Spring 2012

Determining Kanban Size Using Mathematical Programming and Discrete Event Simulation for a Manufacturing System with Large Production Variability

Abigail Michele Gaston
Embry-Riddle Aeronautical University - Daytona Beach

Determining Kanban Size Using Mathematical Programming and Discrete Event Simulation for a Manufacturing System with Large Production Variability

by

Abigail Michele Gaston

B.S. Embry-Riddle Aeronautical University, 2010

A Graduate Thesis Submitted to the

Department of Human Factors and Systems

in Partial Fulfillment of the Requirement for the Degree of

Master of Science in Human Factors and Systems

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Spring 2012

Determining Kanban Size Using Mathematical Programming and Discrete Event Simulation

for a Manufacturing System with Large Production Variability

by

Abigail Michele Gaston

This thesis was prepared under the direction of the candidate's thesis committee chair, Dahai Liu, Ph.D., Department of Human Factors and Systems, and has been approved by the members of the thesis committee. It was submitted to the Department of Human Factors and Systems and has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors and Systems.

THESIS COMMITTEE:

Dahai Liu, Ph.D., Chair

Jason Kring, Ph.D., Member

Massoud Bazargan, Ph.D., Member

MS HFS Program Coordinator

Department Chair, Department of Human Factors and Systems

5/8/12

Associate Vice President for Academics

Acknowledgements

This author would like to express her sincere gratitude and respect to her thesis chair, Dr. Liu. Without his assistance, guidance, and knowledge, this study would not have been possible. Throughout the past three years, Dr. Liu has been an outstanding instructor and mentor, demonstrating great enthusiasm and knowledge toward the courses he taught and also, providing valuable life lessons along the way.

The author would like to express her appreciativeness to her other committee members—Dr. Kring and Dr. Bazargan**.** Dr. Kring motivated the author to pursue a career in Human Factors and continue her endeavor in persuing higher education at ERAU. He has served as an outstanding mentor the past four and a half years. Without the collective knowledge, expertise, and input of Dr. Kring and Dr. Bazargan, the study would not have been possible.

The author would like to express her thanks to the employees of SMA for allowing her to utilize their production metrics and for always providing a quick response to her many questions.

Additionally, the author would like to acknowledge the entire Human Factors and Systems department—faculty, staff, and students. They have made the past five years possible and valuable by providing their experience, knowledge, and unique personalities.

Finally, this author would like to express a special acknowledgement to her parents, Mike and Kathy Gaston. Aside from providing the author the opportunity to pursue higher education, they have instilled such a work ethic that reminds her to always do her best, while enjoying and learning every step of the way.

Table of Contents

List of Figures

List of Tables

Abstract

Author: Abigail Michele Gaston

Title: Determining Kanban Size Using Mathematical Programming and Discrete Event Simulation for a Manufacturing System with Large Production Variability

Institution: Embry-Riddle Aeronautical University

Year: 2012

In order to become more competitive and aggressive in the market place it is imperative for manufacturers to reduce cycle time, limit work-in-process, and improve productivity, responsiveness, capacities, and quality. One manner in which supply chains can be improved is via the use of kanbans in a pull production system. Kanbans refer to a card or signal for productions scheduling within just-in-time (JIT) production systems to signal where and what to produce, when to produce it, and how much. A Kanban based JIT production system has been shown to be beneficial to supply chains for they reduce work-in-process, provide real time status of the system, and enhance communication both up and down stream.

While many studies exist in regards to determining optimal number of kanbans, types of kanban systems, and other factors related to kanban system performance, no comprehensive model has been developed to determine kanban size in a manufacturing system with variable workforce production rate and variable demand pattern. This study used Stewart-Marchman-Act, a Daytona Beach rehabilitation center for those with mental disabilities or recovering from addiction that has several manufacturing processes, as a test bed using mathematical programming and discrete event simulation models to determine

the Kanban size empirically. Results from the validated simulation model indicated that there would be a significant reduction in cycle time with a kanban system; on average, there would be a decrease in cycle time of nine days (almost two weeks). Results were discussed and limitations of the study were presented in the end.

Introduction

Manufacturing Systems

Due to increasing customer demand for manufacturing responsiveness and reduced lead-times, manufacturing environments are reevaluating the design of their processes to meet both customers' quality and delivery expectations and reduce inventory in order to remain competitive. Manufacturers are able to remain aggressive in the market place by increasing and maintaining their responsiveness to customers, increasing facilities' capacities, quality, and productivity, while reducing lead-times and inventory (Mathaisel, 2005). To achieve optimal process flow, with reduced cycle time and waste, many manufacturers are adopting the lean manufacturing architecture model. Lean manufacturing is more than a production technique; it is also a revolutionary way of thinking. The lean manufacturing philosophy is to shorten time between the customer order and delivery by making the product(s) flow through the system without waste or interruption (Liker & Womack, 2004). To realize this philosophy, the lean manufacturing model is focused on the following principles: specify value, identify value stream and eliminate waste, make the value flow, let the customer pull the process, and continuously improve the process (Haque, 2003).

Traditional push manufacturing systems. Traditional manufacturing has a history of utilizing the "Push" method of manufacturing, consisting of a central planning system that starts or "pushes" the initiation of work based on prior forecasts of future demands for orders from the beginning of the manufacturing process to each consecutive stage within the manufacturing system (Ip, Yung, Huang, & Wang, 2002). This method is focused on "batch-and-queue," task-oriented, functionally isolated production, which typically results in excess inventory requirements, parts travel time, and variable process flow (Sharma & Moody, 2001). The push method erroneously schedules what work should be released based on projected demand; nonetheless, the original production schedule is not modified based on the actual conditions of the production. Because push production is dependent on projected demand and forecasts, and said forecasts are hardly ever as accurate as one would like, this reliance degrades the overall system performance. Great risk of a congested line, loss of flexibility to integrate design or engineering changes or specify priority changes, and inaccurate number of customer orders are potential hindrances of push production systems. For example, work can continuously be released in a push system, whether or not the system is congested, only to have to work get stuck somewhere mid-process (Hopp & Spearman, 2001).

Material Requirements Planning. The most popular planning system for push, batch manufacturing environments is Material Requirements Planning (MRP). MRP utilizes the system's bill of materials, inventory records, and company-wide information to calculate the system components and material required to fulfill customer demand (Deleersnyder, Hodgson, Muller-Malek, & O'Grady, 1989; Taylor, 2002; Wong & Kleiner, 2001). Because MRP computes schedules of what should be started, it is a push system (Hopp & Spearman, 2001). MRP makes an effort to drive excess work in process (WIP) levels to zero, by striving to produce and manufacture on an as-needed basis (Taylor, 2002). Furthermore, it enables management to identify the products that were going to be produced (Wong & Kleiner, 2001).

While MRP was originally developed to have promising results for manufacturers to reduce WIP, such results were not delivered. Because managing and adhering to a master schedule is very challenging due to inaccurate or invalid data, manufacturing results are often less than satisfactory. Inventory levels, lead times, and WIP are overestimated, the overall feasibility of the master schedule is not verified, and the system's capacity is ignorant of actual volume (Deleersnyder et al., 1989; Taylor, 2002; Wong & Kleiner, 2001). Also, production often varies and engineering changes are made throughout production, that are not originally accounted for in MRP, affecting production rate and end item quantities and dates (Hastings, Marshall, & Willis, 1982). Whiteside and Arbose (1984) reported that critics believe the implementation of MRP has resulted in a $100 billion mistake across manufacturing environments. Due to MRP's inability to handle variable demands, manufacturers started to look at the schedules from a different perspective. A new method of production, called the pull method, became manufacturers' focus.

Pull manufacturing systems. In contrast to the push method that releases work based on an erroneous schedule, the pull method authorizes the release of the work based on a signal generated by the completion of a job in the system (Spearman, Woodruff, & Hopp, 1990). Release of work is triggered by an outside, (typically) inaccurate schedule in a push system, while the release of work in a pull system is triggered by an internal signal based off the systems' current demand. Essentially, work is "pulled" through the system based on the actual system's demand from the end of production. Figure 1 depicts these release triggers in push and pull production systems. The figure displays the different release of work triggers work between the push and pull production systems. While the push manufacturing method work is released based on a schedule's forecasts, orders, arrivals, and/or upstream information, pull manufacturing method work is released based off the system's status within the process and/or other downstream information (Hopp & Spearman, 2001).

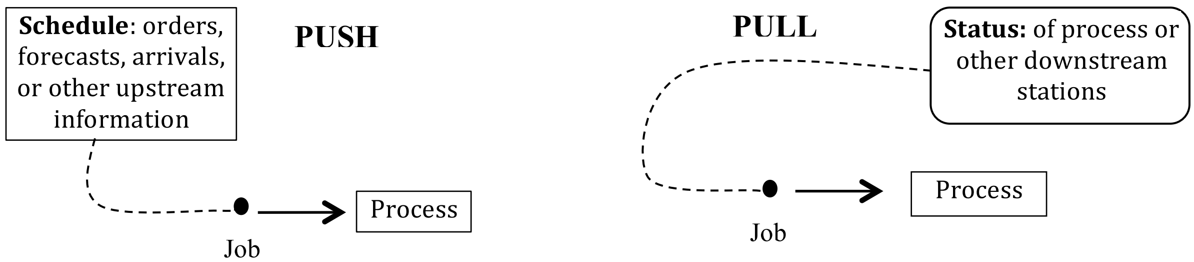

Figure 1. Comparison of release triggers in push and pull production systems. Adapted from *Factory Physics* by Hopp & Spearman, 2001.

The pull method is focused on the concept of just-in-time (JIT) production with the objective to produce what is needed when it is needed (Di Mascolo, Frein, & Dallery, 1996). The pull manufacturing system has shown to greatly enhance quality control, responsibility of personal performance, create an atmosphere of parsimony and frugality and is one of the five principles of lean manufacturing (Hopp & Spearman, 2001; Wood, 1990).

Just-in-Time Production. As a fundamental philosophy of the pull system, JIT production's aim is to have each workstation receive the required materials from preceding workstations precisely as needed, which requires very smoothly operating systems (Hopp & Spearman, 2001; Correia, 2003). The notion of JIT production is straightforward: produce and deliver finished goods just-in-time to be sold, sub-assemblies just-in-time to be assembled into finished goods, fabricated parts just-in-time to enter sub-assemblies, and purchase material just-in-time to become fabricated parts (Presutti, 1988). With this notion, JIT production is heavily dependent on the preceding stage within the production line.

JIT benefits can be attributed to the fact that the line's WIP is controlled. By controlling the system's WIP, the amount of material that needs to be scrapped or reworked is diminished, cutting financial losses. According to Little's Law, the number of items in a system, or WIP, over some time interval, is equal to the arrival rate (λ) times the cycle time (Little, 1961). Symbolically this can be represented as:

$$WIP = CT * \lambda \tag{1}$$

Based on Little's Law, by limiting the WIP there is a reduction in variability in cycle time, while still allowing the system to achieve the same throughput (Marek, Elkins, & Smith, 2001). Little's Law applies to all production line systems, including: single station, a line, and the entire plant (Hopp & Spearman, 2001).

Various techniques have been applied to achieve the philosophy of JIT. For example, inventory is ordered in small quantities, thus removing the buffer stock; preventative maintenance, as opposed to reactive maintenance, occurs on all machines to prevent breakdowns; employees are given greater responsibility to make decisions and correct problems as they occur in each stage; the plant layout is redesigned and excess space reduced for multi-skilled employees to adequately manage a number of machines and processes, to diminish set-up time, and to standardize manufacturing products and processes (Correia, 2003; Ramaswamy, Selladurai, & Gunasekaran, 2002).

The absolute ideals of JIT production in terms of the "seven zeros," which are required to achieve zero inventories, are described in Table 1 below. According to Hall (1983), zero inventories imply a level of perfection that may not be possible to fully realize in the production process. However, the notion of high level excellence is vital because it simulates a mission for continuous improvement through imaginative attention to both the overall task and the minute details.

Table 1

JIT in terms of the seven zeros with description. Adapted from "Factory Physics," by Hopp and Spearmam, (2001).

Seven Zero Ideal	Logic Behind Ideal
Zero defects	Every part should be made correctly the first time to avoid production disturbances
Zero (excess) lot size	Maximum responsiveness is maintained when each workstation is capable of replacing parts one at a time. Goal is to achieve lot size of one
Zero setups	Precondition to achieve lot size of one.
Zero breakdowns	Breakdowns and machine failure will bring production to a halt throughout the line.
Zero handling	No extra moves to and from storage to avoid intermediate pauses.
Zero lead time	Eliminate queue time and processing time per part.
Zero surging	Sudden changes (surges) in quantities or product plan without excess WIP to level changes cause disruptions and delays.

To achieve the seven zeros, which would translate into instantaneous production, is physically impossible; however, the purpose of setting these ideals as goal is to promote an environment of continuous improvement (Hopp & Spearman, 2001).

Kanban Systems. One of the most commonly used methods to implement pull systems and to realize JIT production is achieved via the practice of kanbans (meaning "marker" or "card" in Japanese). The kanban system, which was developed by Toyota production systems, is a multi-stage production system used to manage scheduling and inventory control (Bitran & Chang, 1987; Hopp & Spearman, 2001). While kanbans are utilized to pull work through the system, they are also used to visualize and control in-process inventories (Akturk & Erhun, 1999).

Figure 2 below represents the flow of items and kanbans between a three-stage production system. Process P_i produces items to fill a container, that full container is stored in inventory point I_i with a kanban attached to it, and process P_{i+1} will take the container from inventory point I_i and continue production for process P_{i+1}. When the first piece of a full container in I_{i-1} is used in the production process P_i, the kanban is detached, set aside, and triggers a signal to begin the process in the preceding stage, P_{i-1}. The kanban process typically follows the first-in-first-out (FIFO) rule. So, once the process P_i produces a full container, the kanban that ordered the full container is attached and the container is sent to I_i.

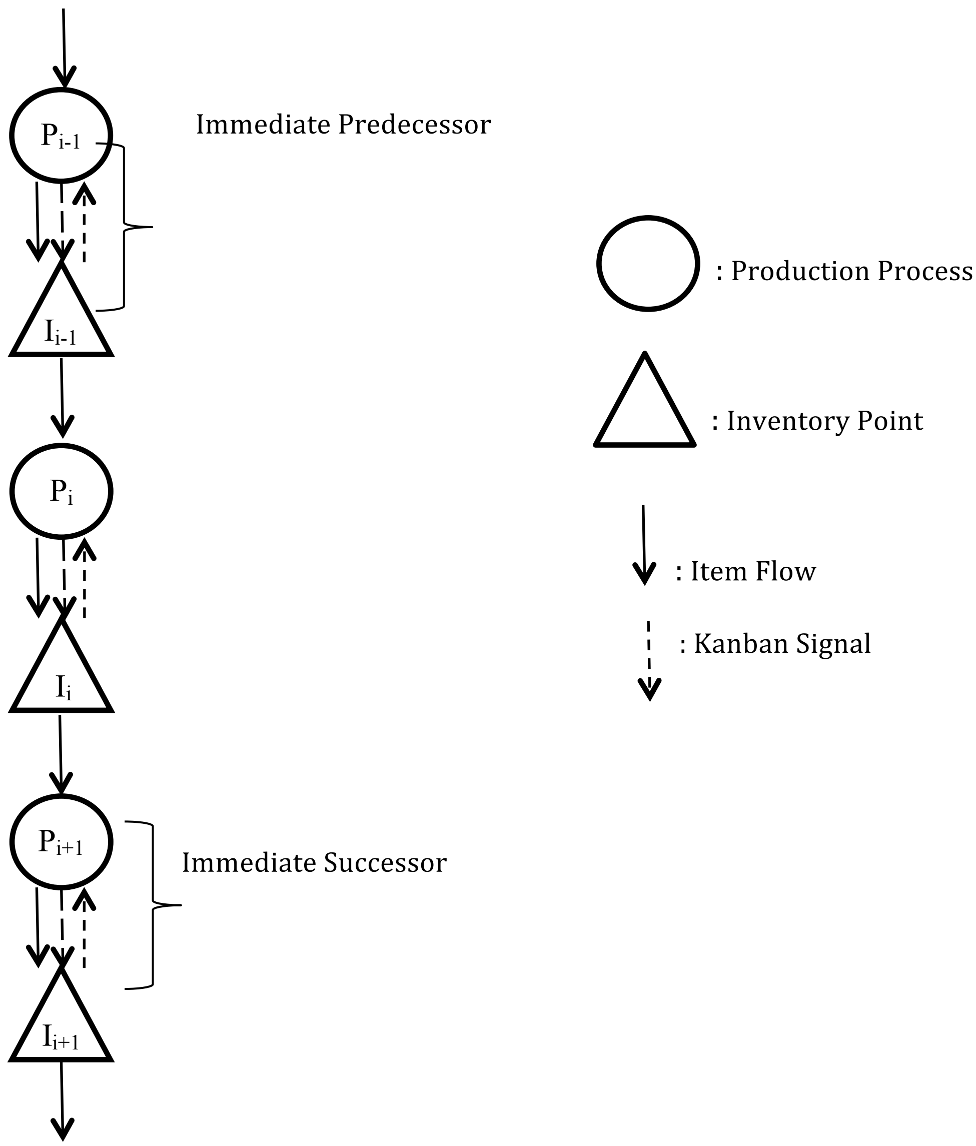

Figure 2. Flow of items and kanbans in a production system. Adapted from "A mathematical programming approach to a deterministic kanban system," by Bitran & Chang, 1987.

In the system seen in Figure 2, there are four important observations to note. First, the total number of kanbans circulating between inventory points and processes is unchanged over time, unless management intervenes. The maximum inventory buildup—WIP—in each inventory point is limited by the number of kanbans circulating between the inventory points and processes. By controlling the number of circulating kanbans and requiring that every full container has an attached kanban, management can be assured that excess inventory buildup will not exceed a certain level. The movement of kanbans between I_i and P_i is triggered by inventory withdrawn from I_{i-1}. In particular, P_{i-1} will produce what is needed to replenish what has been withdrawn from I_{i-1}. Lastly, by circulating kanbans within every stage, all the stages within a production system are linked together. As a result, the production schedule of the final stage is sent back to all preceding stages, because the detached kanban automatically becomes a new order, management does not need additional triggers to order in a preceding stage; these preceding stages can become self-operated (Bitran & Chang, 1987).

