CHALMERS

# DESIGN AND EVALUATION OF FUTURE PRODUCTION SYSTEM CONFIGURATIONS – PROVIDING DECISION-SUPPORT USING DISCRETE EVENT SIMULATION

Master's Thesis in the Master Degree Programme, Production Engineering

DAVID ALIN
ANDREAS ISAKSSON

Department of Product and Production Development
Division of Production Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2010

# DESIGN AND EVALUATION OF FUTURE PRODUCTION SYSTEM CONFIGURATIONS – PROVIDING DECISION-SUPPORT USING DISCRETE EVENT SIMULATION

Master's Thesis in the Master Degree Programme, Production Engineering

DAVID ALIN
ANDREAS ISAKSSON

Department of Product and Production Development
Division of Production Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2010

# DESIGN AND EVALUATION OF FUTURE PRODUCTION SYSTEM CONFIGURATIONS – PROVIDING DECISION-SUPPORT USING DISCRETE EVENT SIMULATION

DAVID ALIN
ANDREAS ISAKSSON



**Examiner**
Björn Johansson
Department of Product and Production Development
Division of Production Systems
Chalmers University of Technology
SE41296 Göteborg
Sweden

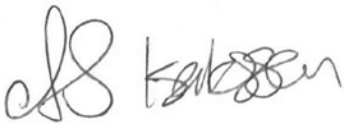
Printed by
Reproservice, Chalmers
2010

# Preface

This report is the result of our Master’s Thesis in Production Engineering. The thesis is being organized by the Department of Product and Production Development at Chalmers University of Technology and performed at an industrial production site in Sweden. We are grateful towards the company for letting us perform our thesis at their factory. This opportunity has given us a lot of insights about an industrial company and we hope that the company has gained some knowledge and inspiration out of this thesis.

We would like to thank people at the company that has helped us in varying ways during our thesis work. Especially, our supervisor at the company, production engineering project leaders for valuable support and discussions, and the persons involved in setting up and making this thesis possible.

We are also thankful towards our supervisor at Chalmers, Anders Skoogh, for help and guidance during the thesis work, and Peter Schill at PMC for initial help with the software used in this thesis.



David Alin

Andreas Isaksson

Gothenburg, June 3 2010

# Abstract

This report is the result of a Master's Thesis in Production Engineering carried out at a company providing components to the heavy vehicles industry. The financial crisis in 2008 caused order levels to drop radically, leaving the company with a huge overcapacity. This led to initial layoffs but the production system also has to be adjusted for the future, lower volumes. The aim of this work is to analyze different production system configurations, developed by the students in collaboration with the Company, and to provide decision-support on how to fulfill long-term goals of the business unit.

Discrete event simulation (DES) is used in industrial development to analyze current and future production systems and is used as the main tool. Production simulation models are developed in order to evaluate different improvement proposals. Two main questions (MQ) that this work should answer have been raised. MQ1 focuses on the flow of main components within the factory while MQ2 concerns the optimization of a machining cell where several machines has been removed in order to increase productivity and save space.

In MQ1, two future production systems are designed and analyzed: One system design based on Kanban for machining of the main component, and another based on the make-to-order principle. Since the whole factory is modeled some simplifications of the simulation models are made due to time limits and avoiding problems with a too complex simulation model. The simulation results shows that it is possible to implement the two production systems, maintain the same productivity and keep promised lead times to customers while primarily lowering work in progress drastically and creating prerequisites for increased leveled production. The recommendations for MQ1 is to transform the current production system into a make-to-order system, implement a better throughput time measuring system and keep a strong focus on changeover time reductions.

In MQ2 several production cell configurations are simulated. The results show that the output levels are similar between the models and what really differentiated them is the throughput time, both per product and per batch. Depending on how many variants that are processed in the cell simultaneously and the planned batch-size, different system configurations are superior. However, the recommendation is to only produce one variant at the time whichever layout proposal is chosen due to the throughput time per batch in that case approximately is halved.

DES is shown to be a valuable tool for evaluating and comparing different production system configurations. Modeling the product flow throughout a whole factory also shows that a simplified DES model is able to produce reliable answers.



Keywords: production simulation, discrete event simulation, production system evaluation, leveled planning, make to order, Kanban, Kaikaku

# Table of Contents

# 1 Introduction

This master's thesis is conducted at a company (from this point on referred to as "the Company") providing components to the automotive industry. The specific production facility of interest incorporates machining, heat-treatment, painting and assembly of a product that once was developed and patented by the company. Since the expiration of the main patent, competition has risen and less expensive copies, manufactured in low-cost countries are now available on the market. The Company has therefore resorted to extensive customization and focus on quality assurance to keep market shares. This has proved functional but also implied higher costs.

Due to the major breakdown of the financial system during 2008, order levels have dropped radically. The Company does not expect the volumes to ever be as high as prior to the drop but assumes the demand to be stable at the new, lower level. The production site have therefore been forced to adjust their capacity after the new prerequisites and thus a drastic reorganization has begun and increased efforts concerning process improvements has been put down.

The company has during several years been working on the implementation and auditing of a continuous improvement process aiming at reducing waste and create robust production systems. The corporation has many production sites located around the world which are following the same guidelines and auditing system. Thanks to this, benchmarking and exchange of experiences are possible and an important part of the process improvement work on a higher level.

At the plant level, Value Stream Mapping (VSM) is used to map flows, identify waste and develop alternative configurations. The VSM gives the Company an idea of the improvement potential but does rely on static calculations and differences between product variants are not considered to any large extent. No dynamic tools, such as production simulation, have been applied to the whole factory flow where product variants, batch-sizes or breakdowns are taken into account.

Parallel to the internal improvement processes the Company also cooperates with researchers and universities when it comes to the development of new improvement tools and methodologies. This master's thesis is part of such a research project aiming at facilitate modeling, simulation and evaluation of production systems.

## 1.1 Purpose

The business unit at the production site that is manufacturing these components has defined a set of goals for the production development within their production facility. Until 2012 a large improvement project will focus on:

- Increasing the productivity
- Reducing lead time
- Reduce used production space

The Company has very little experience from production simulation but is in need of more data in order to take any major decisions on how to reach these goals. This work will assist the Company in the collection of the additional data needed to take the next big step (Kaikaku) in the reorganization and capacity-reduction process.

## 1.2 Goal

The aim of the master's thesis is:

- To develop sufficiently accurate simulation models that enables evaluation and analysis of performance and improvement potential in the production system.
- To identify and measure improvement potential.
- To provide decision-support for the Company regarding the development of a more flexible and productive production process.

## 1.3 Problem definition

Due to the change in the market, the production system must be transformed into an even more flexible and productive process. Currently, the vast amount of different product variants prevents a distinct and stable production flow. Moreover, the different machining and assembly operations are separated with buffers in between, which further make the internal logistics ineffective and the process overly space consuming. The Company is interested in a tool that can be used to evaluate different logistical ideas before implementation. This tool should be applied to the whole production flow and for a more detailed evaluation of one specific machining cell where several machines are removed in order to reduce capacity and production floor usage.

Two main questions (MQ) that this work should answer have been raised and sub-questions (SQ) are used to facilitate the answering of the main question. Main question 1 deals with the main flow of products within the factory while main question 2 concerns the configuration of the more detailed modeling of the machining cell:

MQ1 What must the Company change and/or improve in order to reach the goals of the business unit?

SQ1 How much can the lead time be reduced and productivity increased by a better production flow created through reorganization and new routines on a Kaikaku level?

SQ2 What operations, resources and parameters are the most critical for increased productivity, shortened lead time and reduced usage of production space?

MQ2 How should the specific machining cell be configured in order to perform as good as possible?

SQ1 If only one product variant at a time is processed in the cell, which configuration should be used to optimize output and throughput time?
SQ2 If two product variants are processed simultaneously in partly parallel flows within the cell, how well does such a system perform compared to a single variant system?
SQ3 What are the advantages and disadvantages of the two options from a holistic point of view?

## 1.4 Project focus

**MQ1:** The Company already has a fair understanding of flows and problem areas within each specific operation but is lacking knowledge of how ineffective the logistics between different areas in the factory are and how large the improvement potential is. A mapping of throughput times and work in progress (WIP) is central in this project.

**MQ2:** There is a need of a swift answer to MQ2 which require fast evaluation of the available options. The focus is on evaluating future configurations and identifying advantages and disadvantages.

## 1.5 Factory description

The business unit manages parts of a larger factory run by the Company. Its operations are mainly located in two halls: the machining and tenifering in one hall and the painting and assembly operations are located in another part of the building (see Fig. 1). The specific machining cell that is more thoroughly analyzed is the cell machining component 1.

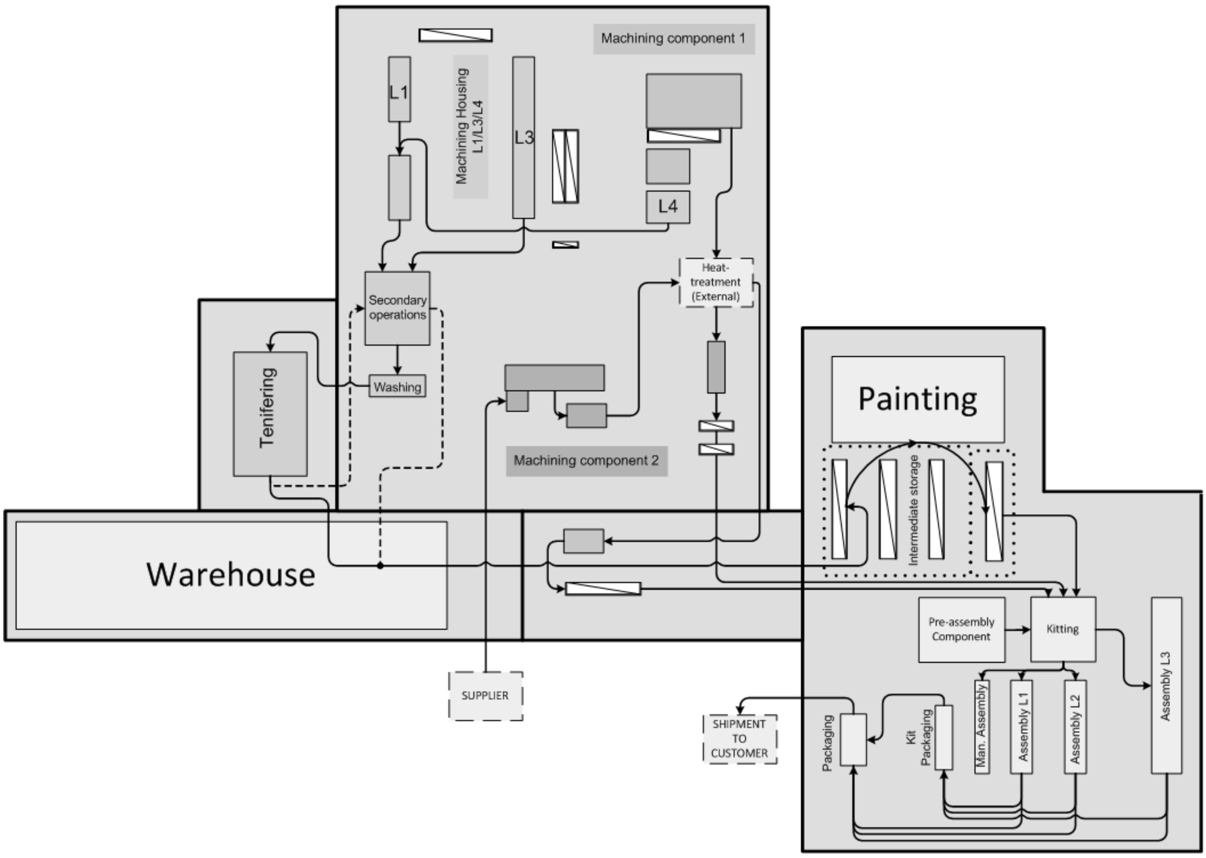



**Fig. 1: Main production flow at the Company. The top hall pictures the machining hall. The bottom hall includes painting and assembly. The products pass through one of the machining lines, secondary operations, washing, tenifering, painting, kitting and assembly.**

## 1.6 Thesis outline

The sections regarding System description and modeling, Results, Analysis and Discussion are specific for each main question. Therefore, after the joint method chapter, the two parts are described separately in order to keep a focus and line of argument. In the end of the report, the conclusions and recommendations for both questions are presented in one chapter.

# 2 Theory

This chapter introduces concepts used in various parts of this project. The arguments and theories presented are mainly based on scientific research in the areas of production and quality management.

## 2.1 Kanban

Kanban is a Japanese word for card or sign and is a scheduling system that prevents overproduction and thus restricts work in progress (WIP) (Jonsson and Mattsson, 2009). It has a planning principle that is the same as for a reorder point system and is used as an alternative idea in the design of MQ1-configurations. The system uses cards that are attached to a container or lot size. Even though physical cards are the most common, electronically Kanban cards can also be used and is preferable when distances are long, for example between a factory and its suppliers. Kanban is suitable in production environments where demand and lead time are stable. If there are a high number of variants that are ordered infrequently, generic Kanban cards can be used. These are not tied to a specific variant and initiates production from a given dispatch list (Jonsson and Mattsson, 2009). The number of Kanban cards needed is calculated from the following formula:

$$\mathrm{N} = \frac{\mathrm{D} \times \mathrm{L} \times (1 + \alpha)}{a}$$

N = number of cards needed,
D = demand per time unit,
L = lead time,
α = safety factor,
a = number of pieces of an item

Since each card represents a container of the lot-size, the WIP level is directly influenced by the number of cards being used. The number of cards should continually be updated as changes in demand occur over time. The goal is eventually to minimize the number of Kanban cards through improvements in the production system.

## 2.2 Leveled production

The first thing that should be done in companies that try to embrace lean production is to level their production (Liker, 2004). Leveled production, or Heijunka as it is called in Toyota Production system (TPS), does not produce according to the actual flow of customer orders. The rate of production is instead kept at a constant level, even if the demand fluctuates. Demand can be either in volume or product variants. Basically, customer orders over a given time are grouped and the volume and product mix are leveled so that the same volume and product mix are made each day.

The main idea of leveled production is to eliminate waste such as overproduction, waiting and inventory. According to Kasul and Motwani (1997) overproduction is the worst since it hides other wastes. One of the advantages of leveled production is that you have the right amount of inventories and staffing. High fluctuations will mean that you always have to keep resources for the highest level of production even if that is not always the case (Liker, 2004). In MQ1, the different proposals might facilitate leveled production to various extents which is an important factor for a holistic analysis.

## 2.3 Manufacturing planning strategies

Many routines are used for operational production planning and the choice of routine highly depends on the character of the demand, the product and the production system. MQ1:s primary experimental parameter is different planning strategies.

The make-to-order (MTO) principle means that products are only produced whenever a customer order exists. It is suitable for companies that have low-volume, high-variety products (Hendry, 1998) or products that are expensive to keep in stock. Make-to-order is characterized by low WIP levels since no semi-finished or finished goods storage is needed. The drawback with this strategy is that the delivery lead time is long and that order backlogs are large (Zaerpour et al., 2008). According to Arnheiter and Maleyeff (2005) adopting lean principles often reduces lead time to the extent that it is possible to introduce MTO and still keep desired delivery lead times.

Another common strategy is assemble-to-order (ATO), which is based on semi-finished items that are either manufactured or purchased into stock (Jonsson and Mattsson, 2009)**.** This strategy is common for companies with standard sub-components that can generate many product variants. Assembly of final products is only generated by customer orders. Since there is no stock of finished goods there always exists a delivery lead time. This lead time is determined by available capacity and checking material requirements. Compared to make-to-order, the delivery lead time is shorter but one of the drawbacks is that the level of WIP is higher. Short delivery lead times can be a qualifier or even an order-winner. Companies that can promise this have a clear competitive advantage but short delivery lead times and high product diversity is hard to combine (Zaerpour et al., 2008). A finished goods inventory gives low delivery lead times but is not suitable for high product diversity since the amount of finished variants is too large. This can be avoided by designing products and process which allows for easy change-to-order where finished products swiftly become customized (Liker, 2004).

## 2.4 Kaikaku

*Kaikaku* is a Japanese word meaning transformation or breakthrough improvement. It is not as well-known as K*aizen* (continuous improvements) but both have the purpose of removing waste (Arnheiter and Maleyeff, 2005). Often, Kaizen is local improvements while Kaikaku is more wide-spread through the organization.

In some situations Kaikaku is more suitable than Kaizen. It can be that a company is caught in a negative spiral and need to change fast, which makes Kaizen unsuitable since things need to happen quickly. In order for companies to stay competitive they need not only rely on continuous improvements but must also make larger improvements from time to time (McAdam, Stevenson and Armstrong, 2000). Kaizen and Kaikaku are to be jointly used in order to gain advantage. Kaikaku should be used as a “kick-start” and from that point on Kaizen should be used to continuously improve and sustain the Kaikaku implementation; else the risk is that it will fall back to the original performance (McAdam, Stevenson and Armstrong, 2000). The desired “kick-start” is evident in both MQ1 and MQ2.

## 2.5 Theory of constraints

Developed by Dr. Eli Goldratt, the Theory Of Constraints (ToC) is a method proven successful for improvement work with bottlenecks (constraints), productivity and lead times in production flows (Jonsson and Mattsson, 2009). Five steps are included (Avraham Y. Goldratt Institute, 2009):

1. Identify the constraint: Find the bottleneck of the system
2. Exploit the constraint: Increase the utilization of the bottleneck, for example by minimizing changeovers and changeover times.
3. Subordinate and synchronize everything else to the above decisions: Do not let any other resource produce at a pace higher than the pace of the bottleneck.
4. Elevate the performance of the constraint: Increase the pace of the bottleneck.
5. If in any of the above steps the constraint has shifted, go back to step 1.

By continually performing this iteration, focus is kept at the key parameter at a specific point in time, and waste-generating actions such as overproduction is disallowed. In practice, this work takes time but in a simulation study the same sequence can be utilized to generate an evaluated stepwise redesign and implementation plan for the organization. The ToC are used as guidelines when increasing volumes in MQ1.

## 2.6 Design of Experiments

Evaluation of the impact of a parameter change or different scenarios can be analyzed through Design of Experiments (DoE) (Sanchez, 2007). Applying DoE estimates both the effect of one input factor for the system response and interaction effects between several factors. According to the author, there are two especially frequent pitfalls in designing these experiments:

- Changing more than one factor simultaneously. This will result in confounding of effects and that neither factor can be identified as having more impact on the response.
- Performing one-factor-at-a-time experiments by starting at a baseline model and testing one parameter at a time. This means that interaction effects are not considered at all.

Instead, carefully chosen full or fractional factorial designs should be used. These designs consider all or a sub-set of the possible combinations of factors at different levels and enables the estimation of interaction effects (Sanchez, 2007). This leads to more efficient and thorough experiments which saves time and resources. DoE is used in MQ2 to distinguish random variation for an output from significant difference between different settings.

## 2.7 Overall Equipment Efficiency

A well-functioning improvement process includes metrics in order to measure its progress. The performance indicators used in many companies are often too carelessly chosen and does not provide an overall view on the efficiency of equipment and other resources (Schmenner and Vollman, 1994). They even argue that many measures, such as single machine utilization, often have an unproportionate impact on decisions.

