POLITECNICO DI MILANO

School of Industrial and Information Engineering

Master of Science in Management Engineering

Academic Year 2017/2018

An improved model for baton zone bumping considering release period, idleness rate of the worker, transfer time and worker efficiency

Supervisor:
Prof. Alberto Portioli Staudacher

Assistant Supervisor:
Eng. Federica Costa

Author:
Zhihui Huang
Matricola:875422

CATALOGUE

List of Figures

List of Tables

List of Nomenclatures

DRC	Dual-resource-constrained
Effi. 100%	Bumping Model with 100% worker efficiency
Effi. 85%	Bumping Model with 85% worker efficiency
Effi. 70%	Bumping Model with 70% worker efficiency
ERP	Enterprise resource planning
FCFS	First come first served
GTT	Gross throughput time
JIT	Just in time
MRP	Material requirements planning
MTO	Make to order
OPT	Optimized production technology
P-240	Bumping Model with order released two times per day
P-480	Bumping Model with order released one time per day
PR	Periodic release
PSP	Pre-shop Pool
SFT	Shop floor throughput time

SMEs	Small and medium-sized enterprises
TOC	Theory of constraints
TT-5	Bumping Model with transfer time equals to five minutes
TT-10	Bumping Model with transfer time equals to ten minutes
TT-15	Bumping Model with transfer time equals to fifteen minutes
TT-30	Bumping Model with transfer time equals to thirty minutes
TT-70	Bumping Model with transfer time equals to seventy minutes
WIP	Work-in-process
WLC	Workload control

ABSTRACT (IN ENGLISH)

Since globalization, small and medium-sized manufacturing companies have grown rapidly and become an inestimable force in the economy. SMEs mainly focus on the make to order (MTO). With the increasing personalization and diversification of customer needs, the competition is becoming more and more fierce. Thus, how to satisfy the customer's need effectively and efficiently becoming a tough problem to the production system of MTO companies.

This thesis aims to introduce relevant concepts and methods proposed by predecessors about the capacity adjustment of the production system in MTO companies. Also, an innovative method, the Bumping Model, which is derived from the Baton Zone Balancing, is introduced to improve the performance of the production system. After defining this model, the corresponding "where rule", "when rule" and "who rule" in a Dual Resources Constrained (DRC) system about worker flexibility are determined. Then, a simulation via the Python language is done to prove the benefits of the Bumping Model compared with the traditional Static Model. Besides, relevant factors such as release period, worker efficiency and transfer time etc. are tested in the model. After analyzing the corresponding results, this thesis tries to find a new improved model for baton zone bumping.

Keyword: Make to order, Worker flexibility, Baton Zone Balancing, Bumping Model

ABSTRACT (IN ITALIAN)

Con l’avvento della globalizzazione, le piccole e medie imprese manifatturiere sono cresciute rapidamente e sono diventate una forza notevole dal punto di vista economico. Le PMI si concentrano principalmente sulla produzione make-to-order (MTO). Con la crescente personalizzazione e diversificazione delle esigenze dei clienti, la concorrenza sta diventando sempre più agguerrita. Pertanto, una delle principali sfide per questa tipologia di aziende risiede nel soddisfare le richieste dei clienti con efficacia e rapidità.

Questa tesi ha lo scopo di analizzare i concetti e le metodologie più rilevanti in merito all'adeguamento della capacità del sistema produttivo nelle imprese che adottano un sistema di produzione MTO. Fra i metodi più innovativi in analisi vi è il Bumping Model, che deriva dal Baton Zone Balancing, ed è atto a migliorare le prestazioni di tali sistemi di produzione. Inoltre, è stata fatta un’analisi delle regole per determinare con precisione il livello di flessibilità dei lavoratori nei sistemi a doppia risorsa vincolata (DRC). Viene successivamente presentata una simulazione tramite il linguaggio di programmazione Python atta a dimostrare i vantaggi del modello Bumping rispetto ai modelli statici tradizionali. Tra l’altro, sono stati testati anche fattori come il periodo di rilascio, l'efficienza del lavoratore e il tempo di trasferimento nel modello. Dopo aver analizzato i risultati corrispondenti, questa tesi si è posta l’obiettivo di trovare un nuovo modello migliorato per quanto riguarda la “baton zone bumping”.