Within any production system, there are stages which consist of certain processes, and together the stages produce an output. In pull systems that utilize a fixed number of kanbans, the cards signal movement of a stage's finished parts that are put into containers and authorize the production of new parts, until an order is complete, which stops the signal of kanbans (Di Mascolo, Frein, & Dallery, 1996). The production activities are connected in such a manner that the stages are linked like a chain to the preceding stage which materialize JIT production (Sugimori, Kusunoki, Cho, & Uchikawa, 1977). This assembly-like fashion of kanban production creates a manual method of harmoniously controlling production and inventory within the manufacturing plant (Akturk & Erhun, 1999). The ultimate goal kanban systems is the conversion of raw materials into finished products, with lead time equal to processing time (Younies, Barhem, & Hsu, 2007).

Benefits of implementing kanban systems. There are many benefits to the implementation of kanban systems. First, it is a relatively simple means to implement communication across multiple stages of the production process (Di Mascolo et al., 1996). Through kanban implementation, production systems can realize the JIT production philosophy with a reduction of inventory and lot sizes, reduction of setup costs, elimination of queues, effective maintenance programs to eliminate production defects entirely, reduction of lead times, collaboration with vendors in terms of planning needs and delivery times, and minimized employee turnover through harmonious management (Younies, Barhem, & Hsu, 2007). Kanban systems limit the amount of WIP, thus reducing cycle time as stated by Little's Law—WIP is equal to the arrival rate times cycle time, creating a direct relationship between the two variables (Marek et al., 2001). The amount of multi-tasking typically required by traditional manufacturing is greatly reduced, which enables work to get done, quicker, with higher quality, and delivered to the customer when necessary (Anderson & Roock, 2011). Paperwork and overhead to run and control the process and inventory is greatly reduced because the kanban automatically triggers a signal to all preceding stages (Bitran & Chang, 1987). The kanban system is robust in the sense that it flexible enough to absorb and adapt to unexpected situations, which would typically require continuous managerial oversight. For instance, if there is a machine breakdown, kanbans are no longer being sent to predecessors, hence preventing the buildup of inventory between stages (Bitran & Chang, 1987). According to Sugimori et al. (1977) kanbans provide a rapid acquisition of facts regarding the continuously changing status of production capacity, operating rate, and manpower of production, and according to Bitran and Chang (1987), a full container, with an attached kanban shall provide the following facts: item name, item number, description of the item, container type, container capacity, kanban identification number, preceding stage, and succeeding stage. These facts enhance managerial knowledge of the continuously changing system status.

An electronic manufacturer transformed their manufacturing process from push to pull via integration of kanbans because they were experiencing excessive waste of materials, unnecessary cost, and long lead times. After six months the company reduced lead times from 180 hr to 60 hr and WIP was reduced by 70%. Additionally, the company reduced overproduction, inventory, and experienced fewer and more quickly resolved stoppages within the process, all of which enhance process efficiency, keep cost to a minimum, and increase production (Lee-Mortimer, 2008). Ramnath, Elanchezhian, and Kesavan (2009) conducted a case study of multiple manufacturers to identify the results of implementing kanbans and found similar results: inventory levels were reduced and material flow was standardized.

Supply chain systems, which is comprised of a series of manufacturing organizations and companies to provide a service to that supply chain, can also benefit from kanban implementation; however, on a much larger scale. One supply chain was experiencing excess inventory due to poor planning, poor purchasing behavior, poor communication, inadequate quality levels, wastage of materials, and uneconomical use of resources and funds. To increase the production efficiency, effectiveness and competency, reduce the amount of wasted materials, time, and effort involved in production processes, a kanban system was implemented. Because the manufacturing facilities were in different locations, there was a significantly greater material flow compared to that at manufacturing plants, and kanban containers were considered as an automated guide vehicle (AGV), car, truck, ship, or train. Using a fixed number of kanbans to transfer materials to subsequent organization in the supply chain and to demand information flows to the preceding organization, the supply chain was able to minimize the total cost of the supply chain system, inventory, wasted labor, and customer service in a supply chain (Wang & Sarker, 2004).

Challenges of implementing kanban systems. While there are many benefits of implementing kanban systems, there are also several challenges. Logistically, the identification of flow lines is one problem area; streamlining the production requires the simultaneous consideration of products (processing requirements and demand pattern) and resources (machines, personnel, and transport) to achieve flow lines that operate around product families with good levels of utilization. To achieve optimal flow lines with minimal extra investment requires special planning and the collaboration of various stakeholders within the process. Problems with flow line loading can occur. Identifying the appropriate amount of work for each stage, also known as the kanban size, for the purpose of avoiding bottlenecks is difficult. Determining the number of kanbans to use in a production system to control the interaction between production and inventory levels also takes extensive planning and coordination (Deleersnyder et al., 1989).

Other challenges that can occur in kanban system implementation deal with getting people to change their mindset on how production systems should run. Systems are very focused on efficiency, which makes people want to contribute work across the statement of work (SOW) to deliver results. Instead kanban systems have individuals focus on the end result and the team, affecting what work the team does first and what work the team puts the most effort into; modifying individuals' behavior from a personal efficiency mindset to a collaborative mindset takes time (Verweij & Maassen, 2011).

Implementing kanban systems in environments with mixed and changing demand, poor quality production, or with a wide variety of products can cause problems. An environment that continuously increases or decreases the parts and processes increases the complexity of the kanban system, which could lead to system breakdowns. System breakdowns and unexpected situations can lead to the process being shut down (Jarupathiru, Cigane, Chotiwankaewmanee, & Kerdpita, 2009).

Types of Kanban Systems. The type of kanban system that is implemented depends of the dynamics and characteristics of the manufacturing environment. While movement of production is authorized by way of kanban cards, these kanban cards can be physical or electronic. Traditionally, kanban cards thought to be a physical, card-stock system that detaches the card to transfer an authorization signal to the preceding stage. Yet, kanbans can also be automated, computer-based systems that rely on a coding and scanning infrastructure. Electronic signals are automatically conveyed across stages via digital kanban display boards, which allow each stage to visualize exactly what they should be working on and the status of surrounding stages (Lee-Mortimer, 2008).

Another variance across types of kanban systems is the count of cards used within each stage. The most commonly used kanban system is the dual-card system (Esparrago, 1988). This type of system uses one card to authorize the movement of one full container of a part at a stage, while a second card authorizes the transfer of one full container of raw material for a part from one stage to the next (Yang, 2000). Also, dual card configurations can control the facility and the transportation out of the facility (Karmarkar & Kekre, 1989). On the other hand, single-card kanban systems combines both push and pull mechanics; parts are made or assembled according to schedule, but replenishments are authorized by signals. In other words, parts are pushed through the system, while work centers pull their supplies (Esparrago, 1988). These cards communicate the need to produce and transfer a container of raw material for a specific part from an upstage stage to the next stage in the process (Yang, 2000).

Kanban systems can also be composed of a single or multiple stages. Single-stage kanbans are characterized by an arrival and departure process. Each batch corresponds to a single kanban. Batches may not begin until the preceding batch has been completed, and a kanban signal has been released (Krishnamurthy & Suri, 2006). CONWIP (CONstant Work in Process), discussed in further detail below, is a specific example of single-stage kanbans. Kanban systems with multiple stages consist of multiple stages in a system that draws upon one another (Karmarkar & Kekre, 1989). The kanban systems discussed thus far have contained multiple stages.

CONWIP. The CONWIP approach is a generalized form of a single-stage kanban system. Like kanban systems, CONWIP relies on signals or cards to authorize the movement of materials through the system based on the system's demand. In contrast to kanban systems, CONWIP systems use a signal set of production cards to pull work at the beginning of the system and traverse the entire production line, whereas kanban systems pull work between every pair of workstations, everywhere in the system, which is depicted in Figure 3 below. CONWIP system cards are assigned to the production line, not a specific part of the line. Material enters the CONWIP system only based on demand, and the raw material receives a card to authorize entrance; the same card used to authorize entrance moves the material through the system and completes production (Spearman et al., 1990; Marek et al., 2001).

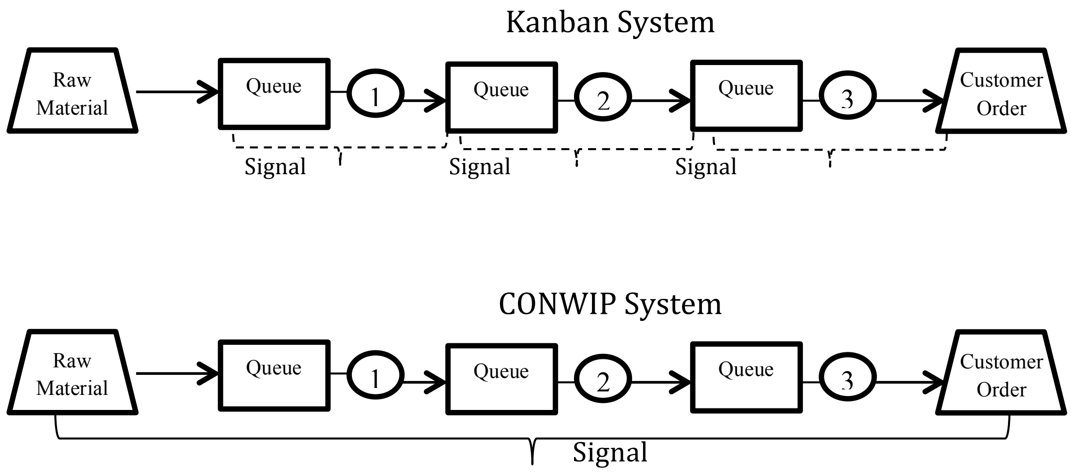

Figure 3. Comparison of kanban and CONWIP system signals. Adapted from "*Factory Physics,"* by Hopp & Spearmann, 2001; "*Understanding the Fundamentals of Kanban and CONWIP Pull Systems using Simulation,"* by Marek et al., 2001.

Previous kanban system studies. Though there are challenges in adopting the kanban system processes, many organizations have greatly benefited from this newer manufacturing culture. While discussing previous kanban studies, it is important to discuss the automobile manufacturer Toyota. Toyota was the first company to conceptualize and integrate kanbans into their production system. Kanbans were used by Toyota to convey signals from one process to the preceding process and to order production of the inventory withdrawn from the subsequent process; the kanbans are always attached to containers with parts. A model of Toyota's flow of parts and kanbans is seen in Figure 4 below. The production is connected in a chain-like manner between processes. Utilizing the chain-like kanban system, Toyota no longer had to rely on a computerized system, for the kanbans reduced the cost of implementing a system to provide production schedule for the processes and suppliers that considered adjustments and alterations because the kanban provided real time control. Like a computerized system, kanbans contained the necessary rapid and precise data for management. Additionally, Toyota saw labor productivity was at its highest among automotive competitors, and workers positively participated in future improvements (Sugimori et al., 1977). For these reasons, other manufacturers look to Toyota as a model for their manufacturing system when implementing kanbans and other lean manufacturing initiatives.

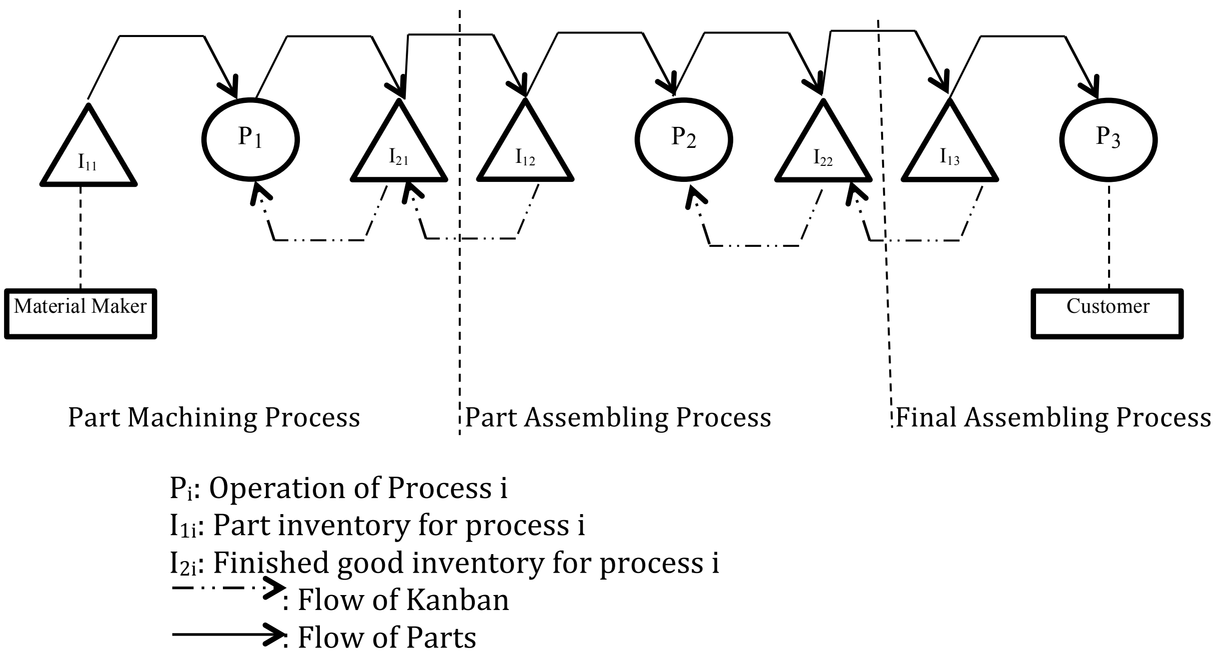

Figure 4. Flow of parts and kanbans in Toyota Production System. Adopted from "Toyota production system and kanban system materialization of just-in-time and respect-for-human system," by Sugimori et al., 1977.

A made-to-order clothing manufacturer, with plants across the world, utilized a push controlled production system for years, which often led to bottlenecks, high work-in-process (WIP), and highly variable lead times—time from product was ordered to delivery. However, upon kanban system implementation, WIP levels dropped greatly, inventory was cut by $1.5 million, and lead times were much more consistent compared to pre-implementation (Billiesbach, 1994).

Mathematical Models

Process designs can be further explained and facilitated through the use of various analogical and mathematical models, which are simplified representations of the real world system relative to assess the likely consequences and outcomes of the various alternative courses of action within the process. Through the use of mathematical models, relationships between various aspects of the system can be uncovered that may have not been apparent before, comparison of multiple solutions can be rapidly and efficiently made to aid in the selection of the best solution, unexplained situations can be explained by indicating cause-and-effect relationships, the type of data to be collected to deal with the system is indicated, future events—such as effectiveness factors, reliability, and maintainability—can be predicted, and risks and uncertainty can be identified. However, it is important to note that the mathematical model shall not be the decision maker but should be utilized as a tool to provide necessary data to support the decision making process (Blanchard & Fabrycky, 2006).

To create and develop a successful mathematical model there are several factors to consider while building it. The model shall properly represent the dynamics of the given system in a way that is simple enough to comprehend and manipulate, yet still adequately yield successful results to the problem at hand. The model itself shall be simplified enough to allow for timely implementation in the system's problem solving. Factors that are most relevant to the problem shall be highlighted, while those factors that are not as important shall be suppressed, with discretion. All relevant factors shall be comprehensive and reliable in terms of results (Blanchard & Fabrycky, 2006).

Mathematical programming. Mathematical programming is one of the most widely used tools in management sciences. The focus is to plan or program the optimized allocation of limited resources among competing activities, under constraints imposed by the nature of the environment. These constraints include: financial, technological, marketing, or organizational considerations. The mathematical models provide guidelines to management in making effective decisions within the state of current information or in seeking additional information if current state is inadequate to reach an appropriate decision (Bradley, Hax, & Magnanti, 1977).

Mathematical programming is the process of formulization and the solution of the constrained optimization problem—where f is a value function of variables $x_1, x_2,\ldots, x_n$ and Ω is the subset of the domain f—of the basic form:

$$\text{Minimize (or maximize) } f(x_{1,\ldots,} x_n) \text{ subject to } (x_{1,\ldots,} x_n) \in \Omega \qquad (2)$$

The function $f(x_{1,\ldots}, x_n)$is known as the objective function, while Ω is often known as the set of feasible solutions and is a subset of the domain of function f defined by equations, called constraints (Jeter, 1986; Snyman, 2005).

Mathematical programming and kanban systems. Because mathematical programming is focused on optimizing processes, while taking system limitations into consideration, there have been many studies on mathematical programming approaches to optimize kanban systems. Previous studies utilize kanban (or container) size, number of kanbans, and safety stock level as the decision variable, while layout, number of time periods, number of items, number of stages, and capacity are system constraint variables (Akturk & Erhun, 1999). These models can be utilized to select the optimal system design, to increase the overall effectiveness of kanban systems, and draw conclusions regarding system dynamics and relationships.

