_k + \eta_k$ (3.2)

where G_k is a n×n matrix, which linearly relates the state at the previous time x_{k-1} to the current state x_k. H_k is a m×n matrix, showing how the system state x_k linearly relates to observation z_k . ε_k and η_k are the noise in state transition and measurement, following multivariate Gaussian distribution. $\varepsilon_k \sim \text{i.i.d } N_n(0, Q_k)$, $\eta_k \sim \text{i.i.d } N_m(0, R_k)$. Assume ε_k and η_k are independent.

Equations (3.3)-(3.7) of KF recursively calculate the prior estimate x_k^-, the optimal posterior estimate x_k^+ and their error covariance matrix P_k^-, P_k^+ where $P_k^- = E[(x_k - x_k^-)(x_k - x_k^-)^T]$, $P_k^+ = E[(x_k - x_k^+)(x_k - x_k^+)^T]$. In specific, Equations (3.3) and (3.4) project the estimate and error covariance of system state from time t-1 to time t. The prior estimate x_k^- and error covariance P_k^- are dependent on the linear operating matrix G_k and the process noise covariance Q_k. After that, with the goal of minimizing the posterior estimate error covariance P_k^+, Equations (3.5) - (3.7) are the optimal procedures to inject observation z_k with the prior estimate. In (3.5), the Kalman gain K_k is calculated optimally. It decides how much the posterior estimate x_k^+ will be adjusted to consider the deviation between the model and the observation $(z_k - H_k x_k^-)$ as in (3.6). The posterior estimate error covariance P_k^+ is then calculated by (3.7).

$x_k^- = G_k x_{k-1}^+$ (3.3)

$$P_k^- = G_k P_{k-1}^+ G_k{}^T + Q_k \quad (3.4)$$

$$K_k = P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k\right)^{-1} \quad (3.5)$$

$$x_k^+ = x_k^- + K_k(z_k - H_k x_k^-) \quad (3.6)$$

$$P_k^+ = (I - K_k H_k) P_k^- \quad (3.7)$$

Kalman Filtering is one of the optimal linear estimators (i.e., among all the estimators that are linear combinations of the measurements) according to the least square criteria. It can also be derived from Bayesian estimation (maximum likelihood criteria) when the initial system state follows Gaussian distribution (Barker et al. 1995). KF has been successfully applied in many areas. As examples in the context of supply chain management, Aviv uses state space model and KF as a general framework to predict the demand in the study of forecasting/inventory replenish policies (Aviv 2003). The application of KF in monitoring the fulfillment of server supply network has been studied in Parmar et al. (2006). However, a notable issue with KF is that it only applies to a linear state-space system shown in (3.1) – (3.2), thus, its application is limited in a complex supply chain which is nonlinear in system transition or measurement.

3.3.2 Extended Kalman Filtering (EKF) for nonlinear system

In a nonlinear state-space model, the functions of state transition equation (g_k) and the measurement equation (h_k) are nonlinear:

State Transition Equation: $x_k = g_k(x_{k-1}, \varepsilon_k)$ (3.8)

Measurement Equation: $z_k = h_k(x_k, \eta_k)$ (3.9)

As an extension, Extended Kalman Filtering (EKF) first linearizes the nonlinear functions before applying similar recursive steps as those in KF. In details, the recursive equations for Extended Kalman Filtering are:

$$x_k^- = G'_k x_{k-1}^+ \tag{3.10}$$

$$P_k^- = G'_k P_{k-1}^+ {G'_k}^T + Q_k \tag{3.11}$$

$$K_k = P_k^- H'^T_k \left(H'_k P_k^- H'^T_k + R_k\right)^{-1} \tag{3.12}$$

$$x_k^+ = x_k^- + K_k(z_k - H'_k x_k^-) \tag{3.13}$$

$$P_k^+ = (I - K_k H'_k) P_k^- \tag{3.14}$$

where the matrix G'_k is the Jacobian matrix of g() evaluated at x_{k-1}^+, H'_k is the Jacobian matrix of h() evaluated at x_k^-.

EKF has become a standard technique used in nonlinear estimation and machine learning applications (Wan and Van Der Merwe 2000). However, the local linearization in EKF may lead to poor error covariance updates and in some cases unstable growth of error covariance matrix (Evensen, 1992).

3.3.3 Ensemble Kalman Filtering (EnKF) for nonlinear system

Instead of linearization, Ensemble Kalman Filtering (EnKF) uses Monte Carlo simulation to approximate the nonlinear state transition function $g_k()$. Unlike KF and EKF where the estimation error is analytically propagated from time t-1 to time t, a group of samples of the system state, called an ensemble, is used to track the evolution of system state in EnKF. The mean and covariance of the ensemble, derived from samples, are taken as state estimation and error covariance.

Let V is the ensemble size, x_{ik} is the state of i^{th} instance in the ensemble at time k. At each time point of data fusion, every instance in the ensemble $x^{+}_{i(k-1)}$ advances to x^{-}_{ik} following $g_k()$. The sample mean is taken as the prior estimation of system state (x^{-}_{k}) and the error covariance is approximated by the sample covariance (P^{-}_{k}), which is used in the calculation of Kalman gain in Equation (3.17). Next, each instance x^{-}_{ik} is updated with the noisy observation z_{ik}. The posterior estimation of system state (x^{+}_{k}) is calculated by the ensemble mean. It is proven that EnKF converges to KF for a linear system as V $\rightarrow\infty$ (Evensen 2006).

$$x^{-}_{k} = \frac{\sum_{i=1}^{V} x^{-}_{ik}}{V} \tag{3.15}$$

where $x^{-}_{ik} = g_k(x^{-}_{i(k-1)},\ \varepsilon_{ik})$, $\varepsilon_{ik} \sim N(0, Q_k)$

$$P^{-}_{k} = \left[\sum_{i=1}^{V}(x^{-}_{ik} - x^{-}_{k})(x^{-}_{ik} - x^{-}_{k})^{T}\right]/(V-1) \tag{3.16}$$

$$K_k = P^{-}_{k} H'_k{}^{T}\left(H'_k P^{-}_{k} H'_k{}^{T} + R_k\right)^{-1} \tag{3.17}$$

where H'_k is the Jacobian matrix of $h_k()$ evaluated at x^{-}_{k}.

$$x^{+}_{k} = \frac{\sum_{i=1}^{V} x^{+}_{ik}}{V} \tag{3.18}$$

where $x^{+}_{ik} = x^{-}_{ik} + K_k(z_k - H'_k x^{-}_{ik} - \eta_{ik})$, ($\eta_{ik} \sim N(0, R_k)$)

$$P^{+}_{k} = \left[\sum_{i=1}^{V}(x^{+}_{ik} - x^{+}_{k})(x^{+}_{ik} - x^{+}_{k})^{T}\right]/(V-1) \tag{3.19}$$

Since it was first proposed in 1994, EnKF has been applied to a number of complex and large-scale systems, including systemquasi-geostrophic ocean model (Evensen 1994), wild land fire simulation (Mandel et al. 2008), reservoir engineering (Krymskaya et al. 2009), atmospheric model (Houtekamer and

Mitchell 2001), estimation model of soil moisture (De Lannoy et al. 2007), just to name a few.

In summary, both EKF and EnKF have successful applications. Considering a supply network with dynamic behaviors, some interesting questions are: will EKF and EnKF achieve comparable performance for calibrating supply chain models, some are highly dynamic, some are stable? How will the calibration frequency impact the performance? How will the calibration data impact the performance? Experiments on a 'beer game' supply chain are conducted attempting to answer the questions.

3.4 Experiments

In this section, a simulation based on discrete time difference equation model and data fusion methodology is developed to study the dynamics in a nonlinear system. The method is tested in the classic 'beer game' supply chain (Sterman 1989). Three experiments are designed to explore 1) chaotic behavior of a supply chain system 2) the impact of data availability on data fusion performance and 3) the impact of data accuracy on data fusion performance.

3.4.1 Game Supply Chain Model

The beer game model (Sterman 1989) is a typical serial supply chain, having four entities - a retailer, a wholesaler, a distributor and a factory. The order flow (right to left in Figure 7) and the product flow (left to right in Figure 7) connect the four autonomous nodes sequentially. The end-customer demand input into the retailer is exogenous. Each level has its own buffer stock (S) to absorb uncertainty in demand. In each level of echelon, six state variables considered are:

- S (Stock): inventory level
- SL(Supply Line): placed orders that have not been received
- B (Backlog): customer orders that haven't been fulfilled
- EOR (Estimated customer Order Rate): estimation of orders from customers per time unit
- OR (Order Rate): orders placed to supplier per time unit
- SR (Shipment Rate): the number of products shipped to customer per time unit

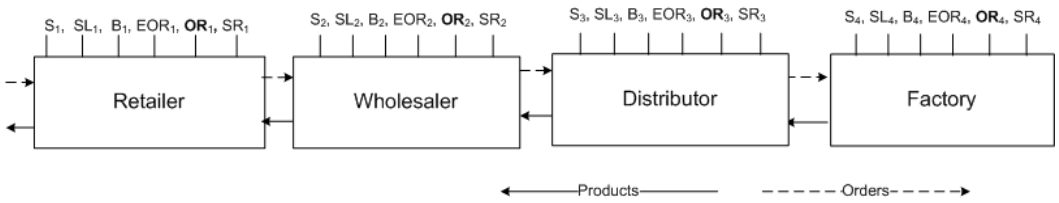


Figure 7. Beer Game supply chain model

Using subscripts 1 to 4 for retailer, wholesaler, distributor and factory respectively, the overall system state X is presented as a vector of 24 states = $[\mathrm{S}_1, \mathrm{SL}_1, \mathrm{B}_1, \mathrm{EOR}_1, \mathrm{OR}_1, \mathrm{SR}_1, \mathrm{S}_2, \dots, \mathrm{SR}_3, \mathrm{S}_4, \mathrm{SL}_4, \mathrm{B}_4, \mathrm{EOR}_4, \mathrm{OR}_4, \mathrm{SR}_4]^{\mathrm{T}}$.

Note Order Rate ($\mathrm{OR_i}$) for each echelon is highlighted in Figure 7 which is used as measurement, that will be discussed in Section 3.4.2.

At time k, each echelon places an order to its supplier, which aims to 1) supply customer demand, 2) compensate for discrepancy in the desired stock level and current stock level, and 3) compensate for discrepancy in the desired supply line and current supply line. The discrete time difference equations for the ordering amount are shown in (3.20) – (3.23) (Hwarng and Xie 2008). Note that DS is Desired Stock and DSL is Desired Stock Level.

$$EOR_{i,k+1} = \theta_i * OR_{i-1,k} + (1-\theta_i) * EOR_{i,k} \quad (3.20)$$

where θ_i is the exponential smoothing parameter

$$DS_{i,k+1} = EOR_{i,k+1} * opl_i + B_{i,k} \quad (3.21)$$

where opl_i is time for order processing lag, $B_{i,k}$ is the backlog

$$DSL_{i,k+1} = EOR_{i,k+1} * ssl_i \quad (3.22)$$

where ssl_i is time for shipment lag

$$OR_{i,k+1} = Max[(DS_{i,k+1} - S_{i,k})\alpha_{S_i} + (DSL_{i,k+1} - SL_{i,k})\alpha_{SL_i} + EOR_{i,k+1} + \varepsilon_{i,k}, 0] \quad (3.23)$$

where ε_k is the disturbance in the order rate, $\varepsilon_k \sim i.i.d.N(0,\sigma^2)$. Parameters $\alpha_S \in (0,1)$ and $\alpha_{SL} \in (0,1)$ decide the compensation ordering amount for the discrepancy in Stock and Supply Line.

Some research (Hwarng and Xie 2008; Larsen et al. 1999) observe that the two parameters (α_S and α_{SL}) are associated with the chaotic behavior of the 'beer game' supply chain. Detailed discussion is provided in Section 3.4.5.

The simulation is built in Matlab with the minimum time step of simulation being a week, the total simulation time being 100 weeks. The experiment parameters are adopted from Sterman (1989) (Table 7). The exogenous customer demand is assumed to be a step function, started from 4 units/week for the first 50 weeks, doubled to 8 units/week for the next 50 weeks (Table 8).

Table 7. Simulation Parameters (i=1,2,3,4)

Parameters	Values
opl_i	3 weeks
ssl_i	1 week
θ_i	0.25

Table 8. End-customer demand

End-customer order time series	4 units/week, where k=1,2,. 50
	8 units/week, where k=51,…, 100

3.4.2 Observation of Beer Game Supply Chain

Let assume that we can only observe the order rate (OR_i) at different levels of the beer game supply chain. The measurement of the system is denoted by Z_k, then,

$$Z_k = (OR_{1,k}, OR_{2,k}, OR_{3,k}, OR_{4,k})^T + \eta_k \quad (3.24)$$

where η_k is observation error, $\eta_k \sim N(0, R)$.

3.4.3 Beer Game Supply Chain Simulation and Emulation Development

A pair of emulation and simulation is implemented to evaluate the prediction accuracy of the simulation model with/without data fusion. Emulation is a random instance of the dynamic model that mimics a 'real world'. Simulation represents the modeling of the system and it is adjusted with the data collected from emulation when data fusion is used.

Simulation is initialized at the system equilibrium (see Table 9). Emulation is initialized at the system equilibrium with a Gaussian noise

$\delta \sim N(0,0.16)$ added to the initial stock level $S_{i,0}$ (i=1,2,3,4). In emulation, the supply chain dynamics is also reflected by the white noise ($\sigma^2 = 0.16$) in the ordering rate $OR_{i,k}$, i =1,2,3,4.

The pair of simulation and emulation is run in parallel. When data fusion is adopted, observations are collected from emulation following (3.24). Simulation is paused and calibrated by the observation (measurement) data using EKF or EnKF. Otherwise, simulation advances without any calibration, which is considered as the benchmark model (without calibration).

Table 9. Initial state of simulation (i=1,2,3,4) (Hwarng and Xie 2008)

State Variables	Values
$S_{i,0}$	12 units
$SL_{i,0}$	4 units
$B_{i,0}$	0
$EOR_{i,0}$	4 units/week
$OR_{i,0}$	4 units/week
$SR_{i,0}$	4 units/week

3.4.4 Performance Metric

Given that emulation represents the "real" operating supply chain system, the differences between the simulation states and the emulation states are considered as an accuracy measurement of the simulation. One measurement is the Mean Square Error (MSE) that is calculated as:

$$MSE_x = \sqrt{\frac{\sum_{t=1}^{T}\left(x_k^{simulation} - x_k^{emulation}\right)^2}{T}} \qquad (3.25)$$

where $x_k^{simulation}$ denotes the simulation state, $x_k^{emulation}$ is the emulation state for time k.

For any of two simulation models, the ratio of their MSE (Equation (3.25)) is called Theil's Inequality Coefficient $U \in (0, +\infty)$ (Makridakis et al. 2008). Let the simulation without calibration be the benchmark model. Theil's U can be used to assess the relative performance of a simulation calibration method (3.26). The comparison based on Theil's U lead to three potential conclusions: (1) When U =1, the performance of calibrated model is equivalent to that of benchmark model; (2) When U <1, calibrated model outperforms benchmark model; and (3) When U >1, calibrated model underperforms benchmark model. The smaller U is, the better calibrated model performs. Let U_{EKF}, U_{EnKF} be Theil's Inequality Coefficient for the models calibrated by EKF and EnKF respectively. If U_{EKF} > U_{EnKF}, we conclude that EnKF outperforms EKF; otherwise, EnKF underperforms EKF.

$$U = \frac{\text{MSE(Calibrated Simulation)}}{\text{MSE(Benchmark)}} \tag{3.26}$$

$$U_{EKF} = \frac{\text{MSE(Simulation Calibrated with EKF)}}{\text{MSE(Benchmark)}} \tag{3.27}$$

$$U_{EnKF} = \frac{\text{MSE(Simulation Calibrated with EnKF)}}{\text{MSE(Benchmark)}} \tag{3.28}$$

Since totally 24 states are studied in the beer game, the average of *U* is used as the overall performance indicator in our study.

3.4.5 Benchmark Models and Associated Chaotic Behaviors

Larsen et al. (1999) and Hwarng and Xie (2008) have conducted interesting study to explore the chaotic behaviors of the beer game model. It has

been found that the simple beer game could behave quite differently with different parameter settings. Intuitively, for highly chaotic setting, the simulation could deviate from the emulation greatly. Therefore, the first step in this study is to explore the characteristics of beer game chaotic behavior so that appropriate benchmark models can be identified.

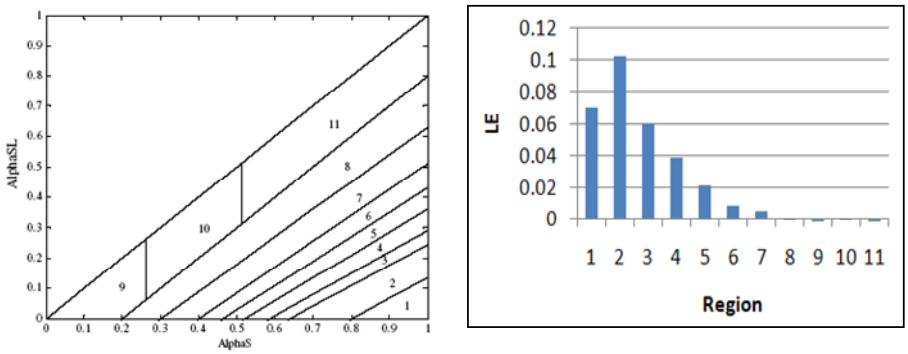
The chaotic level of a system can be measured by Lyapunov Exponent (LE) (Larsen et al. 1999; Hwarng and Xie 2008; Makui and Madadi 2007; Pathak et al. 2010). LE describes the exponential divergence rate of the outputs compared to the difference in initial states. Interpreted in a general way, LE is "the average factor by which an error is amplified within a system" (Makui and Madadi 2007). For a system which can be presented by $x_{k+1} = f(x_k)$, where x_k is the system state, LE (denoted by λ) is derived as in Equation (3.29). If the system is diverging (the distance between pairs of points increases exponentially), LE is positive. If the system is oscillating, LE is 0. If the system converges, LE is negative.

$$|x_{k+1} - x_k| = |x_1 - x_0| \exp(\lambda n), \lambda > 0$$

$$\Rightarrow \quad \lambda = \lim_{n\to\infty} \sum_{t=0}^{n-1} \ln|f'(x_k)| \qquad (3.29)$$

For the 'beer game' model, Hwarng and Xie (2008) analyze the system behaviors given different values of two parameters - α_S and α_{SL} – and, as a result, they divide the parameter spaces into 11 regions (shown in Figure 8(a)). For example, $0.8 \le \alpha_S \le 1$ and $0 \le \alpha_{SL} \le 0.13$ defines region 1. Within each region, the system behaviors (measured by LE) are homogeneous. LE of different parameter regions are collected from Hwarng and Xie (2008) shown in Figure

8(b). It is observed that in general LE decreases from region 1 to region 11, that is, the level of chaotic behavior decreases from region 1 to region 11.