EXECUTIVE SUMMARY

Introduction

Make to order (MTO) is a manufacturing management system where the production of a part only starts when the order is received by the company. Usually, the production system of MTO company faces the interference of many uncertain factors such as the changing needs of external customers, the shortage of internal resources, the imbalance of workload, the bottleneck etc. These factors make the production system very dynamic, versatile and complex. Faced with the challenge, the MTO companies usually adopt a job shop configuration, which allows for more types of product requiring different stations with different routings. However, the increasing trend to implement lean manufacturing has brought many companies to lean streamline the process, moving toward flow shop configurations, where products are processed by the same stations in the same sequence. Lean manufacturing is a popular topic that always discussed by many scholars and practitioners in manufacturing industries. In the book *The lean practitioner's filed book: Proven, Practical, Profitable and Powerful Techniques for Making Lean Really Work*, a new concept of "Baton Zone Balancing" is proven to be an effective technique to ensure pull production and "one-piece flow" for a flow shop. This method improves the utilization of workers and reduces the WIP effectively compared with traditional station balancing.

Based on the trend of adopting the flow shop in MTO companies and the effective line balancing technique (the Baton Zone Balancing), this thesis proposed an innovative method of capacity adjustment in flow shop of MTO companies: The Bumping Model. Then, a simulation is done to prove the benefits of the Bumping Model compared with the traditional Static Model. Furthermore, relevant factors are tested to assess the impact of them on the Bumping Model. After analyzing the corresponding results, this thesis tries to find a new improved model for baton zone bumping.

Following the aim of the thesis proposed above, two research questions were considered in the thesis:

Question 1: What are the contributions of the Bumping Model to performances of the flow shop?

Question 2: How do factors, release period, worker efficiency, transfer time and the processing time, affect performances of the flow shop in the Bumping Model?

Literature review

At first, because the WLC was proved to the effective method to manage and control the production system in the MTO companies, the thesis discusses some concept about it. But we only talk mainly about the output control in the WLC, because the output control is focusing on the capacity adjustment. The workload control is derived from the I/O concept proposed by

Wight in 1970, the task waiting for delivery is first placed in the task pool at the front of a shop floor. According to a certain task delivery mechanism, the tasks in the task pool are gradually and orderly placed into the shop floor to alleviate the phenomenon of congestion and work-in-process accumulation in the production process of the shop floor. Simplifying the production process at the shop floor into a series of short, easy-to-manage processing queues. In general, the job's input to the shop floor is controlled according to the shop's work center capability (output rate) to adjust and maintain WIP at a stable level.

Workload Control integrates two control mechanisms: (a), input control, to regulate the inflow of work to the system; and (b), output control, which uses capacity adjustments to regulate the outflow of work from the system. Much Workload Control research has focused on input control, while output control has been largely neglected. Normally, in the order entry phase, companies prefer to accept all the orders arriving in a period and still deliver them within due dates issued. The reason is that they can use output control methods to adjust their production capacity for a period of time (Kingsman, 2000) or they will try to extend capacity or at most to delay due date. In WLC literature, we find many researchers are focusing on accounting the workload while the description of capacity is simplified to standard output rates (Yuan 2017). However, real production capacity may be very different from the standard one while it depends on the resources in the production system. It means that companies can easily adjust their production capacity by changing these resources. There are three approaches used in the production management:(a), working overtime, (b), relocating workers, and (c), subcontracting.

The workforce in capacity adjustment seems to play an important role. Actually, in a real production system, the configurations of the production system are more complicated than simulation one, the workers and machines will bring an incredible impact on the performance of the company. Many researchers assume station is the minimum unit of a system which is a unique entity with a fixed capacity and it can work for all the time. However, as is known to us, there are machines and operators within stations. Machines need workers to run, while workers need machines and tooling in order to handle jobs and to perform tasks. Compared with operators, machines are more specialized, so they are usually idle or under-loaded for the dynamic demand of the marketing. As a result, the number of operators is frequently lower than the number of machine and station. In this case, some operators would work on different stations, this kind of ability is called worker flexibility which gives the production system a flexible workforce. Usually, a flexible workforce allows the manager of the shop floor to move around to respond to temporarily overload department. Actually, worker flexibility is the most relative topic for it can give more approaches to adjust capacity. However, it is obvious that if a company want to get a flexible workforce, it should invest in the cross-training program to make the workers be multi-skilled (Hopp & Van Oyen, 2004). There are others opinions to support the implementation of worker flexibility, one is proposed by the authors in Malachowski (2015), worker flexibility is considered as a competitive advantage while machines can be bought by company provided the availability of enough capital. Others consider more on the economic view, they stated workers are less expensive than machines and they easily allow switching from the production of one product to another one.