Philipoom, Rees, Taylor, and Huang (1990) developed two integer mathematical models to determine optimal container sizes to signal in multi-item, multi-stage systems to be used in conjunction with kanban signals. The models assumed no system backorders, eliminating stage interdependencies. Minimizing inventory was an objective functions. To develop the model three constraints were identified. First, the cycle time for each maintenance must be greater than or equal to the production time including setups, which is formulated mathematically into the following

$$t_i \geq \sum_{j=1}^{n} (q_{ij}) PT_j + Y_{ij} S, \tag{3}$$

where q_{ij} is the container size in the containers for item *j* processed to machine *i*, t_i is the production cycle time for the *i*th machine at the signal kanban work center, PT_j is the processing time for containers of item *j*, *S* is the setup time, and

$$Y_{ij} = \begin{cases} 1, & if\ q_{ij} > 0 \\ 0, & if\ q_{ij} = 0 \end{cases} \tag{4}$$

The second constraint states each item produced at a work center to be produced on one machine, which is written mathematically as follows

$$\sum_{i=1}^{m} Y_{ij} = 1\text{, for } j = 1 \text{ to } n \text{ items} \tag{5}$$

$$q_{ij} \leq MY_{ij,} \text{ for all } i \text{ and } j \tag{6}$$

The final constraint states that for each machine, demand for each item produced on that machine during the production cycle must equal the container size for that item, which is mathematically expressed as a pair for each item and machine:

$$Q_j \leq d_j t_i + (1 - Y_{ij}) M \tag{7}$$

$$Q_j \geq d_j t_i - \left(1 - Y_{ij}\right)M \tag{8}$$

where

$$Q_j = \sum_{i=1}^{m} q_{ij} \tag{9}$$

And, Q_j is the sum of container sizes for item *j*. Based on these constraints, the following integer mathematical programming model was developed to minimize inventory

$$minimize\ Z = \sum_{j=1}^{n} Q_i \tag{10}$$

A second mathematical model was developed to consider cost minimization, which is as follows

$$\min \sum_{j=1}^{n} \left[C_j \frac{R_j}{Q_j} + C_{H_j} \frac{Q_j}{2} (1 - d_j) PT_j \right] \tag{11}$$

where C_j is the setup cost of item *j*, R_j is the annual demand for item *j*, and C_{Hj} is the is the waiting cost for item *j.* While the results of the models do not provide guidelines on means to implement kanban systems, they provide prudent implications for management to consider: inventory and setup costs, as opposed to simply seeking to reduce the inventory at a minimal level, should be considered, and under certain conditions, a kanban signal system—which triggers the production of larger than normal container sizes with large setup times within a JIT production framework—may be more cost effective than a standard kanban system—which concurrently triggers production of containers.

A similar mathematical model was developed to identify whether kanban systems can operate effectively in unstable manufacturing environments. Moeeni, Sanchez, and Vakharia (1997) proposed a model to implement kanban systems in such environments. Mathematical programming, more specifically the Taguchi function adapted to build quality into the design of products, modeled effects of inherent environmental variations, such as demand, lead times, setup times, processing times, time between breakdowns, and repair times on overall kanban system performance. The design simultaneously considered the number of kanbans, kanban review periods, and container size. The container size was found to be the most vital factor in kanban system performance in an uncertain manufacturing environment.

Karmarkar & Kekre (1989) conducted an analysis to investigate the parametric behavior of kanban systems by using parameters to develop analytical conclusions regarding the behavior. Using a Markovian model, five major results were found. First, container size associated with each kanban card does have a large effect on the kanban system performance. Second, a varying number of cards, with a fixed container size, are analogous to varying the level of base-stock. Next, there is an interaction between the number of cards and the size of containers. Also, in a multistage kanban system, changes in one stage affect the performance of all other stages; for example, increasing the number of kanban cards in one stage increases inventory at subsequent stages, while decreasing them at preceding stages. Lastly, by controlling parameters of a kanban system can optimize the overall behavior of the system.

Limitations of mathematical programming. While mathematical programming is an excellent tool used to optimize processes, there are also several limitations to this method. For example, the mathematical model can greatly increase in complexity when dynamic, stochastic, or nonlinear factors are added (Chandra & Grabis, 2007). Another limitation of mathematical programming is the assumptions that the input data is completely accurate; however, the data are seldom entirely exact. Because the input data are used to verify the accuracy of the model, the model will only be as accurate as the data provided (Chinneck, 2001).

Simulation

Simulation is a general term that refers to the applications and methods that echo the behavior of a real system in order to measure performance, improve operation, or design a system if nonexistent. In situations where direct experimentation is not feasible, simulation is widely used to evaluate the likely outcome of a given decision without changing the operational system itself (Blanchard & Fabrycky, 2006).

Simulation has been around since the 16th century as people originally modeled systems by hand. Today simulation is typically done by computers and software, which over the years has graduated from being very error prone and tedious to being quick, powerful, and flexible. There is a clear improvement in performance/ price ratio of simulation programs than from just a few years prior (Kelton, Sadowski, & Sturrock, 2010). Computer-based simulation is utilized to enhance the understanding of system behavior and logistics when a system is modified and can be fully integrated into complex manufacturing systems and run in real-time (Manuj, Mentzer, & Bowers, 2009; Tavakoli, Mousavi, & Komashie, 2008).

There are different types of simulation models, which can be classified among three dimensions. The first dimension is static versus dynamic. In static models, time does not play a natural role, whereas it plays a role in dynamic; most operational models are dynamic. Another dimension is continuous versus discrete. In continuous models, the system state continuously changes over time, while discrete models change occurs at separate points in time. The third dimension is deterministic versus stochastic.

Deterministic models have no random input, while stochastic models operate with some inputs being random (Kelton et al., 2010).

Discrete event simulation. Discrete event simulation (DES) is a specific type of simulation in which one or more phenomena changes state at discrete events in time, rather than continuously with time. DES offers techniques that can approximate the values of system performance with remarkably small error. These approximations come from data observed on sample paths or sequences generated during simulation that is corresponding to the model of interest. Industries such as engineering, health care, management, military, mathematical, transportation sciences, and manufacturing utilize DES to model and study the behavior of complex systems (Fishman, 2001).

Every discrete event system embodies at least seven characteristics: work, resources, routing, buffers, scheduling, sequencing, and performance. Further explanation of these characteristics is described in Table 2 below.

Table 2

DES Characteristics and Descriptions. Adapted from "Discrete-Event Simulation: Modeling, Programming, and Analysis," by Fishman, 2001.

DES Characteristic	Characteristic Description
Work	Items, jobs, and customers that enter system seeking service
Resources	Equipment, conveyances, and manpower that can provide the service
Routing	Collection of required services, the resources that provide them, and the order in which services are provided
Buffers	Waiting rooms that hold work awaiting service; they may have infinite or finite capacity
Scheduling	Pattern of resource availability, including service times and maintenance
Sequencing	Order in which resources provide service to waiting work (sometimes called queuing discipline)
Performance	Amount of work accomplished by system

Benefits of DES. According to Manuj et al. (2009), over the last several decades, simulation has been ranked as the most popular operations research tool over more traditional tools like queuing theory and mathematical programming. Computer simulation models are able to replicate very complex systems, with high fidelity, and produce comprehensive data analysis. In systems engineering, simulation is primarily used to explore the effects of alternative system characteristics on system performance without physically producing and testing each contending system (Blanchard & Fabrycky, 2006). Simulation is utilized as a decision-making tool in rapid improvement events or "kaizens." Rapid improvement events, or kaizens, consist of examining the current conditions, identifying potential areas of improvement, and implementing proposed changes, and simulation can be crucial in investigating alternative designs (Treadwell & Herrmann, 2005). Simulation is also beneficial when striving to identify and improve system performance, obtain an understanding of cost-service trade off, validate managerial decisions, and evaluate methods to manage supply chains (Bowersox & Closs, 1989; Allen & Emmelhainz, 1984; Min & Zhou, 2002; Manuj et al, 2009).

Limitations of DES. While DES is a useful tool to replicate and modify systems without risk, there are several limitations. First, collecting, analyzing, and preparing the data to be used in a DES model can take extensive time. The data collected or available may not be appropriate for the simulation. Data input is heavily dependent on historical data; yet as the data ages, the results of the model become less accurate. In certain cases, by the time the model is complete, the collected data may be obsolete and cause doubt in the results. Simulation is often used to predict future events. However, because simulation is dependent on historical data, the model is not necessarily reliable. While DES is a cost effective solution to understand system behavior, expertise and time is required to build a reliable model and to keep the model updated (Tavakoli et al., 2008). Also, when building a simulation model, it is necessary to make assumptions based off the collected data and the desired level of detail within the model. Although it is imperative to make assumptions, the more assumptions that are made, thus increasing the simplicity of the model, the farther from the actual system the model behaves (Law, 2006).

DES in JIT manufacturing systems. Because manufacturing environments are continuously focused on delivering results, they do not have the ability to cease production to modify a system in hopes it will provide desired results. To capitalize on the benefits of utilizing DES to model a non-existent system, many manufacturing environments use the method prior to implementing or modifying a system. DES models can be developed for manufacturers to visualize and identify results when lean manufacturing principles are integrated into their existing set up.

In a simulation study conducted by Carlson and Yao (1992), a highly technical production system looked into implementing a JIT production to reduce WIP, rework, space, lead time for product enhancements and improve quality, customer service, and responsive to customer needs. Using DES to determine whether JIT production could achieve these results, the company simulated the existing system, multiple floor layouts, and a JIT assembly system and found JIT production was the best approach for their company.

Gupta and Gupta (1989) developed a simulation model for a multi-stage, multi-line, dual-card kanban system to identify the impact of adjusting the number of kanbans and the size of kanbans on the system's performance. The performance measures of the model were WIP inventory, capacity utilization, and final product shortages. Based on the model, there were several conclusions. First, in order for kanban systems to operate effectively, it is imperative that the timing and quality of suppliers are reliable. To maintain smooth operation, all production stages shall be balanced. Increasing demand variably degrades system performance. Finally, while the number of kanbans is essential to performance of the system, simply by increasing the number of kanbans does not increase the production rate when all other system parameters are held constant.

In another simulation study, conducted by Ardalan (1997), which was purposed to examine the effects of two kanban variables—length of withdraw cycle and type of priority rule— in a dual-card system, on the average customer wait time and total inventory. Withdrawal cycle refers to the time between two consecutive trips to a material handler at a stage to replenish the raw material at the subsequence stage. Priority rules that were studied include: first-in, first-out (FIFO) and shortest processing time (SPT). Through the conduction of the simulation, a number of relationships between variables were identified. First, the FIFO priority rule results in less waiting time compared to SPT, however the effect of priority rule on inventory is minimal. Also, it is possible to reduce the number of kanbans while reducing the total WIP, without drastically increasing customer waiting time. Increasing the number of kanbans in the system, while holding other system factors constant, results in an increase in input and output stock inventories. Yet, the effects of withdrawal cycle on input and output stock inventories were much less than that of the number of kanbans, which implies it may be more appropriate for management to increase the withdrawal cycle rather than decrease the number of kanbans to reduce inventory levels. By reducing inventories, less production related issues occur at one time, which will be easier to solve.

In a similar study, DES revealed the major advantages of implementing a pull system, via kanbans, into an existing system. Arrival rate, rate and location of bottlenecks, and workstation utilization were used to construct the model of the existing system. To model the kanban system, the number and control of kanbans had to be determined. The results were reduced cycle time variability, flexibility to make engineering and design changes, and tighter control of WIP (Marek et al., 2001).

Building a successful model. As previously stated, simulation can be used as a substitute for experimentation of a real-world system design. However, if the model lacks rationale and reliability, any conclusions derived from the model become erroneous and could result in costly decisions. Having a definitive approach to conduct a simulation study is vital to the validation of the model and overall success of the study. Law (2006) has developed such an approach to enhance a simulation's success, seen in Figure 5 below.

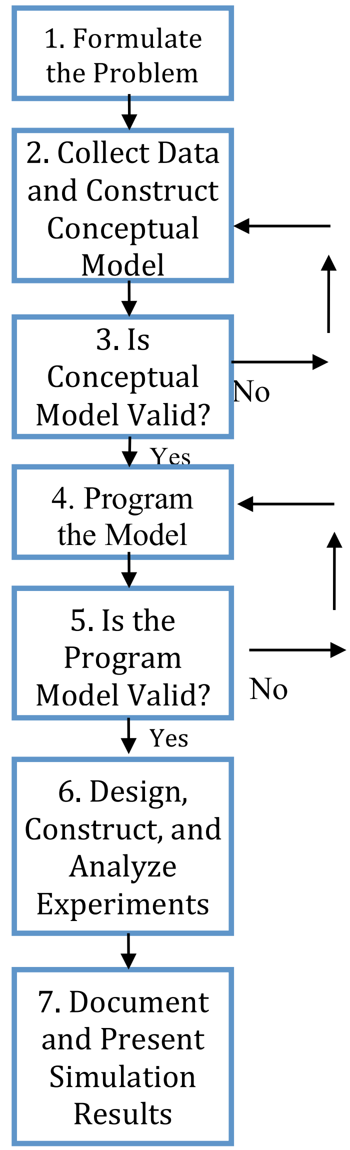

Figure 5. Seven-Step approach for conducting a successful simulation study. Adapted from "How to build valid and credible simulation models," by Law, 2006.

Based on this approach, the initial step is to formulate the problem of interest by identifying the overall objective and scope of the study. Also within this step, the specific question to be answered is identified, system configuration is laid out, and performance measures to be used to evaluate the efficacy of the system configuration are determined (Law, 2006).

Next, all necessary data and information on system layout and operating procedures are to be collected. The data will be utilized to specify system parameters and probability distribution. A conceptual model will be developed with the following details: project objectives, performance measures of interest, data availability, credibility concerns, computer constraints, subject matter expert (SME) opinions, and time and money constraints. Additionally, all model assumptions, algorithms, and data summaries will be documented in the conceptual model. However, a one-to-one correspondence shall not exist between the model and actual system (Law, 2006).

Once the conceptual model is constructed, it shall be validated. A structured walk-through of the model shall be completed before SMEs, analysts, and project managers. Any errors or omissions that are uncovered must be addressed and corrected prior to proceeding to the next stage. With approval of a valid conceptual model, the model is to be programmed into the simulation software, such as Arena version 12 (Law, 2006).

Once the system is modeled, it shall undergo a validation and verification before the experiment begins. If the system is already in existence, the model performance measures are to be compared to the performance measures collected from the actual system, during Step 2; this is called results validation. SMEs and analysts also should review the simulation model's results for rationality. If the results are consistent with how they perceive the system should operate, the model is said to have face validity. During this stage, a sensitivity analysis shall be performed to determine which model features have the greatest effect on performance measures (Law, 2006).

The final two steps in constructing a valid simulation model are to design, make, and analyze the simulation, then document and present the results. The experimentation is performed by adjusting the variables of interest and monitoring the changes in the dependent measures. The conceptual model, detailed description of the computer program, and the results are to be documented. Discussion of the model building and validation process shall be discussed to promote the simulation model's credibility (Law, 2006).

Summary

From the literature review, it can be noted that while manufacturers may have issues in regards to variable lead times, unstable WIP, and system status uncertainty, implementation of JIT production systems, namely a kanbans system, can diminish such issues. Kanbans are able to enhance the overall system performance. However, there are no previous, comprehensive studies that have developed a model to determine the optimal size of a kanban, while simultaneously decreasing the number of kanbans, specifically in a manufacturing environment with a variable workforce. Most studies have focused on optimizing kanban system logistics at a stable manufacturer and by using either DES or mathematical programming. To determine the optimal size of a kanban in a variable manufacturing environment, mathematical models and discrete event simulation will be used. The methods to formulate and validate the model are discussed in further detail below.

Method

Problem Statement

According to Law (2006), in order to build a successful simulation, it is important to first identify the purpose of the study. For manufacturers to remain competitive in their field, meeting customer demands on time is critical. Due to the importance of manufacturers' responsiveness, limiting WIP—jobs that have not arrived at an inventory point—is of great focus (Hopp & Spearman, 2001). The number of kanbans depends on the size of the kanban (i.e., the number of items within the kanban), and together these parameters affect system performance and level of WIP. Also, by minimizing the number of kanbans circulating through the system, the level of inventory is also minimized (Bitran & Chang, 1987). However, very few studies exist that consider kanban size explicitly, specifically in environments with large production and demand variability (Akturk & Erhun, 1999). According to Deleersnyder et al. (1989), identifying the appropriate kanban size for each stage to avoid bottlenecks and reduce WIP is difficult. The aim of this study is to determine the appropriate kanban size for a small manufacturer with a large variability using mathematical programming and DES, which can be utilized as a predictive tool in other manufacturing environment with large production variability for the purpose of reducing the number of kanbans circulating through the system. Determining the appropriate kanban size is dependent upon the time within each production stage.

Stewart-Marchman Production

Data that was utilized to conduct this study was obtained from the small manufacturer Stewart-Marchman Act (SMA) Behavioral Healthcare in Florida. SMA is a non-profit behavioral healthcare rehabilitation center where addiction and mental illnesses are treated. Individuals with mental illnesses or who are recovering from an addiction, who will be referred to as clients from this point forward, are able to maintain a steady job and learn life skills to apply at future jobs. Work that is completed at SMA is contracted from various businesses that generally involve basic assembly of products (Stewart-Marchman-Act Behavioral Healthcare, N.D.).

SMA clients work Monday-Friday on one shift that runs from 9 AM until 1:30PM; thus production runs four and a half hours a day. However, because SMA is a rehabilitation center, clients are not required to show up at 9 AM, and they are able to leave before 1:30PM. Clients are also permitted to take two breaks throughout their shift: one 15 min break in the morning and a 45 min break at lunch (C. Collins, personal communication, January 25, 2012).

Operations flow. One such product assembled at SMA is weight bags that are produced for Sparton Corporation*.* Work orders are placed by Sparton, while SMA gathers necessary material, assembles the product, and delivers the finished product. The bags are constructed out of cloth bags, string, glue, metal pellets, and zip ties. Weight bag assembly is a five-stage push production consisting of: weigh steel pellets in cup, pour pellets into cloth bag, tie bag, glue bag's knot, and zip tie bag around middle. Two quality control stations are integrated into the process to ensure top quality is achieved: after steel pellets are poured into a cup and after the knot is glued (C. Collins, personal communication, January 25, 2012). As seen in Figure 6 below, each stage in the assembly process is dependent upon one another. The figure depicts the process at a high level from the time the Sparton work order is received to weight bag delivery.