(a) (b)

Figure 8. (a) Parameter Region (Hwarng and Xie 2008) (b) Average LE in different parameter regions (derived from Hwarng and Xie (2008))

Next, we further explore the characteristics of the model to identify representative configurations of parameters α_S and α_{SL}. For each region, we randomly locate α_S and α_{SL}, resulting in the values listed in Table 10 (e.g. α_S = 0.95 and α_{SL}=0.05 for region 1).

Each pair of α_S and α_{SL} corresponds to a configuration of the 'beer game' supply chain. Totally 22 configurations are identified upon which both simulation and emulation are executed. The average MSEs of the four echelons are calculated, which results in 4* 22=88 rows of data. A regression analysis is conducted to identify the significant factors impacting system MSE: the region (region 1 to 11), the group of parameter setting and the interaction. As shown in Table 11, the parameter region is the only significant factor.

Table 10. Configurations of the 'beer game' model for each parameter region

Regions	Parameter Set A		Parameter Set B	
	α_S	α_{SL}	α_S	α_{SL}
1	0.95	0.05	0.9	0.02
2	0.9	0.1	0.8	0.05
3	0.85	0.18	0.8	0.13
4	0.8	0.19	0.7	0.12
5	0.75	0.2	0.65	0.12
6	0.7	0.2	0.6	0.12
7	0.65	0.3	0.5	0.13
8	0.65	0.4	0.35	0.12
9	0.2	0.1	0.1	0.03
10	0.4	0.3	0.3	0.15
11	0.6	0.5	0.95	0.8

Table 11. Effect test of regression of MSE on parameter region, parameter set and their interaction

Source	Degree of Freedom	Sum of Squares	F Ratio	Prob > F
Regions	10	15645.005	15.8965	<.0001*
ParameterSet (α_S, α_{SL})	1	23.325	0.2370	0.6280
Region*ParameterSet	10	234.588	0.2384	0.9911

Figure 9 illustrates the MSE at each region. It has similar pattern as that of Figure 8(b). In addition, we observe MSEs in Regions 1 and 2 are high, in Regions 3-7 are medium and in Regions 8-11 are low. Since parameter setting (α_S and α_{SL}) has no significant impact on chaotic performance, we choose Region

1, Region 5 and Region 11 to represent highly chaotic, low chaotic and stable systems in the following comparison study.

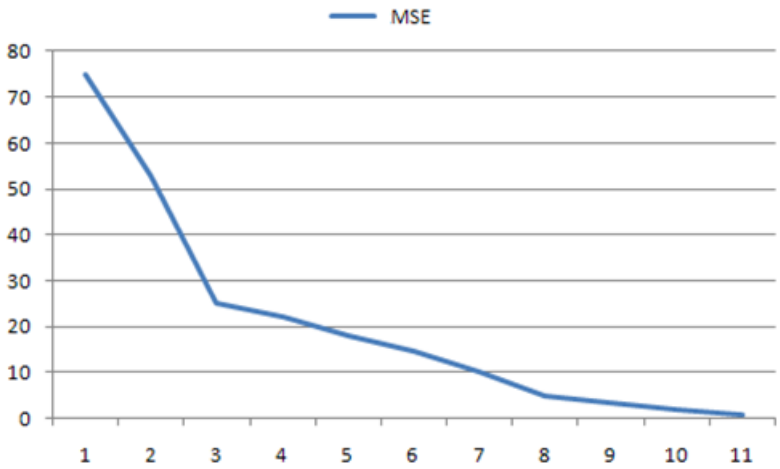


Figure 9. Relationship between system chaos (11 parameter regions) and simulation accuracy (Mean Square Error)

From this experiment, we conclude that the accuracy of simulation is tightly related to the system chaotic level that is measured by LE. For example, MSE for region 1 is much higher than that of region 11. In other words, in a highly chaotic system, the simulation state of the system can deviate significantly from the real system even for a small difference between simulation and real system at the initial stage. Therefore, without fusing real-time data into models, a simulation is prone to error in predicting the system state, especially for a chaotic system. The results of this experiment motivate us to explore the approach to integrate simulation and real time data by data fusion techniques.

3.4.6 Comparison Experiment I – The impact of data fusion frequency (the availability of the data) on data fusion methods

Two data fusion methods – EKF and EnKF are applied to calibrate simulation models with real-time measurements. This experiment is designed to explore the impact of data fusion frequency on the performance of EKF and EnKF. Four levels of frequency are tested - every 1, 4, 8 and 12 weeks. For EnKF, a number of additional experiments are conducted to locate an appropriate ensemble size (V). We set V=400 here as EnKF performs well with reasonable computation expenses for the beer game.

Table 12. Experiment design – the impact of data fusion frequency

Methods	EnKF EKF
Data Fusion Frequency	Every 1,4,8,12 weeks
Parameter Region	Highly chaotic (Region 1) Low chaotic (Region 5) Stable (Region 11)
Observation Error Variance (R)	0.16

For each combination of data fusion technique and parameter setting, 20 pairs of simulation and emulation are run to calculate the mean and variance of Theil's U. The experiment results are summarized in Table 13 and illustrated in Figure 10. In Table 13, the best calibration accuracy is 0.20 indicating 80% performance improvement. In addition, it is observed that there exist apparent advantages of data fusion techniques for dynamic systems (both highly and low chaotic) over stable systems. Especially, when the real data is more frequently

injected to the model (e.g., every 1 week), U_{EKF} is 0.20 and U_{EnKF} is 0.21 for highly chaotic systems, U_{EKF} is 0.55 and U_{EnKF} is 0.58 for stable system. Such advantage diminishes as real time data is less frequently injected. Specifically, the advantage of EKF statistically decreases in comparison with EnKF. Secondly, as the time duration of data fusion interval increases, the calibration accuracy of all three types of systems decreases. However, EnKF statistically outperforms EKF for some cases. For example, in a highly chaotic system, when the data is fused every 8, or even 12 weeks, EnKF is preferred over EKF. The difference between EKF and EnKF for any system under frequent data fusion (e.g., every 1 week, every 4 week) is undistinguishable. Therefore, for the second comparison experiment, we design the experiments on more frequent data fusion scenarios.

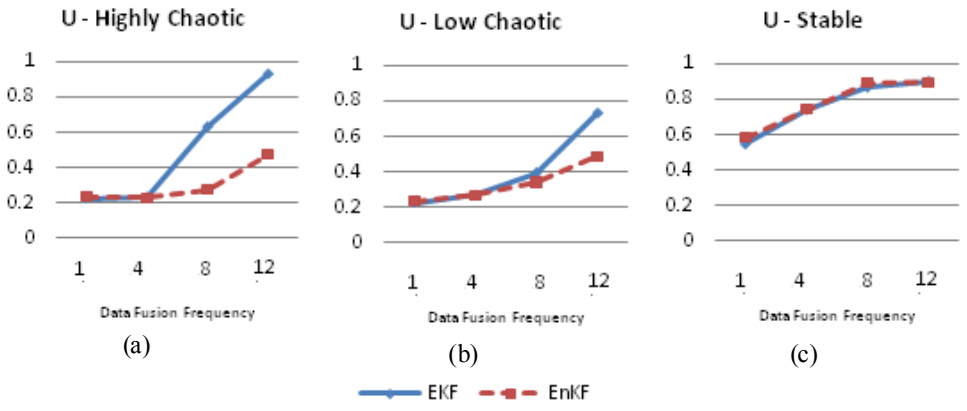


Figure 10. Comparison of EKF and EnKF performance (Theil's U) with varying data fusion frequency

Table 13. Performance comparison between EKF and EnKF with varying data fusion frequency

System Dynamics	Data fusion Frequency	U_{EKF}		U_{EnKF}		Outperformed Method*
		Mean	Standard Deviation	Mean	Standard Deviation	
High chaotic	1	0.20	0.0094	0.21	0.0128	/
High chaotic	4	0.23	0.0227	0.23	0.0227	/
High chaotic	8	0.63	0.0975	0.28	0.0309	EnKF
High chaotic	12	0.93	0.0781	0.47	0.1824	EnKF
Low chaotic	1	0.23	0.0173	0.24	0.0205	/
Low chaotic	4	0.27	0.0310	0.27	0.0277	/
Low chaotic	8	0.40	0.0269	0.34	0.0700	/
Low chaotic	12	0.73	0.0587	0.49	0.1942	EnKF
Stable	1	0.55	0.0615	0.58	0.0862	/
Stable	4	0.74	0.0845	0.74	0.0788	/
Stable	8	0.87	0.0235	0.89	0.1062	/
Stable	12	0.90	0.0126	0.90	0.0732	/

* A method is preferred if and only if its performance is significant better than another method (p value <0.1)

3.4.7 Comparison Experiment II – The impact of measurement error (the quality of data)

It is recognized that the collected data for calibration may not be always complete, accurate. In this experiment, we test the impact of data quality quantified by the measurement error on the performance of both EKF and EnKF (Table 14). Five levels of measurement error is tested – variance R= 2.56, 1.44, 0.16, 0.04, 0.01 (note R=0.16 is used in the first comparison experiment). As discussed earlier,

we specifically focus on the data quality issue under scenarios with frequent data fusion (every 1 and 4 weeks) in this experiment.

Table 14. Experiment design – the impact of measurement error

Methods	EnKF EKF
Data fusion frequency	Every 1, 4 weeks
Parameter Region	Highly chaotic (Region 1) Low chaotic (Region 5) Stable (Region 11)
Observation Error variance (R)	2.56, 1.44, 0.16, 0.04, 0.01

The experiment results are shown in Table 15. There is not statistically difference between EKF and EnKF for accuracy improvement for all three types of system under 1 week data fusion and 4 week data fusion. To better illustrate the performance, the average performance of two methods for different system under 1 week, 4 week fusion interval is shown in Figure 11. The first observation is that more frequent data fusion (e.g., every 1 week) of real time data always outperforms less frequent data fusion (e.g., every 4 week). The performance improvements in both highly and low chaotic systems are greater than that for stable systems. That is, U_{EKF} and U_{EnKF} range from 0.20 to 0.33 for dynamic systems and from 0.42 to 0.88 for stable systems. In addition, with the decrease of observation error, the calibration performance improves for all three types of systems. However, the improvement magnitude for stable system appears to be larger than that of dynamic systems. By average, the improvement (decrease of Theil's U) for stable system changes from 0.78 to 0.46 when data fusion

frequency is ‘every 1 week’ and from 0.87 to 0.72 when data fusion frequency is ‘every 4 weeks’; the improvement for low chaotic system drops from 0.27 to 0.22 (‘every 1 week’) and from 0.31 to 0.27 (‘every 4 weeks’); and the improvement for highly chaotic system is from 0.23 to 0.20 (‘every 1 week’) and from 0.27 to 0.23 (‘every 4 weeks’). This is intuitive that stable system is more sensitive to measurement errors, while in dynamic system, the impact of measurement error is compromised by the dynamics of the system.

Table 15. Performance comparison between EKF and EnKF with varying observation error

System Dynamics	Observation Error Variance (R)	Data fusion frequency =every 1 week			Data fusion frequency =every 4 week		
		U_{EKF}	U_{EnKF}	Average	U_{EKF}	U_{EnKF}	Average
High chaotic	2.56	0.23	0.23	0.23	0.29	0.25	0.27
High chaotic	0.64	0.21	0.22	0.215	0.27	0.26	0.265
High chaotic	0.16	0.20	0.21	0.205	0.23	0.23	0.23
High chaotic	0.04	0.20	0.21	0.205	0.23	0.22	0.225
High chaotic	0.01	0.20	0.20	0.20	0.23	0.23	0.23
Low chaotic	2.56	0.27	0.27	0.27	0.32	0.30	0.31
Low chaotic	0.64	0.25	0.26	0.255	0.30	0.29	0.295
Low chaotic	0.16	0.23	0.24	0.235	0.27	0.27	0.27
Low chaotic	0.04	0.23	0.24	0.235	0.27	0.27	0.27
Low chaotic	0.01	0.21	0.23	0.22	0.27	0.27	0.27
Stable	2.56	0.77	0.79	0.78	0.88	0.86	0.87
Stable	0.64	0.67	0.71	0.69	0.80	0.81	0.805
Stable	0.16	0.55	0.58	0.565	0.74	0.74	0.74
Stable	0.04	0.53	0.59	0.56	0.70	0.72	0.71
Stable	0.01	0.42	0.50	0.46	0.74	0.70	0.72

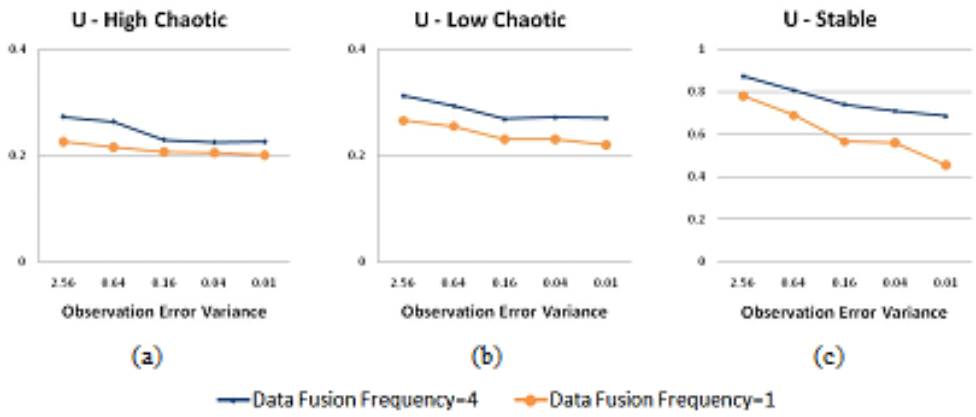


Figure 11. EKF an EnKF performance (Theil's U) under varying observation error variance

3.5 Conclusions and Future Work

The dynamics and uncertainty of supply chain have drawn great attention from both research academics and industry. A precursor for better understanding the system is to robustly and accurately sense the system status. In this chapter, two data fusion techniques: Extended Kalman Filtering and Ensemble Kalman Filtering are studied to calibrate the 'beer game' simulation model. The dynamics of the model is first explored to locate three benchmark systems: highly chaotic, low chaotic and stable system. Theil's Inequality Coefficient, U, over MSE is then used to assess the calibration performance over these three type systems. We conclude data calibration can significantly improve the system estimation accuracy. For highly dynamic system, the improvement can be up to 80%. Considering EKF and EnKF, for the cases when the real time data is less frequently injected for model calibration, EnKF statistically performs better than

EKF. For the cases with more frequent fusion, the calibration performance exhibits no statistical difference between the two methods. This is confirmed from both two comparison experiments. In exploring data quality issue, we observe the less measurement error is, the better calibration is. A stable system is more sensitive to this type of error.

This research demonstrates the value of data fusion to monitor the dynamics of nonlinear supply chain. Based on the availability of the real time data and system chaotic levels, the managers may choose EKF or EnKF. While promising, the computational tradeoff between the two methods has not been explored. Considering a system with n states and m measurements, the computation cost for EKF is $O(n^2 + m^2)$, the computation cost for EnKF is $O(nV + m^2)$ where V is the number of ensembles. Thus, for a large-scale supply chain where n is a large number, the advantage of EnKF over EKF may be apparent. Since the beer game studied has only 24 states, such advantage is not shown. One immediate next study is to explore a more complex supply chain with more state variables. Secondly, we are planning to develop control strategies with the capability of better sensing system states to make superior responses. Thus a fully sense and response framework can be presented.

Chapter 4

AN ONLINE CALIBRATED SIMULATION FOR ADAPTIVE INVENTORY MANAGEMENT

This chapter presents an online calibrated simulation-based approach for adaptive inventory management. The simulation incorporates information on a regular basis using a data fusion method called Ensemble Kalman Filtering. It is able to incorporate real-time noisy data into simulation, to interpret their influence on the correlated variables, and to predict the future state of the supply chain by considering all available information. A case study is shown where information sharing is applied in a multi-echelon supply chain facing non-stationary end-consumer demand. When the shared information does not have noise, a closed-form expression of Minimum Mean Square Error (MMSE) forecast is derived explicitly. Information sharing is beneficial to the node that receives information and all its upstream suppliers. Experiment results show that Calibrated Simulation-based Forecast (CSF) generates comparable result to MMSE forecast. When the shared information has noise, the derived MMSE model is not directly applicable, but CSF is still able to provide reliable forecast to facilitate cost saving. We explore various scenarios under varied levels of observation noise, different information source(s) in the supply chain and different number of nodes providing information. Some managerial insights are gained from this study, which are related to the value of information sharing and the selection of information source(s) in a serial supply chain.