The machines and workers are considered as two constraints of a production system. Thus, a concept called Dual-Resource-Constrained is proposed by scholars. After studying on the topic for several years by scholars, the research can be summarized in the following three rules:

- Where' rules or assignment rules: they define the policy by which a worker is assigned to a particular station.
- 'When' rules or control rules: they define when a worker is available for transfer to another department.
- 'Who' rules: they define which worker has to be transferred to another department.

As we all know that manufacturing is experiencing a dramatic change nowadays. The market is not like the last century; it is gradually dominated by customers. In this background, many manufacturing companies are trying to change the production mode and introducing more suitable managerial philosophy. "Lean" as a famous and proved effective philosophy get more and more attention from companies and the concept called "lean manufacturing" is becoming popular in manufacturing industries.

When talking about lean manufacturing, four principles should be mentioned: (a) flow, (b) pace, (c) pull and (d) zero defects. Thanks to these principles, more and more companies are optimizing their production system. In order to test whether a production system is lean or not, two items on your lines or in the office process are examining. These are excessing inventory and idle time.

From the game shown in the book: “The Lean Practitioner’s Field Book: Proven, Practical, Profitable, and Powerful Techniques for making Lean Really Work”, we witness the improvement when the production system develops from batch to station balancing flow shop and then to baton zone balancing flow shop step by step. Station balancing flow shop has many constraints that make the line always unbalanced, some of them are summarized in following: (a), The major problem with this approach is that people are not robots and we all work at different rates, sometimes a worker will in a low spirit or a worker will become more efficient; (b), if the line needs to run more than one model or type of product, it would also be difficult to keep the labor balanced; (c), because station balancing was designed based on a set number of operators, if an operator is missing, it creates problems; (d), most station balanced lines are sit-down lines, in which the product flow is normally lost; (e), sit-down stations are 10%–30% less efficient than stand-up and walking lines; (f), the next problem with station balanced system is when starting a production flow. In this case, each operator has to sit and wait until enough inventory is in the line to start working. If there is a large amount of labor in a unit, two hours, for example, some operators near the end of the line may sit idle for hours; (g), if we have a fast worker in a station balanced line, he/she will finish the work and sit idle, but we still have to pay the same salary; (h), another problem we confront is where a line is balanced to Takt time, there is hardly situation cycle time is equal to Takt time perfectly; (i), another issue is when the total labor changes because of a different product or even sometimes a slower operator. Rearranging a line means an obvious violation of Lean flow.

Considering the process of Baton zone balancing, they can solve these problems in station balancing effectively, and here the algorithm of it can explain in following way: the entire baton handoff cycle is triggered by the completion of a product at the end of the line. Therefore, it is a true pull system. When the operator at the end of the line (let's call them $operator\ 3$) completes the product and places it in a bin, tray, or hopefully the shipping box, then $operator$ 3 moves back down the line to the operator immediately prior to him ($operator$ 2) and takes his product from him regardless of where it is in the process. It is very difficult to get operators to hand off their parts initially. They want to finish their work at the station first, which will cause $operator$ 3 to be idle until they ($operator$ 2) hands it off. When the product is given to $operator$ 3, $operator$ 2 then moves down to take the part from $operator$ 1, this is considered a handoff or what we call bumping. This bumping down the line causes a baton handoff and bump by each subsequent operator on the assembly line until the operator closest to the beginning of the assembly line no longer has a product to work on. This first operator ($operator$ 1) will then start another product at the beginning of the line. The rule is that the operators on the line must continue to work on the product in their hands until they are bumped or complete a product.

Methodology

In order to investigate the issues presented in the research questions, a simulation model was developed. The Python programming language is used to write the model, and thanks to the external libraries SimPy which provide a simple way to create the simulation environment.

This model simulates a shop floor made of 5 stations, each station includes one machine and one worker. The flow of the shop is directed which means that the job should be processed in fixed routing namely from station 1 to station 5 (a pure flow shop). One pool is allocated to each station, which plays the role to decouples the production flow from the station upstream to the one downstream. In addition, there is a Pre-shop pool in the system plays the role to store orders to ensure the periodical release (PR) in the model.

The workers on the shop floor, according to the concept of Baton Zone Balancing, can move to another station to catch the job. In order to simulate the bumping process, when the worker downstream finish his current job and there is no job in the upstream pool, he will go to the station upstream to catch the current job of the previous worker, the model replace the movement of worker with the mandatory completion of job on the previous station whenever is finished or not. In order to recognize this type of workforce from the tradition one that we will compare with, we called this type of workforce “Bumping”. There are only two kinds of the workforce in our simulation model:

- Static – workers are constrained at their main station and they cannot provide extra capacity to the others.
- Bumping – workers are assigned to a main station in the simulation model, but they can move to the station upstream to catch the job in a practical scene. Here, in order to simulate the algorithm easily, the movement of job replaces that of worker.