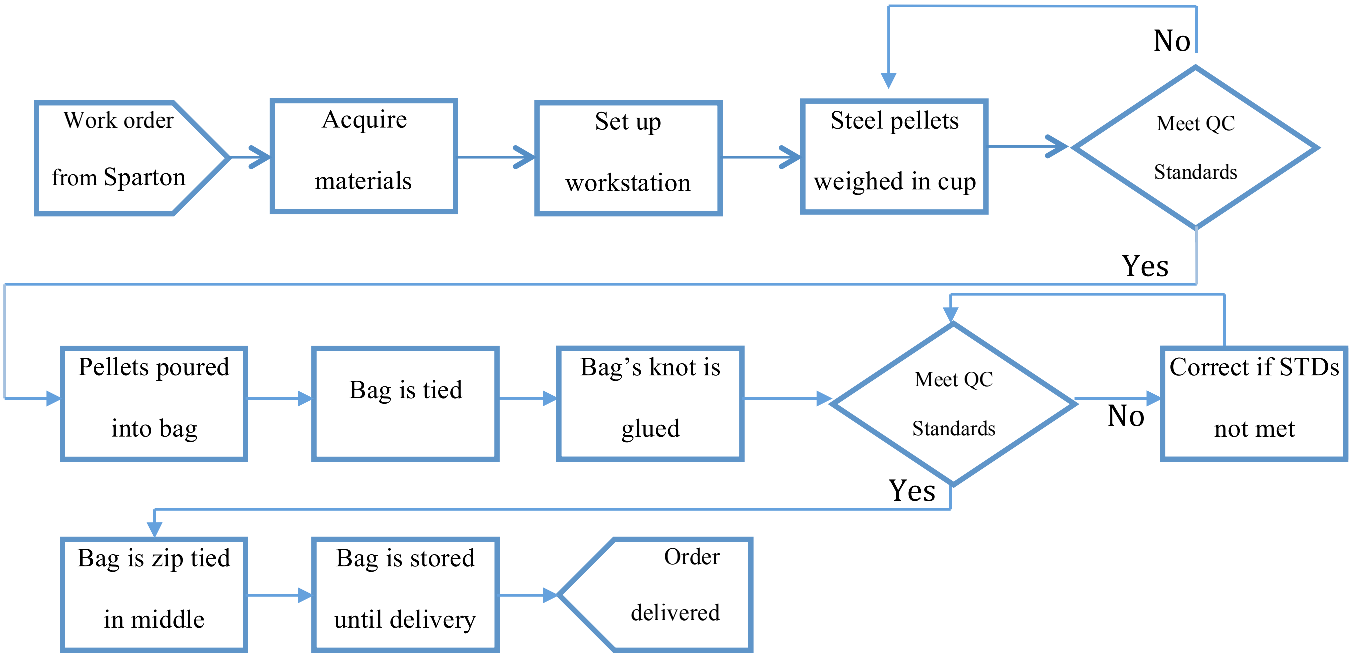

Figure 6. High level functional flow block diagram of weight bag assembly from work order to delivery.

However, the production in SMA has a high degree of variability, including: clients do not show up to work every day, there is a large variation in monthly product demand, and client skill level varies dramatically. For example, where one client may complete 30 units in an hour, another client may only complete two. During an SMA client's shift they are assigned to a work station based on their skill level and the day's attendance. Yet, because SMA clients have a wide variety of skill levels, where one client may complete a task quickly, another client may perform very slowly, there is great risk of high WIP and bottlenecks. Also, SMA receives variable demand from Sparton. For example, one month they may have as little as 700 desired quantities while another month may have up to 3,000 desired quantities. Due to the high production variability, SMA is an appropriate set of data collection for the purpose of this study.

Data collection. The data that was utilized for the study was provided by SMA in an Excel Spreadsheet and based off four and a half years (2007-2012) of weight bag data production metrics. These metrics were utilized to construct a simulation model, validate the models, and experiment the JIT system. Table 3 outlines the input data utilized to construct the model. The input data that has been collected that is valuable for the purpose of the study includes: number of productions completed in each stage (i.e. number of times bag was tied), time to complete given number of weight bags in each stage, number of clients, demand from Sparton, number delivered to Sparton, and time between arrivals.

Table 3

Input data and description

Input Data	Description
Production Stage	Stage (weight steel pellets, pour pellets into bag, tie bag, zip tie bag around middle, and glue bag's knot) in production
Number Completed in Stage	Number of items completed in each stage per day
Hours Worked in Stage	Time client worked in one stage per day
Production Rate	Number of items completed per hours worked in stage (Number Completed in Stage/Hours Worked in Stage)
Number of Clients	Average number of clients in each stage per order
Delivery Rate	Number delivered to Sparton
Time Between Arrivals	Time between Sparton orders in days

Assumptions. In order to develop a credible, verifiable, and valid model and also to simplify the problem, ignoring irrelevant factors, it is essential to make assumptions. Assumptions shall be made based off the collected data and the desired level of detail within each model. According to Law (2006) and Kelton et al. (2010), models shall be only as detailed as they need to be to examine the variable(s) of focus and accurately echo the real-world system of interest. Model assumptions allow for simpler logic and flow. The following assumptions were made to the simulation model to enhance simplicity, while still maintaining a certain level of detail to accurately represent the system of interest for the study:

- Weight bag production will be the only assembly focus of this study. Other productions occur at SMA, however there are few stages within the production and would therefore not provide adequate data for the purpose of the study. Due to the few number of stages in weight bag production, there are limited interactions with other production; therefore any interaction between weight bag production and other SMA productions are irrelevant and will be ignored.
- The model will incorporate two breaks in SMA clients' schedule because clients do have two scheduled breaks throughout their shift. The first break is 15 min and the second is 45 min. These breaks will be factored into the model. If additional breaks are taken, the time is factored into the time to complete a given number of products in each production stage.
- Production at SMA is between the hours of 9 AM and 1:30 PM. Clients are not permitted to work before 9 AM or after 1:30 PM. However, due to the two breaks which total one hour, the model will calculate production from 10 AM to 1:30 PM, for a total of three and a half hours of production per day.

- The number of clients in each stage is based on SMA historical data, where the stage capacity is adjusted based on order size.
- All stages within the production will be modeled as first-in first-out (FIFO) rule.
- While a client is permitted to work at multiple stages within weight bag production, the frequency of such situation is negligible.
- Set-up time and quality control points are carried out by management. Time spend in these stages will be ignored.

Determining Kanban Size using DES and Mathematical Programming

To properly construct a mathematical model to determine kanban size using DES and mathematical programming, the data collected from SMA was utilized in a variety of methods. First is data analysis, then the existing system was modeled using fitted SMA data and constructed in the DES software, ARENA. This model underwent both a validation and verification test. Using the data that was used to construct the model, two Arena add-on programs—Input Analyzer and OptQuest—were used to define parameters and optimize the system's configuration. Based on the output provided by Input Analyzer and OptQuest, a second Arena model with kanbans was constructed to simulate the process with kanbans and to calculate the cycle time with optimal kanbans size.

Arena Simulation. To model and analyze the existing system and the system with kanbans, Arena version 13.90 was utilized. Arena was developed by Rockwell Automation and is a Graphical User Interface (GUI)-based DES tool that enables the user the flexibility to model their system. The software provides an in-depth analysis of simulation results of the current and/or future system(s) modeled without modifying or disturbing the system of interest operational flow (Kelton et al., 2010).

Arena software is based on and includes SIMAN simulation language, which maintains Arena's modeling flexibility and intuitive nature. The software is intuitive in the sense that it allows users to visualize their system by inputting different modules, entities, and resources, with specified parameters, that represent the logical process flow. Also, simulation within Arena provides users with a visual representation of the actual process, which aids in the comprehension of the model's operational flow for those not familiar with computer simulation (Kelton et al., 2010).

Data Analysis, verification and validation of the model. The verification and validation process of a model is critical; otherwise any decisions made with the model may be erroneous (Law & McComas, 2001). The simulation underwent both a validation and verification test, which is described in further detail below.

Data analysis with Input Analyzer. Arena comes with an add-on program, Input Analyzer, which is designed to analyze real-world data, using a goodness-of-fit test, to estimate appropriate parameters and build an expression to be utilized in the Arena model. By analyzing the data and determining the distribution, Input Analyzer allows the user to make the simulation more realistic, and to explore simulations that were not actually observed (Kelton et al., 2010).

This add-on program utilizes a goodness-of-fit test to determine how close the fitted distribution is to the empirical distribution, which is defined by the real-world data (Kelton et al., 2010). Goodness-of-fit tests ask whether the deviations from what would be expected by chance are large enough to lead us to conclude responses are not random (Howell, 2010).

Input Analyzer uses two primary good-of-fit tests: the chi-square test and the Kolmogorov-Smirnov (KS) test. These two standard good-of-fit tests are used to assess if a theoretical fitted distribution is a good match for the data. Chi-square test is best suited for large sample sizes. With this test, a range of the observed data is divided into a discrete number of intervals, and the number of data points under each interval is compared to the fitted distribution's predicted value (Panneerselvam, 2004; Howell, 2010). With observed data denoted as "O" and expected values denoted as "E," the formula for chi square is as follows

$$X^2 = \sum \frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} \tag{12}$$

The KS test is similar to the chi-square test; however it better suited for small samples. The formula is to calculate the KS statistic, D, written as follows

$$D = max|OF_i - EF_i| \tag{13}$$

where OF_i is the observed probability of the *i*th value and EF_i is the expected probability of the *i*th value (Panneerselvam, 2004). With the sample size for this study, a chi-square test was used.

The chi-square test was conducted to test the following the null hypothesis (H_o):

H_o: The random variable data conforms to the distributional assumptions by the given parameter estimations.

H_a: The random variable data does not conform to the distributional assumptions.

If the null hypothesis is not rejected, with a *p*-value of 5%, it is said to be a theoretical distribution. However, if the null hypothesis is rejected, an empirical distribution is used to represents the data instead. The following equation was used in order to specify the continuous, piecewise-linear cumulative distribution function F by first sorting the $X_{(i)}$'s into increasing order, letting $X_{(i)}$ denote the i^{th} smallest of the data set so that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ (Kelton et al., 2010). F is given by:

$$F(x) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \dfrac{i-1}{n-1} + \dfrac{x - X_{(i)}}{(n-1)\left(X_{(i+1)} - X_{(i)}\right)} & \begin{array}{l}\text{if } X_{(i)} \leq x \leq X_{(i+1)} \\ \text{for } i = 1, 2, \ldots, n-1\end{array} \\ 1 & \text{if } X_{(n)} \leq x \end{cases} \tag{14}$$

Verification of the model. Verification of a model refers to the accuracy and correctness in which the model was transformed from the actual system to the model. It is focused on building the model right (Balci, 1997). To test the verification of the model, Kelton et al. (2010) suggests running the test using a variety of scenarios in an attempt to cause an error, otherwise known as debugging. The purpose is to ensure the simulation is accurate with respect to the entity paths and logic. Arena is built with a model verification and debugging tool, which examines in detail the movement of entities through the system; this tool was utilized to verify the simulations' accuracy.

Validation of the model. The validation process determines whether the model is an accurate representation of the system for the particular objectives of the study. If the model is valid, then it can be used to make decisions regarding the system. To determine the model's validity, the most definitive test is to establish that the model's output data closely resembles the output data of the actual system (Law, 2006).

Identifying how "close" the simulation system's output was to the actual system's output can be done by examining the face validity and by using statistical validation. Using face validity, analysts and SMEs review the simulation's output for accuracy and reasonableness; if the simulation's results are perceived to accurately represent the actual system, the simulation is said to have face validity. Statistical validity is conducted to test whether the simulation represents the actual system through statistical validation. This is completed by statistically comparing the results of the simulation's output to the actual system's output. The output compared was the average number of produced weight bags and cycle time. For the models' output to be considered valid, a *t*-test was used to statistically compare both systems' results. This comparison yielded a *p* value. A statistical *p* value greater than a 0.05 alpha level indicates that there is no significant difference between the two outputs; thus, the simulation model is valid (Law & McComas, 2001).

Experimentation

The purpose of this study was to optimize the production of weight bags by minimizing the number of kanbans to determine the appropriate size of kanbans for SMA weight bag production. The performance measure, or dependent variable, of this study was cycle time. To conduct the study, the original model was constructed in Arena using the data obtained from SMA and parameters specified by Input Analyzer. Upon verification and validation of the Arena model, mathematical programming was employed to determine the optimal size of kanbans in a manufacturing environment with production variability based on the following constraints:

Objective:

$$Min \sum_{i=1}^{5} I_i + I_6$$

Subject to:

Percent orders fulfilled within 20 days ≥ 90%

Where:

I is the size of the kanbans

The objective this model is to minimize the sum of the kanban sizes and level of safety stock, while fulfilling 90% of the orders in 20 days or less, to determine the optimal size of kanban. By minimizing the size of the kanban, the level of WIP is reduced; also, based on Little's Law (WIP= CT * λ), by reducing the level of WIP, cycle time is also reduced. The two constraints of the model were implemented to provide an attainable, yet impressive, solution for the model. The mean cycle time of weight bag production at SMA is 26 days. By constraining the cycle time to 20 days or less is a reasonable cycle time goal. While the majority of orders can be completed in 20 days or less, there are order sizes that would not be able to be completed in less than 20 days. For example, some orders sizes are greater than 4000; completing such an order size in less than 20 days would not be feasible. These large order sizes had to be taken into consideration.

OptQuest. Arena is built with an additional add-on program, OptQuest; OptQuest works in sequence when running an ARENA model in quest for an arrangement of input controls that optimize (minimize or maximize) a selected output response. It provides empirical approximation solution to mathematical programming when an analytical solution can't be obtained. OptQuest uses a tabu search and scatter search algorithm, which utilizes a search-based method to find an empirical solution, based on input controls and specified criterion. These search heuristics intelligently move around in the input control space to reliably determine which scenarios to consider in a repeated manner that would lead to an optimal combination of input control variables. With OptQuest, Arena Controls and decision variables are manipulated to seek the optimized solution for the objective specified by the user (Kelton et al., 2010).

With the valid ARENA model and optimized solution for kanban size and level of safety stock determined, a second Arena model was constructed to integrate a kanbans system into the design. The cycle time outputs of the new Arena model were compared with the current system using a single independent sample *t*-test. The *t*-test will test for significant differences between the two models. It is hypothesized that there will be a significant reduction of cycle time from the current system to the system with integrated kanbans. In the following section, the results for the model structure, data analysis, model validation and experimentation are presented.

Results

Model Structure

An Arena simulation model was constructed to replicate the operations flow logic of weight bag production at SMA. First, the weight bag order arrives, and based on that orders size, the capacity (e.g. number of clients) for each stage is determined. The number of clients in each stage is based upon historical data of order size and average number of clients. Table 4 below displays the number of clients per stage based on the order size. Then the order is processed, the weight bags are pushed from the beginning of the system, and go through each of the five stages—fill pellets in cup, pour pellets in bag, tie bag, glue bag, and zip tie middle of bag; once a unit of a weight bag is completed, it proceeds to the following production stage. One unit refers to one completed stage action (i.e. one cup weighted, one cup poured in bag, one knot, one zip tie, or one glue). Then, once the order is complete, the weight bags are then gathered and are shipped to Sparton. The Arena model calculates the cycle time to complete a specified number of orders. The number of weight bags in an order was determined by historical data, from 2007-2012, obtained by SMA. The number of resources (i.e. clients) was determined based on the average number of clients in that stage, depending on the orders' size, from SMA data, as shown in Table 4. Figure 7 illustrates the operations flow of weight bag production.

Table 4

Number of Clients per Stage based on Order Size

Order Size	Stage	Number of Clients
<2000	Fill Pellets in Cup	2
	Pour Pellets in Bag	3
	Tie Bag	3
	Glue Bag Knot	2
	Zip Tie Bag	2
2000-4000	Fill Pellets in Cup	3
	Pour Pellets in Bag	4
	Tie Bag	4
	Glue Bag Knot	3
	Zip Tie Bag	3
4000+	Fill Pellets in Cup	3
	Pour Pellets in Bag	6
	Tie Bag	5
	Glue Bag Knot	3

Zip Tie Bag	3

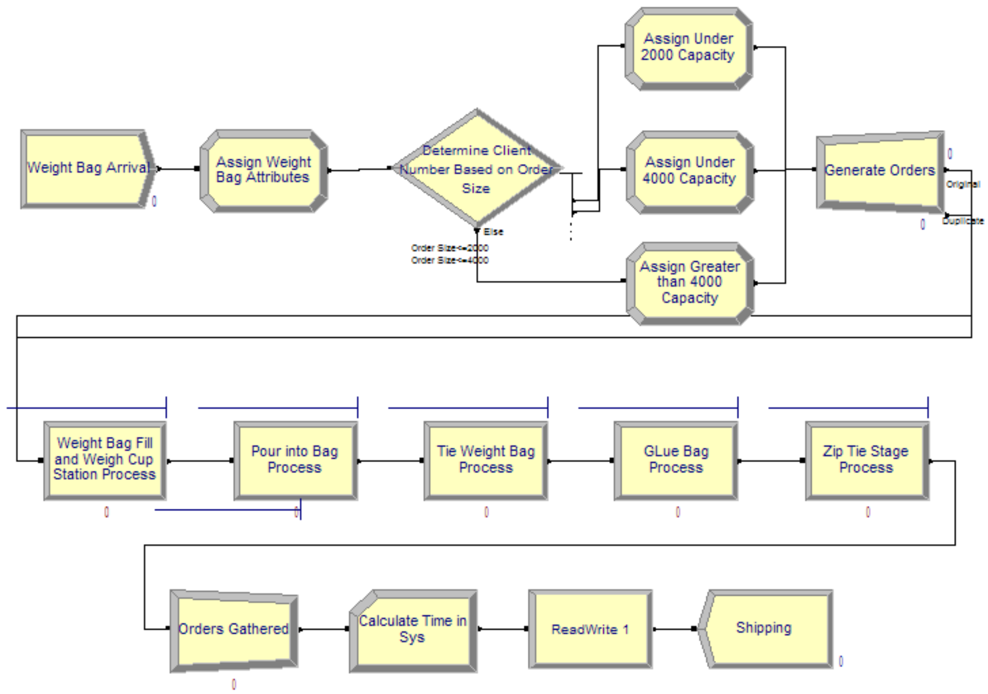

Figure 7. Arena model of current weight bag production process.