4.1 Introduction

To respond to a quickly changing business environment and to develop customer-focused strategies, companies need to manage inventory in an adaptive way. Without the capability of accurately 'sensing' and quickly 'responding' to the evolving consumer demand patterns and unexpected supply conditions, companies are taking the risks of over-stocking and under-stocking. Either one is costly. One well-known lesson is from Cisco Systems, who had to write-off $2 billion of excess inventory of its network infrastructure products when the actual market demands suddenly declined in 2001 (Hau 2004). Besides, the shortage of supply also can be destructive. As an example, Boeing lost $2.6 billion when its suppliers failed to deliver components on time (Wu et al. 2006).

To effectively manage the highly dynamic supply chain and take prompt actions to minimize the striking influence of unexpected disruptions, companies need to manage the uncertainties in the whole supply chain and enable informed decisions in a prompt manner. There are several challenges in monitoring supply chain dynamics and making accurate prediction for inventory management. First, the supply chain system needs to be considered as a whole. Ignoring inter-node interaction may lead to biased prediction results. Hayya et al. (2006) point out that the bullwhip effect measured in a 'sequential-pairs' execution is different from the results of a whole-system simulation, because the former does not consider the cascading stock-out effect in the chain. Second, the supply chain is often under non-stationary uncertainties, that is, the probability distribution of uncertain variables changes over time. In particular, the end-consumer demand is often non-

stationary. A number of factors, including short product cycle times, seasonality, customer buying patterns, may lead to the time-varying dynamics in demand (Neale and Willems 2009). The non-stationary characteristic implies that the modeling of random processes in a supply chain needs to be continuously updated and improved to capture the latest state of dynamics. Third, sometimes we only have partial information (not knowing the exact distributions) of the involved random variables. The combination of these three factors poses grand challenges to researchers. As pointed out by Treharne and Sox (2002), research that considers both non-stationary demand and partial information is relative limited in the literature compared to the study of full information and stationary supply chain.

In this chapter, we propose a simulation-based approach for adaptive inventory control. Although most research in supply chain inventory control use analytical models, we advocate the use of simulation for several reasons. First, simulation is powerful to model large-scale system and to capture interactive behaviors between supply chain players. Second, simulation is flexible to model realistic constraints and heterogeneous configurations. For example, in the study of multi-echelon supply chain systems, while analytical models usually assume that each node uses the same optimal inventory control policies, simulation studies can be used to explore the situation when some echelons derivate from the optimal policy (Son and Sheu 2008). Third, with the recent development of simulation techniques (e.g. data-driven (Tannock et al. 2007), web-based, agent-based, the enrichment of visualization), simulations are faster and more user-

friendly. It could play as a "communicative means" between the analyst and the problem owners (D. J. van der Zee 2005).

However, there are notable challenges in the use of simulation in adaptive supply chain management. Given the fact that analytical methods dominate in practice and in theoretic studies, the foremost question is *will a simulation be reliable and effective in inventory control*? Compared to analytical models, the question that *how can a simulation model incorporate real time information* is less studied. A conventional simulation usually is built with stationary parameters estimated from history data. Since it does not 'learn' from new available data, it is difficult for the simulation to keep up to the changes in real world and really 'sense' the evolving supply chain environment.

In this chapter, we propose a calibrated simulation model which is a hybrid approach taking advantage of both simulation and data fusion methods. The simulation is featured by 1) fusing noisy new information from real-world data that is provided by an emulation model in this study and 2) using a simulation ensemble to estimate both the mean and variance of system state. By data fusion, the simulation model can incorporate noisy information from different nodes (information sources) in a supply chain. The proposed framework is generalized and is able to handle different types of information (e.g. inventory level, demand, lead time). To validate the Calibrated Simulation-based approach, we first compare Calibrated Simulation-based Forecast (CSF) with the analytical derived MMSE forecast under the assumption that the demand information does not have noise. The numerical results show that the performance of CSF is

comparable to MMSE forecast method, which is known to be optimal. Taking advantage of the CSF approach, we are able to explore the benefits of information sharing under more complicated assumptions when the observation noise is non-zero and multiple information sources are available. Some managerial insights are gained regarding the selection of information source(s) in the supply chain.

The remainder part of this chapter is organized as follows. In Section 4.2, we position our work with respect to prior related research. The data fusion method – Ensemble Kalman Filtering (EnKF) is introduced in Section 4.3, and some properties of EnKF are proved to support the framework. In Section 4.4, we introduce the supply chain model and the adaptive order-up-to control policies. The derivation of MMSE forecast and the introduction of the CSF framework are in Section 4.5 and Section 4.6 respectively. Numerical experiments are presented in Section 4.7, followed by the summary and discussion about the future work in Section 4.8.

4.2 Literature Review

The literature in the area of supply chain inventory management is extensive. We focus on reviewing the work that is related to our motivation, model and methods. Forecasting coupled with adaptive inventory control is probably the most common approach for a company to be responsive to external uncertainty such as customer demand. Graves (1999) study the adaptive base-stock policies in N-echelon serial supply chain with ARIMA(0,1,1) end-consumer demand. Our work extends his model to incorporate information lead time and information sharing.

A variety of optimization approaches have been used in adaptive inventory control as well. Treharne and Sox (2002) employ dynamic programming to study adaptive and non-adaptive control of a two-node supply chain. Demand realizations are used to update the system states in a Partially Observed Markov Process model. They show that adaptive policies considering the updates of demand structure outperform non-adaptive control policies. Perea-Lopez et al. (2003) build a Mixed Integer Linear Programming to find the global optimal solution in a multi-product, multi-echelon supply chain (manufacturing and distribution centers) when customer demand is updated daily. Liang and Huang (2006) study a heterogeneous supply chain system where companies may have different inventory policies, such as (R,S), (s,Q) and (s,S). In their study, a centralized agent captures all the system states and Genetic Algorithm is used to find the optimal parameter settings for all supply chain entities.

Some information sharing and collaborative forecasting strategies have been proposed to coordinate the ordering behaviors among supply chain players. For example, Aviv (2003) proposes a time-series framework for Collaborative Planning, Forecasting and Replenishment based on linear state-space models and Kalman Filtering. In this study, since simulation models are considered, Ensemble Kalman Filtering, an approximate method to Kalman Filtering in nonlinear systems, is adopted to consolidate information.

4.3 Data fusion - Ensemble Kalman Filtering

Ensemble Kalman Filtering (EnKF) is an approximate method of the well-known Kalman Filtering (KF) (Welch and Bishop 2001) in a non-linear system.

EnKF originates from Evensen (1994), where Monte Carlo method is used with the Kalman Filtering. The research of EnKF has attracted great attention lately (Evensen 2003, 2006; Krymskaya et al. 2009) and it has been successfully applied in a number of large-scale systems, such as quasi-geostrophic ocean model (Evensen 1994), wild land fire simulation (Mandel et al. 2008) and atmospheric model (Houtekamer and Mitchell 2001).

4.3.1 Ensemble Kalman Filtering

In EnKF, assume x_k is a n-dimension vector, representing the system state at the end of period k; z_k is an m-dimension vector, representing the measurement of x_k. The system transition and the observation process follow (4.1) and (4.2).

State Transition Equation: $x_k = g_k(x_{k-1}, \varepsilon_k)$ (4.1)

Measurement Equation: $z_k = h_k(x_k, \eta_k)$ (4.2)

where $g_k(.)$ is the state transition function, $h_k(.)$ is the measurement function, showing how the observation z_k is related to the current system state (x_k) .

ε_k is the process noise in period k. $\varepsilon_k \sim i.i.d.\, N(0, Q_k)$

η_k is the measurement noise in period k. $\eta_k \sim i.i.d.\, N(0, R_k)$

Ensemble Kalman Filtering (EnKF) uses Monte Carlo method to approximate the nonlinear state transition function $g_k()$. In Monte Carlo method, a set of data points (instances) are sampled to describe the probability density of the system state x_k. Let V is the ensemble size, each instance x_{ik} is an n-dimension state vector. The group of data points $\{x_{ik}\}_{i=1..V}$ is called an 'ensemble'. The statistical prosperities of the state variable x_k can be calculated

from the ensemble. For example, the sample mean $\overline{x_{ik}}$is the unbiased optimal estimator of x_k and it converges to x_k when the ensemble size approaches to infinite due to the Central Limit Theorem and Monte Carlo method.

Similar to Kalman Filtering (see Section 3.3), EnKF recursively follows a two-stage procedure to estimate the system state x_k. Let superscript '-' represent the results after the 'prediction' stage (called prior estimation), '+' represent the results after the 'update' stage (called posterior estimation). x_k and P_k are the estimation and the error covariance matrix of the system state in period k.

At the 'prediction' stage (see (4.3)-(4.4)), each instance in the ensemble, $x^+_{i(k-1)}$, is projected from time k-1 to k by the Monte Carlo approximation of $g_k(\cdot)$. The process noise ε_k is sampled in the transition of individual instances, thus the propagation of noise is implicitly included in the density probability of x^-_{ik}. The sample mean and sample covariance of the ensemble $\{x^-_{ik}\}_{i=1..V}$ are taken as prior state estimation (x^-_k) and prior error covariance matrix (P^-_k).

At the 'update' stage (Equations (4.5)-(4.7)), the prior estimation x^-_k is updated when the measurement z_k is available. The approximate of Kalman Gain, K_k, is calculated in (4.5) and it decides how much the posterior estimation x^+_k will be adjusted to consider the deviation between the prior estimation and the observation $(z_k - H'_k x^-_k)$ in (4.6). The calculation of K_k actually follows the update procedure of Kalman Filtering (see Section 3.3). It ensures that the estimate x^+_k is the optimal estimator in a linear Gaussian system. When the process noise ε_k is not Gaussian, x^+_k is still the best linear estimator (Welch and Bishop 2001). Note that K_k is an approximate of the Kalman Gain in Kalman

Filtering, because its calculation is based on the ensemble error covariance P_k^-, instead of the true state error covariance. However, it will converge to the true Kalman Gain when P_k^- converges to the true state error covariance matrix with increasing ensemble size. Incorporating the latest measurement z_k, x_k^+ and P_k^+ are called posterior estimate and error covariance matrix of the system state.

$$x_k^- = \frac{\sum_{i=1}^{V} x_{ik}^-}{V}, \text{ where } x_{ik}^- = g_k(x_{i(k-1)}^-,\ \varepsilon_{ik}), \varepsilon_{ik} \sim N(0, Q_k) \tag{4.3}$$

$$P_k^- = \left[\sum_{i=1}^{V}(x_{ik}^- - x_k^-)(x_{ik}^- - x_k^-)^T\right]/(V-1) \tag{4.4}$$

$$K_k = P_k^- {H'_k}^T \left(H'_k P_k^- {H'_k}^T + R_k\right)^{-1} \tag{4.5}$$

where H'_k is the Jacobian matrix of $h_k()$ evaluated at x_k^-. The operator $(\cdot)^{-1}$ represents pseudo inverse if the matrix inside is singular.

$$x_k^+ = \frac{\sum_{i=1}^{V} x_{ik}^+}{V}, \text{ where } x_{ik}^+ = x_{ik}^- + K_k(z_{ik} - H'_k x_{ik}^-) \tag{4.6}$$

$$P_k^+ = \left[\sum_{i=1}^{V}(x_{ik}^+ - x_k^+)(x_{ik}^+ - x_k^+)^T\right]/(V-1) \tag{4.7}$$

Equations (4.6) and (4.7) can be written in simpler forms as:

$$\begin{aligned} x_k^+ &= x_k^- + K_k\left(z_k + \frac{\sum_{i=1}^{V} z_{ik}}{V} - \frac{\sum_{i=1}^{V} H x_{ik}^-}{V}\right) \\ &= x_k^- + K_k(z_k - H x_k^-) \end{aligned} \tag{4.8}$$

$$\begin{aligned} P_k^+ &= \frac{\left[\sum_{i=1}^{V}(x_{ik}^+ - x_k^+)(x_{ik}^+ - x_k^+)^T\right]}{V-1} \\ &= \frac{\sum_{i=1}^{V}[(I-K_kH)(x_{ik}^- - x_k^-)(x_{ik}^- - x_k^-)^T(I-K_kH)^T + K_k(\varepsilon_{ik})(\varepsilon_{ik})^T K_k^T]}{V-1} \\ &= (I - K_kH)P_k^-(I - K_kH)^T + K_kR_kK_k^T \\ &= (I - K_kH)P_k^- \end{aligned} \tag{4.9}$$

4.3.2 Some Properties of Ensemble Kalman Filtering

In this section, we prove that the estimation error of system states is monotonically reduced by integrating new measurement data. The resulting lemmas quantify the benefits of new measurements and theoretically support the integrating of new data into an online Monte Carlo simulation using EnKF.

We suppress the subscript k. The superscripts '+' and '-' are omitted when the derivation is applied to both prior estimation and posterior estimation. In this section, let the subscript i represent the i^{th} state variable in the state vector.

With the concerns that the ensemble error covariance matrix P obtained from (4.4) and (4.7) are only approximates instead of the true state error covariance matrixes, we first prove that P converges to the true error covariance matrix when the ensemble size $V \to \infty$.

Lemma 1. The ensemble error covariance P^- and P^+ converge to the true prior and posterior error covariance when the ensemble size $V \to \infty$.

Proof: Let p_{ij} denote the ij element in P, which is the sample covariance of i^{th} state variable and the j^{th} state variable. Let $\overline{x_i}$ is the sample mean of the i^{th} state variable. According to the theory of Monte Carlo Simulation and Central Limit Theorem, $\overline{x_j}$ is an unbiased estimator of the mean of x_i , denoted by $E(x_i)$. Mathematically presented as $\lim_{V\to\infty} \overline{x_i} = E(x_i)$ and $\lim_{V\to\infty} \overline{x_j} = E(x_j)$.

Therefore, we have

$$\lim_{V\to\infty} p_{ij} = \lim_{V\to\infty} E\left[(x_i - \overline{x_i})(x_j - \overline{x_j})\right]$$

$$= E\left[\left(x_i - E(x_i)\right)\left(x_j - E(x_j)\right)\right] \quad (4.10)$$

Since each element of matrix P (p_{ij}) converges to the true error covariance element, matrix P converges to the true error covariance matrix in the limit of an infinite ensemble size. ■

It is worth mentioning that the trace of P (the sum of diagonal elements of P) equals to the sum of square error of state variables.

$$\lim_{V\to\infty} p_{ii} = V(x_i) \tag{4.11}$$

$$\lim_{V\to\infty} p_{ii}\,\mathrm{Tr}(P) = \lim_{V\to\infty} p_{ii} \sum_{i=1}^{i=n} p_{ii} = \sum_{i=1}^{n} V(x_i) \tag{4.12}$$

Where $E(\cdot)$ and $V(\cdot)$ are the expectation and variance of a random variable. $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix.

Next, we prove two lemmas to show the changes in sample error covariance matrix at each update.

Lemma 2. The sum of square error of state variables, $\sum_{i=1}^{n} V(x_i)$, is reduced at each update by $\lim_{V\to\infty} \mathrm{Tr}(P^- H^T (HP^- H^T + R)^{-1} HP^-)$, which is larger than or equal to 0.

Proof: From (4.9), (4.12),

$$\mathrm{Tr}(P^+) = \mathrm{Tr}(P^-) - \mathrm{Tr}(KHP^-) \tag{4.13}$$

$$\sum_{i=1}^{n} V(x_i^+) - \sum_{i=1}^{n} V(x_i^-) = \lim_{V\to\infty} \mathrm{Tr}(KHP^-) \tag{4.14}$$

Substituting (4.5) into (4.14) yields

$$\sum_{i=1}^{n} V(x_i^+) - \sum_{i=1}^{n} V(x_i^-) = \lim_{V\to\infty} \mathrm{Tr}(P^- H^T (HP^- H^T + R)^{-1} HP^-) \tag{4.15}$$

Since P^- and R are covariance matrix, $HP^- H^T + R$ is symmetric and positive-semidefinite. Its inverse $(HP^- H^T + R)^{-1}$ is also positive-semidefinite.

According to the definition of positive semi-definite matrix, there is

$\mathrm{Tr}(\mathrm{P}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP}^{-}) \geq 0.$

So, $\lim_{\mathrm{V}\to\infty}\mathrm{Tr}(\mathrm{P}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP}^{-}) \geq 0$

From (4.11) and (4.12), we get $\sum_{\mathrm{i}=1}^{\mathrm{n}} V(x_i^{+}) \leq \sum_{\mathrm{i}=1}^{\mathrm{n}} V(x_i^{-})$. ■

Lemma 3. The square error in the i^{th} state variable, $\mathrm{V}(\mathrm{x_i})$, is reduced at each update by $\lim_{\mathrm{V}\to\infty}\mathrm{P_i}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP_i}^{-\mathrm{T}}$, which is larger than or equal to 0.

Proof: According to Equation (4.5) and (4.9),

$$\mathrm{p_{ii}^{+}} = \mathrm{p_{ii}^{-}} - \mathrm{P_i}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP_i}^{-} \qquad (4.16)$$

Considering Equation (4.9), we have

$$V(x_i^{+}) - V(x_i^{-}) = \lim_{\mathrm{V}\to\infty}\mathrm{P_i}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP_i}^{-} \qquad (4.17)$$

Because $(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}$ is positive-semidefinite, $\mathrm{P_i}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R}-1\mathrm{HPi}-\geq 0$. That implies $Vxi+\leq Vxi-$. ■

4.4 The Supply Chain Model

4.4.1 Notation

Let subscript k denote time period. Let superscript i represent the number of the node. For convenience, the end-consumer is numbered as node 0 and the external supplier is node N+1.

L^{i} : production/transportation lead time from node i+1 to i, a nonnegative integer

l^{i} : information lead time from node i to i+1, a nonnegative integer

$\mathcal{L}^{i}$: replenishment lead time at node i. $\mathcal{L}^{i} = l^{i} + \mathrm{L}^{i}$

h^{i} : inventory holding cost rate at node i

b^i : backorder cost rate at node i

C_k^i : local cost at node i in period k.