Most researches on worker flexibility are substantially studying on the question: when the workers and which of them(who) will go to which station(where) in order to improve the capacity of a production system, and the researchers have proposed many algorithms to the questions. In fact, the Bumping Model provides an answer to the question. The answer is showed as following:

When the **worker** $\boldsymbol{i}\,(1 < i \leq 5)$ in the shop floor finishes his current job and there is no workload in the pool of the main station and the **worker** $(\boldsymbol{i-1})$ is processing a job in **station** $(\boldsymbol{i-1})$ **(when)**, **worker** $\boldsymbol{i}$ **(who)** will go to catch the job that **worker** $(\boldsymbol{i-1})$ is doing on **station** $(\boldsymbol{i-1})$ **(where)**, otherwise waiting.

About the order release method in the thesis, periodically release (PR) as the common approach in the real world is used. Finally, there are two kinds of flow shop are tested in the thesis:(a), PR + Static workforce and (b), PR + Bumping workforce.

Furthermore, in order to find an improved Bumping Model, some factors really making sense are tested in the thesis, there are: (a), ***release period,*** the thesis test two values of release period: 240 minutes and 480 minutes. They correspond releasing two times a day and one time a day, they are common in the real world; (b), ***worker efficiency,*** as we all know, different level of worker efficiencies means different output rate and also it will give impact on the cost to train the workers. In this sense, it is very meaningful to test the impact of workforce efficiency on the performance of the flow shop in Bumping Model and try to find the best value of the factor to improving the performance of a production system with a low cost. The thesis has tested four values of workforce efficiency (100%, 85%, 70%, and 55%); (c), ***transfer time,*** this factor is unavoidable in real life, if the worker wants to catch a job from the previous worker, he has to take time to go to the previous station and do it. So it is very necessary to consider the transfer time. In this thesis, four values of transfer time are set and tested: 5minutes, 10minutes, 15minutes, and 30minutes; (d), ***processing time***, Portioli (2012) proposed that the processing times of jobs should be computed according to the same stochastic distribution and with the same parameters for all stations for sake of simplicity and in order to avoid the creation explicit bottlenecks in the system. The distribution adopted is a log-normal with mean 30 minutes and variance 900 $minutes^2$, which observations can range between 0 and 360 minutes. But, for the properties of Baton Zone Balancing, we should arrange the workers along the production line according to increasing worker efficiency for it can boost the bumping process. Besides, it is obvious that if we apply the Bumping Model to capacity adjustment, the $worker$ 1 and $worker$ 5 are different with other three workers. So, in order to get some deeper conclusion, the process times of jobs in five stations are modified before the jobs are released to the shop floor, but only the mean processing times in each station are changed. The sum of mean processing time and the variance are not changed. What is most important is that the modification algorithm depends on the results of the Bumping Model in original processing time setting. Aim at testing whether the Bumping Model can benefit from a modified processing time. Therefore, another two experiments (the Bumping Model with original processing time and the Bumping Model with modified processing time) have been done in the thesis.

In conclusions, there are mainly following experiment with different configurations:

Order release method	**Type of workforce**	**Release period**	**Worker efficiency**	**Transfer time**	**Processing time**
PR	Static	480 minutes	100%	0	Original
PR	Static	240 minutes	/	0	Original
PR	Bumping	480 minutes	100%	0	Original
PR	Bumping	240 minutes	100%	0	Original
PR	Bumping	480 minutes	85%	0	Original
PR	Bumping	480 minutes	70%	0	Original
PR	Bumping	480 minutes	55%	0	Original
PR	Bumping	480 minutes	100%	5 minutes	Original
PR	Bumping	480 minutes	100%	10 minutes	Original

PR	Bumping	480 minutes	100%	15 minutes	Original
PR	Bumping	480 minutes	100%	0	Modified

Table 0.1: The main experiments in the thesis

About the code written in the thesis, there are six types of object in the simulation model: ***job, jobs generator, pool, worker, machine, and order releaser,*** but for the reason of the difference between stations in the Bumping Model, $station\ 1$ and $station\ 5$ are different with $station\ 2$, $station\ 3$ and $station\ 4$, therefore, machine in our model are divided into five classes. $machine\ 1$ can only be interrupted by $machine\ 2$ while $machine\ 2$, $machine\ 3$ and $machine\ 4$ can not only be interrupted by others machines but also can interrupt other machines. For $machine\ 5$, it can only $interrupt\ machine\ 4$.