Overall Equipment Efficiency (OEE) is a metric with an extensive scope of use and is deployed in industries world-wide. It joins together resource availability, equipment efficiency and quality issues (see Eq. 1) and due to this uncomplicated summarization it can be compared to other equipment or be used for cross-functional benchmarking purposes (Bamber et al., 2003). OEE is used as a key performance indicator (KPI) in MQ2. In MQ1, the level of detail is too low to perform regular calculations of OEE, however, it is used as a scaling effect on cycle times.

$$OEE = Availability\ (\%) \times Performance\ (\%) \times Quality\ (\%) \qquad \text{Eq. 1}$$

A commonly used definition of the OEE factors is the one of De Groote (1995):

$$Availability \ = \frac{Planned\ production\ time - Unplanned\ downtime}{Planned\ production\ time} \qquad \text{Eq. 2}$$

Where unplanned downtime considers unplanned maintenance, breakdowns, stops due to material shortage, changeover times etc.

$$Performance \ = \frac{Actual\ amount\ of\ production}{Planned\ amount\ of\ production} \qquad \text{Eq. 3}$$

Which shows the ratio of actual output to the possible output using the ideal cycle time and available production time.

$$Quality \ = \frac{Actual\ amount\ of\ production - not\ accepted\ amount}{Actual\ amount} \qquad \text{Eq. 4}$$

Where not accepted amount accounts for products that needs rework or repair.

## 2.8 Layout simulation

Two schools of facility layout simulation exist; the first one advocates optimization of production flows prior to simulation whilst the second argue that the opposite order of actions is more effective (Aleisa and Lin, 2005). According to the first school, when trying to improve a manufacturing facility one must not only try to optimize it on a micro-level by a simulation study. The real savings, looking at a wider perspective, can be achieved by reorganizing the whole plant (Grajo, 1995). However, in many cases total reorganization is not an option and thereby simulation can often be the feasible first step in improving existing production systems (Aleisa and Lin, 2005). These findings makes sense both in MQ1 and MQ2 since one can discuss the best way to approach such a large change in the organization.

# 3 Method

This chapter presents and motivates the choice of methods and tools used in the project.

## 3.1 Simulation of production systems

All too often managers have only good knowledge in their own part of the production system. Simulation can help forming a better overall understanding and give insights of what consequences improvements on a local level has on performance in the whole plant (Kumar and Phrommathed, 2006).

As production systems become more and more automated more focus is put on logistics to cut the total cost of a product (Castino and Watson, 1991). Simulation in logistics has become more frequent and has mainly two purposes; the arrangement of logistic systems dealing with layout planning of physical resources, and the control of logistic processes dealing with the layout of operational control (Kuhn and Schmidt, 1988). The authors also stresses that simulation is a good tool to determine the optimal production layout through minimization of transport costs.

Usage of simulation of manufacturing systems has increasingly been used in industry, especially in the last decade (O'Kane, 2003). Today discrete event simulation (DES) is used as a decision-support tool in many areas such as health care, manufacturing, logistics, and military (Fishman, 2001). DES has its advantages when studying large, complex systems that cannot be solved by other means (Banks et al., 2005) (Chaharbaghi, 1990). Analytical process evaluation alternatives cannot deal with various parameters e.g. scheduling rules and in such cases DES has a clear advantage. DES is also able to mimic dynamics present in the real production system that other tools neglect, for example varying cycle time and machine failures. Simulation makes it possible to test improvement proposals without disturbing the real manufacturing system and tests can be done in a much faster time than it would in reality. The drawbacks among others are that it needs expert knowledge in order to build a model and interpret the results in a good way (Banks et al., 2005). Building a model and maintaining it can also be quite time-consuming, especially since production systems are changing frequently.

In order to evaluate different options of implementations to reach the goals, a discrete event simulation (DES) model is created. Since the production system is very complex and large it is hard or even impossible to solve the task by other means (Banks et al., 2005) (Chaharbaghi, 1990). In this model, a variety of experiments can be performed depicting different scenarios, which then can be evaluated and compared.

In a simulation project, several steps are needed in order to create a trustworthy and useful model. Fig. 2 describes the framework that is used throughout this project in order to properly model the production facility, validate the model and finally use the model as an experimentation tool.

## 3.2 Software

A variety of simulation software is available on the market. The practical differences are generally small although different levels of object-orientation and graphical interfaces exist. Further, each software usually has plug-ins for input data importation, experimentations and exportation of simulation data.

Enterprise Dynamics is a discrete event simulation software developed by Incontrol Simulation Solutions (2010) and the application used in this project. Construction is conducted in a 2D-environment where products, machines, queues, warehouses, operators etc. are positioned. These components are called atoms and are connected to each other preferably using standard input and output channels. Several parameters can be altered for each specific atom: queue disciplines, cycle times, changeover times, breakdowns (mean time to failure/mean time to repair) and routing rules are examples of such parameters. Either a parameter is set to a standardized alternative or it can be a piece of code written by the developer. As the 2D-model develops, a 3D-model with standard elements is rendered in parallel for visualization purposes. The software also has an experiment wizard for experimental configuration and execution. Results from the simulation are summarized in reports that can be exported to spreadsheets.

The main reason for choosing Enterprise Dynamics as the simulation tool is its relatively high level of object-orientation and many standardized components that are useful in rather large simulation models. However, the ability to customize the atoms, by writing code or by programming new atoms, also makes it an efficient tool when developing more detailed models.

## 3.3 Data collection

A simulation model often requires large amounts of input data. To identify, gather, extract and filter data is a time-consuming work and it is vital to know what data is needed in order to avoid even more time lost to these activities (Bengtsson et al., 2009). A useful categorization of data for the identification and gathering process is shown in Table 1 (Robinson and Bhatia, 1995). Available data represents data that is already measured or documented. It can, however, require extensive efforts to produce useful simulation data out of available data. Category B includes parameters that are not monitored but that could be measured and analyzed. Depending on the character of missing data, a useful method for collecting the data should be chosen. Uses of stop watches, visual inspection, frequency studies or video analysis are examples of data collecting methods (Skoogh and Johansson, 2008). Category C mainly consists of parameters that belong to future systems and thus cannot yet be measured. For this category, the modeler often has to rely on estimates of parameter values.

**Table 1: Categorization of data**

| | |
|---|---|
| Category A | Available data |
| Category B | Not available but collectable data |
| Category C | Not available, nor collectable data |

A vast amount of simulation input data is gathered to be able to create true flows of more than 600 product variants. For the model depicting the whole factory, historical machining orders and customer (painting & assembly) orders is extracted from the Company's Enterprise Resource Planning (ERP) system. These orders stretch as far as two years back and include article number, raw material number, reported start and finish date, line where the batch has been machined, product family and batch-size.

Similarly, extensive cross lists are extracted regarding all active article numbers and their specific lists of secondary operations.

Cycle times and changeover times are provided by the production engineering office and work schedules by team-leaders at the different lines and operations. Other, more trivial information such as the number of pallets per line or the number of forklifts is gathered by interviews and walkthroughs with operators, team-leaders and production engineering personnel.

Enterprise Dynamics own application AutoFit is used in those cases statistical distributions are created from raw data. For example, this is used in MQ2 where MTTF/MTTR statistical distributions are generated from raw break-down data.

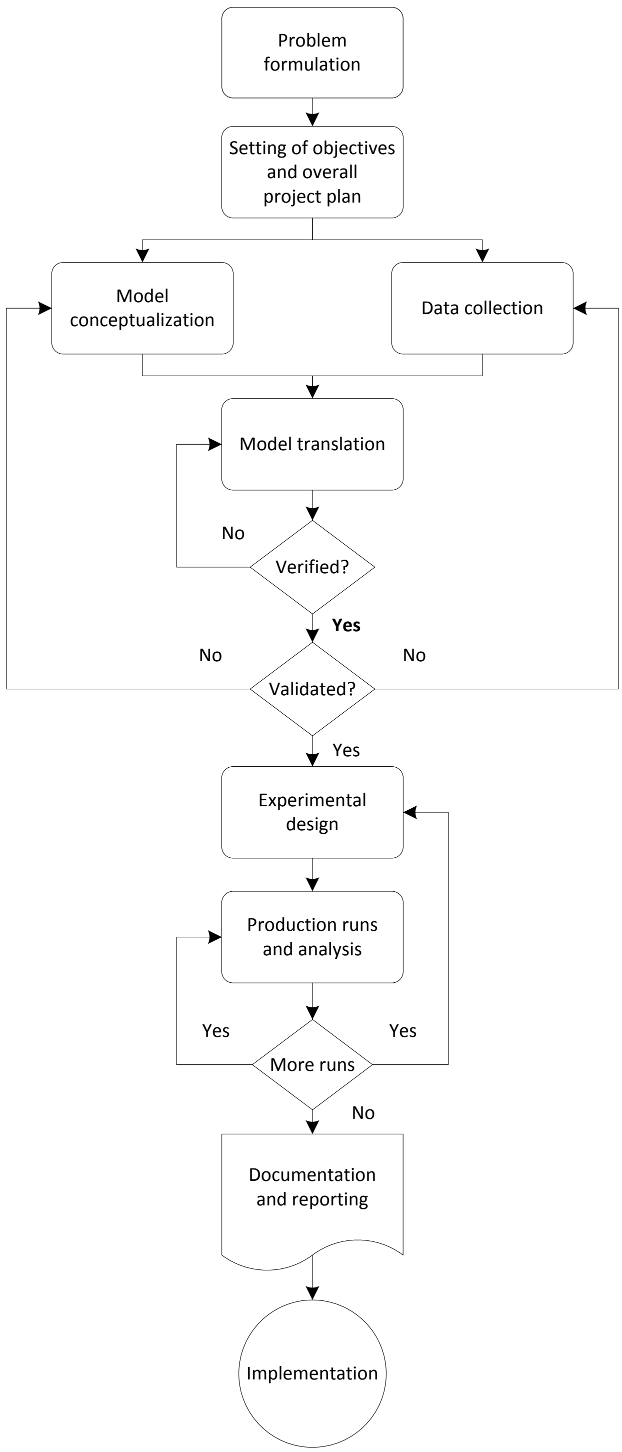



**Fig. 2: Steps in a simulation study according to Banks.**

## 3.4 Model conceptualization

Prior to model building, a conceptual model is developed to document routing rules, work flows, restrictions and interactions within the factory. The conceptual model consists of an overall layout and flow description which is complemented with separate, more detailed explaining texts for each operation.

## 3.5 Simplification of simulation models

Although a lot has been written in research about the importance of simple DES-models (Chwif, Ribeiro Pereira Barretto and Paul, 2000), simplification of simulation models is not often performed (Brooks and Tobias, 2000). Brooks and Tobias (2000) claims that there is a general disbelief that a more complex model will give more accurate results. Simulation models can often be simplified to a great deal without changing the output behavior. On the other hand, if a model is becoming too simple it will lose credibility and the results will be inaccurate. Simplification of models can be done in different ways. Examples include replacing several machines by one, deleting unimportant parts of the model, replacing parts of the model with a random variable etc.

Fig. 3 shows an example provided by Brooks and Tobias (2000) where a line of several machines are combined into a single resource. In case breakdowns are not considered, M3 is constraining the line. The total capacity of the line is 15 (12 in buffers plus 3*1 machine capacity). In order to achieve a correct throughput time M4 needs to be delayed with 7 seconds.

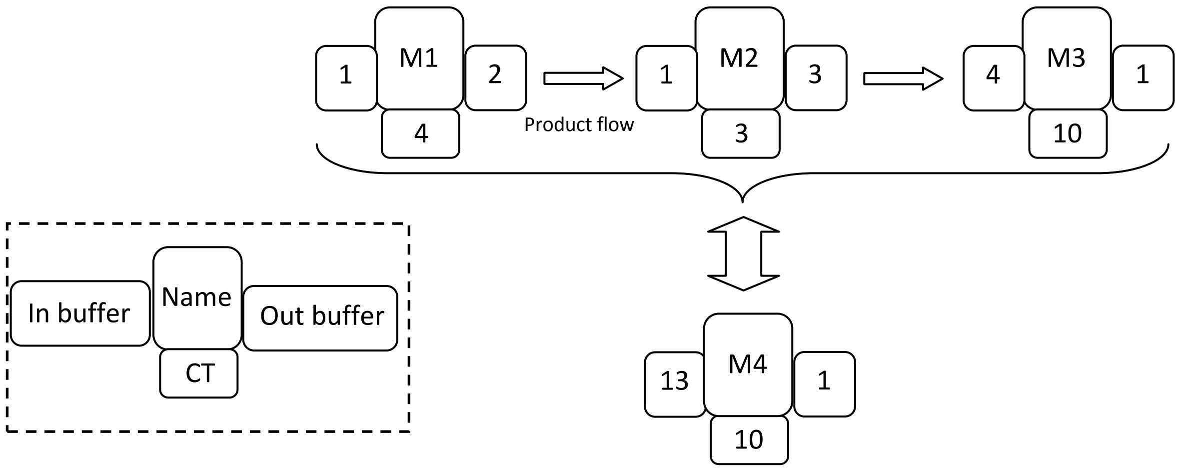



**Fig. 3: Simplification of a line into a single resource**

When simplifying a model it must be done in a proper way. One of many errors that can occur is lot passing, which happens if a part of the model is replaced by a time distribution instead of a constant delay (Rose, 2008). Then a product that is first in a flow can be passed by others that get a shorter time delay and this may not correspond with the real system behavior.

Due to the magnitude of MQ1, the level of detail is kept rather low in that simulation model. For example, individual stations in machining or assembly lines are not modeled separately. Instead, the line is modeled as a block. The MQ2-system is smaller and requires less simplification.

## 3.6 Simulation period

A simulation model can be used to run simulations depicting long periods of time. The simulation period is chosen using a few criteria:

- It must be long enough to be able to show sufficient dynamic behavior for the specific system.
- It must be possible to gather sufficient production data for validation and experimental purposes.
- For modeling purposes, it is preferable if the period does not contain too many differences over time such as holidays, shifts in production scheduling, drastic changes in demand etc.

## 3.7 Verification

In order to avoid verifying the whole model in the end which would be troublesome, the simulations models are firstly built in small sub-systems. This way of working is recommended by Sánchez (2007). These sub-systems are continuously incorporated into the complete systems as soon as they are verified. Sargent (2005) presents a variety of steps useful when verifying a simulation model. Some examples are:

- Studying the animation and control that routings for different variants are the right ones, time schedules works as intended etc.
- Discussing the model behavior with personnel having knowledge of the system.
- Checking that variables as output and throughput time through parts of the model is reasonable.

## 3.8 Validation

The two main questions are quite different which leads to different possibilities in the validation process. A current-state model is developed in MQ1 which enables more validation to be conducted compared.

### *3.8.1* Validation MQ1-model

Validation of the model used to answer MQ1 is performed in three steps:

1. Face validation – By exposing the model for critical review by knowledgeable personnel one get a measure on the reasonableness of the model and deficiencies can be identified (Banks et al., 2005). Banks (2005) further argue that this involvement increases the credibility of the model and facilitates the acceptance of simulation results by managers. Moreover, Sargent (2005) argue that this method gives a subjective result and that it is favorable if the reviewer neither is part of the simulation development team nor is the model sponsor.
2. Turing test – This test is conducted by presenting output data for knowledgeable persons from a real system and from a model of that system in exactly the same format (Banks et al., 2005). If the person cannot distinguish between the different data sets it indicates that the model shows similar dynamics as the real system and thereby to an extent is valid.
3. Statistical input-output validation – When using input data to a simulation model that is identical to the input to the real system during the same conditions, statistics can be used to produce an objective validation result. By conducting a t test of the null hypothesis on the mean difference between system and model output an inadequate model can be identified (Banks et al., 2005). The following example shows how the test is performed:
   The null hypothesis test is set up as:
   $$H_0: \mu_d = 0 \qquad \text{Eq. 5}$$

The alternative hypothesis of no mean difference:

$$H_1: \mu_d \neq 0 \quad \text{Eq. 6}$$

Table 2 shows an example of K runs where $Z_{ij}$ represents the system output corresponding to the j:th input data set. Similarly, $W_{ij}$ represents the model output corresponding to the j:th input data set.

**Table 2: Data needed for the validation process.**

| **Input data Set** | **System output $Z_{ij}$** | **Model output $W_{ij}$** | **Observed difference $d_j$** | **Squared deviation from mean $(d_j\text{-}d_{mean})^2$** |
|---|---|---|---|---|
| 1 | $Z_{i1}$ | $W_{i1}$ | $d_1 = Z_{i1} - W_{i1}$ | $(d_1 - d_{mean})^2$ |
| 2 | $Z_{i2}$ | $W_{i2}$ | $d_2 = Z_{i2} - W_{i2}$ | $(d_2 - d_{mean})^2$ |
| . | . | . | . | . |
| . | . | . | . | . |
| K | $Z_{iK}$ | $W_{iK}$ | $d_K = Z_{iK} - W_{iK}$ | $(d_K - d_{mean})^2$ |

The sample mean difference is calculated as:

$$\overline{d} = \frac{1}{K}\sum_{j=1}^{K} d_j \quad \text{Eq. 7}$$

Using $\overline{d}$, the sample variance is calculated as:

$$S_d^2 = \frac{1}{K-1}\sum_{j=1}^{K} (d_j - \overline{d})^2 \quad \text{Eq. 8}$$

The appropriate statistic to use is the t-statistic:

$$t_0 = \frac{\overline{d} - \mu_d}{S_d/\sqrt{K}} \quad \text{Eq. 9}$$

Assuming a level of significance of $\alpha$ and K-1 degrees of freedom, if $|t_0| < t_{\alpha/2,K-1}$ (found in a t-distribution table) the null hypothesis cannot be rejected. This implies that the model cannot be considered inadequate on the basis of this test and thus it can be considered valid. On the other hand, if $|t_0| > t_{\alpha/2,K-1}$, there is evidently a significant difference between model and system. In this case, the modeler should try to find deficiencies that cause these differences and adjust the model.

### *3.8.2* Validation MQ2-model

Face validation (see section 3.8.1) is used as the main validation method for the model in MQ2. No input/output validation is possible since only future models are modeled. The current system is very complex with a lot of routings and due to the time shortage (see section 1.4) it is chosen not to model the current system.

# 4 System description and modeling MQ1

In this section, a brief description of the current production system and some of its most important characteristics is presented. This is followed by a description of the simulation model developed to mimic the real system.

## 4.1 Current flow

The main production flow at the Company is constituted of machining, heat-treatment and painting of casting housings followed by assembly.

3 additional components of the final product are manufactured in-house in separate production lines. These flows converge with the main flow in a kitting operation prior to assembly (see Fig. 1).

There are three machining lines: Line 1 and Line 3, which are series of numerical controlled (NC) machines connected by conveyers while Line 4 consists of one single NC-machine served by an industrial robot. The production volume is divided into two main product families and the first two lines are each specialized in one family each. Line 4 is more flexible and can machine housings from both families without any major changeovers.

Due to different customer needs a large variety of secondary operations can be performed after the primary machining. The housings from Line 1 and Line 4 must at least pass through the D-hole and Grease-hole operation while these operations are incorporated in Line 3.

The housings are heat-treated through a tenifering process before they reach a storage for semi-finished products. The next operation is the painting which also is the customer order point.

In the machining operations (machining/secondary operations/tenifering), production is triggered by the release of manufacturing orders which are based on preliminary customer orders and forecasts. These orders aim at maintaining a sufficient inventory of semi-finished products at the customer order point (i.e. after the tenifering). From this point, all orders are customer specific. The manufacturing orders have a minimum quantity of 16 and a maximum quantity of 152. Customer orders have no restrictions in quantity but are generally smaller than the manufacturing orders.