$\mathcal{L}^{i(j)}$: replenishment lead time from node i to j (j>i) . $\mathcal{L}^{i(j)} = \mathcal{L}^i + \mathcal{L}^{i+1} + .. \mathcal{L}^{j-1}$

$l^{i(j)}$: information lead time from node i to j (j>i) . $l^{i(j)} = l^i + l^{i+1} + .. l^{j-1}$.

ε_k^i : random noise in the demand process of node i in period k

α^i : moving average coefficient in the demand process of node i

d_k^i : customer demand of node i in period k. It equals to the orders placed by node i-1 in period $k - l^{i-1}$

D_k^i : lead time demand of the next $\mathcal{L}^i$ periods from period k at node i.

$D_k^i = d_{k+1}^i + d_{k+1}^i + \cdots d_{k+\mathcal{L}^i}^i$

Δ_k^i : demand history from period 1 to period k at node i. $\Delta_k^i = \{d_k^i, d_{k-1}^i, \ldots , d_1^i\}$

$\mathcal{E}_k^i$: history of demand noise from period 1 to period k at node i.

$\mathcal{E}_k^i = \{\varepsilon_k^i, \varepsilon_{k-1}^i, \ldots , \varepsilon_1^i\}$

IN_k^i : on-hand inventory or backlog. When it is positive, it is the on-hand inventory at the end of period k; when it is negative, its absolute value is backorders at the end of period k

OR_k^i : orders placed to supplier at node i in period k

SR_k^i : shipment sent to customer at node i in period k

WIP_k^i : work in process at node i in period k

$E(\cdot)$: expectation of the random variable

$V(\cdot)$: variance of the random variable

$STD(\cdot)$: standard deviation of the random variable

4.4.2 Model Description

Considering a simplified N-echelon supply chain in a series, where end-consumer orders arrive at node 1, which will order from node 2, and so on. Let us call node i the immediate 'downstream' node to node i+1, and node i+1 the immediate 'upstream' node to node i. The last node N orders from an outsider supplier which always has sufficient supply. At each node, the received orders will be shipped immediately if there is sufficient on-hand inventory. Unfulfilled orders are fully backlogged. A well-known example of such a serial supply chain is the 'Beer Game' model (Fangruo and Rungson 2000) shown in Figure 12. The supply chain consists of four echelons - Retailer (node 1), Wholesaler (node 2), Distributor (node 3) and Factory (node 4).

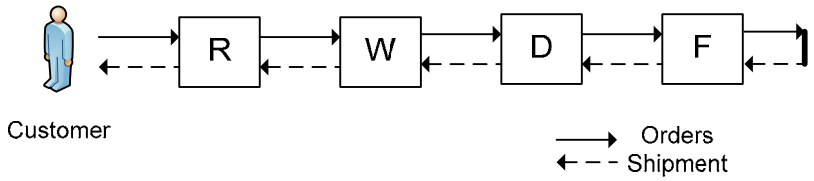


Figure 12. Beer Game supply chain model (N=4)

Similar to the Beer Game model described in Fangruo and Rungson (2000), we assume one product, unlimited capacity, constant lead times, constant cost coefficient in our model. At time period k, the following events happen in sequence at each node i. (1) Orders (d_k^i) that are placed by the customer node i-1 in period k- l^{i-1} is received. (2) The shared information to/from other nodes is sent/ received if information sharing is adopted. (3) The lead time demand (D_k^i) is forecasted and the manager determines this period's replenishment order OR_k^i based on the forecast. (4) The orders that are shipped by supplier in period k-

L^i are received. (5) Products are shipped to customers according to the customer demand. Any demand that cannot be met is backordered. (6) Inventory cost and backlog cost is evaluated.

There are three aspects where our model is different with that of Fangruo and Rungson (2000). First, information sharing strategy is assumed, that is, each node may send/ receive information from other nodes. The customer demand, received at the beginning of each period, could be shared immediately at the same period. Second, the lead time demand forecast needs to be updated at each period because the demand is non-stationary (the probability distribution changes along the time). The order-up-to level is updated accordingly. Third, backorder cost is charged at each stage instead of only being charged at the first echelon. Each node is treated as an independent cost center. Note that the inventory control decisions are distributed at different nodes. Instead of trying to optimize the whole supply chain performance as in Fangruo and Rungson (2000), each node sets the goal to minimize local cost.

4.4.3 End-consumer Demand – ARIMA(0,1,1) model

ARIMA is a general class of time-series model (Box and Jenkins 1994). It is often used to describe the end-consumer demand in a supply chain management. We assume the end-consumer demand $\{d_k^1\}$ is an ARIMA(0,1,1) process, and it is presented as

$$d_1^1 = \mu^1 + \varepsilon_1^1 \tag{4.18}$$

$$d_k^1 = d_{k-1}^1 - (1 - \alpha^1)\varepsilon_{k-1}^1 + \varepsilon_k^1 \tag{4.19}$$

where μ^1 is the process average, α^1 is the moving average coefficient ($0 < \alpha^1 \leq 1$) and the noise series $\varepsilon_k^1 \sim i.i.d\ N(0, \sigma^2)$

Given $0 < \alpha^1 \leq 1$, the process is non-stationary. A larger value of α^1 leads to a less stable process. Considering the initial condition in (4.18) and repeatedly applying (4.19), we get (4.20), where the demand at period k can be divided to two parts 1) the unknown random noise ε_k^1 and 2) the discounted past random noises adding the initial process mean $(\alpha^1 \sum_{t=1}^{t=k-1} \varepsilon_t^1 + \mu^1)$, that is already known before period k. $\alpha^1 \sum_{t=1}^{t=k-1} \varepsilon_t^1 + \mu^1$ implies that the random noise ε_{k-1}^1 shifts the mean of the process by $\alpha^1 \varepsilon_{k-1}^1$. Especially, when $\alpha^1 = 1$, the demand process (4.19) is a random walk.

$$d_k^1 = \varepsilon_k^1 + \alpha^1 \sum_{t=1}^{t=k-1} \varepsilon_t^1 + \mu^1 \tag{4.20}$$

4.4.4 Adaptive Order-up-to Policy

Next, we introduce the adaptive order policy used in the model. The superscript i is omitted because the discussion is applied to each node. Let y_k be the inventory position (= on-hand inventory + outstanding order - backlog) after placing the order in period k, OR_k be the orders placed to the supplier, d_k is the demand received, the balance equation of inventory position is

$$y_k = y_{k-1} - d_k + OR_k \tag{4.21}$$

In an order-up-to policy, a node orders OR_k from its supplier to bring inventory position to the desired 'order-up-to level', denoted by y_k^*. Let $E(D_k|k)$ and $V(D_k|k)$ represent the forecast and forecast error of the lead time demand ($D_k = \sum_{t=1.}^{t=\mathcal{L}} d_{k+t}$) at period k.

$$y_k^* = E(D_k|k) + z\sqrt{V(D_k|k)} \quad (4.22)$$

Using (4.20) and (4.21), the order rate in period k is

$$OR_k = y_k^* - y_{k-1}^* + d_k \quad (4.23)$$

$$OR_k = E(D_k|k) - E(D_{k-1}|k-1) + z\big(V(D_k|k) - V(D_{k-1}|k-1)\big) + d_k \quad (4.24)$$

When any of the forecast $E(D_k|k)$ and its error variance $V(D_k|k)$ changes over time, the desired inventory level y_k^* will change accordingly, which is called the 'adaptive order-up-to policy' (Graves 1999). z is a factor which is decided by the desired service level. Given that local cost $C_k = h \cdot \max(IN_k, 0) + b \cdot \max(-IN_k, 0)$ is a linear combination of on-hand inventory charge and backlog charge, the value of z for the myopic optimal policy is $\Phi^{-1}(\frac{b}{h+b})$, where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative density function of standard normal distribution (Zhang 2004).

A forecasting method with minimum forecast error $\sqrt{V(D_k|k)}$ is the optimal solution to minimize the one-period cost. To prove that, we can derive the expectation of cost at period $k + \mathcal{L}$, which is decided by the orders sent in period k, as

$$E(C_{k+\mathcal{L}}) = [h \cdot Q^+(-z) + b \cdot Q^-(z)]\sqrt{V(D_k|k)} \quad (4.25)$$

where $Q^+(-z) = E[(x-z)^+|x\sim Normal(0,1)]$,

$Q^-(z) = [(z-x)^+|x\sim Normal(0,1)]$, are functions that only depend on z (Zhang 2004).

4.5 Analytical Forecast – with/without information sharing

In the non-information sharing scenario, the manager is unaware of the demand at other nodes and merely observes the orders from its customer. Forecast method only uses local demand history $\Delta_k^i = \{d_k^i, d_{k-1}^i, \dots, d_1^i\}$. In the information sharing scenario, besides local information Δ_k^i, the manager also receives demand information from downstream node j, Δ_k^j. An application example of the information sharing situation is the sharing of Point-of-Sale data (Steckel et al. 2004).

In this section, we present the MMSE forecast for both with and without information sharing scenarios in a supply chain model described in Section 4.4. Here we assume the shared demand information does not have noise and it only comes from one of the downstream node. In the simulation approach (Section 4.6), we will relax these two constraints.

4.5.1 Non-information sharing

It is known that, for an ARIMA(0,1,1) process, the Minimum Mean Square Error (MMSE) forecast is the Single-Exponential Smoothing (SES) method. Graves (1999) shows that, for a multi-echelon serial supply chain, if the end-consumer demand is ARIMA(0,1,1) and each node adopts the adaptive order-up-to policy with SES forecast, the demand at each node is an ARIMA(0,1,1) process.

Without assuming information lead time, information sharing is not beneficial in the supply chain, because each node can get the optimal forecast if it can derive its own ARIMA(0,1,1) process parameters.

We extend Graves's model to incorporate information lead time and find that the demand at each node is still ARIMA(0,1,1) process. According to Proposition A1, the parameters of demand process at node i are:

$$\alpha^{i} = \alpha^{i-1}/(1 + \mathcal{L}^{i-1}\alpha^{i-1}). \tag{4.26}$$

$$\varepsilon_{k}^{i} = \left(1 + \mathcal{L}^{i-1}\alpha^{i-1}\right)\varepsilon_{k-l^{i-1}}^{i-1} \tag{4.27}$$

$$\mu^{i} = \mu^{1} \tag{4.28}$$

By recursively using (4.26) and (4.27), we observe that the demand noise at node i in period k is actually the amplified noise of end-consumer demand delayed by the information lead time $l^{1(i)}$.

$$\varepsilon_{k}^{i} = \left(1 + \mathcal{L}^{1(i)}\alpha^{1}\right)\varepsilon_{k-l^{1(i)}}^{1} \tag{4.29}$$

When $l^{1(i)} = 0$, the result is consist with the findings in Graves (1999).

In Proposition A3, we derive the MMSE forecast of lead time demand and the forecast error variance in this extended model:

$$\mathrm{E}\left(\mathrm{D}_{k}^{i}|\Delta_{k}^{i}\right) = \mathcal{L}^{i}\mathrm{F}_{k+1}^{i} \tag{4.30}$$

where F_{k}^{i} is SES forecast at node i at time k, refer to (4.20) and (4.21)

$$\mathrm{V}\left(\mathrm{D}_{k}^{i}|\Delta_{k}^{i}\right) = \mathrm{V}\left(\varepsilon_{k}^{i}\right)\mathcal{L}^{i}\left\{1 + \alpha^{i}\left(\mathcal{L}^{i} - 1\right) + \frac{\left(\alpha^{i}\right)^{2}\left(\mathcal{L}^{i}-1\right)\left(2\mathcal{L}^{i}-1\right)}{6}\right\} \tag{4.31}$$

4.5.2 Information sharing

We further extend the model to consider information-sharing. Assume node i receives, in addition to local demand, demand information at downstream

node j. It is also assumed that node i know parameter (α^{j}) and μ^{j}, because it either receives the parameter information from node j or derives the values by analyzing history data $\mathrm{d}_{\mathrm{k}}^{\mathrm{j}}$.

When information is shared, the optimal MMSE forecast at node i and forecast variance are, according to Proposition A4:

$$\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}|\Delta_k^i, \Delta_k^j\right) = \mathcal{L}^{\mathrm{i}}\mathrm{F}_{\mathrm{k+1}}^{\mathrm{i}} + \sum_{\mathrm{t=1}}^{\mathrm{t=min}\,(l^{j(i)},\mathcal{L}^i)}\left[\alpha^{\mathrm{i}}\left(\mathcal{L}^{\mathrm{i}} - \mathrm{t}\right) + 1\right]\varepsilon_{\mathrm{k+t}}^{\mathrm{i}} \quad (4.32)$$

$$\mathrm{V}\left(\mathrm{D}_k^{\mathrm{i}}|\Delta_k^i, \Delta_k^j\right) = \mathrm{V}\left(\varepsilon_{\mathrm{k}}^{\mathrm{i}}\right)\mathcal{L}^{i*}\left\{1 + \alpha^{\mathrm{i}}\left(\mathcal{L}^{i*} - 1\right) + \frac{\left(\alpha^{\mathrm{i}}\right)^2\left(\mathcal{L}^{i*}-1\right)\left(2\mathcal{L}^{i*}-1\right)}{6}\right\} \quad (4.33)$$

where $\mathcal{L}^{i*} = max(\mathcal{L}^{\mathrm{i}} - l^{j(i)}, 0)$ and $l^{j(i)}$ is the information lead time from node j to node i.

Comparing (4.30) and (4.32), we see that by observing Δ_k^j, node i is informed of the end-consumer noise $\varepsilon_{\mathrm{k}}^{1}$ earlier by $l^{j(i)}$ periods and it can use the information to improve forecast of lead time demand. As a result, the forecast error $\mathrm{V}\left(\mathrm{D}_{k+1}^{\mathrm{i}}|\Delta_k^i, \Delta_k^j\right)$ is less than $\mathrm{V}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}|\Delta_{\mathrm{k}}^{\mathrm{i}}\right)$ if $l^{j(i)}>0$. Comparing (4.31) and (4.33), we can see that the benefits of information sharing is actually decided by $l^{j(i)}$, which measures how many periods the demand at node j is advanced than the demand at node i. The forecast error becomes 0 when $\mathcal{L}^{\mathrm{i}} \leq l^{j(i)}$.

The orders from node i still follow an ARIMA(0,1,1) process. The parameters for the ordering process at node i (the demand process of its supplier node i+1) are:

$$\alpha^{\mathrm{i+1}*} = \alpha^1/(1 + \mathcal{L}^{1(i)}\alpha^1 + \alpha^1\mathcal{L}^{i*}). \quad (4.34)$$

$$\varepsilon_{\mathrm{k}}^{\mathrm{i}*} = \left(1 + \mathcal{L}^{1(i)}\alpha^1 + \alpha^1\mathcal{L}^{i*}\right)\varepsilon_{\mathrm{k-l^{i-1}}}^{\mathrm{i-1}} \quad (4.35)$$

$$\mu^{i*} = \mu^1 \tag{4.36}$$

Comparing (4.26) - (4.29) with (4.34) - (4.36), we find that the amplification of demand noise (the bullwhip effect) is reduced at node i when information sharing is assumed. It implies that the upstream nodes will benefit from the information sharing by receiving smoother demand signals.

$$\frac{V(d^{i+1*})}{V(d^{i+1})} = \frac{V(OR^{i*})}{V(OR^{i})} = \left(\frac{1+\mathcal{L}^{1(i)}\alpha^1+\alpha^1\mathcal{L}^{i*}}{1+\mathcal{L}^{1(i+1)}\alpha^1}\right)^2 < 1 \tag{4.37}$$

Here we illustrate the analytical results by a numerical example. Assume i=3 (Distributor), j=1 (Retailer), $l^i = l^{i-1} = \cdots = l^1 = 1$, $\mathcal{L}^i = 3$, $\alpha = 0.25$, $V(\varepsilon_k^1) = 10$.

When this is no information sharing,

$$V\left(D_k^i | d_{1..k}^i\right) = (3 \times (1+1.5)^2 + 0.25 \times 2 \times 3 \times (1+1.5)$$

$$+ 0.25^2 \times 2 \times 3 \times \frac{5}{6}) \times 100 = 2281.65$$

When there is information sharing, $\mathcal{L}^{i*} = 3 - 2 = 1$

$$\left(D_k^i | d_{1..k}^i, d_{1..k}^j\right) = V\left(\varepsilon_{k+i}^i\right)\mathcal{L}^{i*}\left\{1 + \alpha^i\left(\mathcal{L}^{i*}-1\right) + \frac{\left(\alpha^i\right)^2\left(\mathcal{L}^{i*}-1\right)\left(2\mathcal{L}^{i*}-1\right)}{6}\right\}$$

$$= (2.5 * 10)^2 = 625$$

The standard deviation of forecast error is reduced by $\frac{\sqrt{2281.65}-\sqrt{625}}{\sqrt{2281.65}} = 52.3\%$

The orders from Distributor are less dynamical. The bullwhip effect is reduced by $1 - \left(\frac{1+\mathcal{L}^{1(i)}\alpha^1+\alpha^1\mathcal{L}^{i*}}{1+\mathcal{L}^{1(i+1)}\alpha^1}\right) = 15.4\%$

4.5.3 Discussion

Equation (4.33) shows that the value of information sharing is related to the information lag between node j and node i ($l^{j(i)}$), and the replenishment lead time of node i ($\mathcal{L}^i$). Equation (4.33) implies that if node i has a very long replenishment lead time ($\mathcal{L}^i > l^{1(i)}$), the most valuable demand information is the Point-of-Sale data. However, if the demand information transmits without delay (lead time information is 0), sharing demand information is redundant. The result is consistent with the findings in Graves (1999) and Gilbert (2005).