Data and Input Analyzer Results

Prior to simulation model construction, verification and validation, the data provided by SMA was analyzed and organized to work within the simulation. First, in each of the five stages, the time—in hours—to complete a unit (number of items completed/ time in stage) was computed and analyzed. Figure 8 depicts the time to complete a part in each of the stages in 2011, where the X-axis represents the days of the year. As seen in this figure, whereas it many take one client to complete one unit in less than three minutes (.05 of an hour), another client, completing the same task, may take up to 30 minutes (.5 of an hour), thus leading to high production variability.

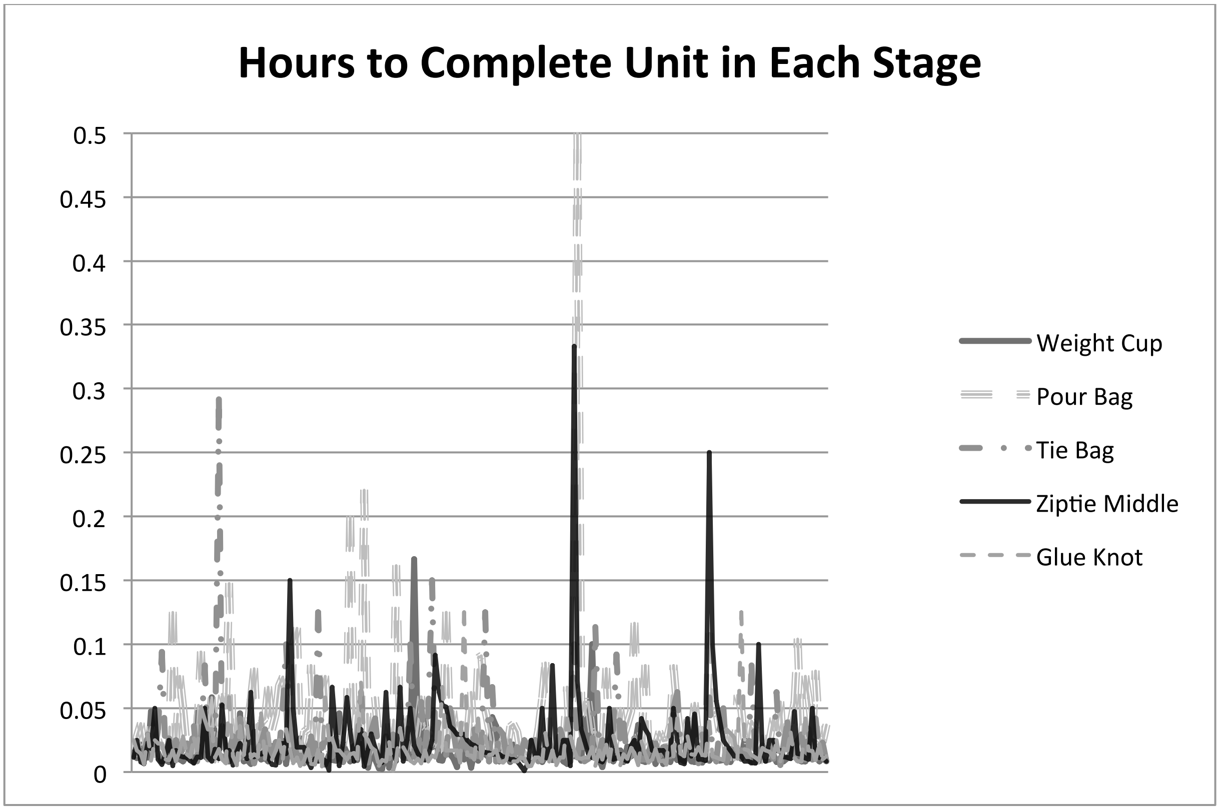

Figure 8. Time (in hours) to complete unit in each stage for weight bag production in 2011.

Arena's Input Analyzer uses a goodness-of-fit test to determine the best distribution function to represent the real-world data to use in the simulation. The data input into Input Analyzer used in the simulation included: time to complete unit in each of the five stages, time between orders, and order size. A chi-square goodness-of-fit test based on an alpha level (α) of 0.05 was performed on the data. A theoretical distribution is valid when $p > 0.05$ because this indicates that the fit distribution data is from the same population as the actual data. However, all data was unable to fit with a theoretical distribution ($p < 0.05$); therefore a continuous empirical distribution was used.

The size of the weight bag order (i.e. number of weight bags) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

$$\begin{gathered}\text{CONT}(0.075,\ 929.999, 0.716,\ 1740.000, 0.821,\ 2550.000, 0.851,\ 3360.000, 0.955, \\ 4170.000, 0.970,\ 4980.001, 0.985,\ 5790.001, 1.000,\ 6600.001)\end{gathered} \tag{15}$$

The distribution of sizes of the weight bag order from Sparton, between 2007-2012, is depicted in Figure 9 below. The cumulative distribution of order size is depicted in Figure 10 below. As seen in the figures below, there is a variable order sizes from Sparton—ranging from less than 500 to greater than 6000.

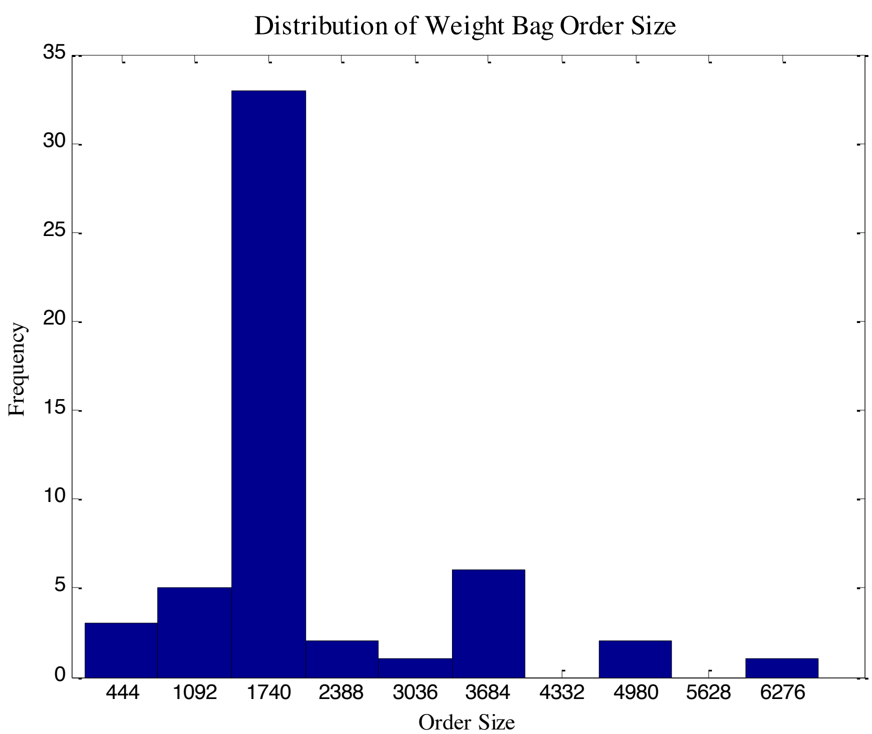

Figure 9. Distribution of order size of weight bags from Sparton.

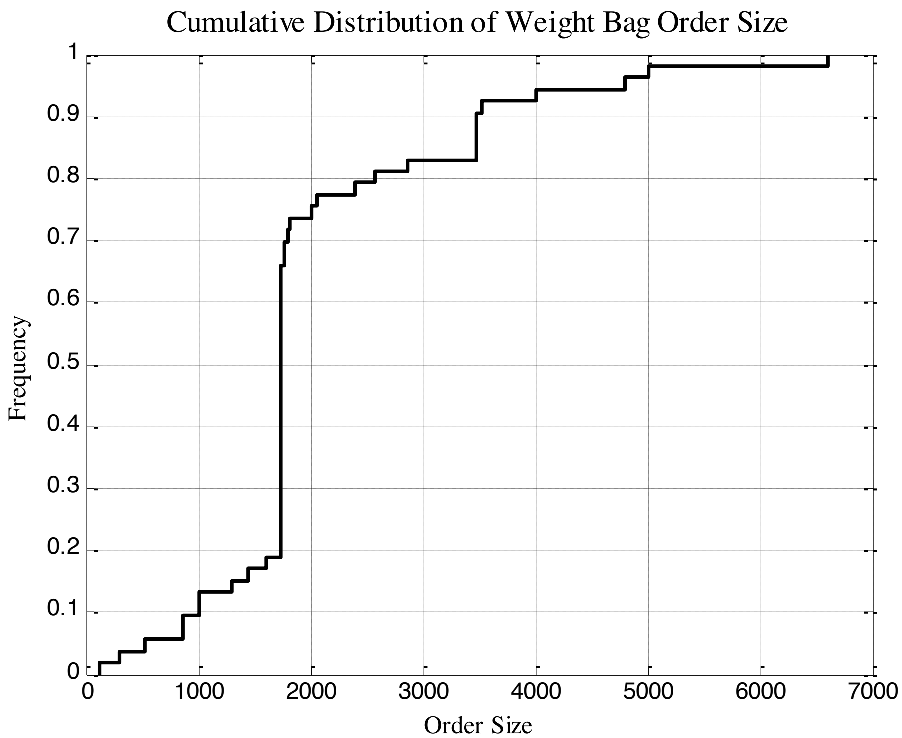

Figure 10. Cumulative distribution of order size for weight bags from Sparton.

The time between weight bag orders (in days) did not fit within a theoretical distribution, p=.026. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

CONT(0.016, 5.500,0.079, 6.500,0.095, 7.500,0.111, 8.500, 0.159, 9.500,0.175, 11.500,0.190, 12.500,0.206, 13.500, 0.238, 14.500,0.349, 15.500, 0.381, 16.500,0.397, 17.500,0.429, 18.500,0.476, 19.500,0.492, 20.500,0.508, 21.500,0.571, 22.500,0.587, 25.500,0.603, 26.500, 0.667, 27.500,0.730, 28.500, 0.746, 29.500,0.762, 31.500,0.778, 32.500,0.810, 33.500,0.825, 35.500,0.873, 39.500,0.889, 42.500,0.905, 48.500,0.952, 61.500,0.968, 78.500,1.000, 96.500) (16)

The distribution of time between orders (in days) from Sparton, between 2007-2012, is depicted in Figure 11 below. The cumulative distribution of time between orders is depicted in Figure 12 below.

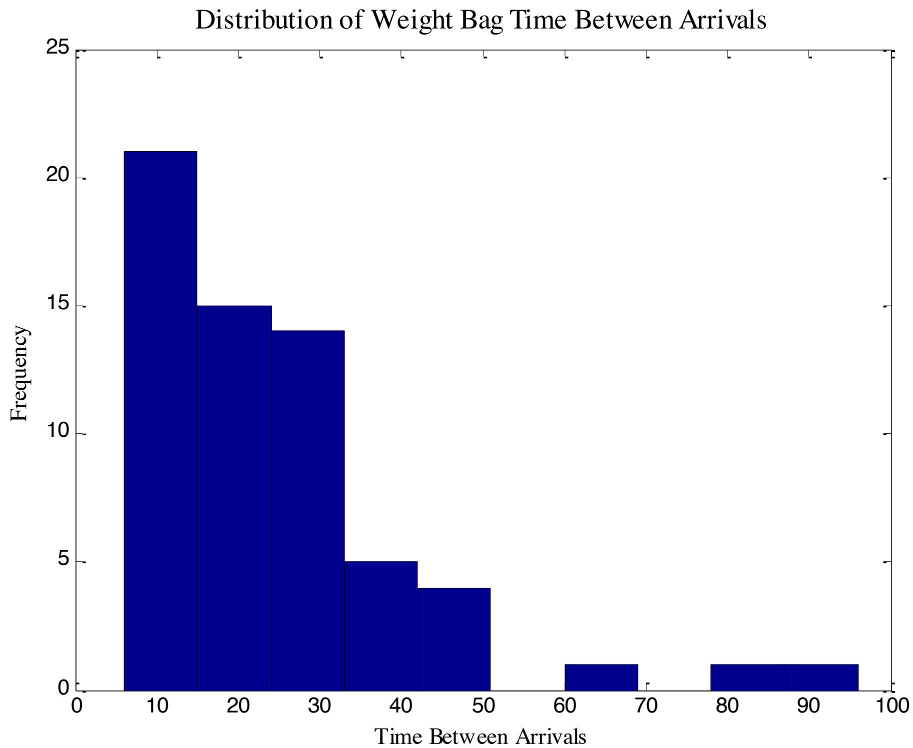

Figure 11. Distribution of time between arrivals from Sparton.

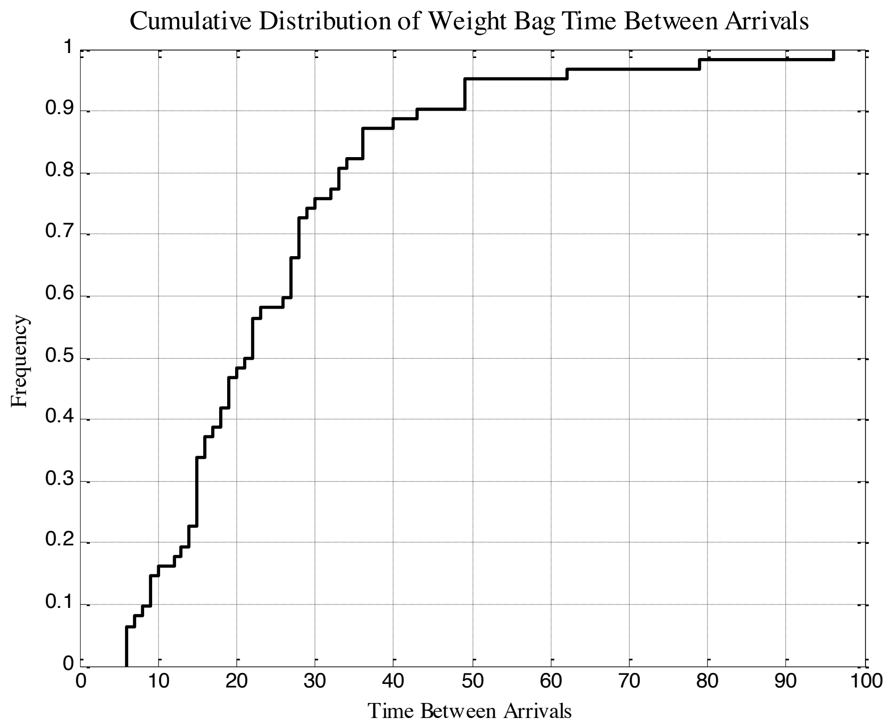

Figure 12. Cumulative distribution of time between arrivals from Sparton.

The time to complete one unit (in hours) for the first stage (pour pellets in cup) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

$$\text{CONT}(0.770,\ 0.022{,}0.929,\ 0.044{,}0.957,\ 0.066{,}0.972,\ 0.088{,}0.981,\ 0.110{,}0.988,\ 0.132{,}0.994,\ 0.154{,}0.997,\ 0.242{,}0.998,\ 0.484{,}1.000,\ 0.550) \tag{17}$$

The distribution of time (in hours) to complete one unit in the first stage, is depicted in Figure 13 below. The cumulative distribution of time to complete one unit in the first stage is depicted in Figure 14 below.

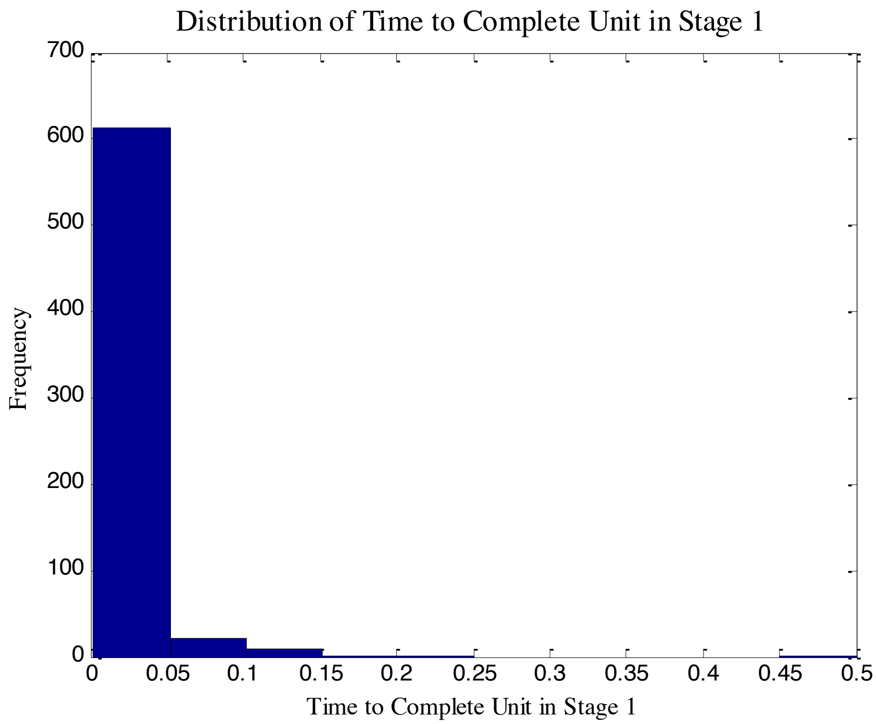

Figure 13. Distribution for the time (in hours) to complete one unit in weigh cup stage.