A customer order triggers painting which in turn triggers the pre-assembly operation connected to the specific customer order. The customer order can, if material is available and the painting has spare capacity, be started as early as six days in advance. When the housing is painted and the pre-assembly is finished, the order is kitted before introduced into one of three assembly lines. A multitude of restrictions are in place regarding the use of the lines. Somewhat simplified: Assembly line 1 and assembly line 2 are rather modern and both families are assembled in assembly line 1. Assembly line 2 and the older assembly line 3 only assemble one product family each.

Subsequent to the assembly lines, the housings are packed and shipped.

## 4.2 Product variants

The Company has a long history of customization of their products in order to respond to the needs of many different customers. This has led to an increasing number of product variants. During the autumn of 2009, the Company used more than 130 unique housing raw materials, more than 300 unique semi-finished articles were manufactured and more than 600 unique articles were assembled to customers. The production system has continually gone through updates in order to keep up with the increasing variant flora without experiencing too many losses to changeovers and capacity restrictions.

## 4.3 Production planning

The current production planning estimates throughput times of less than two days in machining and less than two days in assembly. The lion's share of the production volume (80-20 rule applies) is destined for a few large customers who benefit of a promised lead time of three days and thus makes the production planning in machining dependent on the preliminary orders and forecasts. No detailed planning on an operation level exists in machining which causes orders to overtake each other and arrive to the semi-finished product storage in an incorrect order. In painting and assembly, the Company uses the assemble-to-order planning principle (see section 2.3) but is frequently experiencing a lack of correct semi-finished products due to the sequencing in machining. Moreover, due to the three days promised lead time, customer orders can be cancelled after the point where the corresponding manufacturing order has been started. This leads to overproduction and increased inventory levels of semi-finished products.

## 4.4 Input data

Table 3 shows the data used as input to the simulation models, how they are gathered and its level of detail.

**Table 3: Parameters used as input data.**

| Parameter | Data source | Level of detail |
|---|---|---|
| **Cycle times – Machining lines, D-hole, Grease-hole, pre-assembly and assembly lines** | Cycle time mapping documents | Static in seconds for each machine. |
| **Cycle times – Secondary operations, tenifering, painting and kitting** | Interview | In seconds. Triangular distribution where variation exists. |
| **Forklift speed** | Interview | Static meter/seconds. |
| **OEE – Machining lines and assembly lines** | Database | Manual logging of changeover times, downtime, re-work and scrapping combined with production schedules. OEE value per line and week is transformed into distributions. |
| **OEE – D-hole and Grease-hole** | Interview | Estimations by production engineering personnel |
| **Number of pallets in lines** | Visual inspection | Count |
| **Changeover times – Pre-assembly and secondary operations** | Interview | Static in seconds. |
| **Historical orders** | Database | Actual manufacturing and customer orders from Aug-Dec 2009. |
| **Product variant details** | Database | Complete Bill-of-Operations for each product variant. |
| **Work-shifts, breaks etc.** | Production schedules | Actual shifts and breaks are used. |

## 4.5 Assumptions

In the MQ1-model, an order started at a specific day in the real system starts at the corresponding day in the model. To get accurate transportation times between operations within the factory 2D CAD –drawings are used to position storages and operations according to reality. The simulation model takes into consideration, for every specific variant, its routing and cycle time in each operation.

To restrict the size and complexity of the model of the whole factory, some assumptions and simplifications have been made.

The number of products present in the simulation model simultaneously highly affects the simulation speed of the model. More products demand more processing power and thus the speed is reduced. Therefore, instead of using individual products as the smallest unit, a product atom is considered as one batch of products and it carries information on batch-size, article number etc. which is necessary for the flow control of products. Since products are not sent from one operation to the subsequent operation until all products are finished, this simplification does not have any effect on the flow dynamics.

However, the internal flow through primarily machining and assembly lines must be rectified in order to work accordingly to the real system. Due to the batching simplification, the desired aggregation level stated by the production engineering office and inconsistent reporting of breakdowns for several machines within machining and assembly line operations, each operation is modeled as one or two atoms (see section 3.2). The first atom serves as the restricting bottleneck machine for the operation and the second imitates the accumulated throughput time for the operation. Only one batch at the time can occupy the bottleneck atom while the throughput atom acts as a multi-server where multiple batches can be located with individual arrival and exit points. Similar procedures are described in literature by Brooks and Tobias (2000). The batch-size determines the time in the bottleneck while the throughput time depends on the number of product carriers in each operation. If an operation only consists of one machine the throughput time atom is omitted.

Furthermore, no breakdowns, slack or quality deficiencies are modeled in detailed. To compensate for this an OEE-factor is introduced in order to prolong the cycle time in the bottleneck atom. The Company provides historical OEE-data for each operation which is condensed into statistical distributions. In order to increase the level of dynamics, the cycle time is adjusted for every single batch. Definitions of bottleneck times and throughput times are shown in Eq. 10 and Eq. 11 respectively.

$$Time_{Bottleneck,i} = Batch - size_i \times \frac{Cycle\ time_{Bottleneck}}{OEE_i} \quad i = Batch - number \qquad \text{Eq. 10}$$

$$Time_{Throughput} = Number\ of\ pallets_{Operation} \times Cycle\ time_{Bottleneck} \qquad \text{Eq. 11}$$

## 4.6 Simulation period

Following the criteria in section 3.4 the chosen period is 20 weeks during late 2009. During this period the Company had regained parts of the production volume that was lost during the financial crisis and had a stable, slowly increasing demand. The period allows for in total approximately 100 daily and 20 weekly readings per simulation run. No larger holidays occurred during this period nor did any changes in the time schedules such as additional work shifts. Further, variations in volumes or the proportion of different product families can be altered using the order lists from this period as a base.

## 4.7 Simulation model

The finished model is shown in Fig. 4. It shows the machining operations at the top and the assembly operations at the bottom. The painting operation and its storage of semi-finished products is seen in the centre.

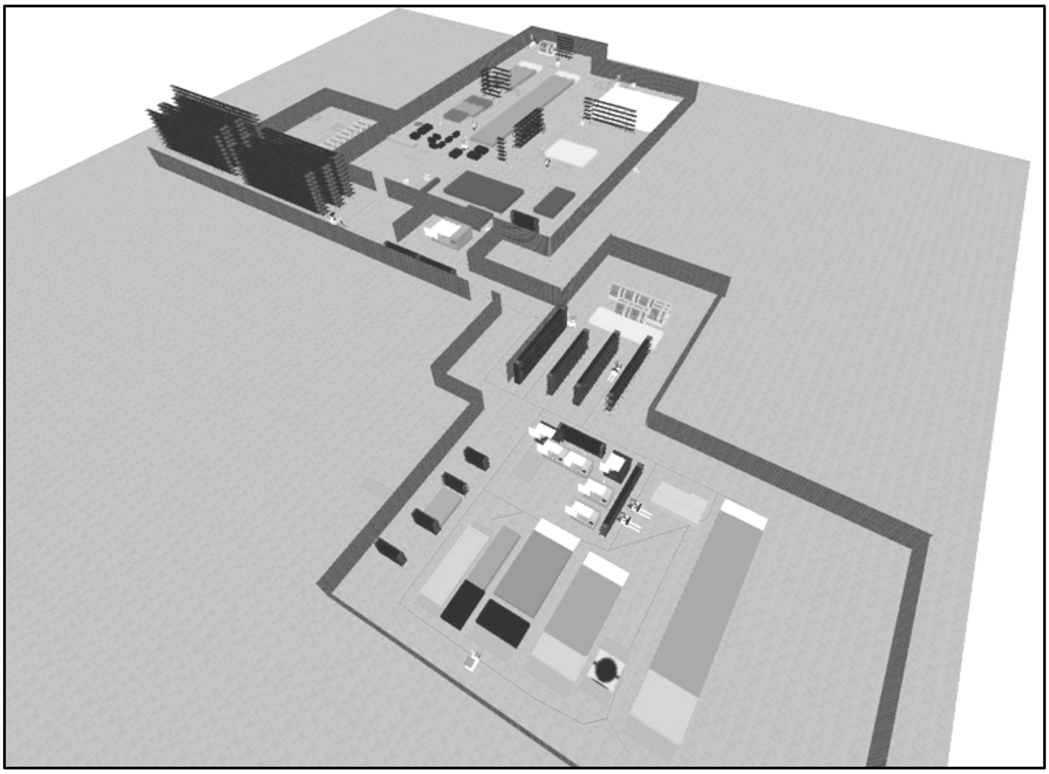


**Fig. 4: 3D-overview of the simulation model. The machining lines are located at the top while the assembly operations are located in the bottom of the picture.**

# 5 Results MQ1

This chapter presents the results from the validation process and experimental execution.

## 5.1 Validation results

Prior to experimentation, a thorough validation phase is vital in order to ensure a sufficiently valid simulation model. Face validation, Turing test and input/output validation is used for this purpose.

### *5.1.1* Face validation

By conducting virtual walkthroughs of the simulation model, explaining flows and presenting input and output data for production engineering personnel, details are introduced, removed or tweaked. By iteration of these steps the face validity of the model and the resemblance to the real system is increasing.

### *5.1.2* Turing test

A Turing test is prepared and conducted first with the Senior Specialist of Operations and secondly with the Master Planner at the Company.

The test consists of three parts:

1. **Daily output for one week.** Eight graphs from randomly chosen weeks during the simulation period are presented whereof three are taken from the real production system while the remaining five are produced by the model. The detailed notion of both test persons is that the real system usually has large outputs on Tuesdays and Fridays and they base their choices on this. The result is that they manage to distinguish one out of the three correct graphs each, however not the same (Number 2 and 5 respectively, see Fig. 5).

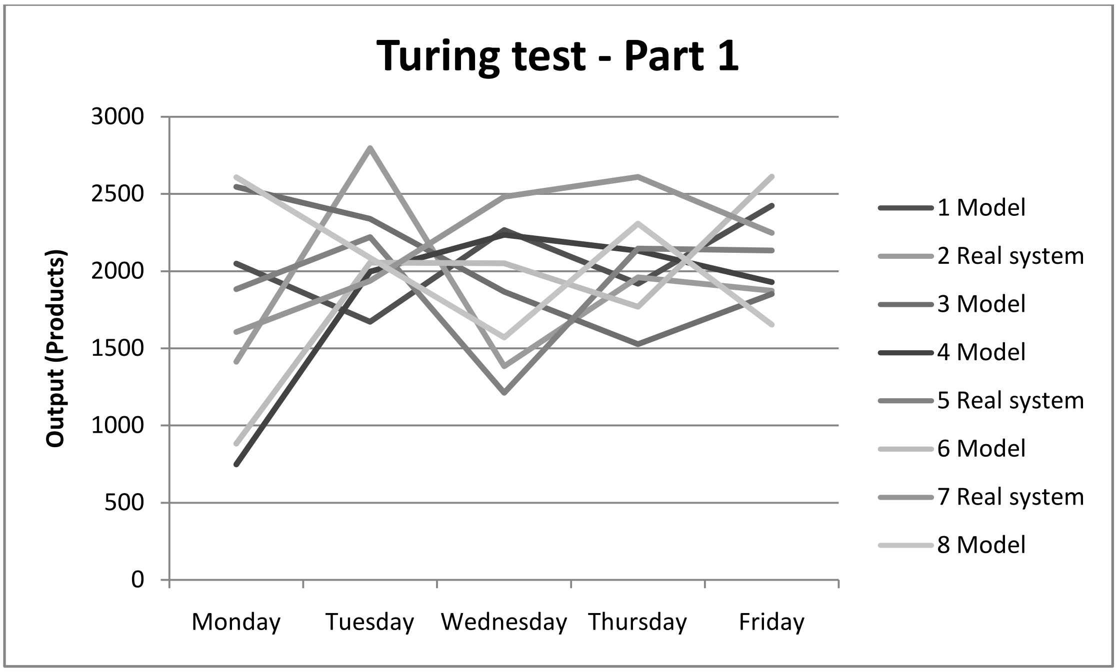



**Fig. 5: Turing test - Part 1: Output per day from different datasets.**

2. **Daily output during the whole simulation period**. Four data sets of approximately 100 data points are presented. One of the sets is the historical daily output from the system while three are produced by the model. By finding single extreme data points both experts chose the correct data set. Table 4 shows an extract of the data set where the output have fallen drastically on a Thursday, probably due to a breakdown somewhere in the system.

**Table 4: Data set showing output per day for a week.**

| | **Model** | **Model** | **Real system** | **Model** |
|---|---|---|---|---|
| **Monday** | 1102 | 1190 | 2401 | 1889 |
| **Tuesday** | 1627 | 2002 | 2125 | 1866 |
| **Wednesday** | 2028 | 2331 | 1607 | 2211 |
| **Thursday** | 2020 | 2325 | **919** | 1829 |
| **Friday** | 1668 | 1786 | 2294 | 2529 |

3. **Daily output during the whole simulation period.** Four graphs visualizing the same points that was presented in part 2. The three simulated output graphs show significantly less variance and thereby one of the experts directly pinpoints the correct graph. The second expert is reluctant to accept that the real system has a higher variance and chooses a graph produced by the model.

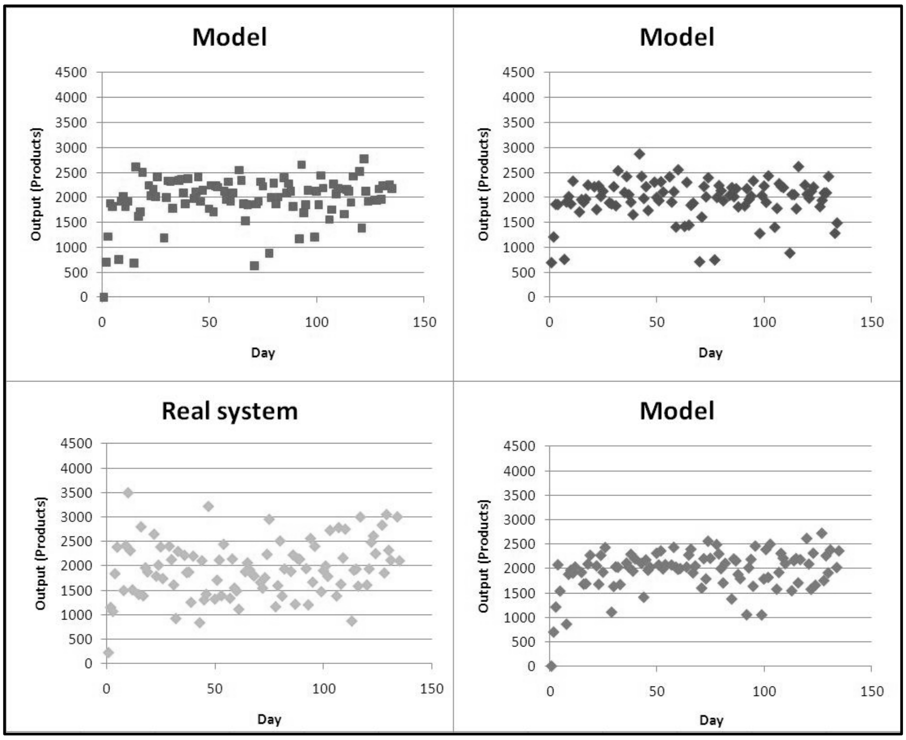



**Fig. 6: Turing test - Part 2: Output per day during the whole simulation period.**

The Turing test result shows that at a random week basis, experts have problems distinguishing model output from system output. When exact numbers are presented it does seem easier to find anomalies that reveal the real system and when looking at graphs over the whole period it clearly shows that the model output has a smaller dispersion. The reason for the decreased variance is most certainly the choice not to model breakdowns which would impose more stress on the system.

### *5.1.3* Input/Output validation:

Using the 20 weekly output values from system and model with identical input data (orders/capacity), a t-test of the null hypothesis of no mean difference is set up:

$$H_0: \mu_d = 0$$

The alternative hypothesis is the one of significant difference:

$$H_1: \mu_d \neq 0$$

Observed difference and the squared deviation from mean are computed in Table 5.

**Table 5: Comparison between system and model output.**

| Input data Week | System output $Z_{ij}$ | Model output $W_{ij}$ | Observed difference $d_j$ | Squared deviation from mean $(d_j-d_{mean})^2$ |
|---|---|---|---|---|
| 1 | 6732 | 5601 | 1131 | 962410 |
| 2 | 11214 | 8625 | 2589 | 5950672 |
| 3 | 9425 | 9621 | -196 | 119785 |
| 4 | 10564 | 10323 | 241 | 8308 |
| 5 | 9346 | 9472 | -126 | 76162 |
| 6 | 9391 | 10740 | -1349 | 2246551 |
| 7 | 8870 | 10447 | -1577 | 2982011 |
| 8 | 8966 | 10406 | -1440 | 2528418 |
| 9 | 7611 | 10468 | -2857 | 9040395 |
| 10 | 9481 | 9798 | -317 | 217482 |
| 11 | 10140 | 9314 | 826 | 457179 |
| 12 | 8576 | 10037 | -1461 | 2592825 |
| 13 | 9594 | 9701 | -107 | 65844 |
| 14 | 9760 | 9792 | -32 | 32843 |
| 15 | 9871 | 9662 | 209 | 3543 |
| 16 | 10690 | 10025 | 665 | 265637 |
| 17 | 11101 | 9500 | 1601 | 2106562 |
| 18 | 10881 | 10273 | 608 | 209787 |
| 19 | 12162 | 10233 | 1929 | 3164930 |
| 20 | 6586 | 3928 | 2658 | 6289562 |

The sample mean difference $\overline{d}$ and its variance $S_d^2$ are calculated according to Eq. 7 and Eq. 8 respectively:

$$\overline{d} = \frac{1}{K}\sum_{j=1}^{K} d_j = 149.7$$

$$S_d^2 = \frac{1}{K-1}\sum\nolimits_{j=1}^{K}(d_j - \overline{d})^2 = 2069521.4$$

The statistic t can be computed as shown in Eq. 9.

$$t_0 = \frac{\overline{d} - \mu_d}{S_d/\sqrt{K}} = \frac{149.7 - 0}{1438.6/\sqrt{20}} = 0.465$$

With the significance level $\alpha = 0.05$ and $K - 1 = 19$ degrees of freedom, the critical value is found in a t-distribution table:

$$t_{\alpha/2,K-1} = t_{0.025,4} = 2.09$$

Since $|t_0| = 0.465 < 2.09$, the null hypothesis cannot be rejected and no significant difference between system and model output is detected.

By performing these tests it is shown that the model to a large extent performs and behaves similar to the real production system and should be adequate to use for future experimentation.

## 5.2 Experimental design

Since the aim of this work is to see how the rather ambitious goals of the business unit can be reached (see section 1.1), large reorganizations is called for. In order to identify and measure some of the improvement potential in the system two modified models are developed: One Kanban model based on the Company's VSM:s and one Make To Order model designed by the modelers. The validated model is from this point on named the Base model. During the first quarter of 2010, production volumes have continued to rise and the weekly output has increased by approximately 65% compared to the simulation period. All three models are therefore, in addition, transformed into high volume (HV) models to be able to stress-test the different ideas.