Results and conclusions

1. The flow shop in the Bumping Model of MTO company surpasses that in the Static Model

The results confirm the beneficial effects of Bumping Model on each system performance evaluated in the study (GTT, SFT, average tardiness, lateness and tardy etc.). The GTT decreases by 40% and SFT decreases by 44% compared with the Static Model. The average lateness, tardiness and tardy reduce more dramatically around by 200%, 96%, and 91% respectively. And we found in the thesis the reason for this huge improvement is the little increments of workers' utilization and around 50% decrement of the average workload in the flow shop during the simulation process. Thus, there are many reasons implement the Bumping Model in the flow shop to improve its performances. This methods of capacity adjustment can easily reduce the WIP and improve the utilization of human resources even if it is not so much. Besides, from the results of the simulation, the idleness rate of $worker\ 1$ is extremely higher than those of other workers, this phenomenon is corresponding to the attribute of a Baton Zone Balancing line which always put the worker in low efficiency in the start of the line. Totally, the thesis gives us a satisfactory answer to question 1 that the Bumping Model brings a distinct improvement of the production system. So it is meaningful for the manager to implement this kind of capacity adjustment in real life.

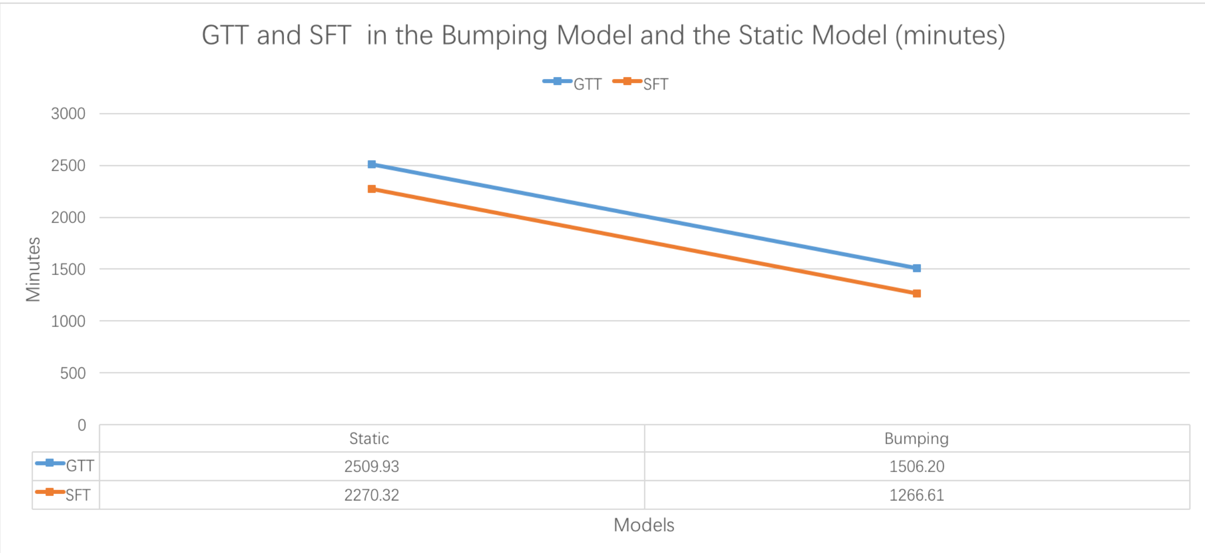


Figure 0.1: GTT and SFT in the flow shop with the Bumping Model and the Static Model

In the end, even if the Bumping Model can improve the performance of the manufacturing system, attaining a flexible workforce means more spend of money and time in the practical world. How to trade off the improvement of performance and the cost spent to training is more realistic to managers. Besides, although the average total workload of the flow shop in the Bumping Model reduces, the variance increases a lot, which will make it tough to predict the WIP during the production process.

2. ***An improved flow shop in the Bumping Model should be in a shorter release period, proper worker efficiency (85%) on the bumping work, lower transfer time and a decreasing trend of processing time along the production flow.***

Impact of worker efficiency on the bumping work: When trading off the cost of training and system's performance, study on worker efficiency is more meaningful. With the information drawn from the simulation results, following conclusions are attained on worker efficiency:(a), the lower worker efficiency brings increasing GTT and SFT; (b), the worker efficiency has a huger influence on SFT than GTT. (c), with the stable decreasing of worker efficiency, the GTT and SFT increase in a higher and higher speed. (d), the Tardiness, Lateness and Tardy increase in the same speed with the decrease of the worker efficiency. When we go deeper to find the reason for these conclusions, we find that the idleness rates of workers decrease and the average workload in the flow shop system increase, while it is interesting that the standard deviation of workload in the system decreases when worker efficiency changes from 100% to 85%, but increases when worker efficiency changes from 85% to 70% and then to 55%. It means that the WIP become more stable when worker efficiency is equal to 85% compared with other values of worker efficiency, which gives the manager a signal to lower the efficiency properly to keep a stable WIP when implementing the Bumping Model. Finally, considering the cost and performances of the system, we propose that the most suitable value of worker efficiency in Bumping Model is 85%.