Equation (4.37) shows that the bullwhip effect at node i is reduced by information sharing. As we know that forecast is a factor to cause the bullwhip effect (Chen et al. 2000), the reduction of the bullwhip effect can be understood as a result of that the lead time covered by the forecast is reduced from $\mathcal{L}^i$ to $\mathcal{L}^{i*}$. However, this conclusion may not apply to other supply chain structure or demand patterns. For example, Zhang (2005) finds the delayed demand information (instead of advanced demand information) dampens the bullwhip effect in for AR(1) demand.

4.6 Calibrated Simulation based Forecast

In this section, we introduce the concept and implementation of a calibrated simulation.

4.6.1 Simulation Model – Initialization and Execution

One important feature of the calibrated simulation is that the model states and parameters are allowed to have uncertainty. Their distribution information is contained in the V parallel simulation instances, called an 'ensemble'. The initial

ensemble is sampled at the beginning of simulation. Each simulation instance is advancing independently, yet updated collectively when new information is available using EnKF. The approach that uses multiple simulation instances collectively to represent the system state is called 'multisimulation' (Mitchell and Yilmaz, 2008). With the ensemble, we can predict not only the mean, but also estimate the covariance of performance matrix.

The time line of the calibrated simulation is illustrated in Figure 13.

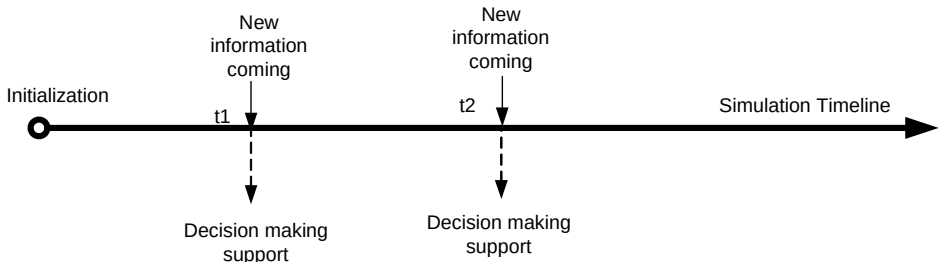


Figure 13. The execution timeline of a calibrated simulation

Initialization: Suppose the mean of initial state and the its error covariance are x_0 and P_0. We generate V samples of the initial state, denoted them by $\{x_{i0}\}$, i=1…V.

Execution: The simulation is paused when 1) real time information is available and the simulation needs to be calibrated, or 2) a decision making is needed. These are separated activities and can be considered as sequential events in the simulation lifetime (Figure 13). If the two activities are requested at the same time point, simulation calibration has a higher priority since data fusion can improve the model accuracy (see Lemma 2 and 3) to better support the decision making.

4.6.2 Simulation Calibration

The calibration requires an explicit definition of a system state vector x_k. The state vector doesn't need to include all the simulation states and parameters. Three basic rules should be considered when constructing the state vector:

- Constant variables, controlled variables that don't have uncertainty may NOT be included. It is simply because that their error variances are already 0 and will not benefit from update.
- Variables that are not correlated to any observed state variables may NOT be included. These variables will not benefit from update according to Lemma 2.
- The uncertain initial state or parameters, even they are constant in real world, should be included. With continuous integration of new data, the calibrated simulation will improve the estimation of the unknown parameters to finally reach its true value. This kind of problem is called parameter estimation and one example is provided in Welch and Bishop (2001).

Note that it will not be harmful if more variables are included than necessary; however, the computational cost for the calculation of covariance matrix will increase with the size of state vector.

For the specific supply chain model we studied (Figure 12), there are two types of entities in a supply chain: end-consumer and supply chain nodes. For this system, we design a 22-dimension state vector. The first two state variables are used to track the AMIMA end-consumer demand. A description of how to present

ARIMA demand in a state space model is provided by Aviv (2003). For an ARIMA(0,1,1) process, the state vector has two variables $Y_k = (y_1\ y_2)^T$ and the model is,

$$Y_k = \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix} Y_{k-1} + \begin{pmatrix} 0 \\ \epsilon_t \end{pmatrix},\ \epsilon_t \sim N(0, \sigma^2) \tag{4.38}$$

Where ϵ_t is the inherent noise in the time-series, α is the moving average coefficient.

This expression is equivalent to (4.18) and (4.19) when $d_k^1 = (1\ 1)Y_k + \mu$.

Besides, the state vector contains the states for each supply chain node. Each node has five state variables, $X^i = (IN_k^i, WIP_k^i, OR_k^i, SR_k^i, E(D_k^i))^T$. As a result, the whole system vector is a 22-dimension vector shown in (4.39)

$$x = \begin{bmatrix} Y \\ X^1 \\ X^2 \\ X^3 \\ X^4 \end{bmatrix} \tag{4.39}$$

Since EnKF does not set any constraint on the transition function g(), a Monte Carlo simulation can be used to replace the nonlinear g() function to advance the state vector. The sampling of process noise is included in the execution of individual simulation instances. Given that the simulation has advanced to period k and get the prior estimation of system state x_k^-, at the time point of simulation calibration, the real-world information is treated as measurement and H_k and R_k need to be determined. The prior covariance matrix of the system state (P_k^-) is estimated from the states of the simulation ensemble in period k. Then, the K_k matrix is calculated. The observations and simulation

predictions are compared and their differences $(z_k - Hx_{ik}^- - \eta_{ik})$ are used to determine how much the prior simulation state x_{ik}^- should be adjusted. Each simulation instance $x_{ik}^-$is then updated according to (4.6) to get x_{ik}^+.

Lemma 4. If we have m set of observations with observation functions H_j and observation errors R_j, j = 1.. m. We define two ways of data fusions. By Method 1, we call (4.5) – (4.7) recursively (without using (4.3) -(4.4)) until all the m observation sets are integrated. By Method2, we define $H = (H_1^T, \ldots H_m^T)^T$ and $R = \begin{pmatrix} R_1 & 0 \ldots & 0 & 0 \\ 0 & R_2 & \ldots & 0 \\ 0 & \ldots & \ldots & 0 \\ 0 & 0 & \ldots & R_m \end{pmatrix}$ and apply (4.5) – (4.7) one time. It turns out that these two methods generate the same result.

Proof: Refer to Willner et al. (1976).

Lemma 4 indicates that we can either integrate the observation m times, each time using a 1 by n observation matrix H, or using an m by n matrix to fusing information, if their observations errors are independent.

4.6.3 Simulation-based prediction

As mentioned in Section 4.4.2, for the adaptive order-up-to control, the desired order-up-to level (y_k^*) is decided by the forecast of lead time demand and the forecast error.

$$y_k^* = E(D_k|k) + z\sqrt{V(D_k|k)}$$

Where $z = \Phi^{-1}(\frac{b}{h+b})$, $\Phi^{-1}(\cdot)$ is the inverse of the cumulative density function of standard normal distribution.

Note $E(\mathrm{D_k}|\mathrm{k})$ and $V(\mathrm{D_k}|\mathrm{k})$ are conditional on previous information from period 1 to k-1. With the Calibrated Simulation approach, we can use the sample mean and mean variance of simulation results to predict $E(\mathrm{D_k}|\mathrm{k})$ and $V(\mathrm{D_k}|\mathrm{k})$, because all the available information has been integrated into the ensemble.

$$E(\mathrm{D_k}|\mathrm{k}) = \sum_{j=1}^{j=V}\left(D_k^i\right)_j \ /V \tag{4.40}$$

$$V(\mathrm{D_k}|\mathrm{k}) = \frac{\sum_{j=1}^{j=V}[\left(D_k^i\right)_j - E(\mathrm{D_k}|\mathrm{k})]}{V-1} \tag{4.41}$$

4.7 Experiments

Three experiments are presented in this section. In the first experiment, Calibrated Simulation-based Forecast (CSF) and analytical MMSE forecast are compared. The second experiment uses CSF to evaluate the value of information when observation noise is present. We also investigate the question-which downstream node provides the maximum value in their demand information? It is followed by exploring the scenarios with multiple information sources in the third experiment.

The supply chain model studied in these experiments has been introduced in Section 4.4.

4.7.1 Experiment I - Analytical Forecast vs. Calibrated Simulation-based Forecast

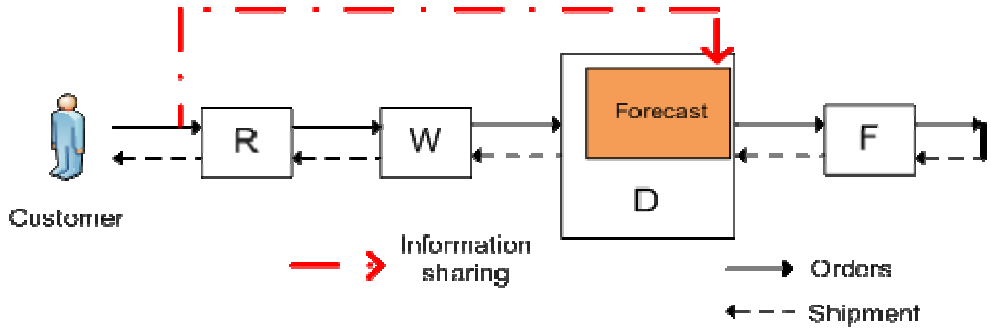


Figure 14. Supply chain model – information sharing between Retailer and Distributor

In this experiment, we study the scenario when there is non-noisy information sharing between a pair of supply chain nodes. Without generality, it is assumed the demand information is shared between Retailer and Distributor. That is, Distributor receives not only the local demand information, but also the demand information at Retailer at the beginning of each time period. There are two prediction options Distributor to choose from to forecast lead time demand $E(D_k|k)$ and estimate the forecast error $V(D_k|k)$:

- MMSE: Minimum Mean Square Error forecast. It is derived in Section 4.5. The calculation follows (4.32) and (4.33).
- CSF: Calibrated Simulation Forecast. It is introduced in Section 4.6.3. The calculation follows (4.40) and (4.41).

For comparison, a 'benchmark' model is implemented, when no information is shared and Distributor uses local information to make replenishment decision.

The configurations of the supply chain, Ensemble Kalman Filtering and the parameters for inventory control policies at different echelons are summarized in Table 16 - Table 18. For each forecast method and the benchmark model, 20 replicates are run, each with a different random seed to generate consumer demand. Each run has 250 periods.

Table 16. Supply chain configurations

Parameters for the End-consumer ARIMA(0,1,1) process	$\sigma_1 = 10, \alpha_1 = 0.25$
Unit backlog cost ($/unit)	$b_1 = b_2 = b_3 = b_4 = 5$
Unit on-hand inventory cost ($/unit)	$h_1 = 1, h_2 = 0.75, h_3 = 0.5, h_4 = 0.25$
Service level = h/(h+b)	0.84, 0.9, 0.93, 0.96 for node 1 to 4 respectively
Replenishment Lead Time	$\mathcal{L}^1 = \mathcal{L}^2 = \mathcal{L}^3 = 3, \mathcal{L}^4 = 2$
Information Lead Time	$l^1 = l^2 = l^3 = l^4 = 1$

Table 17. The configurations of Ensemble Kalman Filtering

The standard deviation of observation error	0
Ensemble size (V)	100

Table 18. Ordering policies and parameters at Retailer, Wholesaler, Distributor and Factory

	Non-information sharing (Benchmark)		Information sharing (MMSE, CSF)	
Parameters	α of SES	$STD(D_k\|k)$	α of SES	$STD(D_k\|k)$
Retailer	0.250	21.93	0.250	21.93
Wholesaler	0.143	34.82	0.143	34.82
Distributor	0.100	47.76	N/A	N/A
Factory	0.0769	47.76	0.0909	40.69

Three types of metrics are evaluated to compare the performance of different forecast methods. Note that i=1,2,3,4 for R, W, D, F

- Local cost

$$\text{Cost\_i} = \sum_{k=1}^{T}[h^i \max(IN_k^i, 0) + b^i \max(-IN_k^i, 0)]$$

Total supply chain cost

$$\text{CostTotal} = \sum_{i=1}^{4}\sum_{k=1}^{T}[h^i \max(IN_k^i, 0) + b^i \max(-IN_k^i, 0)]$$

- Stock-out time

$$\text{StockOut\_ i} = \#\text{ of stock-out periods at node i/ simulation time}$$

- Fill rate

$$\text{FR\_i} = \frac{\text{Sum(orders immediated filled at node i)}}{\text{sum(orders received at node i)}}$$

The experiment results are included in Table 19 (a) – (c). All the metrics are normalized by the performance of the benchmark model to reduce the impact of different demand seeds. It is observed that MMSE forecast and CSF have no significant differences in all performance matrices ($p<0.05$).

Table 19. (a) Cost _ MMSE and CSF Comparison

	MMSE		CSF		Method Comparison (P-value)
	mean	stddev	mean	stddev	MMSE vs. CSF
Cost_total	0.837	0.039	0.839	0.038	0.638
Cost_1 (R)	0.984	0.021	0.984	0.020	0.971
Cost_2 (W)	0.980	0.031	0.981	0.031	0.921
Cost_3 (D)	0.528	0.030	0.530	0.030	0.756
Cost_4 (F)	0.855	0.009	0.862	0.014	0.144

Table 19. (b) Stockout Time _ MMSE and CSF Comparison

	MMSE		CSF		Method Comparison (P-value)
	mean	stddev	mean	stddev	MMSE vs. CSF
Stockout_1 (R)	0.981	0.022	0.980	0.024	0.791
Stockout_2 (W)	0.957	0.068	0.957	0.077	0.588
Stockout_3 (D)	1.201	0.450	1.251	0.490	0.717
Stockout_4 (F)	1.085	0.257	1.124	0.214	0.460

Table 19. (c) Fill Rate _ MMSE and CSF Comparison

	MMSE		CSF		Method Comparison (P-value)
	mean	stddev	mean	stddev	MMSE vs. CSF
FR_1 (R)	1.001	0.002	1.001	0.002	0.908
FR_2 (W)	1.001	0.003	1.001	0.002	0.971
FR_3 (D)	1.001	0.001	1.001	0.001	0.596
FR_4 (F)	1.000	0.001	1.000	0.001	0.637

Since there is no significant difference between two forecast methods, we take the average of cost saving of the two methods to represent the supply chain performance under information sharing. Compared with the benchmark model (non-information sharing), the supply chain under information sharing has significant ($p<0.05$) saving in total supply chain cost (Cost_total), Distributor Cost (Cost_D) and Factory Cost (Cost_F). The costs at the Retailer and Wholesaler are reduced slightly, but not statistically significant. In Figure 15, we see the distributor's cost is reduced by 52.8%, which is close to the derived value 52.3% in the example in Section 4.5.2. The reduction of demand noise at Factory

is estimated to be 15.3%, which is consistent with the analytical result shown in Table 19.

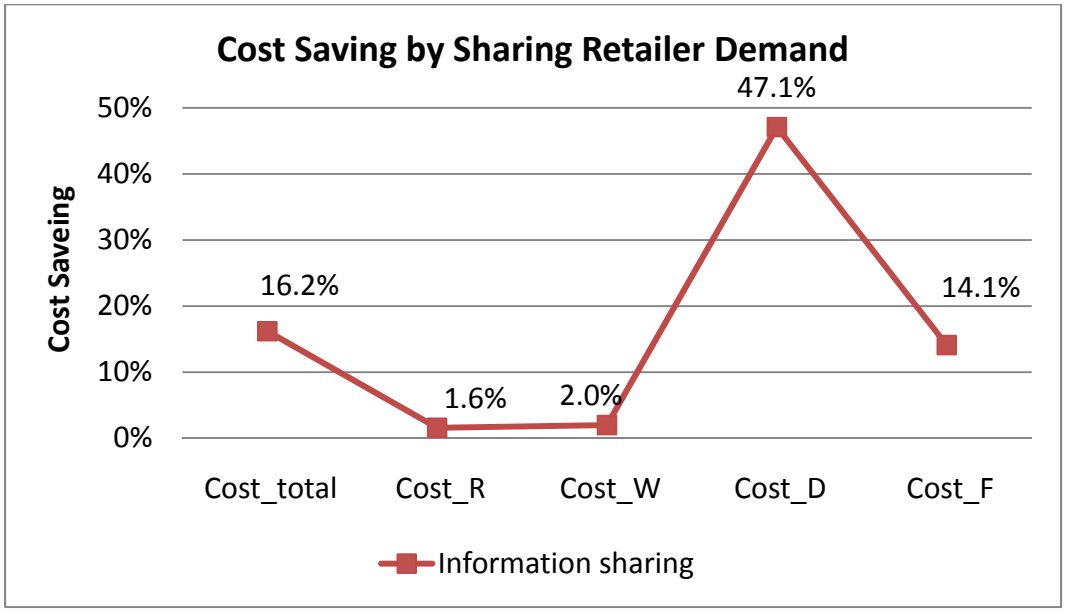


Figure 15. Cost comparison of information sharing vs. non-information sharing

Table 20. Reduction in demand forecast MSE at Factory

	Analytical result	Numerical experiment Result	
		MMSE	CSF
Mean Square Error (MSE)	27.5	27.24	27.46
Percentage of Reduction in MSE	15.3%	16.2%	15.5%

Note: the MSE of non-information sharing is 32.5

Through the above comparison study, we conclude that the analytical model introduced in Section 4.5 is correct and the performance of CSF is close to the optimal MMSE forecast in this specific case.

4.7.2 Experiment II – Information sharing with noisy data from different locations

The second experiment is aimed to demonstrate the performance of CSF under noisy information. The supply chain model is similar to that in Experiment I, except that Factory is the node receiving demand information.