The Kanban model (based on the Company's VSM:s) differs from the Base model in these areas:

- The first modification concerns the removal of the storage for semi-finished products prior to the painting operation.
- The First-In-First-Out (FIFO) principle applies for batches coming from the tenifering into painting.
- The product variant flora is reduced in the machining lines by grouping together certain article numbers and relocation of customization operation to the secondary operations. Mainly, the D-hole operation is separated from machining line 3 and the Grease-hole operation is omitted in machining line 3. The D-hole machine from line 3 joins the existing D-hole machine and a spare Grease-hole machine is taken into use due to capacity reasons.
- Two smaller storages are introduced after the machining lines, one for each product family. The storage is refilled by the machining lines using Kanban cards with fixed batch-sizes for the most frequent variants. The number of Kanban cards is chosen according to the theory in section 2.1. There are a total of ten generic Kanban cards for each product family that is used for less frequent variants. A generic Kanban only triggers production of exactly the number of products that is needed for a customer order.
- The customer order point is changed from the painting to after the machining lines, however, the exact same customer orders as in the Base model is used. If material is available in the Kanban storage, the customer order starts, material is withdrawn from the storage, and the order moves directly into secondary operations. If there is insufficient material, machining of a new pallet of products is triggered. A customer order that triggers a

generic Kanban is also considered started and need to wait for the material to pass through the machining lines.

- Since no historical data can be used for deciding which machining line a batch passes through, a new routing principle is used.
- The old, assembly line 3 is removed and instead, an additional shift is introduced for assembly line 1 and assembly line 2.
- Assembly line 2 can assemble both product families. This improvement project is already initiated.

The Make-To-Order model (designed by the modelers) differs from the Base model in these areas:

- The storage for semi-finished products prior to the painting operation is removed.
- The customer order point is changed so that customer orders start when they enter the machining line and thus transforms it into a make-to-order system (see section 2.3). The manufacturing order for the machining operation is abolished.
- Since no historical data can be used for deciding which machining line a batch passes through, a new routing principle is used.
- The FIFO-principle applies for batches coming from the tenifering into painting.
- Assembly line 3 is removed and instead, an additional shift is introduced for assembly line 1 and assembly line 2.
- Assembly line 2 can assemble both product families.

When modifying the models into HV-models the Base model acts as the template for needed adjustments and already implemented improvements. The adjustments needed to get the Base model to produce 165% (with the assistance of Theory of Constraints, see section 2.5) of the initial volume are then, where applicable, transferred onto the other two models to achieve a Kanban HV-model and a Make To Order HV-model. Modifications that are implemented are not analyzed in detail but are considered necessary to achieve equal behavior to the real system with higher volumes. The implementation of some of them is already in place in the real system.

The modifications that are made to the Base model are:

- Batch-sizes in machining and for customer orders are scaled by a factor 1.65.
- The cycle time in painting is reduced to 10 seconds. It was previously 13 seconds.
- The cycle times correctional OEE-factor is removed in the machining lines, in secondary operations and in the assembly lines.
- The cycle times in assembly line 1 and assembly line 2 is updated to 22.9 seconds instead of 25.5 seconds and 28.5 seconds depending on product family.
- An extra shift is introduced in assembly line 1 (not applicable in Kanban and Make-To-Order models).

### *5.2.1* Definition of measurements

- **Output** – The total amount of products being produced during a period. Output per week and total output during the simulation period is being measured. Since there only exists a certain amount of orders there exists a maximum output level.
- **Product throughput time** – Available production time that a product is present in the system (breaks, nights and weekends are **excluded**). It is divided into three measurement points:

1. The first one is the throughput time in the machining and tenifering processes. In the Base and the Make To Order model the time starts when a batch enters one of the three machining lines. In the Kanban model the time starts when the customer order is started. The throughput time ends when the batch ends up in the painting storage (FIFO-buffer in Kanban and Make To Order models).
2. The second throughput time is through the painting and assembly process, starting when an order is started in painting and ends when the order is sent to shipping.
3. The third throughput time is the total time a product spends in the whole production system. This includes the throughput time in machining and tenifering, the time it spends in the semi-finished storage, and the time in painting and assembly. Material in the semi-finished storage can consist of material with the same article number that has been in production system for different times. To deal with the fact that a customer order receives material with different throughput times in machining the throughput time for an order is weighted according to the product quantities from the different manufacturing orders. For example if half of the semi-finished products for a customer order has a throughput time of 10 hours and the other half a throughput time of 20 hours the weighted throughput time for that order is 15 hours.

- **Customer order throughput time** – The total time in days, **including** not only production time but also breaks, nights and weekends, from customer order start until shipping.
- **Work in progress** – The work in progress (WIP) is defined as the total number of products that are in the production system simultaneously. It is calculated by adding together the size of all batches present in the production system.
- **Free time** – In the simulation models there is a certain amount of orders available to start. Therefore it is not possible to stress-test the model and see what it can produce under a given time. Instead, improved performance is measured as increased free time. Free time is measured as the percentage of available production time that a specific resource is idle. A resource is idle when it is available (not down) and is being starved. The reason for using free time instead of utilization is that it shows true performance improvement. Decreased utilization in a resource indicates *either* that the usage of the resource is less *or* that it is being blocked a larger portion of the time.

## 5.3 Simulation results

In this chapter results from the simulation runs are presented and explained. The models are compared with respect to each of the different performance indicators.

### *5.3.1* Output

The total output from the Base model in 20 weeks is 1% less than from the real production system. Since the simulation models releases production orders according to the real historical production schedule the output was expected to be very close (see Fig. 7 and Fig. 8). The other two models are also capable of producing the amount the real system produces. The simulation models performance are not as accurate when looking per week, there is an average deviation of 11% in output per week. Due to that the start of the simulation period occurs directly after the summer holiday, no products are initially present in the system and thereby reduces the output for the first week. Similarly, the last week before Christmas only includes two days which explains the low output for that week.

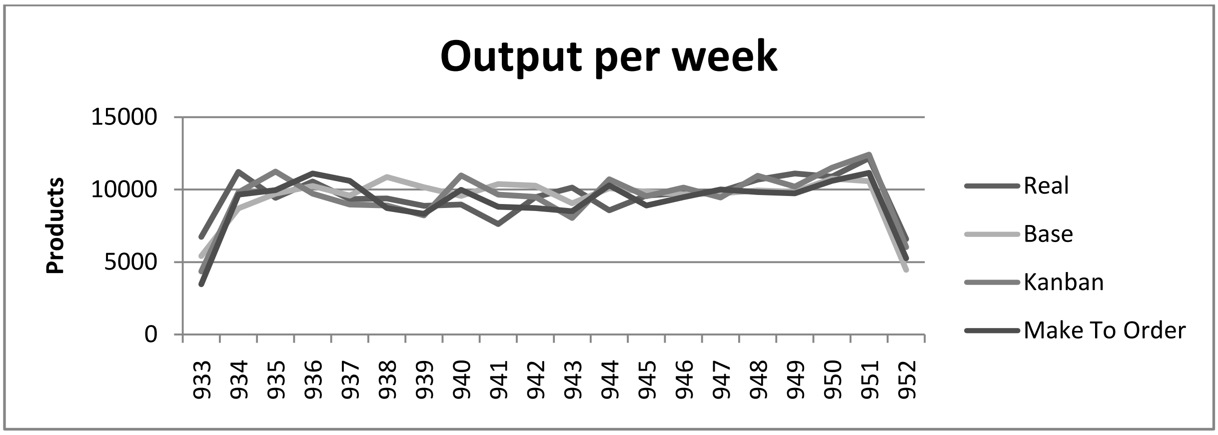



**Fig. 7: Production output per week.**

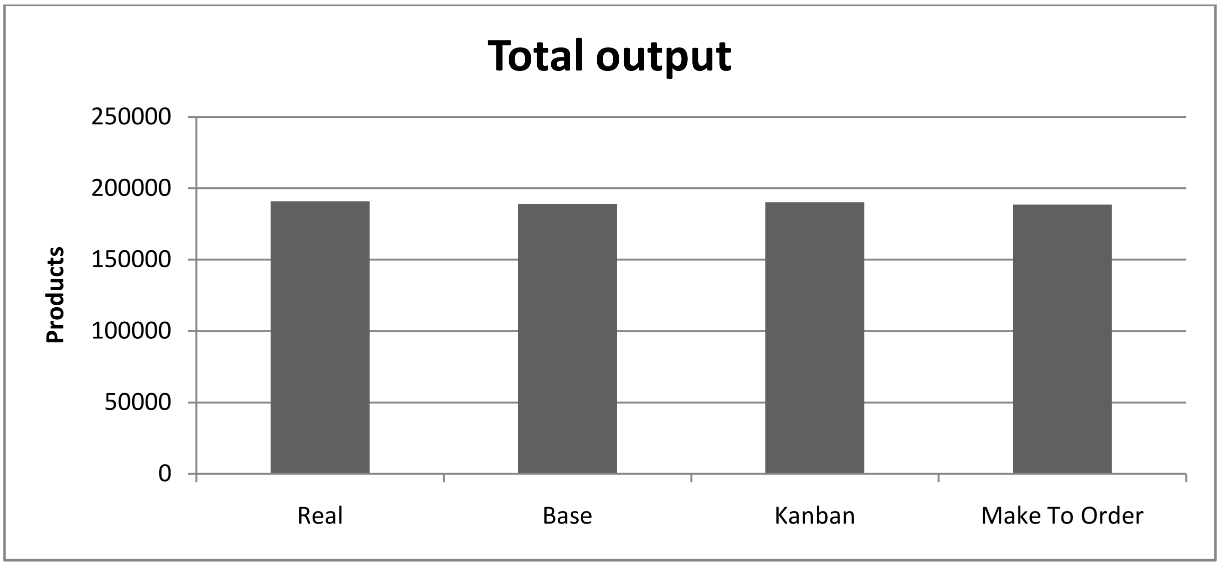



**Fig. 8: Total production output during the simulation period**

### *5.3.2* Product throughput time

Table 6 shows the averages of three throughput times. In the Base model there are common for low-volume products to stay in the semi-finished storage for a long time due to the mismatch between manufacturing orders and customer orders. Therefore, the average production throughput time is quite long. The fact that the Kanban and Make To Order models have eliminated the semi-finished storage reduces the production throughput time significantly.

In the Make To Order and Kanban model, the storage between tenifering and painting is removed. Since batches no longer freely can exit the tenifering process but have to wait until capacity is available in the painting, the throughput time in machining is increasing. The exclusion of the D-hole operation from machining line 3 further increases the throughput time for the Kanban model since many batches now need to pass through another queue and an additional operation.

In assembly, the average throughput time is longer for the Base model since batches that are painted during the evening shift have to wait until the morning to be assembled. This leads to a build-up of products and queues each morning into the assembly.

**Table 6: Average product throughput time through machining, assembly and production.**

| Model | Machining (hr) | Assembly (hr) | Production (hr) |
|---|---|---|---|
| **Base** | 6.79 | 7.47 | 119.34 |
| **Kanban** | 8.88 | 4.09 | 13.78 |
| **Make To Order** | 7.52 | 4.39 | 12.81 |

### *5.3.3* Customer order throughput time

All models manage to manufacture the average batch in less than 36 hours. One would expect the Make To Order model to have a much longer customer order throughput time than the Base model since the customer order point is farther back in the production system i.e. at the start of the machining line compared to at the start in painting. However, this is not the case. The reason for the small differences is that the Kanban and Make To Order models has two shifts in the assembly operations, the semi-finished storage is removed and in the Base model orders risk to stay in the storage after the painting overnight since the assembly only has one active work-shift while the painting runs on two shifts. The already mentioned separation of the D-hole operation from machining line 3 also shows as prolonged customer order throughput time in the Kanban system.

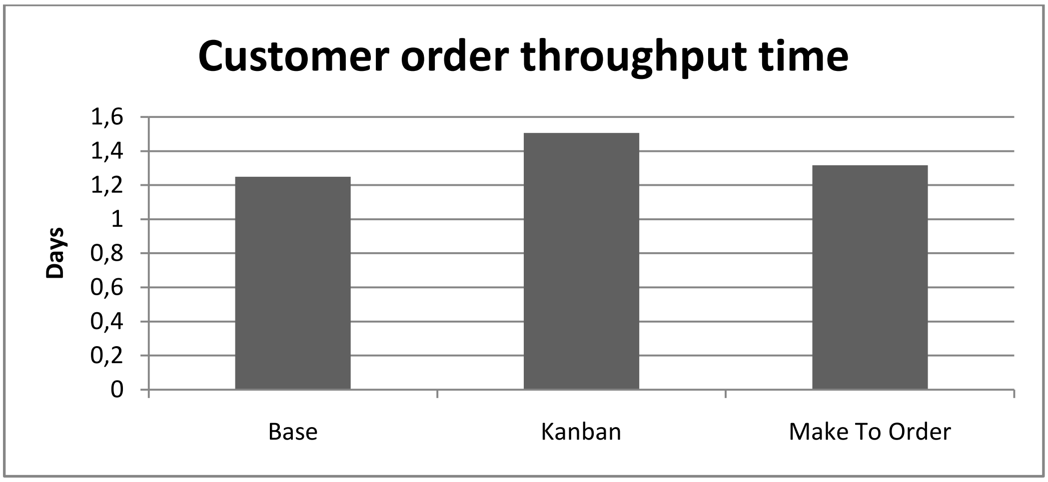



**Fig. 9: Average customer order throughput time for the different models.**

### *5.3.4* Work in progress

The Base and Kanban models both have an initial WIP since they include storages of semi-finished products which are not present in the Make To Order model (see Fig. 10). Due to overproduction, the Base model's WIP is ever increasing. The WIP for the Kanban model is decreasing slightly in the beginning as the Kanban storage is not completely full. It does reach a stable level after approximately one week. In the Make To Order model, there are very low levels of WIP since no intermediate storage exists.

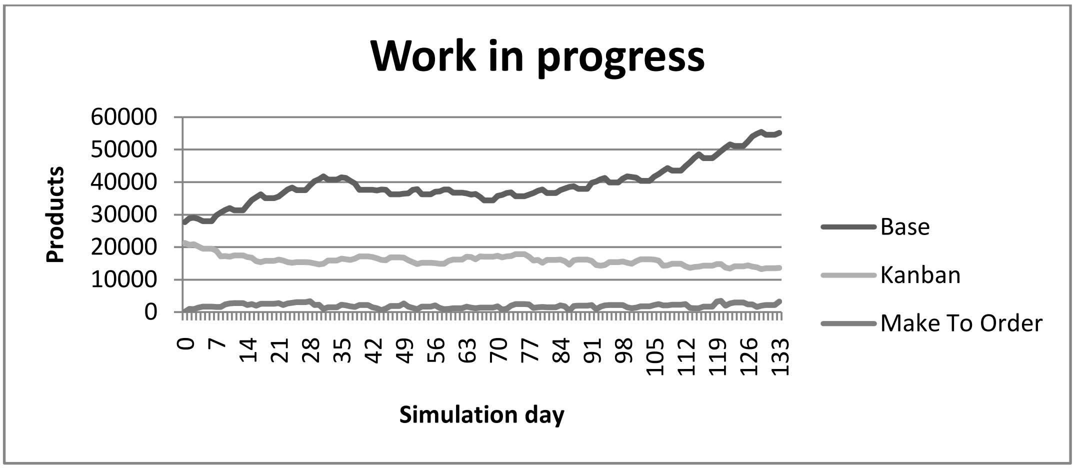



**Fig. 10: Work in progress for each model. The three planning principles show large differences.**

### *5.3.5* Free time

In Fig. 11, the free time of all operations that the housings are passing through is included. The total amount of free time in the machining lines is slightly increased, especially when introducing the Make To Order model. However, the significant increase in free time for machining line 4 does not have a major effect on other parameters since it manufactures a rather small proportion of the total production volume.

It must be taken into consideration that the increased amount of free time in assembly line 1 and 2 mainly are due to the second work-shift that are introduced in the Kanban and Make To Order model.

The tenifering and painting is not restricting each other in the Base model since there is a large buffer in between and the fact that there is different types of orders passing through the operations i.e. manufacturing orders and customer orders. In the Kanban and Make To Order model however, the two operations are closely connected, only customer orders are passing through the operations and they show similar levels of idle time.

D-hole 2 and Grease-hole 2 is only present in the Kanban scenario which explains the lack of data for these resources for the two other models and the increase of free time in D-hole 1 and Grease-hole 1 respectively for the Kanban model (see Fig. 11). The similar case is evident for assembly line 3 which is not present either in the Kanban or Make To Order model.

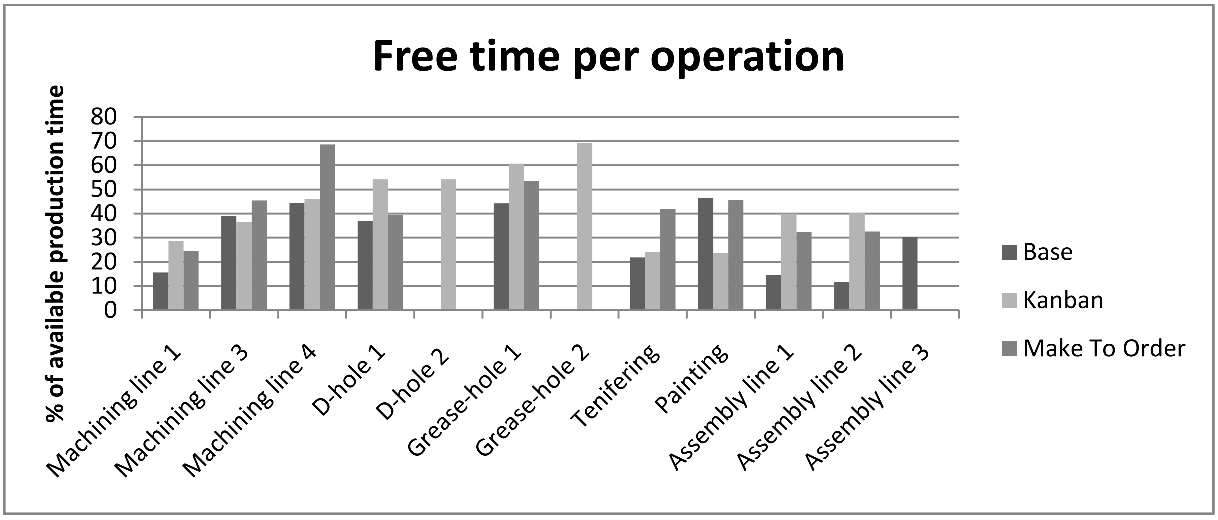



**Fig. 11: Percentage of free time in each operation for the three models.**

### *5.3.6* Output HV

Fig. 12 shows the output levels when testing the models in a high-volume scenario. Compared to the normal volume scenario the models are not as good in producing according to the output of the real production system. The Base model produces 4% less than the real production system. The Make To Order model performs very similar to the Base model while the Kanban does not really reach the same levels.

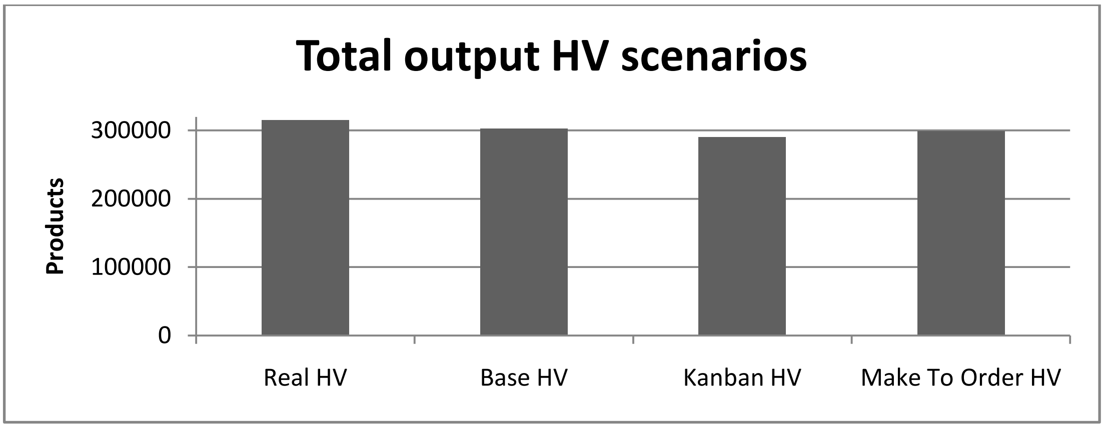



**Fig. 12: Output levels from the high-volume scenario models.**

### *5.3.7* Product throughput time HV

Table 7 shows the three different average throughput times for the high-volume models. The times are slightly longer for all models, due to the longer processing times but otherwise no major differences compared to normal volumes (see Table 6) can be identified.