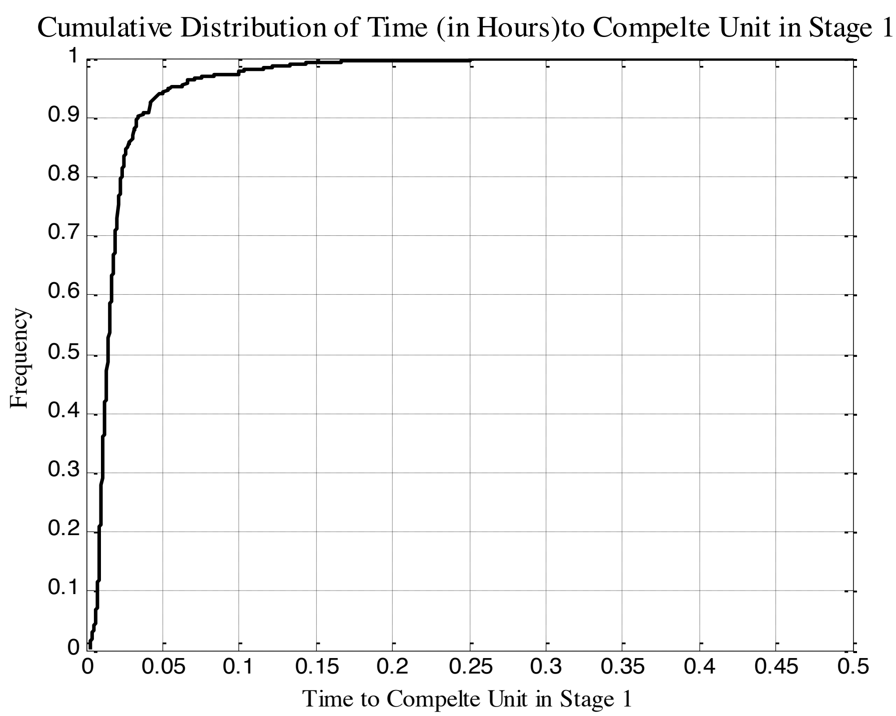

Figure 14. Cumulative distribution for the time (in hours) to complete one unit in weigh cup stage.

The time to complete one unit (in hours) for the second stage (pour pellets into bag) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

$$\begin{gathered}\text{CONT}(0.150,\ 0.016,0.521,\ 0.032,0.777,\ 0.049,0.899,\ 0.065,0.934,\ 0.081,0.955, \\ 0.097,0.973,\ 0.113,0.986,\ 0.129,0.988,\ 0.146,0.992,\ 0.178,0.996,\ 0.194,\ 0.997, \\ 0.243,0.998,\ 0.372,0.999,\ 0.485,1.000,\ 0.550)\end{gathered} \tag{18}$$

The distribution of time (in hours) to complete one unit in the second stage, is depicted in Figure 15 below. The cumulative distribution of time to complete one unit in the second stage is depicted in Figure 16 below.

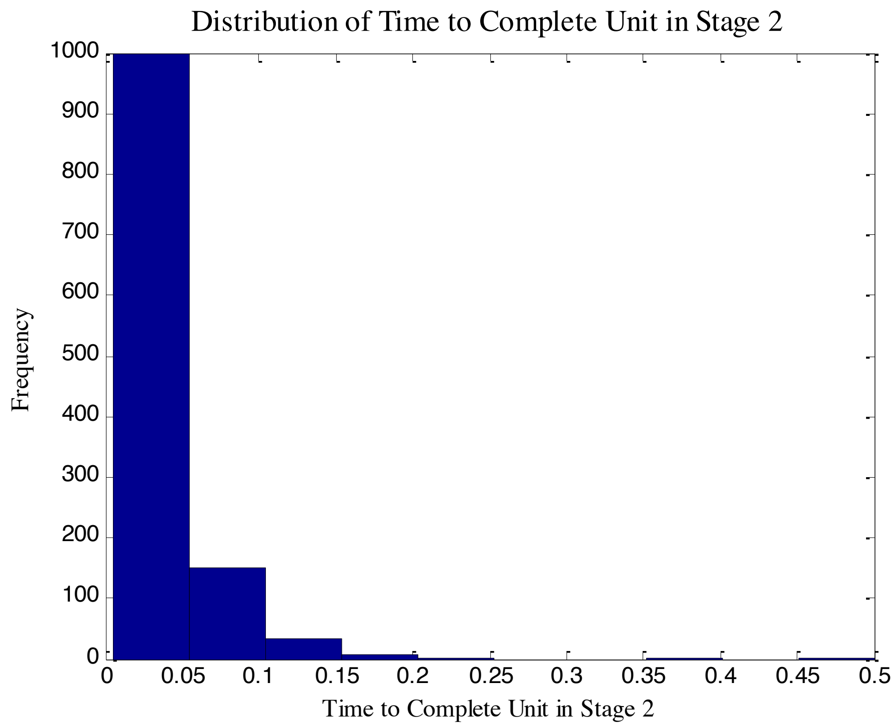

Figure 15. Distribution for the time (in hours) to complete one unit in pour pellets in bag stage.

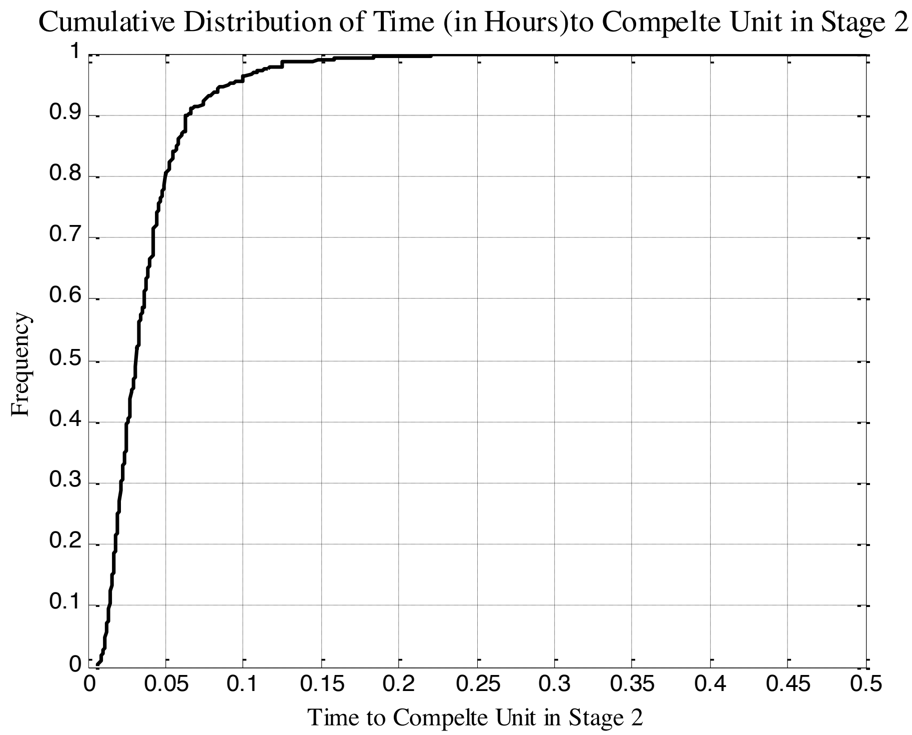

Figure 16. Cumulative distribution for the time (in hours) to complete one unit in pour pellets in bag stage.

The time to complete one unit (in hours) for the third stage (knot the bag) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

CONT(0.289, 0.012,0.677, 0.025,0.861, 0.037,0.915, 0.049,0.948, 0.062,0.969, 0.074,0.974, 0.086,0.980, 0.099,0.984, 0.111,0.985, 0.124,0.991, 0.148,0.993, 0.161,0.996, 0.247,0.997, 0.284, 0.998, 0.346,0.999, 0.371,1.000, 0.420) (19)

The distribution of time (in hours) to complete one unit in the third stage, is depicted in Figure 17 below. The cumulative distribution of time to complete one unit in the third stage is depicted in Figure 18 below.

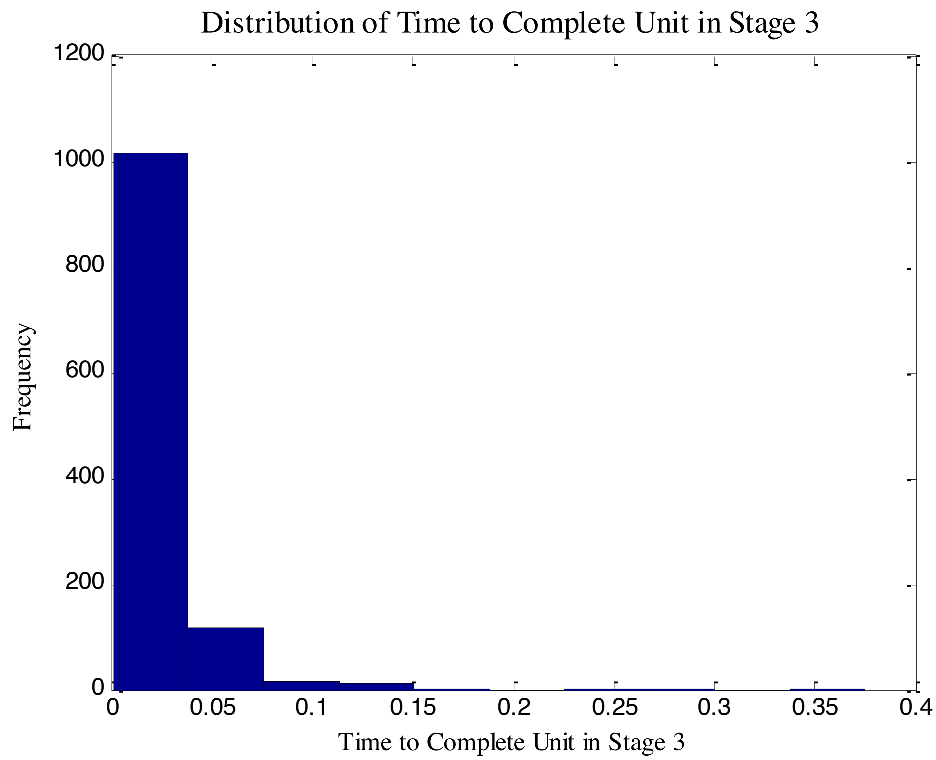

Figure 17. Distribution for the time (in hours) to complete one unit in tie bag stage.

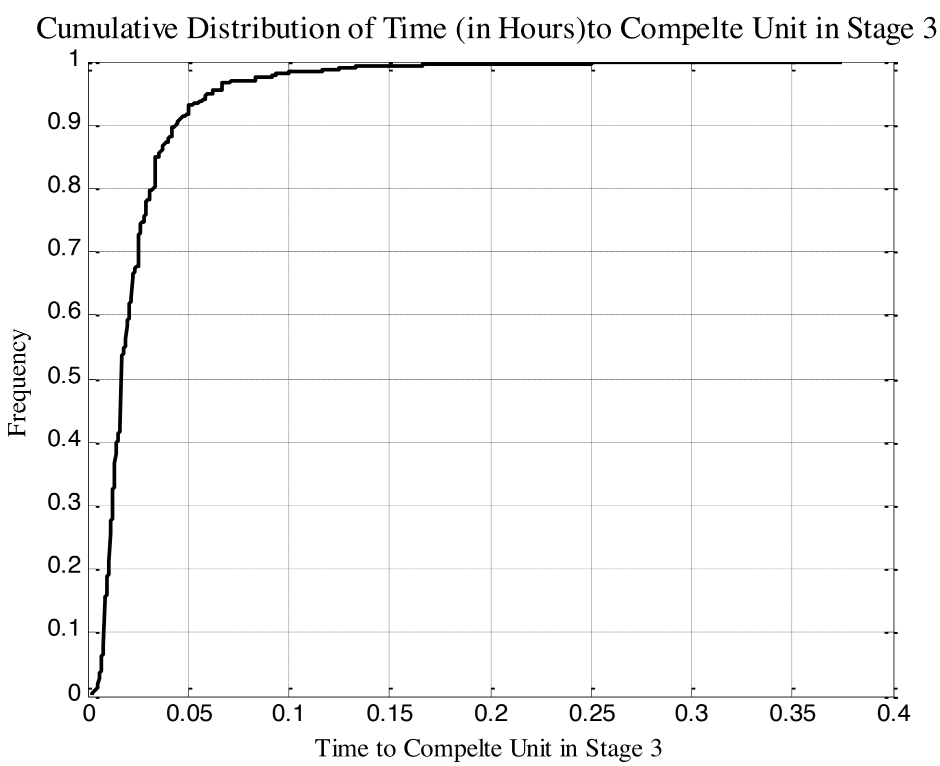

Figure 18. Cumulative distribution for the time (in hours) to complete one unit in tie bag stage.

The time to complete one unit (in hours) for the fourth stage (glue bag's knot) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

$$\text{CONT}\,(0.406,\ 0.017, 0.766,\ 0.033, 0.859,\ 0.050, 0.909,\ 0.067, 0.945,\ 0.083, 0.955,\ 0.100, 0.968,\ 0.117, 0.983,\ 0.133, 0.985,\ 0.167, 0.987,\ 0.183, 0.989,\ 0.200, 0.991,\ 0.217, 0.993,\ 0.250, 0.995,\ 0.267, 0.996,\ 0.300, 0.997,\ 0.317, 0.998,\ 0.367, 0.999,\ 0.500, 1.000,\ 0.550) \qquad (20)$$

The distribution of time (in hours) to complete one unit in the fourth stage, is depicted in Figure 19 below. The cumulative distribution of time to complete one unit in the fourth stage is depicted in Figure 20 below.

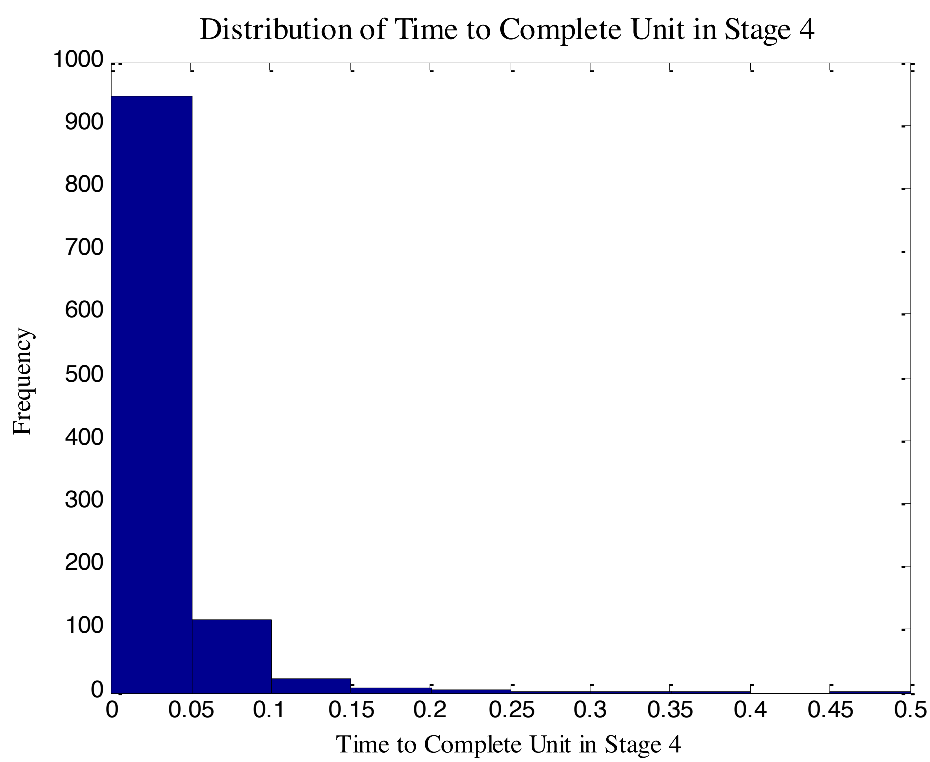

Figure 19. Distribution for the time (in hours) to complete one unit in glue knot stage.

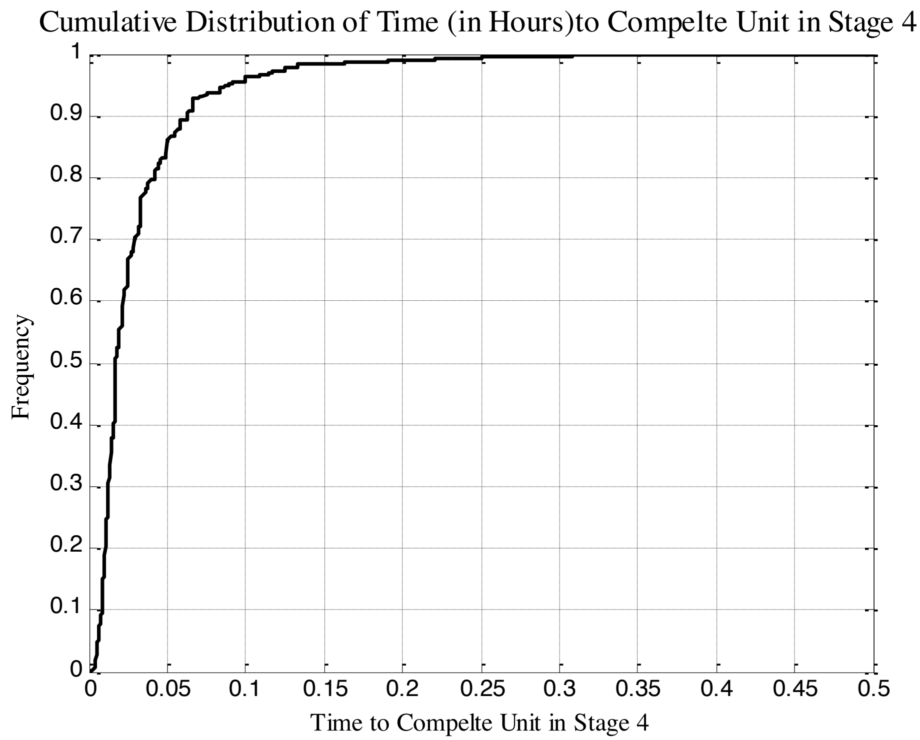

Figure 20. Cumulative distribution for the time (in hours) to complete one unit in glue knot stage.

The time to complete one unit (in hours) for the fifth stage (zip tie middle) did not fit within a theoretical distribution, $p<.005$. A continuous empirical distribution was used to fit the data. The following formula was used to represent the data:

$$\begin{gathered}\text{CONT}(0.576,\ 0.014,0.880,\ 0.028,0.953,\ 0.042,0.961,\ 0.056,0.978,\ 0.070,0.989,\\ 0.084,0.991,\ 0.098,0.994,\ 0.112,0.997,\ 0.196,0.998,\ 0.308,1.000,\ 0.350)\end{gathered} \tag{21}$$

The distribution of time (in hours) to complete one unit in the fifth stage, is depicted in Figure 21 below. The cumulative distribution of time to complete one unit in the fifth stage is depicted in Figure 22 below.