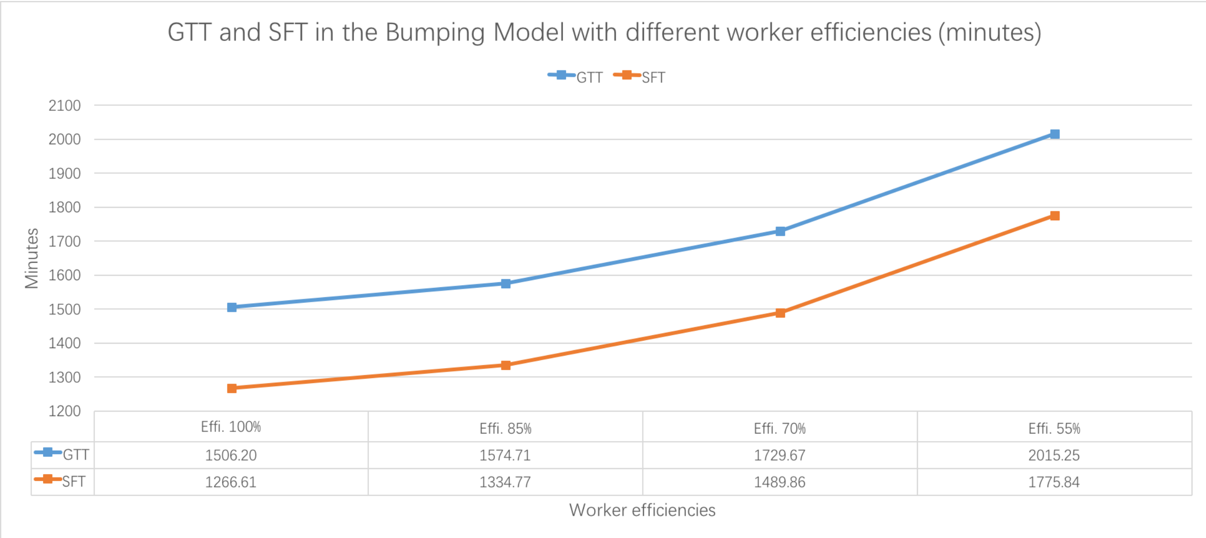


Figure 0.2: GTT and SFT of the flow shop in the Bumping Model with different levels of worker efficiency

Impact of release period: Periodical release (PR) is a common order release method, the release period as the important parameter in PR is tested in the thesis. The results illustrate that GTT and SFT decrease by 45.8% and 48.3% respectively when orders release two times a day (P-240) in the Bumping Model, they decrease by separately 40% and 44.2% when orders release one time a day (P-480) in the Bumping Model compared with the Static Model. Thus, the conclusion is the Bumping Model benefits from the shorter release period. And when calculating the decreasing rates of GTT and SFT in Bumping Model from P-480 to P-240, we find the decreasing rates of GTT and SFT are 17.15% and 10.97%. The results let us believe that shorter release period gives a more improvement to GTT which represents the lead time of an order. However, when we are finding the reasons for this improvement, it surprises us that the average idleness rate of workers doesn't change when releasing orders two times a day, even if GTT and SFT are decreasing. So the utilization of workers is not the reason, actually, the lower average workload in the flow shop system gives us the answer. Compared with releasing one time a day in the Bumping Model, the average workload in the system decreases by around 20%. Besides, the workload load in the flow shop is more stable during the simulation process and its standard deviation decrease by about 53% (from 1044 minutes to 487 minutes). Also, the percentage of time worker spend on the bumping work increase. These phenomena show the bumping process is good for reducing the WIPs and facilitating the flow of orders in the production system. Finally, it is clear that a shorter release period in Bumping Model brings higher improvement.