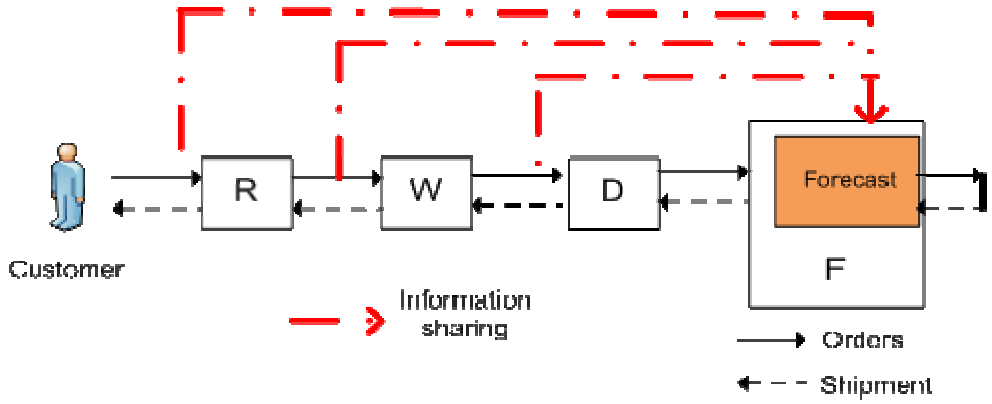


Figure 16. Supply chain model – information sharing between Factory and downstream nodes

The experiment design considers three factors – information source, the error in observations and the replenishment lead time at Factory. Other supply chain settings follow Table 16 and Table 17. In total, 45 experiment scenarios are designed (Table 23). For each scenario, 20 replicates of experiments are conducted to calculate the mean and variance of supply chain performance. Knowing that cost is positively related to forecast error (see Section 4.4.3), we use the cost saving at Factory as the metric to evaluate the performance of adaptive control. Three benchmark models with Factory replenishment lead time= 1, 2, 3 without information sharing are run to enable the calculation of cost saving.

Table 21. Experiment design – the impact of information source, observation error when Factory replenishment lead time = 1, 2, 3

Factors	Levels
Information Source (IS)	Retailer (j=1), Wholesaler (j=2), Distributor (j=3)
Observation error	Standard Deviation = 0, 5, 10, 20, 30
Replenishment lead time of Factory	$\mathcal{L}^4 = 1,2,3$

The experiments results are plotted in Figure 17 – Figure 19.

In the case when the replenishment lead time of Factory is long ($\mathcal{L}^4$ =3) shown in Figure 17, when STD(observation error)=0, the cost saving increases as the information source moves from Distributor to Wholesaler to Retailer, being 28.7%, 52.0% to 100% respectively. When the observation error increases, it is not surprised to see that the cost saving reduces. Interestingly, we find that the drop of cost saving is accelerated when the information provider moves downstream. The increase of observation error has the largest impact when Retailer's demand is shared and the least impact when the Distributor's demand is shared. As a result, the demand sharing from Retailer provides the highest cost saving when STD(observation error)<15 and Wholesaler becomes the most valuable information source when STD(observation error)>15.

Similar impacts of observations errors are found when the replenishment lead time in Factory is medium or short (Figure 18 or 19). When the Factory

replenishment lead time is 1 and there is no observation error, Factory is able to accurately predict its lead time demand by analyzing the demand of any downstream node when the observation error is 0. When observation error increases, the information from Distributor isthe best information source since it is most robust to observation errors.

We can gain some insight about why observation noise is more influential on downstream demand information than on upstream demand information from the update procedure of EnKF. In Lemma 2, we see that the reduced sum of square error is $\mathrm{Tr}(\mathrm{P}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP}^{-})$. In our case, H is a row vector because there is only one observation variable.

$$\mathrm{Tr}(\mathrm{P}^{-}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}^{-}\mathrm{H}^{\mathrm{T}}+\mathrm{R})^{-1}\mathrm{HP}^{-}) = \mathrm{Tr}(\mathrm{P}^{-}\mathrm{H}^{\mathrm{T}}\mathrm{HP}^{-})/(\mathrm{p}_{\mathrm{ii}}^{-}+\mathrm{R})^{-1} \qquad (4.42)$$

where $\mathrm{p}_{\mathrm{ii}}^{-}$ is the error variance of prior estimation of the ith state variable. R is the variance of observation error.

From (4.42), it is easy to see that the impact of R is relative to the value of $\mathrm{p}_{\mathrm{ii}}^{-}$. Since the noise in demand is amplified from downstream node to upstream node because of the bullwhip effect, $\mathrm{p}_{\mathrm{ii}}^{-}$ increases from Retailer to Wholesaler to Distributor. Therefore, the observation error R is less influential from Retailer to Wholesaler to Distributor. In other words, the demand information is less sensitive to the observation noise when the information source moves up the supply chain.

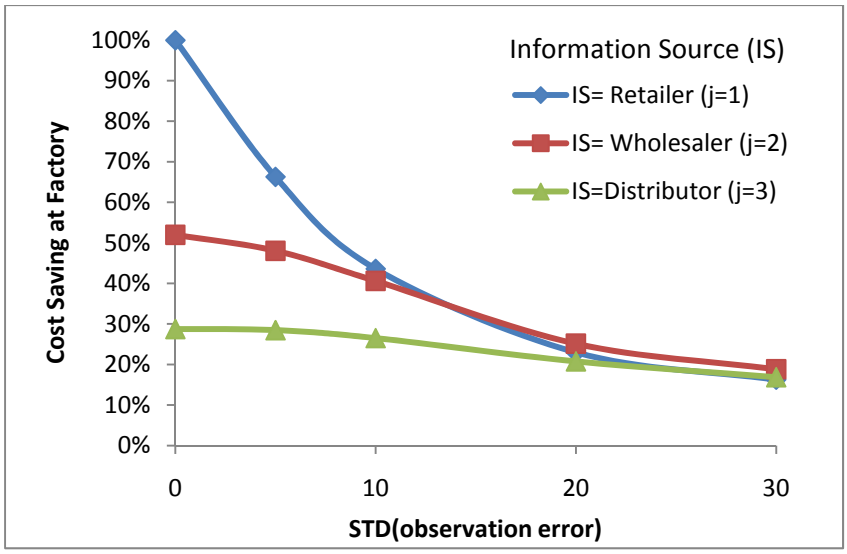


Figure 17. Cost saving under varying observation errors _ long Factory replenishment lead time (L^4=3)

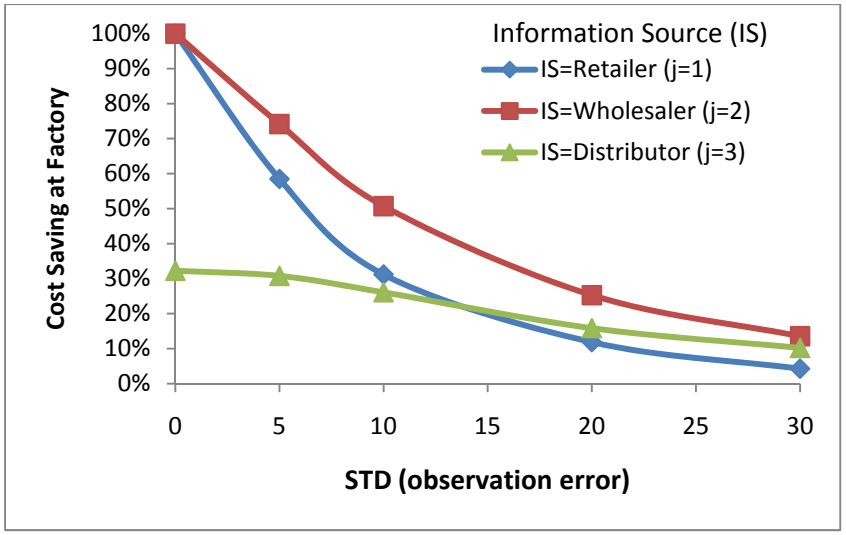


Figure 18. Cost saving under varying observation errors _ medium Factory replenishment lead time (L^4=2)

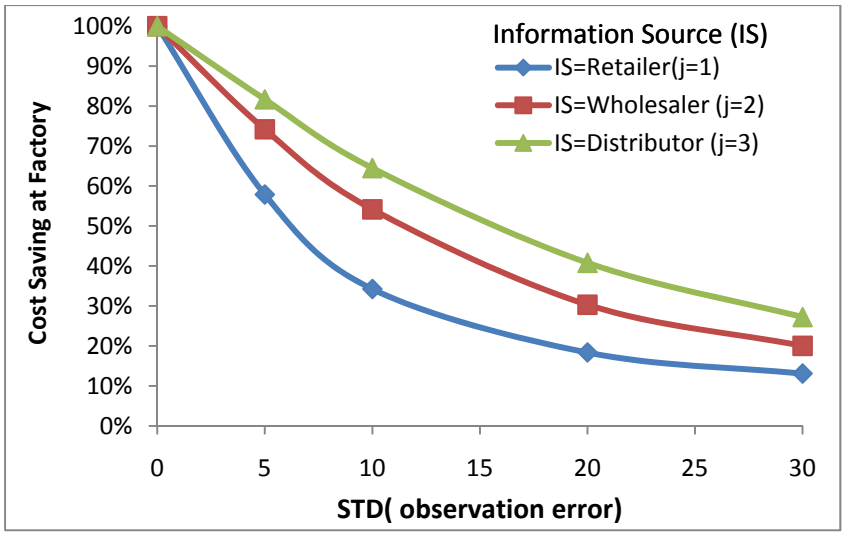


Figure 19. Cost saving under varying observation errors_ short Factory replenishment lead time (L^4=1)

In sum, we find there are two factors determining the values of information sharing in the studied supply chain. First, information source that is at the downstream side of the supply chain (closer to end-consumer) brings more advanced demand information to benefit the prediction of lead time demand. Second, the information provider that is at the upstream side of the supply chain (closer to Factory) is more robust to observation noise. Such tradeoff makes different information source is preferred when the replenishment lead time change (Figure 17 to Figure 18).

4.7.3 Experiment III– Single and Multiple Information Sources

In the last experiment, we investigate scenarios when multiple information sources exist. The questions include *how to integrate multiple items of data? Will more data brings more benefits?*

A similar information sharing scenario is assumed to as in Experiment II, that is, Factory receives downstream information and uses Calibrated Simulation-based Forecast to predict its future demand. However, in this experiment, the information may come from multiple nodes and the combinations of information sources are considered (Table 22). The observation error is fixed as STD(observation error)=10 and the replenishment lead time $\mathcal{L}^4 = 2$.

Table 22. Experiment design – single and multiple information sources

Factors	Levels
Information Source(s)	R, W, D R +W, W+ D, R+D R+W+D

Using the update schema of EnKF, it is flexible to incorporate more than one information source. When there is more than one measurement used in calibration, the observation is a vector instead of a scalar variable. Lemma 4 in Section 4.6.2 indicates that it is equivalent to integrate each information source one at a time (using a row vector H) or to integrate them together (using the H matrix). In our implementation, we use the latter method.

In

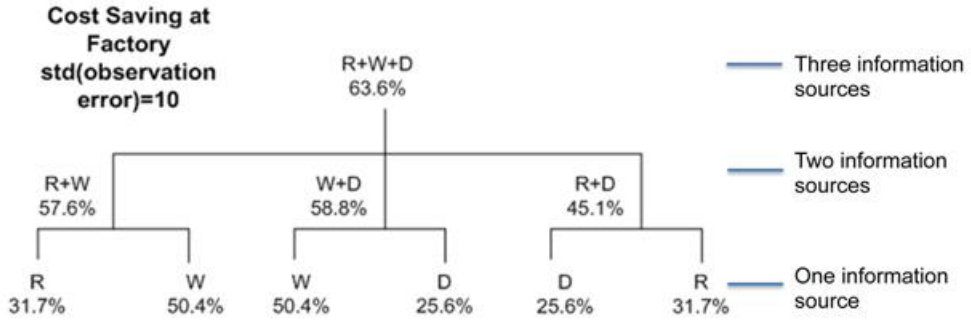


Figure 20, a tree plot shows the result when one or multiple information sources are used. At the lowest level, only one information source is adopted. As a result, Wholesaler is the best information source. At the middle level, two information sources are assumed. At the highest level, Factory uses the demand information from all the three downstream nodes. We find that the value at a parent node in the tree is no less than the value at any of its children, which implies that using additional information always improves the cost savings.

An interesting finding is that the combination of the two top individual information sources (W+D) doesn't provide the best information combination. Instead, W+R delivers more cost saving than W+D. This finding can be explained by the correlation between the information provided at different nodes.

Table 23. Cost saving at Factory (single and multiple information sources)

Information Sources(s)	Average Cost Saving at Factory	Standard deviation of Cost Saving at Factory
R	30.7%	5.9%
W	50.4%	4.5%
D	25.6%	3.9%
R+W	57.6%	3.2%
W+D	58.8%	2.9%
R+D	45.1%	3.2%
R+W+D	63.6%	2.8%

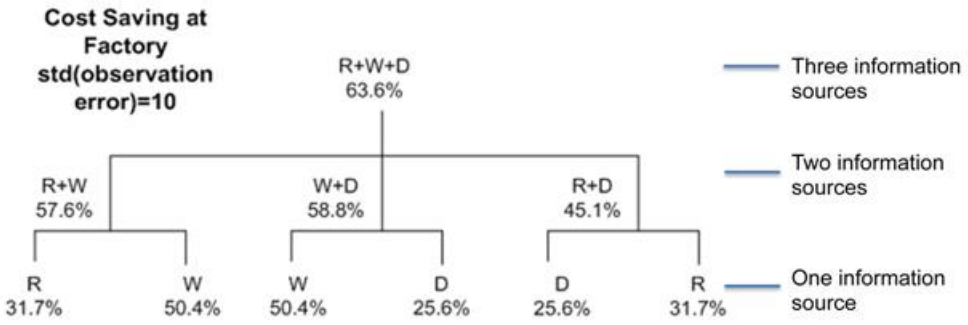


Figure 20. Cost saving at Factory (single and multiple information sources)

4.8 Summary and Future Research

In this chapter, we present the use of Calibrated Simulation in adaptive inventory control. We first evaluate the correctness of Calibrated Simulation-based Forecast (CSF) by comparing it with MMSE forecast when the shared information is not noisy. After the CSF approach is validated, we use CSF to explore scenarios when shared information has noise and when there are more than one information sources available.

Managerial insights are gained about the information sharing and adaptive inventory control in a serial supply chain with ARIMA (0,1,1) demand. First, information sharing downstream demand reduces the cost at the node that receives the information, and reduces the bullwhip effect of the node. There are two factors influence the selection of information provider. (1) When there is no observation noise, the benefit is conditional on the difference of the information lead time from the information provider to information receiver and the information receiver's replenishment lead time. A closed-form formula to

calculate the reduction in forecast error given downstream demand information is provided. (2) When the shared information is noisy, the downstream observation data is more sensitive to noise than upstream data. This is mainly because the demand noise are amplified through each echelon (the bullwhip effect), that makes the observation errors become less disturbing. When choosing the information provider, both factors take effect. The dominant strategy is changing depends on the lead time of Factory.

Second, incorporating data from multiple sources in the supply chain helps reduce the forecast variance in CSF, thus reduces the cost. Manager should notice that the combination of top information provider doesn't deliver the best value as a team.

We should mention that the simulation approach has its disadvantages compared to analytical methods. In implementing Ensemble Kalman Filtering, the computation cost grows high considering that we need to run V simulation in parallel with the physical system (emulation). However, we argue that since the simulation of each instance is independent and could be executed in parallel. The newly developed techniques such as distributed simulation, grid computing, could alleviate such pressure.

Last but not least, the integrated framework of simulation and a nonlinear data fusion method is flexible to extend to more complicated supply chain scenarios. This approach has the potential to predict the dynamical behaviors of the supply chain nodes from a system view by integrating different sources of noisy information. It can be used to provide real-time information for decision

making in adaptive supply chain management. In this study, we use it as a prediction tool. However, this framework could be coupled with other optimization techniques, such as simulation-based optimization, to find optimal control in real time.

Future work will be to further develop the simulation technique (e.g. sampling approach, guided information collection) and apply this approach in solving more complicated supply chain problem.

Chapter 5

CONCLUSIONS AND FUTURE RESEARCH

An integrated framework of online simulation and data fusion is proposed and developed to facilitate the study of adaptive management of large-scale complex supply chain systems. An online simulation is regularly calibrated by fusing noisy real world data to track the dynamics of a supply chain. This framework enables monitoring early signals of exceptional events and provides more accurate information for decision makers to manage a supply chain in a prompt manner.

The framework can be used to analyze the system-level behaviors of a nonlinear and complex supply chain system. It has several advantages over conventional simulation approaches. (1) It is able to take advantage of noisy information. Kalman Filtering-based data fusion provides an "optimal" way to integrate all available information taking care of the observation error to assure the deficit data will not harm the results. The model is flexible enough to incorporate various sources of information, such as expert opinion, marketing and sales information. (2) The simulation is continually validated by the process of data fusion. The deviation between simulation and emulation is reduced, in other words, a calibrated simulation usually has higher prediction accuracy than a traditional simulation model.

The proposed framework can be used to 'sense and respond' to the changing environment of a large-scale and complex supply chain. In Chapter 4, we have seen that this integrated approach is able to include more realistic

constraints (e.g. observation noise) to reveal different aspects of decision making. Potentially, the ability of the framework to incorporate multiple information sources and to consider the overall dynamics of the system will lead to better supply chain performance (e.g. cost saving).

In future research, this framework will be investigated to solve practical supply chain problems. One potential use of the calibrated simulation is to incorporate dispersed information to provide a comprehensive view of the risks in a company's supply system. Given the fact that there is a large amount of information available across a supply chain (e.g. RFID, transaction data, manufacturing data), a flexible and dynamic model like the calibrated simulation will help companies to translate information into knowledge that can lead to effective actions. In addition, several technical problems regarding to the framework implementation need be further addressed, such as how to reduce the computational cost, and how to handle discrete type of state variables.

BIBLIOGRAPHY

Intel News Release. Intel fourth-quarter revenue $9.7 billion. Intel News Release. 2007 http://www.intel.com/pressroom/archive/releases/20070117corp.htm) (accessed October 11, 2010)

Alfonso, S., R. Luis, et al. "Stability Analysis of the Supply Chain by Using Neural Networks and Genetic Algorithms." in Proceedings of the 39th conference on Winter simulation. Washington D.C., IEEE Press. 2007

Aviv, Y. "A Time-Series Framework for Supply-Chain Inventory Management." Operations Research 51, no. 2, 2003: 210-227.