**Table 7: Throughput times for high-volume scenarios.**

| | Machining (hr) | Assembly (hr) | Production (hr) |
|---|---|---|---|
| **Base HV** | 7.27 | 8.73 | 121.20 |
| **Kanban HV** | 9.43 | 4.24 | 14.42 |
| **Make To Order HV** | 8.90 | 5.11 | 15.21 |

### *5.3.8* Customer order throughput time HV

The customer order throughput time in the Kanban HV and Make To Order HV models are very similar (see Fig. 13). In the Kanban HV model, the time is slightly increased due to longer processing times. In the Make To Order HV model, the longer processing times also forces more batches to remain in the system more than one night which also prolongs the customer order throughput time. Since the Base HV model has an extra work-shift in assembly and thereby always has *some* capacity available when the painting is running, its average throughput time is actually slightly decreased.

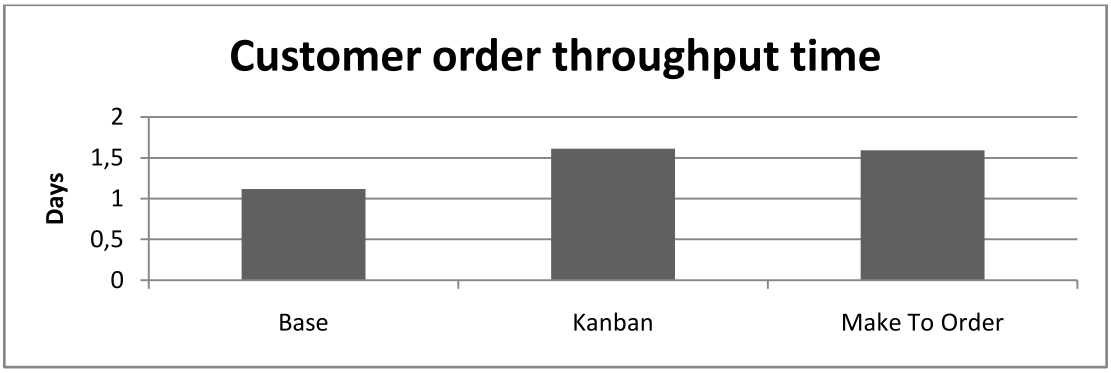



**Fig. 13: Customer order throughput time for high-volume scenarios.**

### *5.3.9* Work in progress HV

As expected, increasing the batch-sizes give higher WIP levels (see Fig. 14). The total accumulated WIP in the HV Base model is 1.65 times higher than it is in the original Base model and thereby indicates that it is proportionate to the increased volume. It takes longer until the Kanban model's levels are stabilized and in the Make To Order model, all batches are no longer finished by the end of the day.

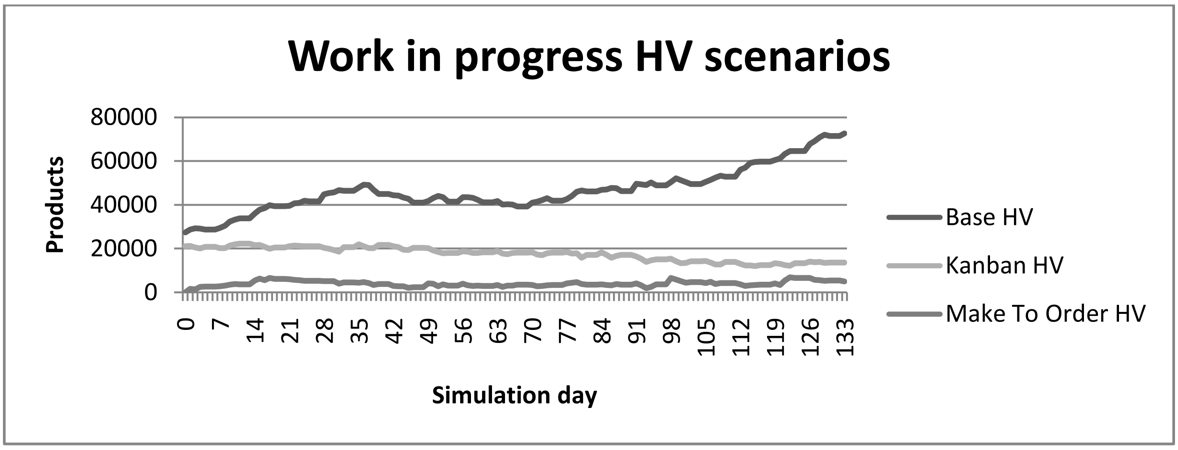



**Fig. 14: Work in progress levels for each HV-model.**

# 6 Analysis MQ1

All three models have the ability to produce the volumes manufactured during the simulation period in the real system. The two alternative models produces slightly less than the Base model due to the customer order start window of 6 days that allows customer orders to be started a few days earlier than planned if material is available and the painting operation has spare capacity. This shows in the final simulation week where the Kanban and Make To Order models have problems finishing all orders before the simulation ends.

Apart from the above mentioned technical difference, no further output problems are experienced for the two new configurations in the HV scenarios.

Even though the customer order point is moved backwards, The Make To Order and Kanban models allows for start of customer orders closer than 3 days to planned delivery date. This makes the production planning independent of preliminary orders since even the most challenging orders, the 3 days lead time orders, can be produced after the order cancelling point has been passed. This is enabled by the embracement of lean principles, which corresponds with the research of Arnheiter and Maleyeff (2005). When moving the customer order point backwards, more smaller batches will be present in the machining operations since customer order generally are smaller than the current manufacturing orders.

To implement changes always requires resources, planning, time and investments to some extent. A comparison of the changes in the two alternative configurations is presented in the experimental design section in 5.2. The changes needed for the Kanban system includes more physical rearrangements but also a more thorough reorganization of the ERP-system to handle all "new" variants such as Kanban article numbers etc.

The WIP graphs clearly show the problem of the current production planning system. During the simulation period, the Base model's inventory levels are ever increasing. This happens due to two reasons: First, the manufacturing orders of semi-finished products are based on preliminary customer orders which can be cancelled and thus leaving material to sit in the semi-finished product storage. Second, extra semi-finished products are produced partly due to lot size restrictions and partly in order to cover for scrapped material. However, the control of how much such extra material that exists in the system is poor. This is because the same storage is used both for the active orders and for spare material.

The Kanban model does have a stable WIP due to the pull principle. Most of the WIP in this model is due to the Kanban storage of machined housings, but no build-up is present. This confirms the statement by Jonsson and Mattsson (2009) that Kanban restricts WIP and reduces overproduction. For the Make To Order model, WIP levels are very low since only customer specific material is present in the production system at any given time. In this model, no inventories are affecting WIP and there exists no build-up of material.

Scrapping of products is a fact and it has to be considered even though the Company continuously is working with process improvements. In the current system products that are scrapped at a late stage can be replaced by semi-finished products in the storage located prior to the painting. Since the new proposals both suggests earlier customer order points they are more likely to experience delivery difficulties due to quality issues.

In the HV scenarios, the WIP levels for the Kanban and Make To Order model are naturally slightly above the original models (not HV). It is, however, the amount of active products that is increasing while for example the Kanban storage is kept at similar levels as in the first scenario. The flaws of the Base model configuration are clarified when increasing the production volumes. Looking at absolute numbers, the accumulation of WIP in the HV Base model is very close to 1.65 times the WIP accumulation in the original Base model. This indicates that the higher inventory levels does not even facilitate customer order starts and further points out the mismatch between the planning of manufacturing orders in machining and the actual customer orders.

# 7 Discussion MQ1

Using discrete event simulation, two alternative production system configurations is evaluated in order to collect more data on how layout and organization changes affect the performance of the system (see section 1.1). A Base model is developed for validation purposes and as a current state reference. The first experimental simulation model introduces a Kanban system for semi-finished products and moves the customer order point to an earlier phase. The second experimental model uses the make-to-order principle and thereby only customer orders exist in the factory.

The use of simulation as an improvement tool can be questioned. As described in section 3.1, simulation is a very powerful tool for providing production data in current and future configurations and analysis of complex systems. However, to model a whole factory will take time and resources which otherwise could have been used to quicker identify problems and come up with more direct solutions. For example, the simulation team identified faults in the production system during the input data gathering for which there was no time to further analyze causes and effects. This was mainly due to time restrictions following the decision to develop simulation models and thereby having to set up few, large, and less detailed experiments instead.

A simulation model whose purpose is to depict whole factories quickly becomes too complex if the level of detail is too high. Instead, as Brooks and Tobias (2000) argue, simple models can often be just as adequate depending on the purpose. Simplifications made in the developed models concerning batches instead of products, lines instead of separate stations and scaling of orders to mimic changes in volumes have been necessary in order to keep the complexity at a reasonable level. It has also helped in reducing the time needed for input data gathering, modeling and validation. Although, by simplification, the model is delimited and some problems regarding answering more detailed questions were experienced.

Simulation in a Kaikaku context is interesting in several ways. Since the reorganization is intended to be large (see section 2.4), many questions on details remain unanswered which makes the accuracy of such simulation models low. However, as stated by Aleisa and Lin (2005), simulation can be a very powerful tool as the second step when the overall layout is optimized, more detailed knowledge exists and certain parameters must be estimated or optimized (see section 2.8). The level of detail provided by the VSM:s, the production system data of the current state and complementary estimations and assumptions did function very well in developing simulation models of future states that could generate useful information.

The model was validated using a set of methods (see section 5.1). It would have been preferable to use more than daily output levels in this process since one wants to capture as much of the dynamic behavior as possible in the model. A comparison of the actual throughput time of each batch to that of the corresponding batch manufactured by the simulation model would have added another dimension to the validation and experimentation. This was not possible since no such logging was conducted. Furthermore, it makes it very hard to tell if the throughput times in the models are longer or shorter than the real throughput times and when the lead time goal stated in section 1.2 is reached.

When converting the original models into HV models, many of the cycle time OEE-correctional factors had to be removed since the output was too much restricted. This shows that the use of the factor was practical when there was a need to change the system performance quickly. However, it is quite troublesome that the actual improvements were not sufficient to raise the performance of the models. Some reasons for this might be that the manual assembly line that were very scarcely used during the original simulation period now is utilized more frequently, that breaks in the assembly line are managed by overlapping so that the line never comes to a full stop and that no overtime work anywhere is modeled. The results should thereby be considered with the HV Base model as a reference and alternative HV models' results relative to that of the HV Base model.

The Kanban and Make To order configurations implies manufacturing of smaller batches in the machining operations. As described in section 2.3, make-to-order is useful in low-volume/high-variety production which requires longer lead times (Hendry, 1998) (Zaerpour et al., 2008). At the Company, the production is generally high-volume/high-variety together with short lead times. In order to keep the desired production pace and service level to customers, further efforts must therefore be put into lowering of changeover and throughput times.

The earlier in the production system the customer order point is located, the more sensitive the production system becomes to quality deficiencies. To manage this without losing too much in delivery accuracy, a very small safety stock of semi-finished products could be introduced. However, this safety stock should **only** be used when products are scrapped, not due to bad planning which only would disguise planning problems. To keep inventories at an absolute minimum, it is also possible to only hold high-frequency variants in this safety storage. An alternative would be to keep a finished product storage and change-to-order when needed (Liker, 2004). This would require an ability to customize the products when they are already assembled e.g. stamping that in today's system is performed in the pre-assembly area must be performed on finished products. Due to the large number of raw material, this option might not be feasible for the Company.

Simulation has shown that the average customer order throughput time in all configurations is well shorter than the critical 3 days. A strict make-to-order might cause uneven flows as volumes might be high one week and low the next one resulting in peaks where available capacity is not sufficient or having to keep an overcapacity for those rare occasions (Liker, 2004). These problems exist at the Company today and the fluctuating volumes can be seen in Fig. 15.

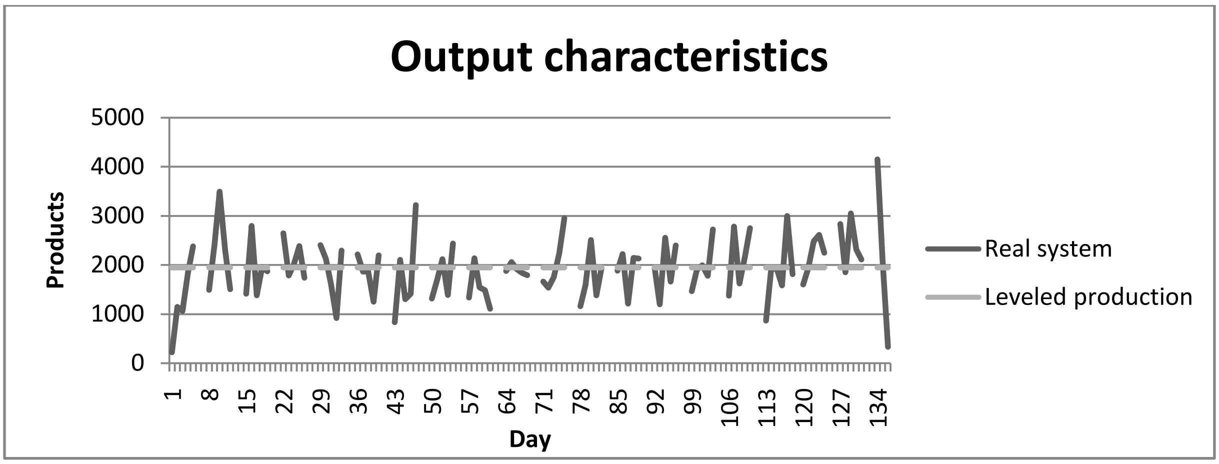



**Fig. 15: Produced units per day in reality compared to a leveled production**

Orders having longer promised lead time than 3 days can be used to even out the daily and weekly production volumes to get a more leveled production and to more effectively use resources (see section 2.2). Generally, an order should not be started until it is transformed into a firm order that cannot be cancelled i.e. the promised lead time prior to the delivery date. However, depending on fluctuations in demand, it can be started as late as the actual throughput time allows i.e. less than 2 days prior to the delivery date (see Fig. 9). For example, an order with a lead time of 5 days can either be produced early or be temporarily put aside for a couple of days due to leveled production control reasons.

# 8 System description and model MQ2

This chapter describes the current flow in the specific machining cell that is to be modified. Furthermore it presents the new layout proposals, a mapping of input data sources, assumptions and delimitations and finally the two main models.

## 8.1 Current flow

This specific cell that exists in the factory today is responsible for machining one of the components that is later to be assembled together with the housings. The total process includes the following sequence; one turning operation, one milling operation and last a reaming operation. Today the workload is divided in two cells. In total, the first cell consists of four turning machines, three milling machines and one reamer. The second cell consists of one turning machine and one milling machine and uses the same reamer as mentioned in the first cell. Accumulating conveyors are transporting materials to and from operations within each cell.

Two in-buffers, called Bulk1 and Bulk2, have the task of feeding the first cell with raw material. When needed, operators tilt a part of the content in a pallet, containing raw material, onto a shaking table. This table aligns the raw material properly and places them on a conveyor belt that feeds the cell. The second cell is fed by an industrial robot that picks material from a pallet. Since the capacity of a pallet in the feeding stations is 1200, orders above this level needs two pallets of raw material.

The quantity of a production order can vary from about 600 up to a maximum of 2150 products. The upper number is based on restrictions at the external heat-treatment facility. Most efficient for their operation is deliveries of 2150 batches but smaller batches is also possible. Variants standing for approximately 80% of the order volume have a quantity of 2150.

The current production system has a few major drawbacks. The conveyor system of the first cell is rather complex, containing a lot of switches between different conveyor tracks giving the system a quite low reliability. On top of this the system is said to be hard to maintain properly. The conveyor system also works as a buffer and are therefore very long to hold many products. This leads to a lot of WIP and long throughput times.

## 8.2 New layout proposals

Due to the drawbacks of the current production system two new layout proposals has been defined and are presented below. The machines that are to be kept in the future layout are three turning machines, three milling machines and one reamer. These machines as well as the conveyor system already exist and no new machines are needed.

### *8.2.1* Proposal 1

This layout proposal has emerged from operators working in the cell. It mainly consists of two flows that are fed by one Bulk each and finally merges in front of the reamer.

**Left flow** (see Fig. 16)**:** This flow has a parallel changeover of machines. Products are transported from Bulk1 to either Lathe1 or Lathe2 depending on their availability. Once a product has been turned it is transported to either Milling1 or Milling2 depending on certain cycle time-based routing rules.

**Right flow** (see Fig. 16)**:** The second flow has a serial changeover. Products are transported from Bulk2 into Lathe3, after which they are transported to Milling3.

This layout, due to separated flows, has the ability to run one or two variants at the same time, which the other proposal has not (see section 8.2.2). Between different product variants changeover in the lathes and reamer needs to be performed. When changing variant, all products from the previous batch must pass through the lathes before new raw material is entering the conveyor system. This also applies for Proposal2 (see Fig. 17). The need of changeover has a major impact on how material should be fed from each flow into the reamer. When running one variant, the reamer is fed alternately by products from each flow. In order to minimize changeover in the reamer when running two variants, buffering is performed in the end of each flow and only one variant at a time is allowed to enter the reamer. When the conveyor that is being blocked has reached its capacity, changeover is performed in the reamer and the other flow blocked. After the changeover products from the blocked flow are released and allowed to enter the reamer. This procedure is repeated so that one flow alternately buffers and the other one is feeding the reamer.

The right flow is intended to run high-volume variants since Lathe3 has a longer cycle time for other variants. The left flow on the other hand has the same performance independent of variant.

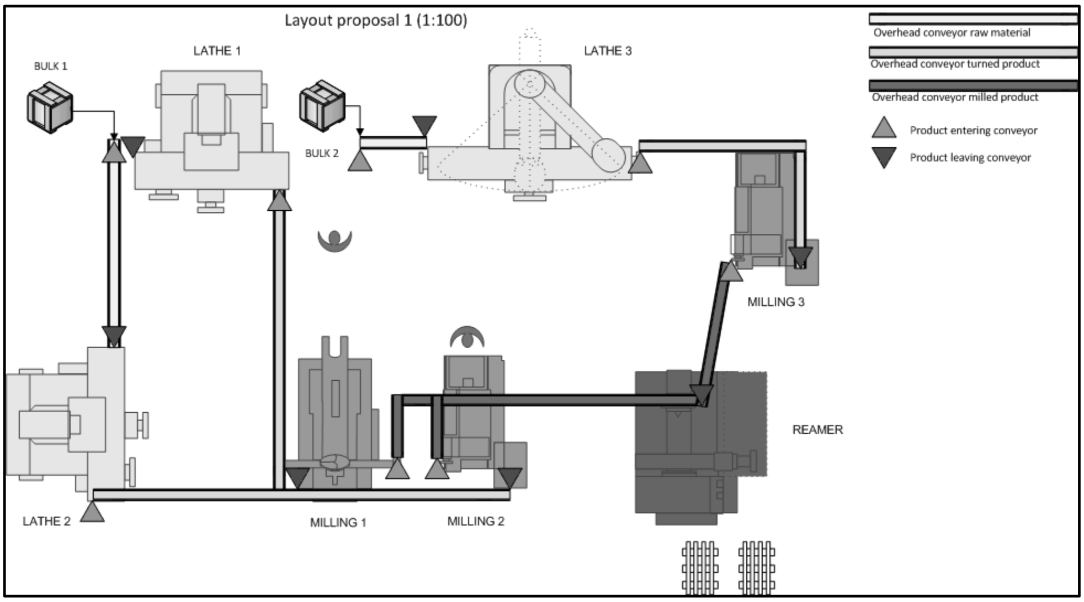



**Fig. 16: Layout proposal 1 - Operators**

### *8.2.2* Proposal 2

This layout proposal has emerged from the production engineering office. One Bulk is feeding the flow (see Fig. 17) with products. Meanwhile the other Bulk can be prepared to switch whenever the other Bulk is empty. The flow has a parallel structure meaning that products that are fed from one Bulk are transported to one of the lathes and afterwards to one of the milling machines and finally into the reamer. Since there is one common flow of materials, only one variant can be run at the same time.