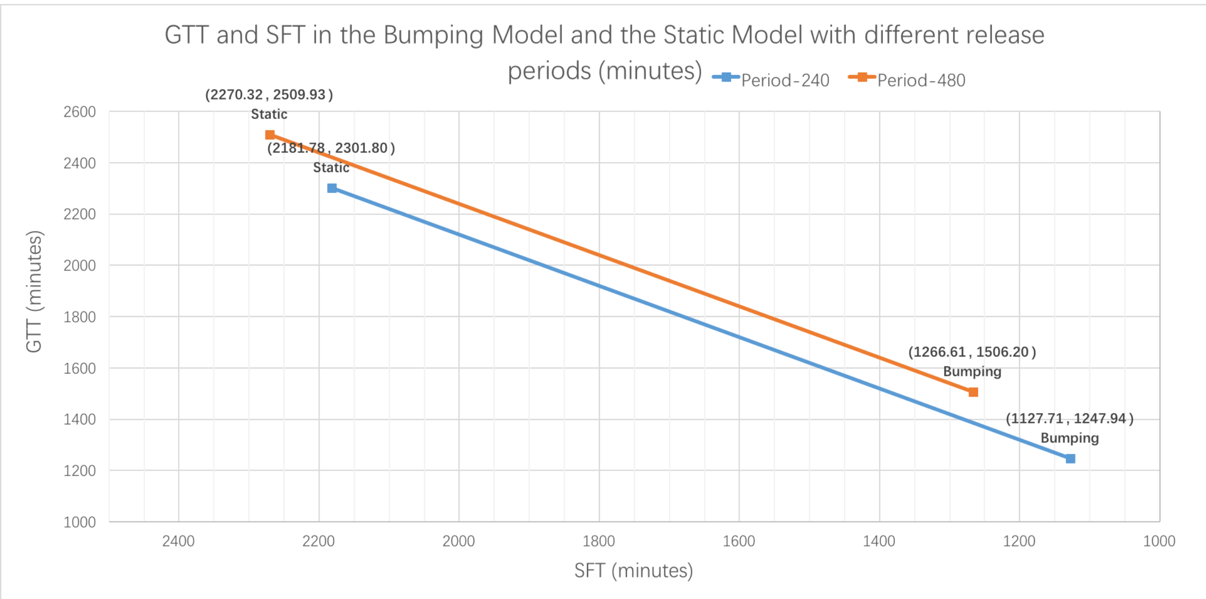


Figure 0.3: GTT and SFT of the flow shop in the Bumping Model and the Static Model with different release period (240 minutes vs. 480 minutes)

Impact of transfer time: The thesis actually proposed four opinions on the transfer time in the Bumping Model:(a), GTT and SFT increase when introducing the concept of transfer time; (b), the increments of GTT and SFT are very stable when rising the value of transfer time; (c), an increasing transfer times give the flow shop an increasing WIPs and a decreasing percentage of time worker spend on the bumping work; (d), only the utilization of $worker\ 1$ increases while those of all of the other workers decrease. In addition, based on the linear relationship between GTT, SFT and transfer time, we get the maximum value of transfer time which will remove the benefit of Bumping Model. It is around 70 minutes. However, when we give this value to transfer time in the Bumping Model and simulating, there are still decrements of GTT and SFT compared with the Static Model, about 8% and 9% respectively. We believe that even if the long transfer time, the variance of the processing time of a job is enough bigger to get the benefit from the bumping process. In other words, there is more unpredictable order in real life, therefore, a bumping process definitely can give an improvement to the production system. The proper transfer time just can release the impact but can’t remove it. Finally, about the fluctuation of total workload of the system in the Bumping Model with different values of transfer time, it doesn’t experience any change.