Barker, A. L., D. E. Brown, et al. "Bayesian Estimation and the Kalman Filter." Computers and Mathematics with Applications 30, no. 10, 1995: 55-77.

Blackhurst, J., T. Wu, et al. "Network-based Approach to Modeling Uncertainty in a Supply Chain." International Journal of Production Research 42, no. 8, 2004: 1639–1658.

Blackhurst, J., T. Wu, et al. "A Systematic Approach for Supply Chain Conflict Detection with a Hierarchical Petri Net Extension." Omega 36, no. 5, 2008: 680-696.

Box, G. E. P. and G. M. Jenkins. Time Series Analysis: Forecasting and Control, Upper Saddle River, NJ, USA: Prentice Hall PTR, 1994.

Buckley, S., Ettl, M., Lin, G., Wang, K-Y. "Sense and Respond Business Performance Management." Supply Chain Management on Demand – Strategies, Technologies, Applications, Springer Berlin Heidelberg, 2006, 287-311.

Cerpa, N. and J. M. Verner. "Case study: The Effect of IS maturity on Information Systems Strategic Planning." Information & Management 34, no. 4,1998: 199-208.

Chen, F., Z. Drezner, et al.. "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead times, and Information." Management science 46, no. 3, 2000: 436-443.

Chen, Y., J. Fowler, et al. "An Adaptive Distributed Simulation Framework for a Server Fulfillment Supply Chain." IEEE Conference on Automation Science and Engineering, 2006: 649-655.

D. J. van der Zee, J. G. A. J. v. d. V. "A Modeling Framework for Supply Chain Simulation: Opportunities for Improved Decision Making." Decision Sciences, 36, 2005: 65-95.

Darema, F. Dynamic Data Driven Applications Systems: New Capabilities for Application Simulations and Measurements, Atlanta, GA, United States, Springer Verlag, Heidelberg, D-69121, Germany, 2005.

De Lannoy, G. J. M., P. R. Houser, et al. "State and Bias Estimation for Soil Moisture Profiles by an Ensemble Kalman Filter: Effect of Assimilation Depth and Frequency." Water Resources Research 43, no. 6, 2007: W06401.

Evensen, G. "Using the Extended Kalman Filter with a Multilayer Quasi-geostrophic Ocean Model." Journal of Geophysical Research 97, no. C11, 1992: 17905-17924.

Evensen, G. "Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model using Monte Carlo Methods to Forecast Error Statistics." Journal of geophysical research 99, no. C5, 1994: 10143-10162.

Evensen, G. "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation." Ocean dynamics 53, no.4, 2003: 343-367.

Evensen, G. Data Assimilation: The Ensemble Kalman Filter, Springer-Verlag New York, 2006.

Fangruo, C. and S. Rungson. "The Stationary Beer Game." Production and Operations Management 9, no. 1, 2000: 19.

Fleisch, E. and C. Tellkamp. "Inventory Inaccuracy and Supply Chain Performance: a Simulation Study of a Retail Supply Chain." International Journal of Production Economics 95, no. 3, 2005: 373-385.

Fordyne, K. "New Supply Chain Management Applications Provide Better Customer Service: Serious Gets Exciting." IBM Microelectronics, 2001: 15-19.

Fujimoto, R., M. Hunter, et al. “Ad Hoc Distributed Simulations.” in Proceedings of the 21st International Workshop on Principles of Advanced and Distributed Simulation, 15-24. IEEE Computer Society Washington, DC, USA. 2007.

Fujimoto, R., D. Lunceford, et al. Grand Challenges for Modeling and Simulation, Enabling Technologies for Simulation Science VI, Orlando, FL, USA, SPIE, 2002

Gav, B. "A Gas Supply Disruption Case Study - the Varanus Island Explosion." 2008. http://peakenergy.blogspot.com/2008/07/gas-supply-disruption-case-study.html(accessed October 11, 2010)

Gilbert, K. "An ARIMA supply chain model." Management Science 51, no. 2, 2005: 305-310.

Graves, S. C. "A Single-item Inventory Model for a Nonstationary Demand Process." Manufacturing & Service Operations Management 1, no. 1, 1999: 50-61.

Gunasekaran, A. and E. W. T. Ngai "Information Systems in Supply Chain Integration and Management." European Journal of Operational Research 159, no. 2, 2004: 269-295.

Haeckel, S. H. Adaptive enterprise: Creating and leading sense-and-respond organizations, Harvard Business School Press, 1999.

Hanley, J. A. and B. J. McNeil. "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve." Radiology 143, no. 1, 1982: 29-36.

Hau, L. L. "The Triple-A Supply Chain." Harvard Business Review 82, no. 10, 2004: 102.

Hayya, J. C., J. G. Kim, et al. "Estimation in Supply Chain Inventory Management." International Journal of Production Research 44, no. 7, 2006: 1313-1330.

Hendricks, K. B. and V. R. Singhal. "The Effect of Supply Chain Glitches on Shareholder Wealth." Journal of Operations Management 21, no. 5, 2003: 501-522.

Hendricks, K. B. and V. R. Singhal. "Association Between Supply Chain Glitches and Operating Performance." Management Science 51, no. 5, 2005: 695-711.

Houtekamer, P. L. and H. L. Mitchell "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation." Monthly Weather Review 129, no. 1, 2001: 123-137.

Hunter, M. P., R. M. Fujimoto, et al. “An Investigation of Real-time Dynamic Data Driven Transportation Simulation.” In Proceedings of the 38th conference on Winter Simulation, Monterey, CA, USA: IEEE, 2006: 1414 - 1421

Hwarng, H. B. and N. Xie. "Understanding Supply Chain Dynamics: A Chaos Perspective." European Journal of Operational Research 184, no. 3, 2008: 1163-1178.

Jain, V., S. Wadhwa, et al. "Revisiting information systems to support a dynamic supply chain: issues and perspectives." Production Planning & Control 20, no. 1, 2009: 17.

Johnson, B. "Turning Supply and Demand Uncertainty into Competitive Advantage". 2010. http://www.gsb.stanford.edu/scforum/login/documents/teradata2010-summary.pdf (accessed October 11, 2010)

Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." Journal of basic Engineering 82, no. 1, 1960: 35-45.

Koyuncu, N., S. Lee, et al. "DDDAS-based Multi-fidelity Simulation for Online Preventive Maintenance Scheduling in Semiconductor Supply Chain." in Proceedings of the 39th conference on Winter simulation, IEEE Press. 2007: 1915-1923.

Krymskaya, M. V., R. G. Hanea, et al. "An Iterative Ensemble Kalman Filter for Reservoir Engineering Applications." Computational Geosciences: 13, no. 2, 2009: 235-244.

Labarthe, O., B. Espinasse, et al. "Toward a Methodological Framework for Agent-based Modelling and Simulation of Supply Chains in a Mass Customization Context." Simulation Modelling Practice and Theory 15, no. 2, 2007: 113-136.

Larsen, E. R., J. D. W. Morecroft, et al. "Complex Behaviour in a Production-distribution Model." European Journal of Operational Research 119, no. 1, 1999: 61-74.

Lau, H. C. W., I. K. Hui, et al. "Monitoring the Supply of Products in a Supply Chain Environment: a Fuzzy Neural Approach." Expert Systems 19, no. 4, 2002: 235-243.

Li, G., P. Ji, et al. "Modeling and Simulation of Supply Network Evolution Based on Complex Adaptive System and Fitness Landscape." Computers & Industrial Engineering 56, no. 3, 2009: 839-853.

Li, S. and B. Lin. "Accessing Information Sharing and Information Quality in Supply Chain Management." Decision Support Systems 42, no. 3, 2006: 1641-1656.

Li, Z., A. Kumar, et al. "Supply Chain Modelling-a Co-ordination Approach." Integrated Manufacturing Systems 13, no. 8, 2002: 551-561.

Liang, W. Y. and C. C. Huang. "Agent-based Demand Forecast in Multi-echelon Supply Chain", Decision support systems. 42, no. 1, 2006: 390-407.

Liu, R., A. Kumar, et al. "A Formal Modeling Approach for Supply Chain Event Management." Decision Support Systems 43, no. 3, 2007: 761-778.

Low, M. Y. H., K. W. Lye, et al. "An agent-based approach for managing symbiotic simulation of semiconductor assembly and test operation." in Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, 85 -92. The Netherlands, International Conference on Autonomous Agents, 2005.

Low, M. Y. H., S. J. Turner, et al. "Symbiotic Simulation for Business Process Re-engineering in High-tech Manufacturing and Service Networks." in Proceedings of the 39th conference on Winter simulation, 568-576. IEEE Press, 2007.

Lu, C. and X. W. Li "Supply Chain Modeling using Fuzzy Sets and Possibility Theory in an Uncertain Environment." in The Sixth World Congress on Intelligent Control and Automation, 3608 - 3612. Dalian, China: WCICA 2006.

Makridakis, S., S. C. Wheelwright, et al. Forecasting Methods and Applications, Wiley India Pvt. Ltd., 2008.

Makui, A. and A. Madadi. "The Bullwhip Effect and Lyapunov Exponent." Applied Mathematics and Computation 189, no. 1, 2007: 35-40.

Mandel, J., L. S. Bennethum, et al. "A Wildland Fire Model with Data Assimilation." Mathematics and Computers in Simulation 79, no. 3, 2008: 584-606.

Mason-Jones, R. and D. R. Towill. "Information Enrichment: Designing the Supply Chain for Competitive Advantage." Supply Chain Management: An International Journal 2, no. 4, 1997: 137-48.

Mitchell, B. and L. Yilmaz. "Symbiotic Sdaptive Multisimulation: An Autonomic Simulation Framework for Real-time Decision Support under Uncertainty." In ACM Transactions on Modeling and Computer Simulation 19, no. 1, 2008: 1-31.

Mustafee, N., S. J. E. Taylor, et al. "Distributed simulation with COTS simulation packages: a case study in health care supply chain simulation." in Proceedings of the 38th conference on Winter simulation, 1136 – 1142. Winter Simulation Conference, 2006.

Neale, J. J. and S. P. Willems. "Managing Inventory in Supply Chains with Nonstationary Demand." Interfaces 39, no. 5, 2009: 388-399.

Ouyang, Y., J. E. Zhang, et al. "Dynamic Data Driven Application System: Recent Development and Future Perspective." Ecological Modelling 204, no. 1-2, 2007: 1-8.

Parmar D., W. T., Fowler J., Callarman T., and Hargraden V. "An Integrated Framework for Responsive Supply Network.": The 16th International Conference on Flexible Automation and Intelligent Manufacturing. Limerick, Ireland, FAIM, 2006

Pathak, S., M. McDonald, et al. "A Framework for Designing Policies for Networked Systems with Uncertainty." Decision Support Systems 49, no. 2, 2010: 121-131

Pathak, S. D., D. M. Dilts, et al. "On the Evolutionary Dynamics of Supply Network Topologies." Engineering Management, IEEE Transactions on 54, no. 4, 2007: 662-672.

Paul, H., M. L. Young, et al. "Utilizing Simulation to Evaluate Business Decisions in Sense-and-Respond Systems." in Proceedings of the 36th conference on Winter simulation, 1205 – 1212. Washington, D.C., Winter Simulation Conference, 2004.

Perea-Lopez, E., B. E. Ydstie, et al. "A Model Predictive Control Strategy for Supply Chain Optimization." Computers & Chemical Engineering 27, no. 8-9, 2003: 1201-1218.

Qi, X., J. F. Bard, et al. "Supply Chain Coordination with Demand Disruptions." Omega 32, no. 4, 2004: 301-312.

Reichhart, A. and M. Holweg. "Creating the Customer-Responsive Supply Chain: a Reconciliation of Concepts." International Journal of Operations & Production Management 27, no. 11, 2007: 1144-1172.

Riddalls, C. E., Bennett, S., Tipi, N. S. "Modelling the Dynamics of Supply Chains." International Journal of Systems Science 31, no. 8, 2000: 969-976.

Sarimvels, H., P. Patrinos, et al. "Dynamic Modeling and Control of Supply Chain Systems: A Review." Computers & Operations Research 35, no. 11, 2008: 3530.

Schwartz, J. "Next Generation Supply Chain Management: Control, Optimization, and System Identification." United States - Arizona, Arizona State University. Ph.D.,2008.

Son, J. Y. and C. Sheu. "The Impact of Replenishment Policy Deviations in a Decentralized Supply Chain." International Journal of Production

Economics 113, no. 2, 2008: 785-804.

Steckel, J. H., S. Gupta, et al. "Supply Chain Decision Making: Will Shorter Cycle Times and Shared Point-of-Sale Information Necessarily Help?" Management Science 50, no. 4, 2004: 458-464.

Sterman, J. D. "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." Management Science 35, 1989: 321-339.

Swaminathan, J. M., S. F. Smith, et al. "Modeling Supply Chain Dynamics: A Multiagent Approach." Decision Sciences 29, no. 3, 1998: 607-632.

Tang, C. S. "Perspectives in Supply Chain Risk Management." International Journal of Production Economics 103, no. 2, 2006: 451-488.

Tannock, J., B. Cao, et al. "Data-driven Simulation of the Supply Chain-Insights from the Aerospace Sector." International Journal of Production Economics 110, no.1-2, 2007: 70-84.

Taylor, D. and D. Brunt. Manufacturing operations and supply chain management: the lean approach, Cengage Learning, 2001.

Taylor, S. J. E., L. Behli, et al. “Investigating Distributed Simulation at the Ford Motor Company.” The 9th IEEE International Symposium on Distributed Simulation and Real-Time Applications, 2005: 139-147.

Terzi, S. and S. Cavalieri. "Simulation in the Supply Chain Context: a Survey." Computers in Industry 53, no. 1, 2004: 3-16.

Treharne, J. T. and C. R. Sox. "Adaptive Inventory Control for Nonstationary Demand and Partial Information." Management Science 48, no. 5, 2002: 607-624.

Unver, A. “Observation Based PDE Models for Stochastic Production Systems.” United States -- Arizona, Arizona State University. Ph.D.,2008.

Waltz, E. and J. Linas. Multisensor data fusion. Boston, Artech House, 1990.

Wan, E. A. and R. Van Der Merwe. “The Unscented Kalman Filter for Nonlinear Estimation.” Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000: 153 - 158

Webster, J. "Networks of Collaboration or Conflict? Electronic Data Interchange and Power in the Supply Chain." The Journal of Strategic Information Systems 4, no.1, 1995: 31-42.

Welch, G. and G. Bishop. “An Introduction to the Kalman Filter.” 2001. http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf (accessed October 11, 2010)

Wilding, R. "The Supply Chain Complexity Triangle: Uncertainty Generation in the Supply Chain." International Journal of Physical Distribution and Logistics Management 28, no. 8, 1998: 599-616.

Willner, D., C. B. Chang, et al. “Kalman Filter Algorithms for a Multi-Sensor System.” Decision and Control including the 15th Symposium on Adaptive Processes, 1976 IEEE Conference on, 1976: 570 – 574.

Wu, L. S. Y. "Business Planning under Uncertainty: Quantifying Variability." The Statistician 37, no. 2, 1988: 141-151.

Wu, L. S. Y., J. R. M. Hosking, et al. "Business Planning under Uncertainty." International Journal of Forecasting 8, 1992: 545-557.

Wu, T., J. Blackhurst, et al. "A Model for Inbound Supply Risk Analysis." Computers in Industry 57, no. 4, 2006: 350-365.

Wu, T., J. Blackhurst, et al"Methodology for Supply Chain Disruption Analysis." International Journal of Production Research 45, no. 7, 2007: 1665-1682.

Wu, T. and P. O'Grady. "An Extended Kalman Filter for Collaborative Supply Chains." International Journal of Production Research 42, no. 12, 2004: 2457-2475.

Wu, T. and J. Blackhurst. "A Modeling Methodology for Supply Chain Synthesis and Disruption Analysis." International Journal of Knowledge-based and Intelligent Engineering Systems 9, no. 2, 2005: 93-105.

Xiao, T., G. Yu, et al. "Coordination of a Supply Chain with One-manufacturer and Two-retailers under Demand Promotion and Disruption Management decisions." Annals of Operations Research 135, no. 1, 2005: 87-109.

Zeng, F., S. J. Turner, et al. “Symbiotic Simulation Control in Supply Chain of Lubricant Additive Industry.” Distributed Simulation and Real Time Applications, 2009. DS-RT '09. 13th IEEE/ACM International Symposium on: 165-172.

Zhang, X. "The Impact of Forecasting Methods on the Bullwhip Effect." International Journal of Production Economics 88, no.1, 2004: 15-27.

Zhang, X. "Delayed Demand Information and Dampened Bullwhip Effect." Operations Research Letters 33, no. 3, 2005: 289-294.

APPENDIX A

DEMAND FORECAST WITH/WITHOUT INFORMATION SHARING IN AN N-ECHLON SUPPLY CHAIN WITH ARIMA (0,1,1) END-CONSUMER DEMAND

This appendix provides the derivations of Minimum Mean Square Error (MMSE) forecast of demands for the entities in a serial supply chain, when the end-consumer demand process is ARIMA(0,1,1). The detailed description of the supply chain model is in Section 4.4.

From the node that is closest to end-consumer (downstream) to the node that is farthest to end-consumer (upstream), the entities in the serial supply chain are numbered from 1 to N. For each node i, we assume the following notations.

L^i : production/transportation lead time from node i+1 to i, a nonnegative integer

l^i : information lead time from node i to i+1, a nonnegative integer

$\mathcal{L}^i$: replenishment lead time at node i. $\mathcal{L}^i = l^i + L^i$

d_k^i : demand of node i in period k. It equals to the orders placed by node i-1 in period $k - l^{i-1}$

D_k^i : lead time demand of the next $\mathcal{L}^i$ periods from period k at node i.