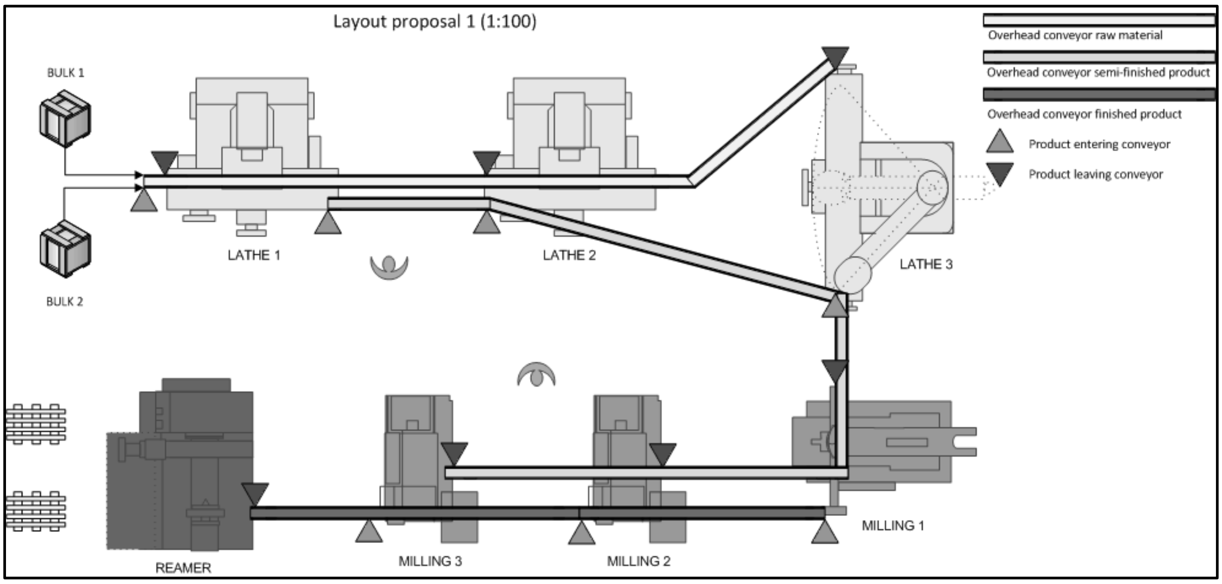



**Fig. 17: Layout proposal 2 - Production Engineering Office**

## 8.3 Input data

The data presented in Table 8 in describes the character of input data sets for the simulation models in MQ2 and Table 9 presents the cycle times used in the different machines.

**Table 8: Input data sources MQ2.**

| Parameter | Data source | Level of detail |
|---|---|---|
| **Cycle times** | Cycle time mapping documents | Static in seconds for each machine. |
| **Conveyor speed** | Stopwatch measurements | Static meter/second |
| **Mean Times Between Failure/ Mean Times To Repair** | Database | Manual logging of breakdown data per machine or line, converted into statistical distributions. |
| **Changeover times** | Interview | Dependent on operator's skill. Triangular distribution used. |
| **Tool changes** | Interview | Dependent on operator's skill. Frequency dependent on machine type. Triangular distribution used. |
| **Historical orders** | Database | Actual historical orders from Aug-Dec 2009. |
| **Material handling times** | Interview | Operator and situation dependent. Triangular distribution is used. |
| **Work-shifts, breaks etc.** | Production schedules | Actual shifts and breaks are used. |
| **Scrap rates** | Interview | Static in percentage. |

**Table 9: Cycle times for resources in the cell.**

| Resource | Cycle time (s) |
|---|---|
| **Lathe1** | 47.9 |
| **Lathe2** | 43.8 |
| **Lathe3** | 34.8/42 (high/low vol. variants) |
| **Milling1** | 43.4 |
| **Milling2** | 38.5 |
| **Milling3** | 24 |

## 8.4 Delimitations

- The simulation model consider incoming raw material until the last machine operation in the cell and all the operations in between.
- No allowances for personal needs are deducted from the operator performance. Personal needs are the time for going to the restroom etc. (Freivalds and Niebel, 2003).

## 8.5 Assumptions

- Production orders and belonging raw material is always available when needed.
- Recent, real orders which represent the current product variant flora and true production quantities are used in the simulation.
- The feeding stations are always sending a new product each 13:th second.
- To make the simulation model behave as the real system, routing rules has been defined in both proposals. These include re-routing when machines are down, not sending too many products to a single machine etc.
- The models are initially run with two operators. Further tests are performed with 1, 2 or 3 operators to ensure that an appropriate workforce is chosen.

## 8.6 Simulation period

To achieve credible and statistically correct results, each model is run over a period of time several times. During experimentation, the run time is set to 890 hours which represents 12 weeks of planned production time when breaks, nights and weekends have been deducted. 10 replications are made for each model to be able to establish confidence intervals and further reduce deviations due to unfortunate random events. An example of such an event would be rare, very long production stops due to machine failures, which affect a single run significantly and thus render that specific run unsuitable to represent the behavior of the model.

## 8.7 Simulation models

One main model for each proposal is developed. Proposal 1 is shown in Fig. 18 and Proposal 2 is shown in Fig. 19.

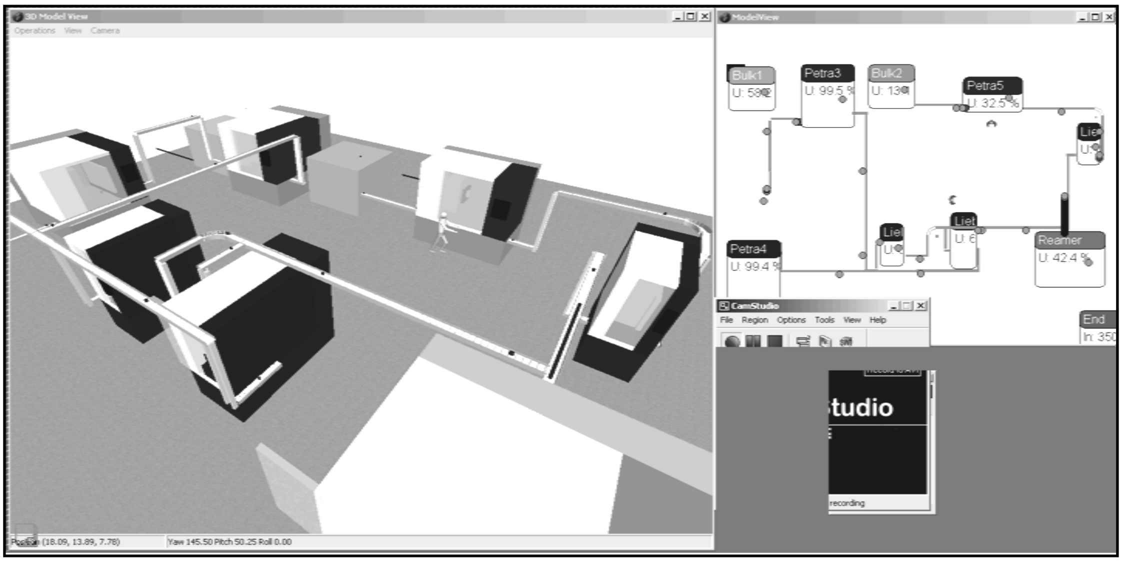



Fig. 18: Simulation model of Proposal 1. 3D-view to the left and 2D-view to the right.

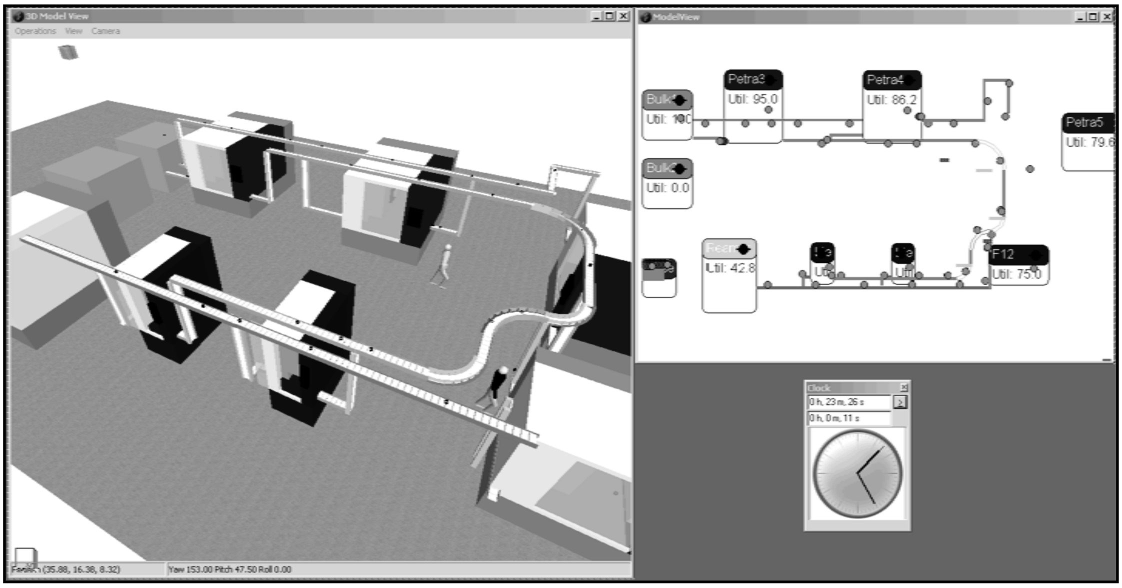



Fig. 19: Simulation model of Proposal 2. 3D-view to the left and 2D-view to the right.

# 9 Results MQ2

This chapter presents the results from the validation process and experimental execution.

## 9.1 Validation results

Face validation is performed in a meeting with the operators working in the cell and representatives from the production engineering office. The simulation model is introduced and an explanation of how it works and the assumptions made are presented. The model performance are also presented and checked for reasonableness.

## 9.2 Experimental design

The first indications from the simulation runs show that output levels are very similar in both main proposals. The throughput time per piece however, is approximately five times longer in Proposal 1 where two product variants are produced at the same time and buffers are needed due to changeovers in the reamer. Moreover, Lathe 3 which is specialized at machining high volume variants has a low utilization due to blocking by Milling 3. It is decided, by discussion with the Company that three modified models should be introduced with Proposal 1 as base:

1. SplitOrders model – Instead of producing two variants simultaneously, only one order is started at a time. The order is split and run in both flows, reducing the need for changeover in the reamer to less than once per batch.
2. MillSwitch model –The milling machine Milling 1 has a cycle time almost half of the other two milling machines. In Proposal 1, this machine is placed in the flow with two lathes and two milling machines even though it is the only individual milling machine able to match the cycle time of Lathe 3. In this model, Milling 1 and Milling 3 switches place in an attempt to mitigate blocking of Lathe 3.
3. SplitOrdersMillSwitch model – Combining the single variant sequencing from the SplitOrders model and the milling operation layout from the MillSwitch model.

Running two variants at a time is only possible in Proposal1 Base, Proposal1 MillSwitch and Proposal2 Base.

To investigate the impact of smaller batch-sizes, experiments with a batch-size of 600 are also introduced. The smaller batch-size was chosen after discussion with the Company.

Design of Experiments (see section 2.6) is performed to analyze the effects on the performance indicators when combining the three models and to reveal interaction effects. The design matrix in Table 10 is used to conduct the experiments. Given three factors and two levels 8 configurations are needed ($2^3$) to achieve a full-factorial design.

**Table 10: Configuration of conducted experiments.**

<table>
<tr><td rowspan="9">A=SplitOrders<br>B=MillSwitch<br>C=Batch-size<br>600</td><th>Config.</th><th>A</th><th>B</th><th>C</th><th>A x B</th><th>A x C</th><th>B x C</th><th>A x B x C</th><td rowspan="9">-1 = Not active<br>1 = Active</td></tr>
<tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td></tr>
<tr><td>2</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr>
<tr><td>3</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>1</td></tr>
<tr><td>4</td><td>1</td><td>1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr>
<tr><td>5</td><td>-1</td><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>1</td></tr>
<tr><td>6</td><td>1</td><td>-1</td><td>1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td></tr>
<tr><td>7</td><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>1</td><td>-1</td></tr>
<tr><td>8</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table>

### *9.2.1* Definition of measurements

In order to evaluate the performance of the different models, key performance indicators are identified. These are:

- **Total output** – The number of products that are finished in the reamer during the simulation.
- **Output per hour** – The number of products that are finished in the reamer each individual hour during the simulation.
- **Throughput time per product** – The average time a product spends in the production cell. This gives valuable information on WIP levels and the sensitiveness/responsiveness of the configuration proposals.
- **Overall equipment efficiency (OEE)** – A performance measure that takes into account availability, performance and quality factors. The OEE is calculated as described in section 2.7:
  $$Overall\ Equipment\ Efficiency = Availability \times Performance \times Quality$$
  Based on the cycle times, the turning operation is the bottleneck of the future cell and thus its OEE is indicating the capacity of the whole cell.
- **Operator workload** – The share of the production time when the operators are busy serving machines, inspecting products, handling material etc.

## 9.3 Simulation results

### *9.3.1* Output

The results show that the output is similar in all the models (see Fig.20). There are two models that are slightly better than the rest, namely Proposal1 Base and Proposal1 MillSwitch. Common for these two is that they do not split orders in half into the two flows in the cell (SplitOrders). This means that the batch in a flow is kept large and fewer changeovers are performed in the turning and milling machines. On the other hand, more changeovers are performed in the reamer but this to do not influence the output in a negative way as the changeover especially in the turning and milling does.

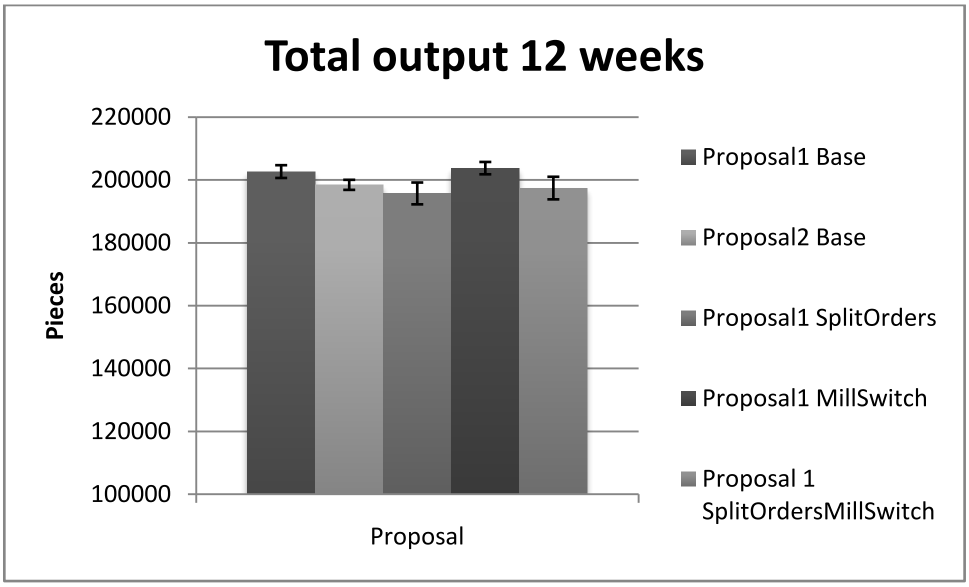



**Fig.20: Total cell output for the five different proposals.**

According to the output per hour results in

Fig.21, the Proposal1 Base model seems to have the most even 10 period moving average but both Proposal2 and the split orders model have a smaller variance (see Fig. 22).

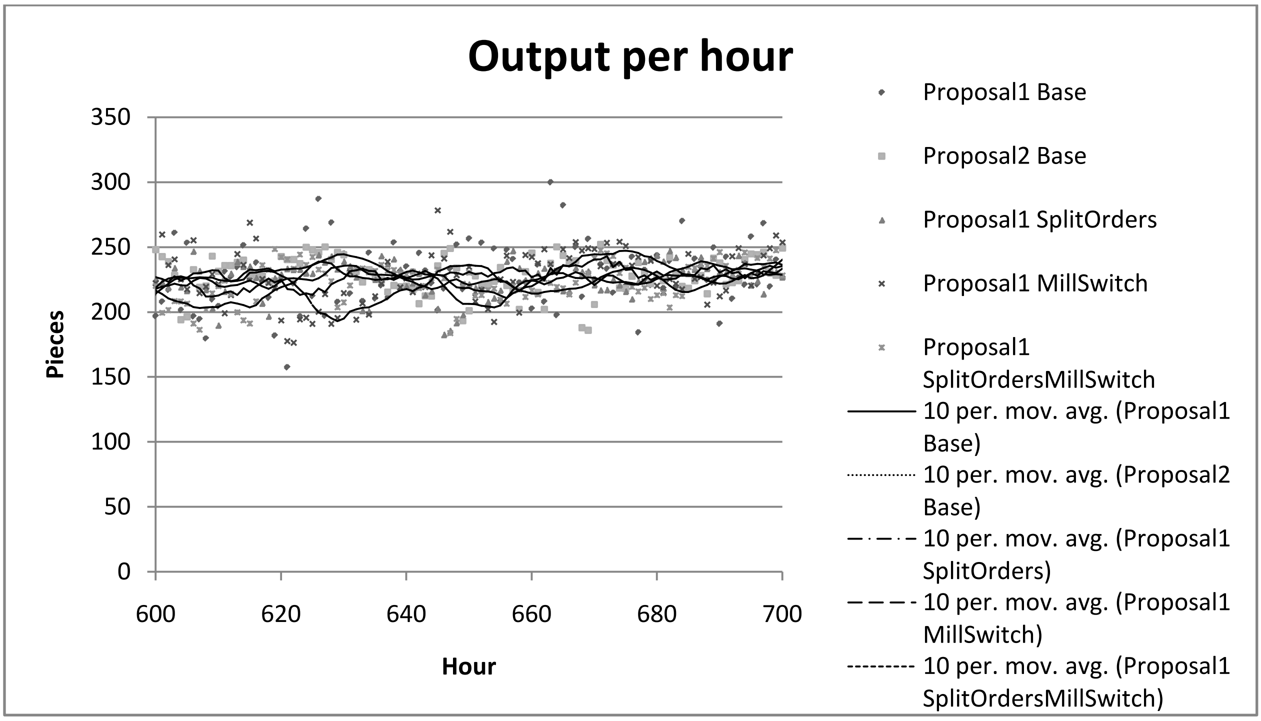



**Fig.21: Cell average output per hour and 10 period moving average during a part of the simulation period.**

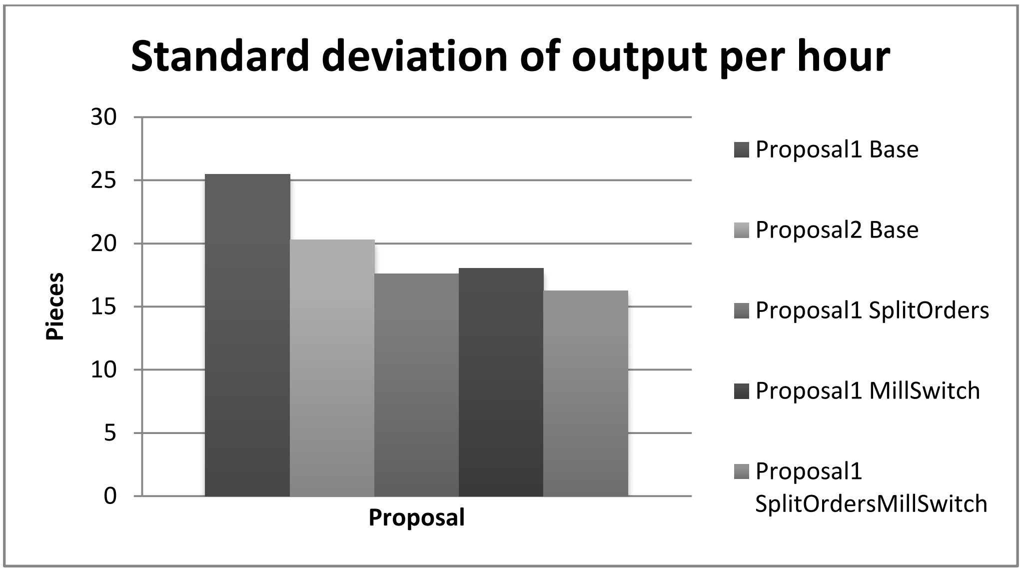



**Fig. 22: Comparison of the standard deviation, output per hour**

### *9.3.2* Throughput time per piece

The simulation results show that Proposal2 have a much shorter throughput time than Proposal1, approximately 60 minutes less per piece (see Fig.23). Proposal1 has a throughput time of nearly 85 minutes per piece invoked by its need of buffers compared to 25 minutes per piece in Proposal2. The important result is the drastic decrease in throughput time in the SplitOrders model. It can also be noted that positioning the milling operation as in the MillSwitch and in the SplitOrdersMillSwitch reduces the throughput time by more than 50%. In Fig. 10 this can be seen by comparing the before and after values of proposals where MillSwitch has been introduced. This means that it is possible to achieve a shorter throughput time in Proposal1 than in Proposal2 by combining the SplitOrders and MillSwitch options.