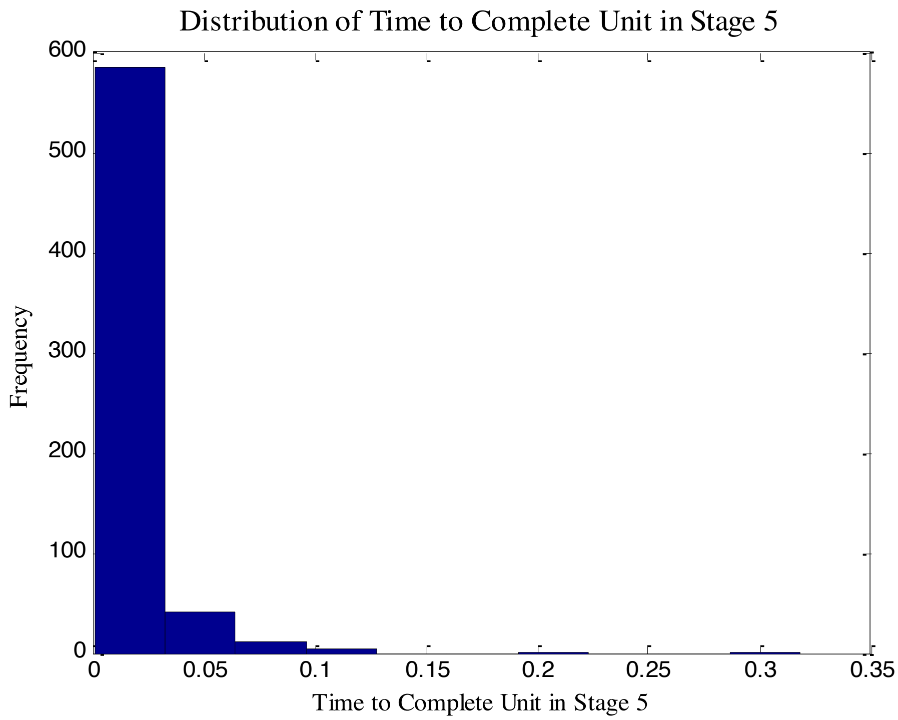

Figure 21. Distribution for the time (in hours) to complete one unit in zip tie middle stage.

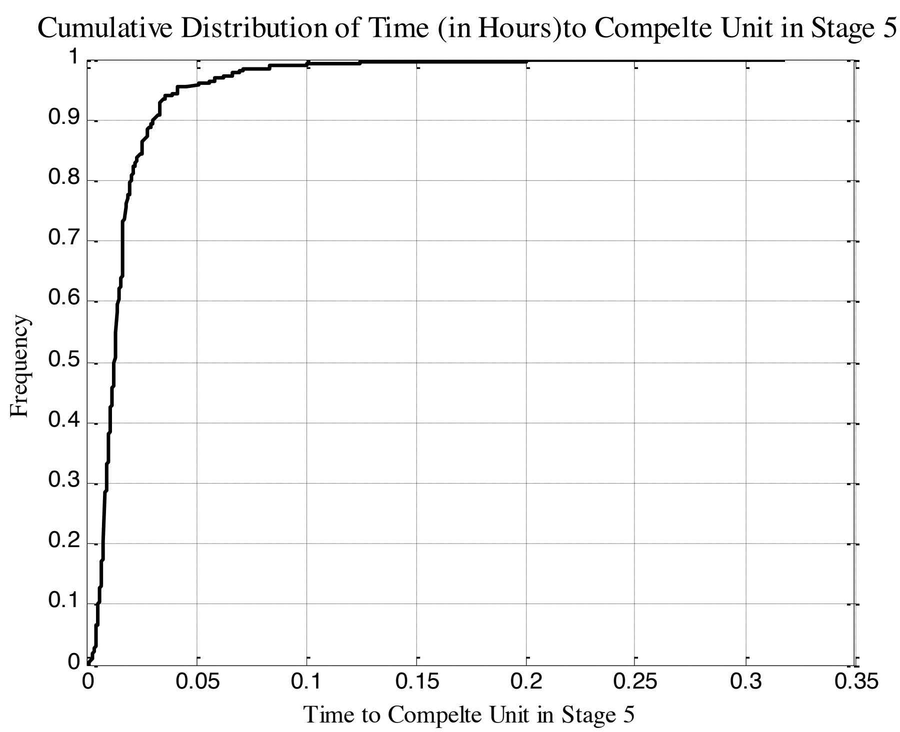

Figure 22. Cumulative distribution for the time (in hours) to complete one unit in zip tie middle stage.

As seen in Figures 13- 22, there is high production rate variability amongst clients. While the majority of clients many be able to complete one unit in under three minutes, there are outliers who require a longer time to complete the same act.

Once the empirical data was organized and formatted for the model, it was input into their respective process locations. Next, the model went through a verification process, which consisted of debugging to ensure simulation accurateness. The model was verified by running the simulation through multiple scenarios and ensuring entity paths and logic were correct. Once the model was verified, it was statistically validated using the system's cycle time.

Model Validation

Validation of the model was conducted by statistically comparing the average cycle time—in days—from the simulation model to the cycle time from historical data at SMA. Three actual orders, of different sizes, were modeled in the simulation to compare cycle time. The average number of clients in each stage during that order was modeled as resources for their respective stage. Table 5 displays the stage capacity (e.g. number of clients in each stage) based on the average number of clients in the stage for that order. The Arena model ran for 50 replications which yielded 50 samples (n = 50) of the average time order cycle time. This data set was compared to the actual observed cycle time. Table 6 displays the descriptive statistics for the data samples used in this study.

Table 5

Average Number of Clients in each Stage

Order Size	Stage	Number of Clients
4740	Fill Pellets in Cup	3
	Pour Pellets in Bag	5
	Tie Bag	5
	Glue Bag Knot	3
	Zip Tie Bag	3
1666	Fill Pellets in Cup	2
	Pour Pellets in Bag	4
	Tie Bag	2
	Glue Bag Knot	2
	Zip Tie Bag	2
1861	Fill Pellets in Cup	2
	Pour Pellets in Bag	4
	Tie Bag	3
	Glue Bag Knot	3
	Zip Tie Bag	3

Table 6

Descriptive statistics for the independent-samples t-test of cycle time

Order Size	Actual Cycle Time	Model Cycle Time	t	df	Sig. (2-tailed)	Mean Difference
4740	33	33.043	.558	49	.580	.04304
1666	15.5	15.590	1.092	49	.280	0.095
1861	13	12.972	-.488	49	.627	-0.028

Three independent-samples t-tests were conducted in order to statistically analyze the simulations and SMA's actual cycle time to each other. The results of the independent-samples t-test showed there was not a significant between the order size of 4740 groups, $t(49) = .558$, $p = 0.580$, the order size of 1666 groups, $t(49) = 1.092$, $p = 0.280$, and the order size of 1861 groups, $t(49) = -.488$, $p = 0.627$. These results from the independent-samples t-tests indicate that the Arena model is a valid representation of the actual SMA cycle time. Once the model was determined to be valid, the model was modified with an integrated kanban system, and mathematical programming was used to determine the optimal size of kanbans for weight bag production.

Kanban Simulation Model

To determine whether cycle time would be reduced in weight bag production by integrating a kanban system, a second Arena model was constructed. To model the production system with kanbans, first, the first kanbans was seized, or held, prior to the first process in the system. Upon complete of the first process, the second kanban was seized, then the first kanbans was released, and the second process in the production began. This system continued throughout all processes in production. Because pull systems—including kanbans systems—release of work is authorized from actual system demand, the final kanbans is released based on the level of safety stock. If the safety stock is less than a predetermined level of safety stock, the final kanban will be released, which will continue to authorize work to the subsequent stage, and production will continue. However, if the safety stock is equal to the order size, production will be held until the next order, thus producing only what is needed when it is needed. Figure 17 below illustrates the operations flow of weight bag production with integrated kanbans to release work across processes.

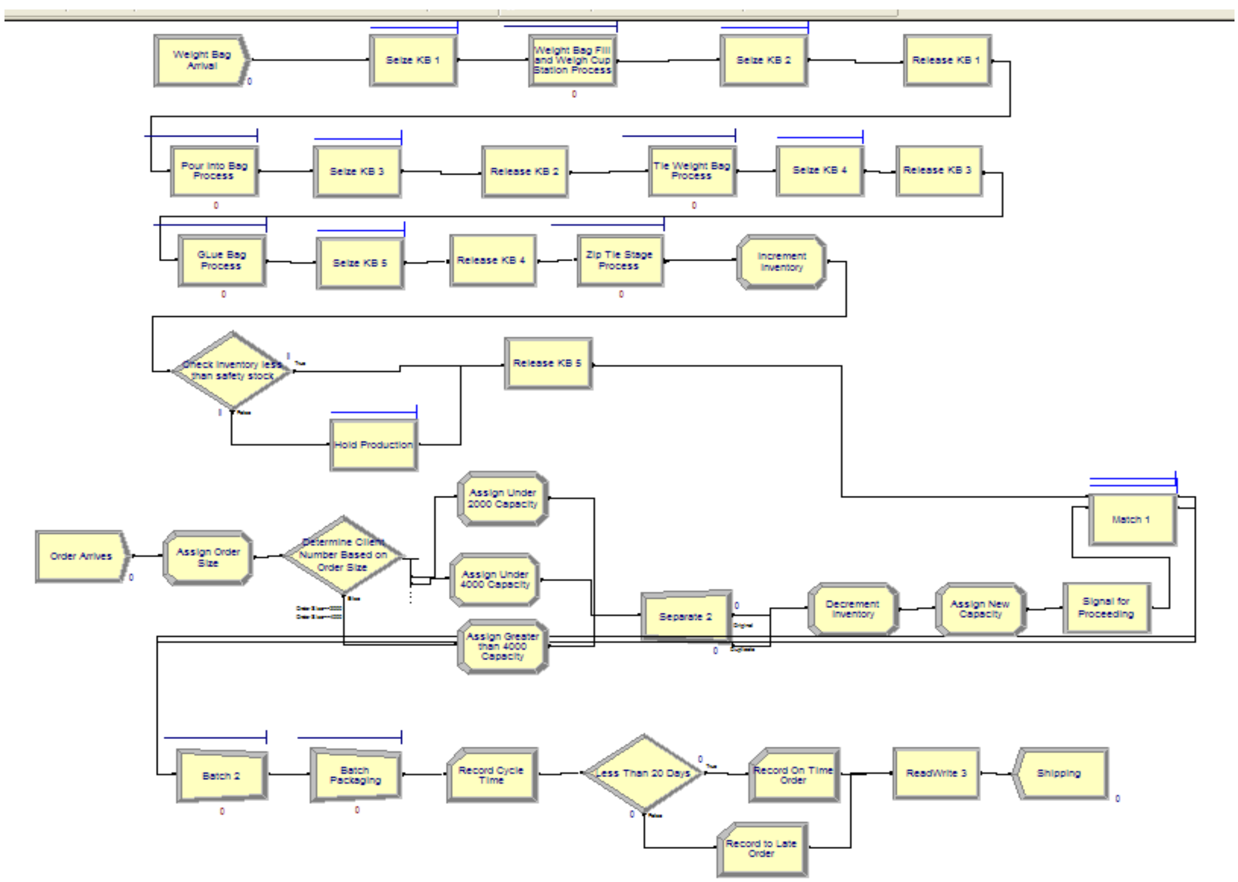

Figure 17. Arena model of weight bag production with kanban system.

Kanban size optimization and sensitivity analysis. The objective of mathematical programming using OptQuest was to find an optimal size of kanbans for each stage and level of safety stock in order to reduce the overall production cycle time. A constraint—

90% of the order cycle time shall be completed in less than or equal to 20 days—was implemented into OptQuest to find the minimum size of kanbans and safety stock based on the fitted order sizes and time to complete a unit in each stage. Table 7 below depicts the mathematical programming solution from OptQuest to optimal kanbans size and level of safety stock to reduce SMA's production cycle time.

Table 7.

Optimal Kanban Size and Level of Safety Stock based on Sensitivity Analysis

Stage	Optimal Size
Fill Pellets in Cup	20
Pour Pellets in Bag	10
Tie Bag	10
Glue Bag Knot	10
Zip Tie Bag	10
Safety Stock	500

Cycle time results. To determine whether SMA weight bag production would benefit, with respect to cycle time, by implementing and integrating a kanban system, the current weight bag production system and the weight bag production with kanbans models' cycle time output were statistically compared. The first models compared were based on the empirical distributions of time between arrivals and order size. Table 8 displays the cycle time means and standard divinations of the two systems. Figure 23 graphically depicts the cycle time means of the two systems.

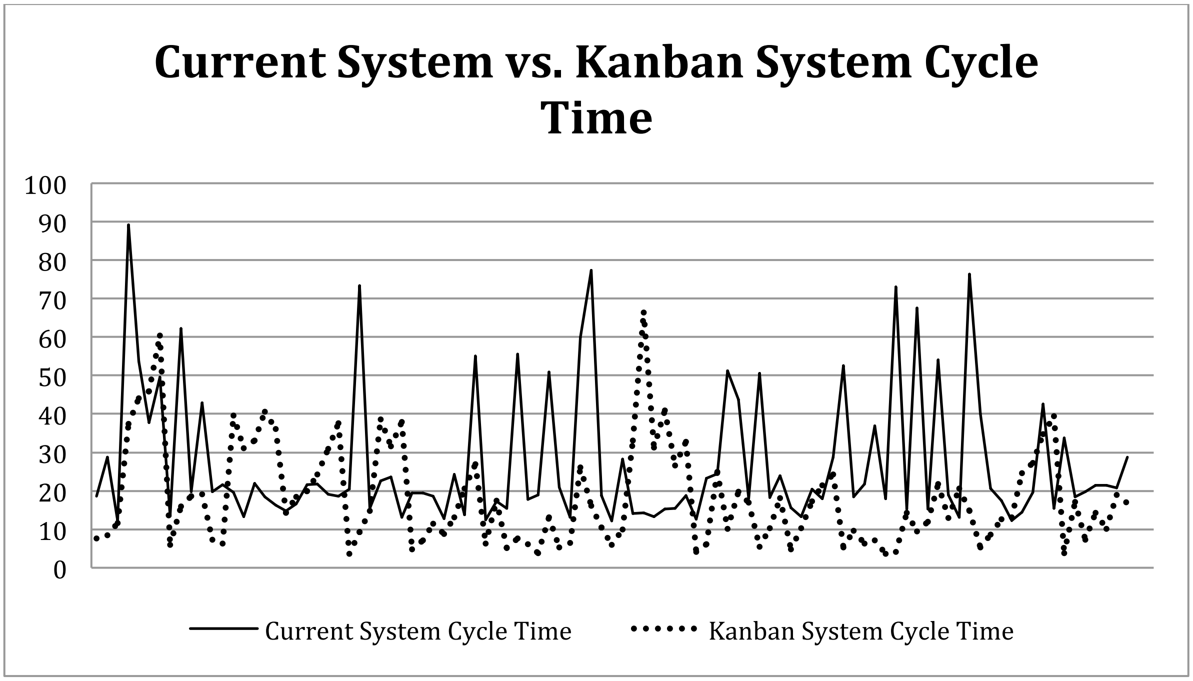

Figure 23. Graph of cycle time means of current and kanbans systems.

Table 8.

Cycle Time Statistics based on Empirical SMA Data

System	N	Mean	Std. Deviation	Std. Error Mean
Current System	100	26.9036	17.95014	1.79501
Kanban System	100	17.5852	12.38261	1.23826

An independent-samples *t*-test was conducted in order to statistically analyze the two groups' cycle time to each other. The results of the independent-samples *t*-test showed there was a significant difference between the two systems, where the kanbans system had a significantly lower cycle time, $t(198) = 4.273$, $p = 0.000$.

Additional independent-samples *t*-tests were conducted using the data—order size and number of clients per stag—used to validate the current system model. Figure 24 graphically depicts the cycle time means of the systems for different order sizes. Results of these cycle time descriptive statistics are displayed in table 9 below.

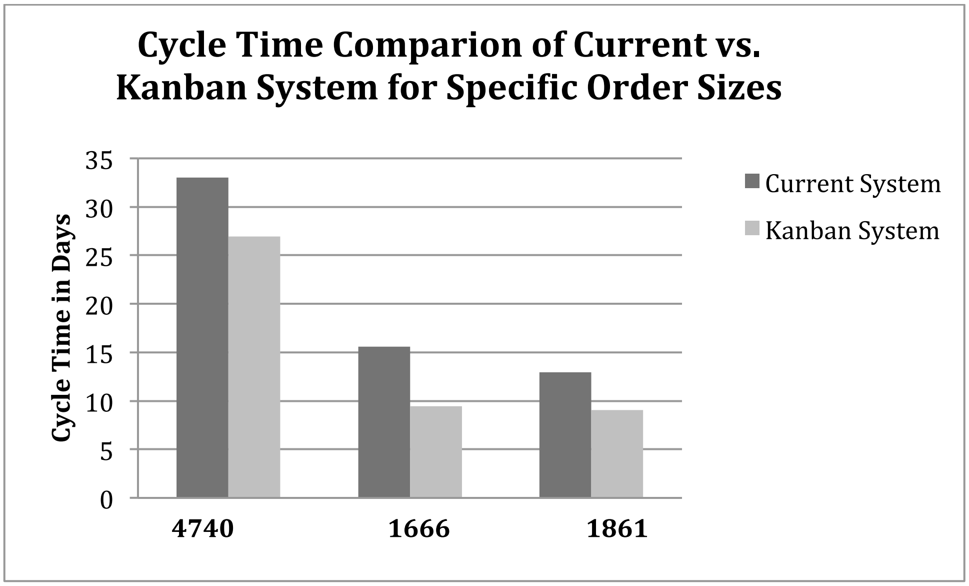

Figure 24. Graph of cycle time means of current and kanbans systems for specific order sizes.

Table 9.