$D_k^i = d_{k+1}^i + d_{k+1}^i + \cdots d_{k+\mathcal{L}^i}^i$

OR_k^i : orders placed to supplier at node i in period k

Proposition A1: For a node, if 1) the demand from its customer is an ARIMA(0,1,1) process (Equations (A.1) – (A.2)); and 2) the node uses the Single-Exponential Smoothing (SES) method (Equations (A.3) – (A.4)) in the adaptive order-up-to policy, the demand at its supplier is an ARIMA(0,1,1) process with parameters $\alpha^{i+1} = \frac{\alpha^i}{1+L^i\alpha^i}$, $\mu^{i+1} = \mu^i$ and the random noise $\varepsilon_k^{i+1} = \big(1 +$ *Liaiεk−lii* .

$$d_1^i = \mu^i + \varepsilon_1^i \tag{A.1}$$

$$d_k^i = d_{k-1}^i - \left(1 - \alpha^i\right)\varepsilon_{k-1}^i + \varepsilon_k^i \tag{A.2}$$

Where $\varepsilon_k^i \sim i.i.d.\ N\,(0, (\sigma^i)^2)$

$$F_1^i = \mu^i \tag{A.3}$$

$$F_{k+1}^i = \alpha^i * d_k^i + (1 - \alpha^i) * F_k^i \tag{A.4}$$

Proof: From (A.1) and (A.2), it is easy to derive

$$d_k^i = \varepsilon_k^i + \alpha^i \sum_{t=1}^{t=k-1} \varepsilon_t^i + \mu^i \tag{A.5}$$

From (A.3) and (A.4), it is easy to derive

$$F_{k+1}^i = \alpha^i \sum_{t=1}^{t=k} \varepsilon_t^i + \mu^i \tag{A.6}$$

Substituting k with k+1 in (A.5) and subtracting it from (A.6), we get

$$F_{k+1}^i - d_{k+1}^i = \varepsilon_{k+1}^i \tag{A.7}$$

$$V\left(F_{k+1}^i - d_{k+1}^i\right) = V\left(\varepsilon_{k+1}^i\right) = (\sigma^i)^2 \tag{A.8}$$

Equation (A.8) implies that the forecast error variance of F_{k+1}^i is time-invariant, that is, $V(D_k|k) = V(D_{k-1}|k-1)$. Consequently, the order placed by the adaptive order-up-to policy (4.24) is

$$\begin{aligned} OR_k^i &= \left(E(D_k|k) - E(D_{k-1}|k-1)\right) + \ d_k \\ &= \mathcal{L}^i\left(F_{k+1}^i - F_k^i\right) + \ d_k^i \end{aligned} \tag{A.9}$$

From (A.5), (A.6) and (A.9), it is easy to derive (A.10).

$$\begin{aligned} OR_k^i &= \ \mathcal{L}^i\left(F_{k+1}^i - F_k^i\right) + \ d_k^i \\ &= (1 + \ \mathcal{L}^i\alpha^i)\varepsilon_k^i + \alpha^i \sum_{t=1}^{t=k-1} \varepsilon_t^i + \mu^i \end{aligned} \tag{A.10}$$

Compared with the general expression of the ARIMA(0,1,1) model in (A.5), the order sent from node i (A.10) is actually an ARIMA(0,1,1) process. Since the information lead time between node i and node i+1 is assumed to be l^i.

$$d_k^i = OR_{k-l^i}^i = \varepsilon_{k-l^i}^{i+1} + \alpha^{i+1} \sum_{t=1}^{t=k-1} \varepsilon_{t-l^i}^{i+1} + \mu^{i+1} \tag{A.11}$$

Where

$$\alpha^{i+1} = \frac{\alpha^i}{1+\mathcal{L}^i\alpha^i} \tag{A.12}$$

$$\varepsilon_k^{i+1} = \left(1 + L^i\alpha^i\right)\varepsilon_{k-l^i}^i \tag{A.13}$$

$$\mu^{i+1} = \mu^i \tag{A.14}$$

■

Proposition A2: In an ARIMA (0,1,1) demand model, the history of noise $\mathcal{E}_k^i = \{\varepsilon_k^i, \varepsilon_{k-1}^i, \dots, \varepsilon_1^i\}$ and the history of demand $\Delta_k^i = \{d_k^i, d_{k-1}^i, \dots, d_1^i\}$ contain the same information, therefore, they can be derived from each other.

Proof: Equations (A.1) – (A.2) show that $\{d_k^i\}$ can be derived from $\{\varepsilon_k^i\}$. By rewriting (A.1) and (A.2) to (A.15) and (A.16), it is easy to see that $\{\varepsilon_k^i\}$ can be derived from $\{d_k^i\}$ as well.

$$\varepsilon_1^i = d_1^i - \mu^i \tag{A.15}$$

$$\varepsilon_k^i = d_k^i - d_{k-1}^i + \left(1 - \alpha^i\right)\varepsilon_{k-1}^i \tag{A.16}$$

■

Proposition A3: Assume the end-consumer demand of the supply chain is ARIMA(0,1,1) and all the downstream nodes of node i (node 1 to i-1) use the SES forecast and adaptive order-up-to policy. For the i^{th} node in a serial Supply Chain, given the local demand history, the MMSE forecast of the lead time demand (D_k^i) is $\mathcal{L}^i F_{k+1}^i$, where $\mathcal{L}^i$ is the replenishment lead time of node i and F_{k+1}^i is the SES forecast.

Proof: According to Proposition A2, the expectation of the lead time demand given local information Δ_k^i is

$$\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}\middle|\Delta_k^i\right)=\mathrm{E}\left(\mathrm{d}_{\mathrm{k}+1}^{\mathrm{i}}\middle|\Delta_k^i\right)+\mathrm{E}\left(\mathrm{d}_{\mathrm{k}+2}^{\mathrm{i}}\middle|\Delta_k^i\right)+\ \ldots+\mathrm{E}\left(\mathrm{d}_{\mathrm{k}+\mathcal{L}^i}^{\mathrm{i}}\middle|\Delta_k^i\right)$$

$$=\ \mathrm{E}\left(\mathrm{d}_{\mathrm{k}+1}^{\mathrm{i}}\middle|\mathcal{E}_{\mathrm{k}}^{\mathrm{i}}\right)+\mathrm{E}\left(\mathrm{d}_{\mathrm{k}+2}^{\mathrm{i}}\middle|\mathcal{E}_{\mathrm{k}}^{\mathrm{i}}\right)+\ldots+\mathrm{E}\left(\mathrm{d}_{\mathrm{k}+\mathcal{L}^i}^{\mathrm{i}}\middle|\mathcal{E}_{\mathrm{k}}^{\mathrm{i}}\right)\qquad(A.17)$$

Since the end-consumer demand of the supply chain is ARIMA(0,1,1) and all its downstream nodes use the SES forecast and adaptive order-up-to policy, according to Proposition A1, the demand at node 1 to i all follow the ARIMA(0,1,1) model. Mathematically, the demand at node i can be represented as

$$d_k^i=\varepsilon_k^i+\alpha^i\sum_{t=1}^{t=k-1}\varepsilon_t^{\mathrm{i}}+\mu^i\qquad(A.18)$$

where μ^i is the initial process average for the demand at node i, α^i is the moving average coefficient ($0<\alpha^i\leq 1$) and the noise series is $\varepsilon_k^i\sim i.i.d\ N\left(0,(\sigma^i)^2\right)$

In period k, given the information $\mathcal{E}_{\mathrm{k}}^{\mathrm{i}}=\{\varepsilon_k^{\mathrm{i}},\varepsilon_{k-1}^{\mathrm{i}}\ldots\ \varepsilon_1^{\mathrm{i}}\}$, it is easy to see that the MMSE forecast of future demand d_{k+j}^i is

$$\mathrm{E}(d_{k+j}^i|\mathcal{E}_{\mathrm{k}}^{\mathrm{i}})=\alpha^i\sum_{t=1}^{t=k}\varepsilon_t^{\mathrm{i}}+\mu^i,\ \text{for all j>0}\qquad(A.19)$$

Compared with (A.6), Equation (A.19) actually equals to the SES demand forecast , denoted by $\mathrm{F}_{\mathrm{k}+1}^{\mathrm{i}}$. Hence, we have

$$\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}\middle|\Delta_k^i\right)=\mathcal{L}^i\mathrm{F}_{\mathrm{k}+1}^{\mathrm{i}}\qquad(A.20)$$

The variance of forecast error is derived as

$$\mathrm{V}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}|\Delta_k^i\right)=V\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}-\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}\middle|\Delta_k^i\right)\right)$$

$$=\mathrm{V}\left(\sum_{\mathrm{t}=1}^{\mathrm{t}=\mathcal{L}^{\mathrm{i}}}[d_{k+t}^i-E\left(d_{k+t}^i\middle|\mathcal{E}_{\mathrm{k}}^{\mathrm{i}}\right)]\right)$$

$$= \mathrm{V}\left(\sum_{\mathrm{t}=1}^{\mathrm{t}=\mathcal{L}^{\mathrm{i}}} \left[\alpha^{\mathrm{i}}\left(\mathcal{L}^{\mathrm{i}} - \mathrm{t}\right) + 1\right]\varepsilon_{\mathrm{k+t}}^{\mathrm{i}}\right)$$

$$= \mathrm{V}\left(\varepsilon_{\mathrm{k}}^{\mathrm{i}}\right)\mathcal{L}^{\mathrm{i}}\left\{1 + \alpha^{\mathrm{i}}\left(\mathcal{L}^{\mathrm{i}} - 1\right) + \frac{\left(\alpha^{\mathrm{i}}\right)^{2}\left(\mathcal{L}^{\mathrm{i}}-1\right)\left(2\mathcal{L}^{\mathrm{i}}-1\right)}{6}\right\} \quad \text{(A.21)}$$

■

Proposition A4: Assume the end-consumer demand of the supply chain is ARIMA(0,1,1) and all the downstream nodes of node i (node 1 to i-1) use the SES forecast and adaptive order-up-to policy. For the i^{th} node in a serial Supply Chain, the MMSE forecast of the lead time demand ($\mathrm{D}_{\mathrm{k}}^{\mathrm{i}}$) is

$$\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}} | \Delta_k^i, \Delta_k^j\right) = \mathcal{L}^{\mathrm{i}}\mathrm{F}_{\mathrm{k+1}}^{\mathrm{i}} + \sum_{\mathrm{t}=1}^{\mathrm{t}=\min\left(l^{j(i)}, \mathcal{L}^{i}\right)}\left[\alpha^{\mathrm{i}}\left(\mathcal{L}^{\mathrm{i}} - \mathrm{t}\right) + 1\right]\varepsilon_{\mathrm{k+t}}^{\mathrm{i}} \quad \text{(A.22)}$$

$$\mathrm{V}\left(\mathrm{D}_{k}^{\mathrm{i}} | \Delta_k^i, \Delta_k^j\right) = \mathrm{V}\left(\varepsilon_{\mathrm{k+i}}^{\mathrm{i}}\right)\mathcal{L}^{i*}\left\{1 + \alpha^{\mathrm{i}}\left(\mathcal{L}^{i*} - 1\right) + \frac{\left(\alpha^{\mathrm{i}}\right)^{2}\left(\mathcal{L}^{i*}-1\right)\left(2\mathcal{L}^{i*}-1\right)}{6}\right\}$$

where $\mathcal{L}^{i*} = max(\mathcal{L}^{\mathrm{i}} - l^{j(i)}, 0)$ (A.23)

Proof: Recursively applying (A.13), we get

$$\varepsilon_k^i = \left(1 + \mathcal{L}^{i-1}\alpha^{i-1}\right)\varepsilon_{k-l^{i-1}}^{i-1}$$

$$= \left(1 + \mathcal{L}^{i-1}\alpha^{i-1}\right)\left(1 + \mathcal{L}^{i-2}\alpha^{i-2}\right)\varepsilon_{k-l^{i-1}-l^{i-2}}^{i-2}$$

$$= \cdots = \prod_{t=1}^{t=i-1}(1 + \mathcal{L}^{t}\alpha^{t})\,\varepsilon_{k-l^{j(i)}}^{1} \quad \text{(A.24)}$$

where $l^{j(i)} = l^{i-1} + l^{i-2} + .. l^{j}$

Since $\left(1 + \mathcal{L}^{i}\alpha^{i}\right) = \frac{\alpha^{i}}{\alpha^{i+1}}$, $\prod_{t=1}^{t=i-1}(1 + \mathcal{L}^{t}\alpha^{t}) = \frac{\alpha^{j}}{\alpha^{i}}$ (A.25)

Substitute (A.25) in (A.24), we get

$$\varepsilon_k^i = \frac{\alpha^j}{\alpha^i}\,\varepsilon_{k-l^{j(i)}}^{j} \quad \text{(A.26)}$$

Because $\mathrm{d}_{\mathrm{k}}^{\mathrm{i}} = \varepsilon_{\mathrm{k}}^{\mathrm{i}} + \alpha^{\mathrm{i}}\sum_{\mathrm{t}=1}^{\mathrm{t}=\mathrm{k}-1}\varepsilon_{\mathrm{t}}^{\mathrm{i}} + \mu^{\mathrm{i}}$, $\mathrm{F}_{\mathrm{k+1}}^{\mathrm{i}} = \alpha^{\mathrm{i}}\sum_{\mathrm{t}=1}^{\mathrm{t}=\mathrm{k}}\varepsilon_{\mathrm{t}}^{\mathrm{i}} + \mu^{\mathrm{i}}$, there are

$$\mathrm{d}_{\mathrm{k+1}}^{\mathrm{i}} = \mathrm{F}_{\mathrm{k+1}}^{\mathrm{i}} + \varepsilon_{\mathrm{k+1}}^{\mathrm{i}}$$

$$d^{i}_{k+2} = d^{i}_{k+1} + (\alpha^{i} - 1)\varepsilon^{i}_{k+1} + \varepsilon^{i}_{k+2} = F^{i}_{k+1} + \alpha^{i}\varepsilon^{i}_{k+1} + \varepsilon^{i}_{k+2}$$

$$d^{i}_{k+3} = d^{i}_{k+2} + (\alpha^{i} - 1)\varepsilon^{i}_{k+2} + \varepsilon^{i}_{k+3} = F^{i}_{k+1} + \alpha^{i}(\varepsilon^{i}_{k+1} + \varepsilon^{i}_{k+2}) + \varepsilon^{i}_{k+3}$$

…

Finally, we sum the daily demands, d^{i}_{k+j}, to get the demand over the lead time, D^{i}_{k}, as

$$D^{i}_{k} = d^{i}_{k+1} + d^{i}_{k+2} + .. + d^{i}_{k+\mathcal{L}^{i}}$$

$$= \mathcal{L}^{i} F^{i}_{k+1} + \sum_{t=1}^{t=\mathcal{L}^{i}} [\alpha^{i}(\mathcal{L}^{i} - t) + 1]\varepsilon^{i}_{k+t}$$

From (A.26), there is

$$D^{i}_{k} = \mathcal{L}^{i} F^{i}_{k+1} + \sum_{t=1}^{t=\mathcal{L}^{i}} [\alpha^{i}(\mathcal{L}^{i} - t) + 1] \frac{\alpha^{j}}{\alpha^{i}} \varepsilon^{j}_{k+t-l^{j(i)}} \quad \text{(A.27)}$$

When there is information shared between node i and the downstream node j, node i observes the demand at j. That means, at period k, the random noise $\varepsilon^{j}_{k+t-l^{j(i)}}, \ldots \varepsilon^{j}_{k}$ are known.

Case I: When $l^{j(i)} < \mathcal{L}^{i}$, the unknown part in (A.27) is

$$\sum_{t=1+l^{j(i)}}^{t=L^{i}} [\alpha^{i}(\mathcal{L}^{i} - t) + 1]\varepsilon^{i}_{k+t} .$$

Therefore,

$$E(D^{i}_{k} | \Delta^{i}_{k}, \Delta^{j}_{k}) = \mathcal{L}^{i} = \mathcal{L}^{i} F^{i}_{k+1} + \sum_{t=1}^{t=l^{j(i)}} [\alpha^{i}(\mathcal{L}^{i} - t) + 1]\varepsilon^{i}_{k+t}$$

$$V(D^{i}_{k} | \Delta^{i}_{k}, \Delta^{j}_{k}) = V\left(\sum_{t=1+l^{j(i)}}^{t=\mathcal{L}^{i}} [\alpha^{i}(\mathcal{L}^{i} - t) + 1]\varepsilon^{i}_{k+t}\right)$$

$$= V(\varepsilon^{i}_{k+i})\mathcal{L}^{i*}\left\{1 + \alpha^{i}(\mathcal{L}^{i*} - 1) + \frac{(\alpha^{i})^{2}(\mathcal{L}^{i*}-1)(2\mathcal{L}^{i*}-1)}{6}\right\}$$

where $\mathcal{L}^{i*} = \mathcal{L}^{i} - l^{j(i)}$

Case II: When $l^{j(i)} \geq \mathcal{L}^i$,

$$\mathrm{E}\left(\mathrm{D}_{\mathrm{k}}^{\mathrm{i}} | \Delta_k^i, \Delta_k^j\right) = \mathcal{L}^{\mathrm{i}} \mathrm{F}_{\mathrm{k}+1}^{\mathrm{i}} + \sum_{\mathrm{t}=1}^{\mathrm{t}=\mathcal{L}^i} \left[\alpha^{\mathrm{i}}\left(\mathcal{L}^{\mathrm{i}} - \mathrm{t}\right) + 1\right] \varepsilon_{\mathrm{k}+\mathrm{t}}^{\mathrm{i}}$$

$$\mathrm{V}\left(\mathrm{D}_k^{\mathrm{i}} | \Delta_k^i, \Delta_k^j\right) = 0$$

Combining Case I and Case II, we get Equation (A.22) and (A.23).■