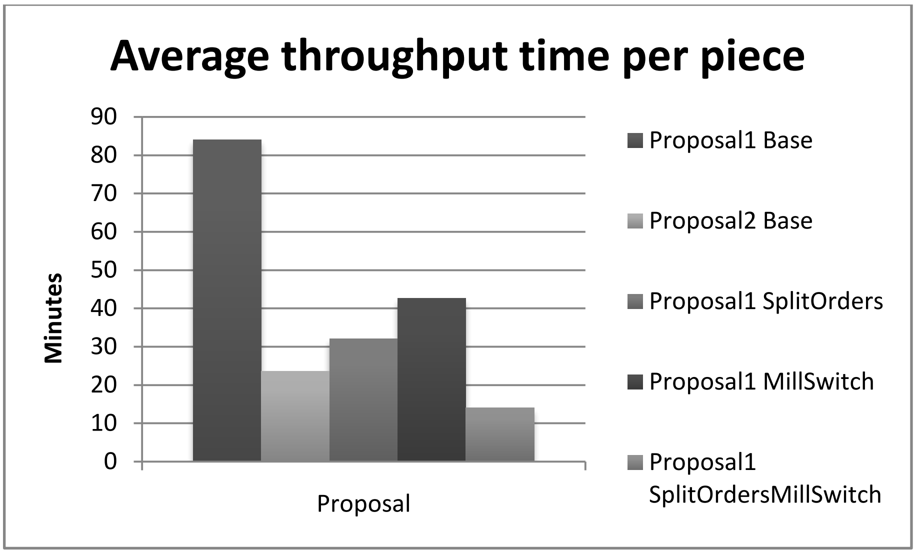



**Fig.23: Throughput time through the production cell for each piece.**

### *9.3.3* Throughput time per batch

Throughput time per batch is calculated as follows:

$$Batch\ throughput\text{-}time = (Batch\text{-}size\ \cdot Inter\text{-}arrival\ time) + Average\ throughput\text{-}time$$

The throughput time per batch has an impact on the agility of the system since a batch can be sent to heat treatment earlier if all pieces are finished faster. The best model, Proposal1 SplitOrdersMillSwitch, finishes two maximum-sized batches in 17.18 hours while two work-shifts are 16.5 hours in total. Improving the system even further would give an advantage of being able to produce two maximum batch-size orders in one day.

The batch throughput times assuming maximum batch-sizes are visualized in Fig.24. The thinner lines represent having two variants in the system simultaneously. Proposal2 is the most time efficient model with its low buffer levels and single variant sequence. Proposal1 base model is the worst due to the required changeovers and buffering and the fact that batches takes longer to process.

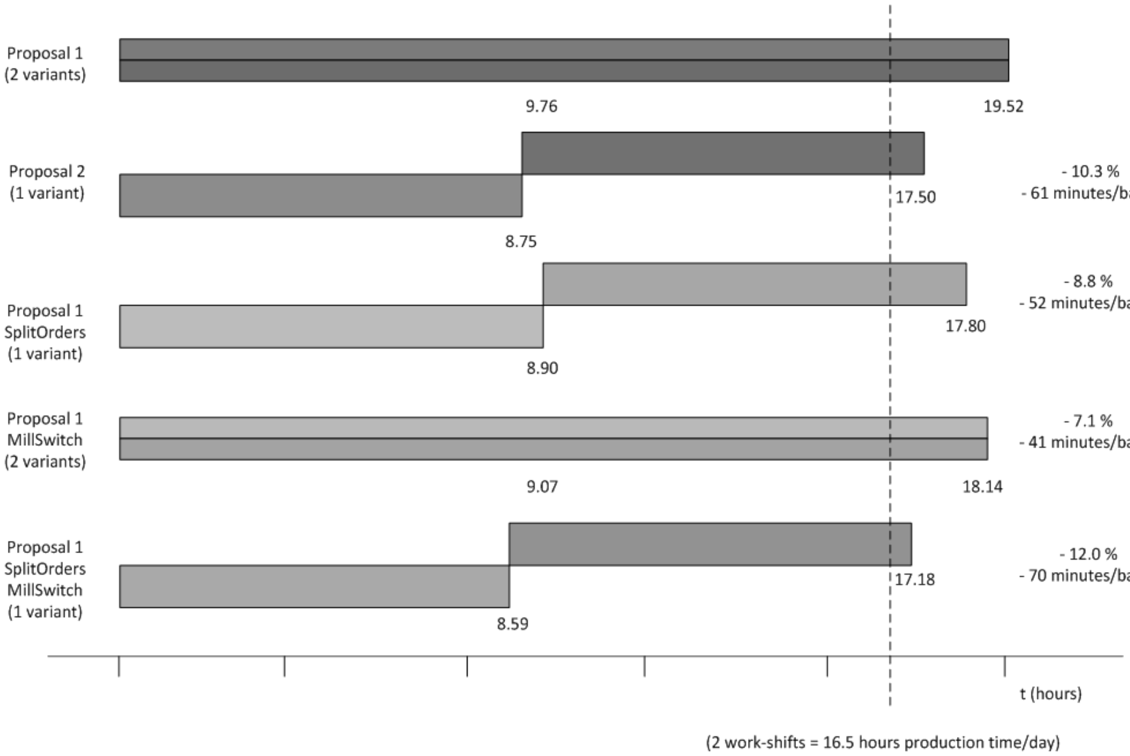



**Fig.24: Throughput time for batches.**

### *9.3.4* Overall Equipment Efficiency

The OEE and output values are strongly correlated which is seen in a comparison of Fig.20 and Fig. 25. This is due to the nature of the OEE calculation where higher output increases the performance of the cell and thereby gives a higher OEE value as long as availability and quality is kept at similar levels.

Proposal 1 is less efficient when running one variant at a time. This is shown in Fig. 25 where SplitOrders and SplitOrdersMillSwitch have lower OEE than the rest of the models. The MillSwitch option does not have any visible impact on the OEE.

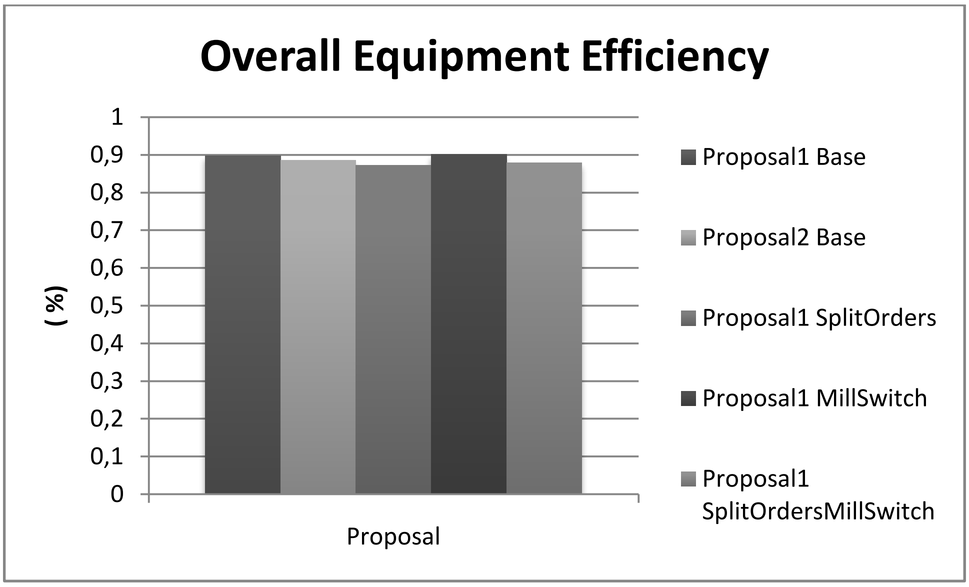



**Fig. 25: OEE results for each proposal**

### *9.3.5* Number of operators

As shows in Table 11 the OEE level does not increase much when having three operators instead of two. On the other hand, one operator significantly decreases the OEE level since one operator cannot keep up with the workload and machine waiting time increases.

**Table 11: Number of operators affecting the OEE level. Tests are performed in the Proposal1 Base model.**

| Operators | OEE |
|---|---|
| 1 | 82.64 % |
| 2 | 89.09 % |
| 3 | 89.81 % |

The workload using two operators is measured to ensure that two operators are enough. Overall, all Proposal1 models require higher operator support compared to Proposal2 but all four models manage with 2 operators (see Fig. 26). No allowances for personal needs (bath room visits etc.) is included, neither are response delays on operator calls (if the operator is free).

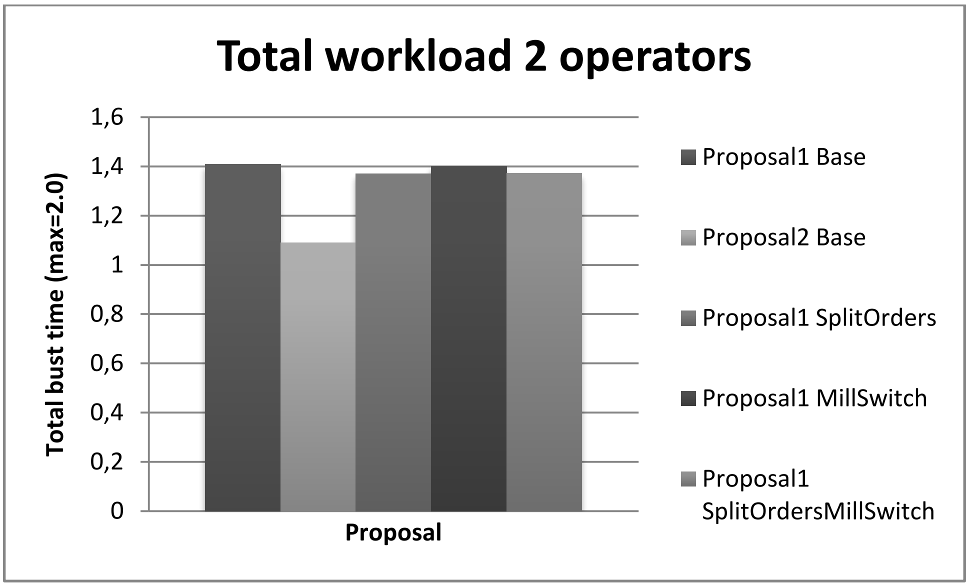



**Fig. 26: Total workload for 2 operators**

## 9.4 Design of Experiments

In order to investigate main effects and gather more knowledge about interaction effects between different factors in the Proposal1 configurations, a full factorial analysis is performed. The results from earlier runs are combined with complementary runs and presented in Table 12.

**Table 12: Results from DoE**

| Proposal1 | Output (products) | OEE (%) | Throughput time (s) | Workload (operators) |
|---|---|---|---|---|
| **Base** | 202697 | 89.8 | 5037 | 1.40 |
| **SplitOrders** | 195747 | 87.2 | 1926 | 1.37 |
| **MillSwitch** | 203797 | 90.1 | 2554 | 1.40 |
| **SplitOrders x MillSwitch** | 197454 | 87.9 | 839 | 1.37 |
| **Base Batch-size 600** | 198009 | 87.7 | 5102 | 1.42 |
| **SplitOrders x Batch-size 600** | 187121 | 84.0 | 877 | 1.40 |
| **MillSwitch x Batch-size 600** | 199710 | 88.3 | 2448 | 1.42 |
| **SplitOrders x MillSwitch x Batch-size 600** | 184506 | 82.8 | 947 | 1.40 |

Overall, the effect of using a smaller batch-size has a negative effect on the output and the OEE (see Table 12). Regarding the throughput time it is only the SplitOrders proposal that is reduced drastically. This means that there is no performance advantage of using smaller batches in the future system. But if it is desired, the output will be slightly lower (see Table 12).

Analysis and interpretation of contrasts acquired by DoE is quite complex and many assumptions must be made. Overall, main effects of factors are stronger and more trustworthy than interaction effects and the interaction effects usually get weaker and weaker the more factors that are involved. The normal distribution curve shows where contrasts should be situated if no effect can be identified. Contrasts that fall outside the bell-shape are considered as factors that have a significant effect on the response measure.

### *9.4.1* Output

SplitOrders and Batch-size 600 have a negative impact on output levels while other factors do not have any significant effect (see Fig. 27). The only positive contrast is achieved by the MillSwitch factor even though it is extremely weak.

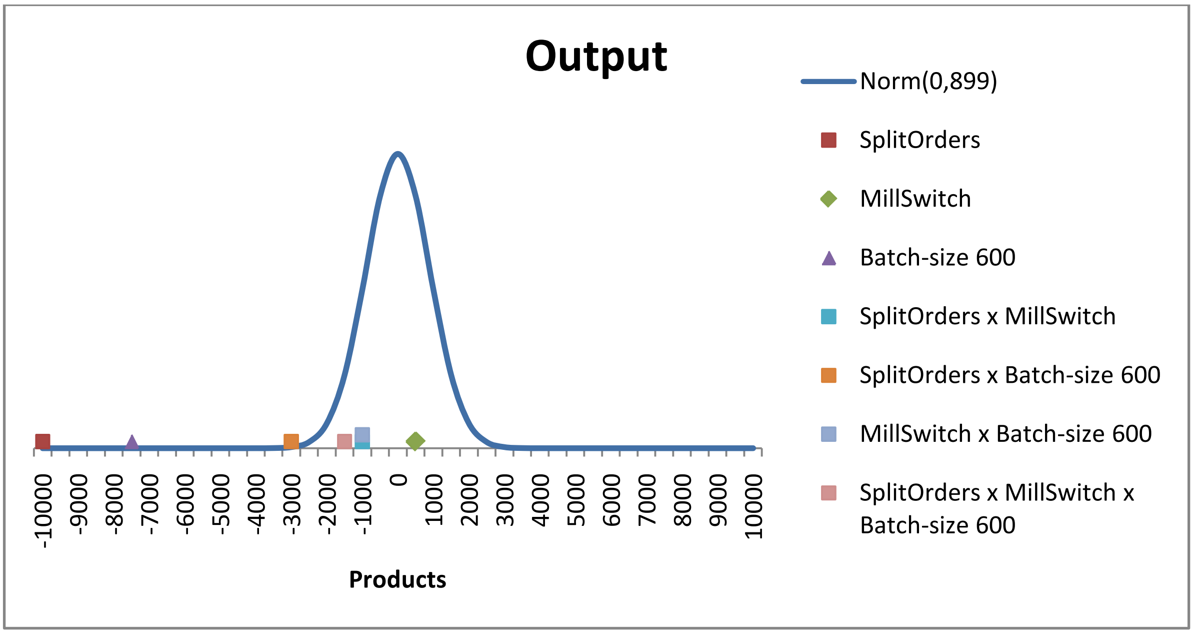



**Fig. 27: Main and interaction effects on output**

### *9.4.2* OEE

The OEE results are very similar to the output results and a strong correlation exists. SplitOrders and Batch-size 600 have strong negative main effects on OEE while the main effect of the MillSwitch factor seems to be slightly positive.

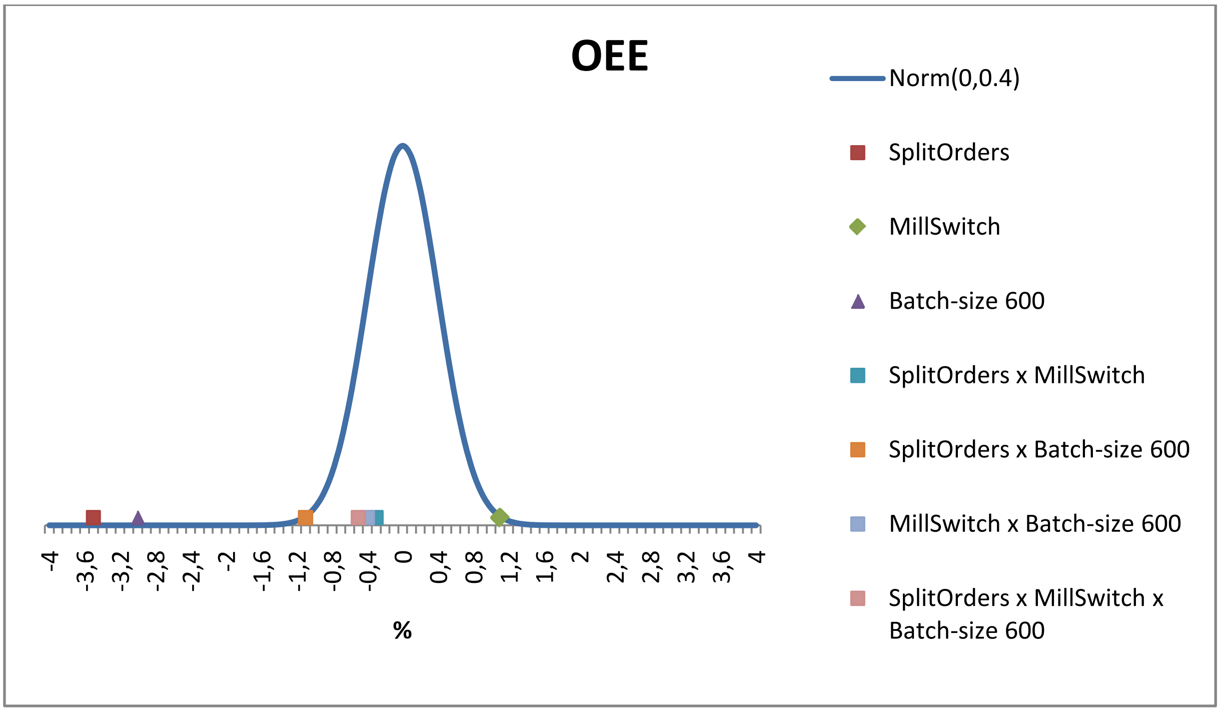



**Fig. 28: Main and interaction effects on OEE**

### 9.4.3 Throughput time

As shown in Fig. 29, the SplitOrders factor has a very strong contrast on the throughput time. The estimated effect is a decrease of approximately 2600 seconds if this option is implemented. The MillSwitch contrast is also strong and its effect on throughput time cannot be neglected. Since both of these contrasts are so strong, their interaction effect is affected negatively and its contrast does not mirror the absolute value results. MillSwitch x Batch-size 600 is further decreasing the throughput time.