Cycle Time Statistics of Current and Kanbans Systems based on Historical SMA Data

	Order Size	N	Mean	Std. Deviation	Std. Error Mean
Kanban System CT	4740	30	26.9282	5.40760	.98729
Current System CT			33.043		
Kanban System CT	1666	30	9.422	2.55899	.46721
Current System CT			15.59		
Kanban System CT	1861	30	9.0318	.11920	.02179
Current System CT			12.972		

These independent-samples *t*-tests were conducted in order to statistically analyze the order sizes of the two groups' cycle time to each other. The results of the independent-samples *t*-test for an order size of 4740 showed there was a significant difference between the two systems, where the kanbans system had a significantly lower cycle time, $t(29)$=-

6.150, p = 0.000. The results of the independent-samples t-test for an order size of 1666 showed there was a significant difference between the two systems, where the kanban system had a significantly lower cycle time, $t(29)$=-13.202, p = 0.000. Results of the independent-samples t-test for an order size of 1861 showed there was a significant difference between the two systems, where the kanbans system had a significantly lower cycle time, $t(29)$=-181.881, p = 0.000.

In the following section, results are discussed, including cycle time, validation of the simulation model, and mathematical programming. Limitations encountered in this study are also discussed. Suggestions on areas of future research are given at the end of the section.

Discussion

The purpose of this study was to determine kanban size for a manufacturing system with variable production rates in order to reduce cycle time. Mathematical programming was utilized to determine the optimal kanban size based on historical SMA data, kanban system layout, constraints—complete orders within 20 days, 90% of the time—and objective—minimize the sum of kanbans and safety stock. Mathematical programming was conducted to determine the optimal, yet feasible, solution for the system. Because the model is non-linear and not able to provide an analytical solution, an empirical solution, based on the Tabu Search, was applied through OptQuest. The results of the optimization model were input into the kanban system model with the obtained kanban and safety stock capacity. Statistical tests were conducted to assess the effects of kanbans on cycle time for different order sizes.

The first statistical test compared the cycle time of the two systems using the empirical distribution of time between arrivals and order size. Results of the test indicated that kanban system's cycle time was 9.4 days, almost two weeks, less than the current system.

Three additional statistical tests were conducted using the data to validate the current system model, including order size and number of clients per stage. Order sizes of 4740, 1666, and 1861 indicated a significant reduction in days with a kanban system, specifically 6.14 days, 6.2 days, and 3.9 days respectively.

The significant reduction in cycle time can be attributed to the kanban system's ability to limit WIP and eliminate bottlenecks. Through observation of the current system model, it appeared there was a large bottleneck at the second stage—pour pellets in bag—and third stage—knot bag—for units took longer, on average, to be completed within these stages as compared to the other stages. With the kanban system, the first stage—pour pellets in cup—was not continuously sending units to the second stage; instead, the first stage had to wait for authorization from the second stage to replenish the inventory stock, thus eliminating the bottleneck at stages two and three.

Findings of this study were consistent with previous studies of manufacturing systems that implemented a kanbans system. For example, an electronic company reduced lead times from 180 hr to 60 hr and WIP by 70% (Lee-Mortimer, 2008). Also, a clothing manufacturer, whom often experienced bottlenecks, high work-in-process (WIP), and highly variable lead times, noted greatly reduced WIP levels, and consistent lead times upon kanban system implementation (Billiesbach, 1994).

By demonstrating that cycle time and WIP can be reduced in a manufacturing environment without stationary production rates or demand and with large variablities, there are great implications for such manufacturers. Jarupathiru et al., (2009) discussed the challenges of integrating kanban systems into systems with varying demand and production rate, stating system breakdowns and unexpected situations are more likely to occur. However, this study demonstrated the positive impact of kanban systems in such manufacturing environments.

The results of this study are of practical use to SMA production and other manufacturers with variable production. Cycle time is a key component to manufacturers' success; by reducing cycle time, in a parsimonious fashion such as the use of kanbans, manufacturers are able to remain competitive to their customers. However, there are practical challenges of implementing this type of system at SMA. For example, implementing a JIT system into a current pull production environment requires a change in culture from the clients, which has shown to be difficult in cultures with highly functional adults (Verweij & Maassen, 2011). Implementing a JIT system in a workforce that is mentally challenge could potentially lead to periodic system breakdown. There are also issues due to training; training would require time away from production, which would increase lead time for a period.

Limitations of the Study

There were several limitations of this study. The first limitation is in regards to the quality of data provided by SMA. Errors existed throughout the SMA data files, including repeats of orders, inaccurate order sizes, order dates and delivery dates. These errors were corrected for by deleting duplicate order dates and calculating time between arrivals. Also, while data from 2007 to 2012 was used to build the model, additional data points would enhance the validity of the model.

Other limitations of the study were due to assumptions made to build the model. First, only one SMA manufacturing process was modeled. Several other manufacturing processes occur at SMA. However, weight bag production has more production stages; therefore it was more practical to utilize weight bag production. While these processes have variable production rates, like weight bag production, order sizes and demand differ. Kanban size for these other processes may not produce the same results obtained in this model. Also, while these other processes were ignored for the model, there is the potential that other processes could interfere with weight bag production.

Another assumption made was that production was 3.5 hours, from 10 AM to 1:30 PM; however, SMA production is from 9AM to 1:30 PM, with two breaks totaling one hour. The scheduled breaks were not taken into consideration in order to simplify the simulation model construction; however, there is potential for decreased model validity because clients may work slower when approaching their break time or it may take them longer to warm-up and continue producing at their previous rate after a break. Despite the model being statistically valid, there is opportunity for enhanced validity.

Additionally, quality control points were not integrated into the model. No data was provided in regards to the number of weight bag units needing to be reworked. To compensate for this lack of information, the time to complete a unit per stage took into account additional time spent on rework. This information would have provided additional information in regards to the benefits of kanban implementation.

Future Direction of the Study

This study was intended to provide a manufacturing system, with high production variability, insight on determining kanban size for the purpose of decreasing production cycle time. A model was constructed to compare the current system and the current system with integrated kanbans to simulation and compare the operations flow. Other small manufacturing facilities could follow the same process and utilize a similar kanban production flow—seize kanban 1, process 1, seize kanban 2, release kanban 1, process 2, etc. The model developed in this study provides a general approach to determine kanbans size. Empirically, it is applicable to other manufacturers, with or without variable production, who could utilize the model to determine kanban size and identify benefits by updating time to complete units in a stage, time between arrivals, order size, and stage capacity. Although other manufacturers have different process flow, they do have similar high level operational flow where an order arrives, units go through a set number of stages, and the order is shipped.

Future studies using this model could alter the model to test different manufacturing philosophies for JIT systems. For example, a study could be conducted to determine the impact of implementing a CONWIP system, generalized form of a single-stage kanban system (Spearman et al., 1990; Marek et al., 2001). Perhaps a single stage kanban, at the end of production, would enhance the efficiency of the manufacturing system with variable production. Also, determining the optimal number of kanbans, could reduce the production cycle time. For example, simulation and mathematical programming could find it is most optimal to have a kanban after every other stage.

Staffing considerations could also be explored for future studies. For example, determining the optimal number of clients per stage, using mathematical programming, could benefit the SMA production.

Additional studies could observe SMA weight bag production flow to count the number of reworks throughout process. This additional information could provide additional insight on the potential reduction of errors by implementing a kanban system because kanban systems have shown to eliminate production defects entirely (Younies, Barhem, & Hsu, 2007).

Conclusions

This study utilized discrete event simulation and mathematical programming to determine the optimal size of a kanban for a real world manufacturing system with variable production rates for the propose of reducing cycle time. The model used historical data from a manufacturing system—SMA weight bag production—to create a validated simulation model of the actual system. A second simulation model was constructed, based on the current system's operational flow, with integrated kanbans into the system. Mathematical programming was used to determine the optimal kanban size, based on current system performance and given parameters and constraints. Results of mathematical programming, specifically optimal kanban size and level of safety stock, were used in the kanban system model and specified as the respective resources' capacity, and the effectiveness of the kanbans based JIT system is verified.

As stated previously in the introduction section there has been limited research on determining kanban size for a production system with variable production, specifically for the manufacturing system SMA. It was unclear as to whether a kanban system would benefit such a system with variable production, but results indicated that on average, there would be a significant reduction in cycle time of nine days.

Discrete event simulation and mathematical programming were shown to be powerful tools in determining kanban size for the manufacturing process to reduce cycle time. However, there are many complex factors that influence the production scheduling; often it is not feasible to solve the optimization for the JIT kanbans system analytically, such as in this studies case. Empirical solutions based on simulation results are of particular value. Similar future studies for manufacturers can follow the same process to optimize their process flow. A model such as this will be advantageous to industry to better schedule clients, plan floor layouts, and determine different lean engineering parameters.

References

Akturk, M. S., & Erhun, F. (1999). An overview of design and operational issues of kanban systems. *International Journal of Production Research,* 37(17), 3859-3881.

Allen, M. K., & Emmelhainz, M. A. (1984), Decision support systems: an innovative aid to managers. *Journal of Business Logistics*, 5 (2), 128-143.

Anderson, D., & Roock, A. (2011). An agile revolution: why kanban is catching on in Germany and around the world. *The Journal of Information Technology Management,* 15(1), 6-11.

Ardalan, A. (1997). Analysis of local decision rules in a dual-kanban flow shop. *Decision Sciences,* 28(1), 195-211.

Balci, O. (1997). Verification, validation, and accreditation of simulation models. *Proceedings of the 1997 Winter Simulation Conference, 1997,* 135-141, Washington D.C.

Billiesbach, T. (1994). Simplified flow control using kanban signals. *Production and Inventory Management Journal,* 35(2), 72-75.

Bitran, G. & Chang, L. (1987). A mathematical programming approach to a deterministic kanban system. *Management Science,* 33(4), 427-441.

Blanchard, B. S. & Fabrycky, W. J. (2006). *Systems engineering and analysis: Fourth edition*. New Jersey: Pearson Prentice Hall.

Bowersox, D. J. and Closs, J. D. (1989), Simulation in logistics: a review of present practice and a look to the future. *Journal of Business Logistics*, 10 (1), 133-48.

Bradley, S. P., Hax, A. C., & Magnanti T. L. (1977). *Applied mathematical programming.* Minnisota: Addison-Wesley Publishing Company.

Carlson, J. G., & Yao, A. C. (1992). Mixed-model assembly simulation. *International Journal of Production Economics*, 26(1), 161-167.

Chandra, C., & Grabis, J. (2007). *Supply Chain Configuration: Concepts, Solutions, and Configurations.* United States: Springer.

Chinneck, J. W. (2001). Introduction to Linear Programming. In *Practical Optimization: A Gentle Introduction* (2). Retrieved from http://www.sce.carleton.ca/faculty/chinneck/po/Chapter2.pdf

Correia, C. (2003). From just in case to just in time (and back again?). *Accoutancy SA,* 14-15.

Deleersnyder, J., Hodgson, H., Muller-Malek, H., & O'Grady, P. (1989). Kanban controlled pull systems: An analytical approach. *Management Science,* 35(9), 1079-1091.

Di Mascolo, M., Frein, Y., & Dallery, Y. (1996). An analytical method for performance evaluation of kanban controlled production systems. *Operations Research,* 44(1), 50-64.

Esparrago, R. A.,Jr. (1988). Kanban. *Production and Inventory Management Journal, 29*(1), 6-10.

Fishman, G. (2001). *Discrete-Event Simulation: Modeling, Programming, and Analysis.* New York: Springer-Verlag.

Gupta, Y. P., & Gupta, M. C. (1989) A system dynamics model for a multi-stage, multi-line dual-card JIT-kanban system. *International Journal of Production Research*, 27(2), 309-352.

Hall, R. W. (1983). *Zero Inventories.* Homewood, IL: Dow-Jones Irwin.

Haque, B. (2003). Lean engineering in the aerospace industry. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* 217(10), 1409-1420.

Hastings, N. A. J., Marshall, P., & Willis R. J. (1982). Schedule based M.R.P.: An integrated approach to production scheduling and material requirements planning. *The Journal of the Operational Research Society,* 33(11), 1021-1029.

Hopp, W., & Spearman, M. (2001). *Factory Physics: Foundations of manufacturing management*. New York, NY: Irwin McGraw-Hill.

Howell, D. C. (2010). *Statistical Methods for Psychology: Seventh Edition.* Belmont, CA: Wadsworth Cengage Learning.

Ip, W. H., Yung, K. L., Huang, M., & Wang, D. (2002). A CONWIP model for FMS control. *Journal of Intelligent Manufacturing,13*(2), 109-117.

Jarupathiru, S., Cigane, A. P., Chotiwankaewmanee, T., & Kerdpita, C. (2009). Supply chain efficiencies through e-kanban: A case study. *International Journal of the Computer, the Internet and Management,* 17(1), 55.1-55.4.

Jeter, M. (1986). *Mathematical Programming: An Introduction to Optimization*. New York, New York: Marcel Dekker.

Karmarkar, U. S. & Kekre, S. (1989). Batching policy in kanban systems. *Journal of Manufacturing Systems.* 8(4), 317-328.

Kelton, W. D, Sadowski, R. P., & Sturrock, D. T. (2010). *Simulation with Arena: Fifth edition*. New York, NY: McGraw Hill.

Krishnamurthy, A., & Suri, R. (2006). Performance analysis of single stage kanban controlled production systems using parametric decomposition. *Queuing Systems, 54*(2), 141-162.

Law, A. M. (2006). How to build valid and credible simulation models. *Proceedings of the 2006 Winter Simulation Conference*, *2006*, 58-66. Monterey, CA.

Lee-Mortimer, A. (2008). A continuing lean journey: an electronic manufacturer's adopting of kanban. *Assembly Automation,* 28(2), 103-112.

Liker, J. K., & Womack, J. P. (2004). *Becoming Lean: Inside Stories of U.S. Manufacturers.* New York, NY: Productivity Press.

Little, J. D. C. (1961). A proof for the queuing formula: L=λW. *Operations Research,* 9(3), 383-387.

Manuj, I., Mentzer, J. T., & Bowers, M. R. (2009). Improving the rigor of discrete-event simulation in logistics and supply chain research. *International Journal of Physical Distribution & Logistics Management*, 39(3), 172 - 201

Marek, R., Elkins, D. A., & Smith, D. R. (2001). Understanding the fundamentals of kanban and CONWIP pull systems using simulation. *Proceedings of the 2001 Winter Simulation Conference, 2001,* USA, 921-929.

Mathaisel, D. (2005). A lean architecture for translating the aerospace maintenance, repair, and overhaul (MRO) enterprise. *International Journal of Productivity and Performance Management,* 54(8), 623-644.

Min, H. & Zhou, G. (2002), Supply chain modeling: past present and future. *Computers & Industrial Engineering,* 43(2), 132-50.

Moeeni, F., Sanchez, S. M., & Vakharia, A. J. (1997). A robust design methodology for kanban system design. *International Journal of Production Research*, 35(10), 2821-2838.

Panneerselvam, R. (2004). *Research Methodology.* India: Prentice-Hall.

Philipoom, P.R., Rees,L.P., Taylor,B.W. III, & Huang, P.Y. (1990) A mathematical programming approach for determining workcentre lotsizes in a just-in-time system with signal kanbans. *International Journal of Production Research*, 28(1), 1-15.

Presutti, W. D. (1988). Just-in-time manufacturing and marketing-strategic relation. *The Journal of Business & Industrial Marketing, 3*(2), 27-35.

Ramaswamy, N. R., Selladurai, V., & Gunasekaran, A. (2002). Just-in-time implementation in small and medium enterprises. *International Journal of Productivity and Performance Management, 51*(2), 85-90.

Ramnath, B., Elanchezhian, C., & Kesavan, R. (2009). Inventory optimization using kanban system: A case study. *IUP Journal of Business Strategy, 6*(2), 56-68.

Sharma, A., & Moody, P. E. (2001). *The perfect engine: How to win in the new demand economy by building order with fewer resources,* The Free Press, New York, NY.

Snyman, J. A. (2005). *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms.* New York, NY: Springer.

Spearman, M., Woodruff, D., & Hopp, W. (1990). CONWIP: a pull alternative to kanban. *International Journal of Production Research,* 28(5), 879-894.

Stewart-Marchman-Act Behavioral Healthcare. (n.d.) *About SMA*. Retrieved February 5, 2012, from *http://smabehavioral.org/about-sma.html*.

Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa S. (1977). Toyota production system and kanban system materialization of just-in-time and respect-for-human system. *International Journal of Production Research,* 15(6), 553-564.

Tavakoli, S., Mousavi, A., & Komashie, A. (2008). A generic framework for real-time discrete event simulation (DES) modeling. *Proceedings of the 2008 Winter Simulation Conference, 2008,* Austin, TX, 1931-1938.

Taylor, L. (2002). An integration analysis of material requirements planning, just-in-time, and the theory of constraints. *Academy of Strategic Management Journal*, 1(1), 109-121.

Treadwell, M., & Herrmann, J. (2005). A kanban module for simulating pull production in Arena. *Proceedings of the 2005 Winter Simulation Conference*, *2005,* Orlando, FL, 1413-1417.

Verweij, D., & Maassen, O. (2011). Kanban at an insurance company in the Netherlands. *The Journal of Information Technology Management,* 10(9), 18-22.

Wang, S., & Sarker, B. R. (2004). A single-stage supply chain system controlled by kanban under just-in-time philosophy. *The Journal of Operational Research Society,* 55(5), 485-494.

Whiteside, D. & Arbose, J. (1984). Unsnarling industrial production: Why top management is starting to care. *International Management*, 20.

Wong, C. M., & Kleiner, B. H. (2001). Fundamentals of material requirements planning. *Management Research News*, 24(3/4), 9 – 12.

Wood, S. (1990). Tacit skills, the Japanese management model and new technology. *Applied Psychology,* 39(2), 169-190.

Yang, K. K. (2000). Managing a flow line with single-kanban, dual-kanban or conwip. *Production and Operations Management, 9*(4), 349-366.

Younies, H., Barhem, B., & Hsu, C. E. (2007). A review of the adoption of just-in-time method and its effect on efficiency. *Public Administration and Management, 12*(2), 25-46.