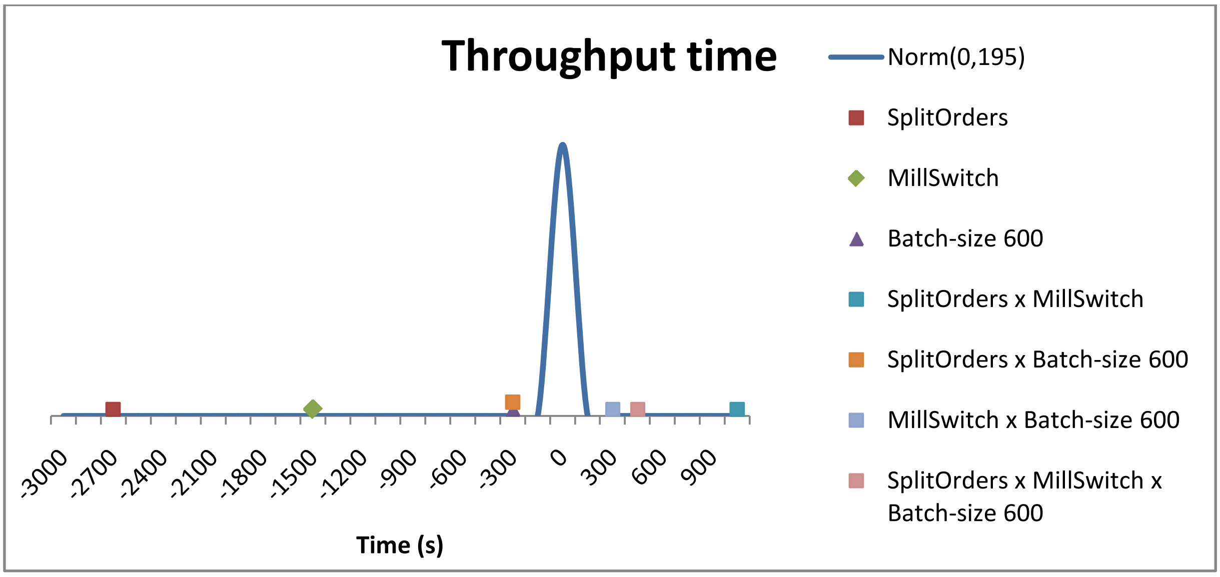



**Fig. 29: Main and interaction effects on lead time**

### 9.4.4 Operator workload

Fig. 30 shows very clearly that the SplitOrders factor reduces operator workload while smaller batch-sizes increase the workload. The interaction effect between the two factors show that it still require slightly more work to uphold a system where both factors are applied. As the MillSwitch contrast is located in the centre of the normal distribution it is assumed that it has no main effect on the workload of operators.

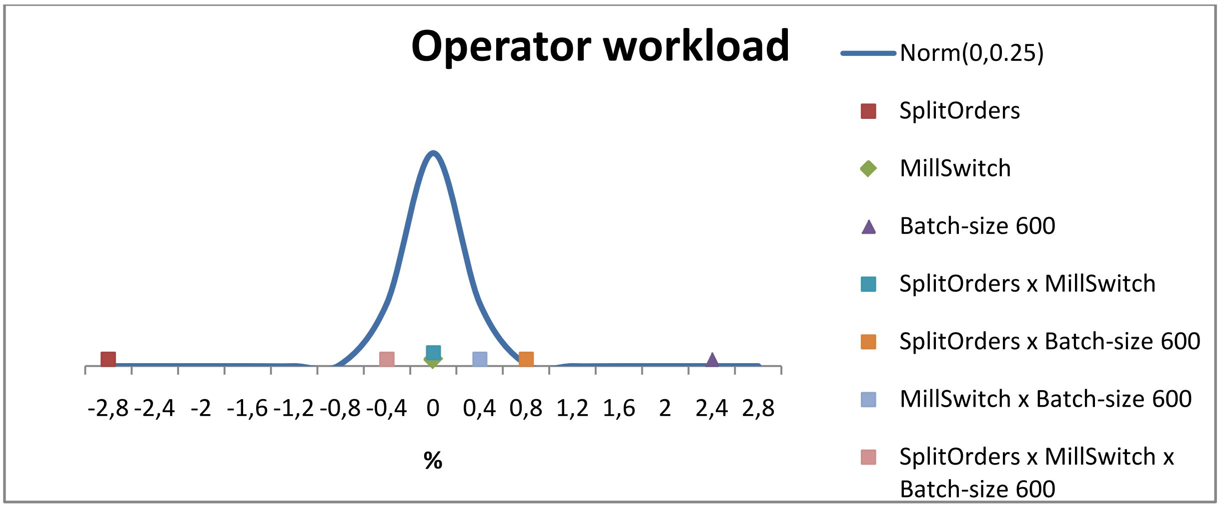



**Fig. 30: Main and interaction effects on operator workload**

# 10 Analysis MQ2

The results show that there is not a large difference in the output and OEE levels between the models. Instead, throughput time per product is the factor that actually differentiates the models. However, since the throughput time per batch depends mostly on the batch-size, this is the parameter to change if the system should become more agile.

As shown in Fig.23, Proposal1MillSwitch and Proposal1 SplitOrders individually reduces the throughput time by more than half compared to Proposal1 Base model. The throughput time reduction in SplitOrders is caused by single-variant processing and that the buffers in front of the reamer are superfluous. In Proposal1 Base and SplitOrders model Milling3 is a bottleneck in the right flow since it has a much higher cycle time than Lathe3 (see Table 9). The build-up of products on the conveyor system in front of Milling3 increases the average throughput time. In the MillSwitch model Milling1 and 3 are being switched. This gives a more well-balanced cell, WIP is decreased and the average throughput time is shortened.

A combination of SplitOrders and MillSwitch achieves an even lower throughput time, approximately halving the throughput time again. This was done even though low volume orders were run in the right flow and the average cycle time in Lathe3 thereby is increased. This means that the increase in cycle time in Lathe 3 (see Table 9) does not have any major impact on the throughput time. What really matters in terms of reducing throughput time is the lack of buffering in front of the reamer and that switching milling machines creates a more balanced cell.

The fact that the operators themselves have designed one of the proposals (Proposal1 Base) is an important thing to take into consideration when choosing the final proposal. If one of the proposal originating from Proposal1 Base is chosen operators may feel that they contribute to their own work environment and there may also be a much smaller resistant to change from certain individuals.

If the company in the future is planning to decrease the batch-size this will an effect on the proposals performance. Since Milling3 is a bottleneck in the right flow (except MillSwitch) large amount of products accumulates in front it and the throughput time increases drastically. When using a smaller batch-size this behavior is not present giving reduced throughput times.

# 11 Discussion MQ2

The purpose of this study was to find the best configuration for the new cell. In collaboration with the company several possible cell configurations was simulated in order to find the best configuration.

In short, the results show that the output and OEE levels are quite similar for all models. There is, on the other hand, a larger difference in throughput time between the two base models. By modifying Proposal 1 Base, the throughput time levels are reduced to similarly low levels as experienced in the Proposal 2 model. DoE confirmed previous results and showed interactions effects and important characteristics for each Proposal1 option with reduced batch-sizes (see Table 12 and section 2.6).

One assumption made was that raw material always is available when needed in the production cell. This may not always be the case. Some operators are more skilled in planning in advance and fetches new raw material to the in-buffer before the in-buffer is empty. Other operators may not be as proactive or are occasionally busy with other tasks. Since this is such an infrequent event, occurring approximately once each fifth hour, it is decided that the assumption still holds. Another assumption made was that orders are always available, meaning that the cell always has work to do. This is not exactly accurate compared to the current or the estimated future state but since the simulation models are using the same assumption this do not affect the relative difference between models in an erroneous way.

Accurate breakdown data is important in order to achieve a good simulation model. In this case breakdowns are manually logged and data points are quite few for certain machines. These possible uncertainties make it a questioning of its validity legitimate. Calculated averages for MTTR and MTTF were early discussed with the Company and it is concluded that the numbers are reasonable since breakdowns in some machines are, in reality, very rare.

The two assumptions mentioned above plus the fact that operators, if available, instantly walk to perform their tasks in the simulations can be reasons for somewhat high output and OEE values.

In the simulation model a lot of time is spent in programming routing rules. These rules are for example how re-routing should be performed if a machine has a breakdown. These rules have been designed in collaboration with the production engineering office. In order for the future system to perform in accordance with the simulation results it is vital that such routing will exist in the future real system.

Since modeling is done of a future system there is no exact knowledge of how the conveyor systems will look like or its lengths. The production engineering office knew the approximate layout of the new cell so discussions were held to ensure that the conveyor system became credible and as realistic as possible. Even though, there will be deviations between models and the future implementation that might have impacts on the results.

# 12 Conclusions and recommendations

The following recommendations are one step in order to increase productivity, shorten lead times and save production space. The simulation models have proven a useful tool for evaluation of production system configurations and if additional time was available even more scenarios could have been tested.

**MQ1**

A transformation of the current production system into a make-to-order system requires less effort than a similar implementation of a system using the Kanban system such as described in the report. No reduced output level or drastically longer throughput times is caused by any of the alternative configurations even if volumes are increased.

The Company can reduce the product lead time from four days to approximately 36 hours. It is shown that both alternative configurations can produce equal number of products as the current system. However, due to delimitations in the modeling, no single metric is established that can estimate the total productivity improvement.

The amount of work in progress is approximately halved when using Kanban while the inventory-based work in progress is completely removed in the make-to-order configuration. Assuming three days promised lead time to customer, neither of the make-to-order and the Kanban configuration has any problems finishing customer orders on time. On the contrary, the average customer order throughput time is less than two days. Together with new planning principles, this enables leveled-planning and more efficient product flows since no ambiguity exists regarding the destiny of products. This leads to the recommendations to initiate a project with the purpose to shift the current system into a make-to-order system.

Until improvements have been made that makes quality problems extremely rare, a small storage for semi-finished products could be kept to cover for scrapped products. Also, due to the customized batch-sizes, even more focus on changeover reduction will be important to facilitate the change of planning principle.

Since the throughput time is not measured detailed enough in the real system, too large safety margins are used in planning which has led to high work in progress and too long throughput times in today's system. Further, it makes it impossible to measure lead time reductions and follow up improvement work. Therefore, the Company should implement throughput time logging in the ERP-system.

When this system is in place, apart from the economical benefits of e.g. lower inventories and less overtime work, bottlenecks will also be more easily identified (step 1 in TOC) which leads to a better prioritization in the continuous improvement work.

The most critical parameters to consider for future improvements are:

- Abolishment of overproduction
- Improved production planning tools
- Changeover time removal or reduction in machining
- Further work with continuous improvements

**MQ2**

The recommended proposal differs depending on whether the Company is planning on running the cell with one or two variants. In both cases, alternative versions of Proposal1 are the best:

If only *one* product variant at a time is processed in the cell *Proposal1 SplitOrdersMillSwitch* is the recommended proposal, while processing *two* product variants *Proposal1 MillSwitch* is to be recommended. Two-variant processing causes somewhat higher output but on the other hand a much longer throughput time per product and should therefore be avoided.

If the Company plans on reducing the batch-sizes, the best proposal, for a *batch-size of 600*, are *Proposal1 SplitOrders* for one-variant processing, and for two variants the same proposal as with large batch-sizes i.e. *Proposal1 MillSwitch*.

It is preferable to produce one variant at the time since individual batches finish quicker and thus can be sent to heat-treatment earlier. This would have a positive effect on the total lead time of the components. The risk of choosing Proposal1 is to run two variants simultaneously too often due to the *possibility* to do that and reluctance to perform the same changeover in all three lathes. This scenario completely erodes the positive effect on total lead time and thus makes it highly advisable *not* to run two variants at a time. In Proposal2, two-variant processing is not possible and thereby, instead, guarantees this effect. Since Proposal1 is the operators' proposal the implementation of the change is probably easier.

A summary of the best proposals and their performance is presented in Table 13.

**Table 13: Showing the best proposal for different scenarios.**

| Batch-size | Number of variants | Best Proposal | Output (products) | Throughput time (s) |
|---|---|---|---|---|
| **2150** | 1 | SplitOrdersMillSwitch | 197454 | 839 |
| | 2 | MillSwitch | 203797 | 2554 |
| **600** | 1 | SplitOrders | 187121 | 877 |
| | 2 | MillSwitch | 199710 | 2448 |

# 13 References


Aleisa, E.E. and Lin, L. (2005) 'For effective facilities planning: Layout optimization then simulation, or vice versa?', Proceedings of the 2005 Winter Simulation Conference, Buffalo, 1381-1385.

Arnheiter, E.D. and Maleyeff, J. (2005) 'The integration of lean management and Sig Sigma', *The TQM Magazine*, pp. 5-18.

Avraham Y. Goldratt Institute, L. (2009) 'The Theory of Constraints and its Thinking Processes'.

Bamber, C.J., Castkta, P., Sharp, J.M. and Motara, Y. (2003) 'Cross-functional team working for overall equipment effectiveness (OEE)', *Journal of Quality in Maintenance Engineering*, vol. 9, no. 3, pp. 223-238.

Banks, J., Carson, J.S., Nelson, B.L. and Nicol, D.M. (2005) *Discrete-Event System Simulation*, 4th edition, Upper Saddle River, New Jersey: Pearson Prentice Hall, 0-13-129342-7.

Bengtsson, N., Shao, G., Johansson, B., Lee, T.Y., Leong, S., Skoogh, A. and Mclean, C. (2009) 'Input Data Management Methodology For Discrete Event Simulation', Proceedings of the 2009 Winter Simulation Conference, Austin, Texas, 1335-1344.

Brooks, R.J. and Tobias, A.M. (2000) 'Simplification In The Simulation Of Manufacturing Systems', *International Journal of Production Research*, vol. 38, no. 5, pp. 1009-1027.

Castino, F. and Watson, L. (1991) 'Computer Simulation: It's use in planning material handling systems', *Logistics Information Management*, vol. 4, no. 1, pp. 8-17, Available: 0957-6053.

Chaharbaghi, K. (1990) 'Using Simulation to Solve Design and Operational Problems', *International Journal of Operations & Production Management*, vol. 10, no. 9, pp. 89-105, Available: 0144-3577.

Chwif, L., Ribeiro Pereira Barretto, M. and Paul, R.J. (2000) 'On Simulation Model Complexity', Proceedings of the 2000 Winter Simulation Conference, Orlando, Florida, 450-455.

De Groote, P. (1995) 'Maintenance performance analysis: a practical approach', *Journal of Quality in Maintenance Engineering*, vol. 1, no. 2, pp. 4-24.

Fishman, G.S. (2001) *Discrete-event simulation: modeling, programming, and analysis*, Berlin: Springer-Verlag, 0-387-95160-1.

Freivalds, A. and Niebel, W.B. (2003) *Niebel's Methods, Standards, and Work Design*, 12th edition, New York: McGraw-Hill Higher Education.

Grajo, E.S. (1995) 'Strategic layout planning and simulation for lean manufacturing a LAYOPT tutorial', Proceedings of the 1995 Winter Simulation Conference, Washington, 510-514.

Hendry, L.C. (1998) 'Applying world class manufacturing to make-to-order companies: problems and solutions', *Internation Journal of Operations and Production Management*, vol. 18, no. 11, pp. 1086-1100.

Incontrol Simulation Solutions (2010) *ENTERPRISE DYNAMICS Simulation Software*, [Online], Available: http://www.incontrolsim.com/index.php/Products/ [23 May 2010].

Jonsson, P. and Mattsson, S.-A. (2009) *Manufacturing, Planning And Control*, London: McGraw-Hill Education, 978-0-07-711739-9.

Kasul, R.A. and Motwani, J.G. (1997) 'Successful implementation of TPS in a manufacturing setting: a case study', *Industrial Management and Data Systems*, vol. 7, pp. 274-279.

Kuhn, A. and Schmidt, R. (1988) 'Simulation of logistic systems', *Logistics Information Management*, vol. 1, no. 1, pp. 47-52.

Kumar, S. and Phrommathed, P. (2006) 'Improving a manufacturing process by mapping and simulation of critical operations', *Journal of Manufacturing Technology Management*, vol. 17, no. 1, pp. 104-132.

Liker, J.K. (2004) *The Toyota Way - 14 Management Principles from the World's Greatest Manufacturer*, New York, New York: McGraw-Hill, 0-07-139231-9.

McAdam, R., Stevenson, P. and Armstrong, G. (2000) 'Innovative Change Management in SMEs: Beyond Continuous Improvement', *Logistics Informantion Management*, vol. 3, no. 13, pp. 138-149.

O'Kane, J.F. (2003) 'Simulation as an enabler for organizational excellence', *Measuring business excellence*, pp. 12-19.

Robinson, S. and Bhatia, V. (1995) 'Secrets of Successful Simulation Projects', Proceedings of the 1995 Winter Simulation Conference, Arlington, Virginia, 61-67.

Rose, O. (2008) 'Benefits and Drawbacks of Simple Models for Complex Production Systems', in *Managing Complexity: Insights, Concepts, Applications*, Berlin/Heidelberg: Springer.

Sánchez, P.J. (2007) 'Fundamentals of Simulation Modeling', Proceedings of the 2007 Winter Simulation Conference, Washington DC, 54-62.

Sanchez, S.M. (2007) 'Work smarter, not harder: Guidelines for designing simulation experiments', Proceedings of the 2007 Winter Simulation Conference, Piscataway, 84-94.

Sargent, R.G. (2005) 'Verification and Validation of Simulation Models', Proceedings of the 2005 Winter Simulation Conference, Arlington, 130-143.

Schmenner, R.W. and Vollman, T.E. (1994) 'Performance Measures: Gaps, False Alarms and the "Usual Suspects"', *International Journal of Operations & Production Management*, vol. 14, no. 12, pp. 58-69.

Skoogh, A. and Johansson, B. (2008) 'A Methodology for Input Data Management in Discrete Event Simulation Projects', Proceedings of the 2008 Winter Simulation Conference, Miami, Florida, 1727-1735.

Zaerpour, N., Rabbani, M., Gharehgozli, A.H. and Tabakkoli-Moghaddam, R. (2008) 'Make-to-order or make-to-stock decision by a novel hybrid approach', *Advanced Engineering Informatics*, pp. 186-201.