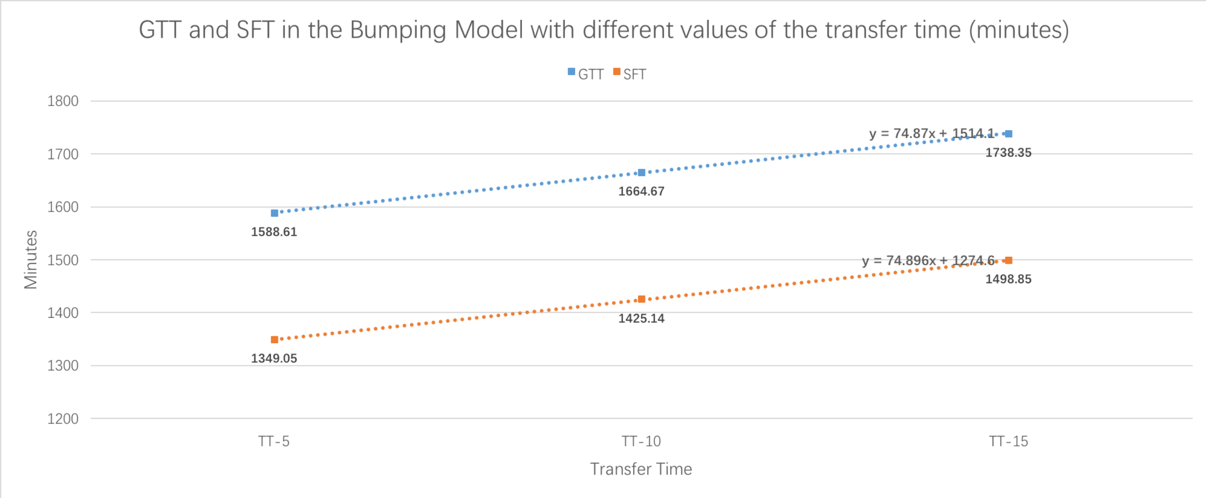


Figure 0.4: GTT and SFT of the flow shop in the Bumping Model with different values of transfer time

Combinations of the three factors: After testing the three factors in a combined way, two conclusions following are attained:(a), the Static Model is always worse than the Bumping Model; (b), any two factors will relieve the impact of another factor on the performances of the system in the Bumping Model. For these conclusions, we can explain in this way, at first, the release period is the most direct factor to affect the frequency of bumping, lower release period means more bumping processing. The transfer time and the worker efficiency are not the direct reasons to influence the times of bumping process, in contrast, they just influence the time worker spending on the bumping process and then influence the frequency of bumping process. So, we can say that the release period is the most important factors among the three factors, and it will influence the impact of other factors on the performance of the Bumping Model.

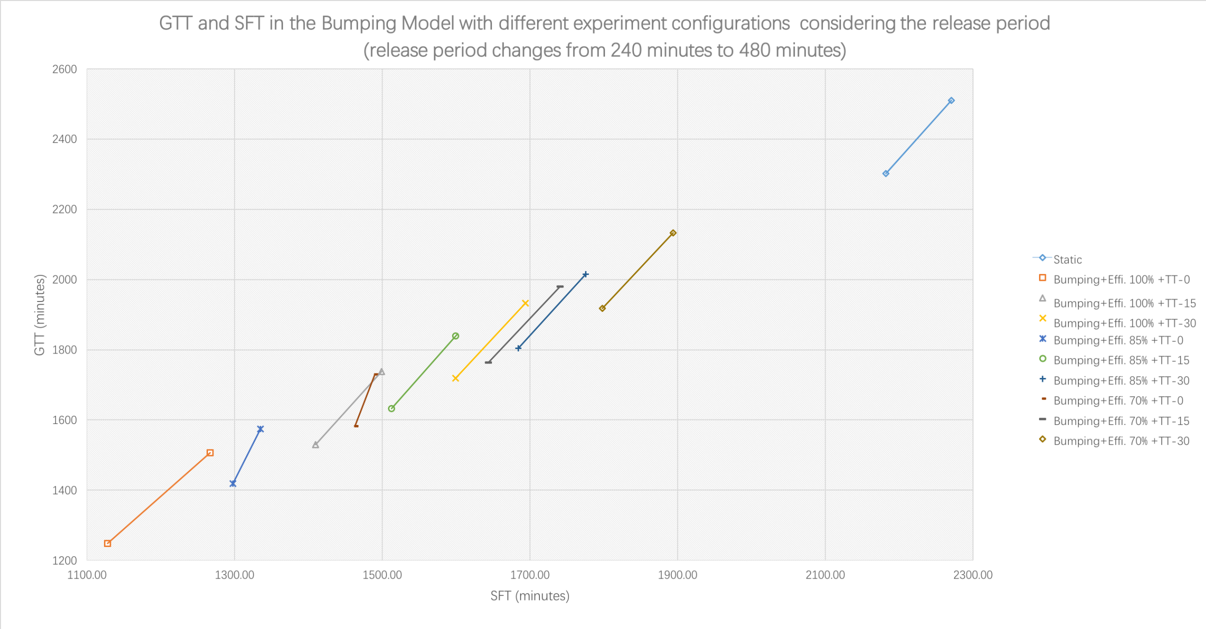


Figure 0.5: GTT and SFT of the flow shop in Bumping Model with different combinations

Impact of the processing time: The thesis proposed a modified approach to the processing time which is just an assumption and an attempt. After simulating, the results give us anticipating results, the GTT and SFT decrease because of the low WIPs. So, we can conclude that implementing the Bumping Model give us more room to improve the performances of a production system if we can rearrange the production flow. Precisely, we can say that the Bumping Model benefit from the decreasing trend of processing time along the production flow.

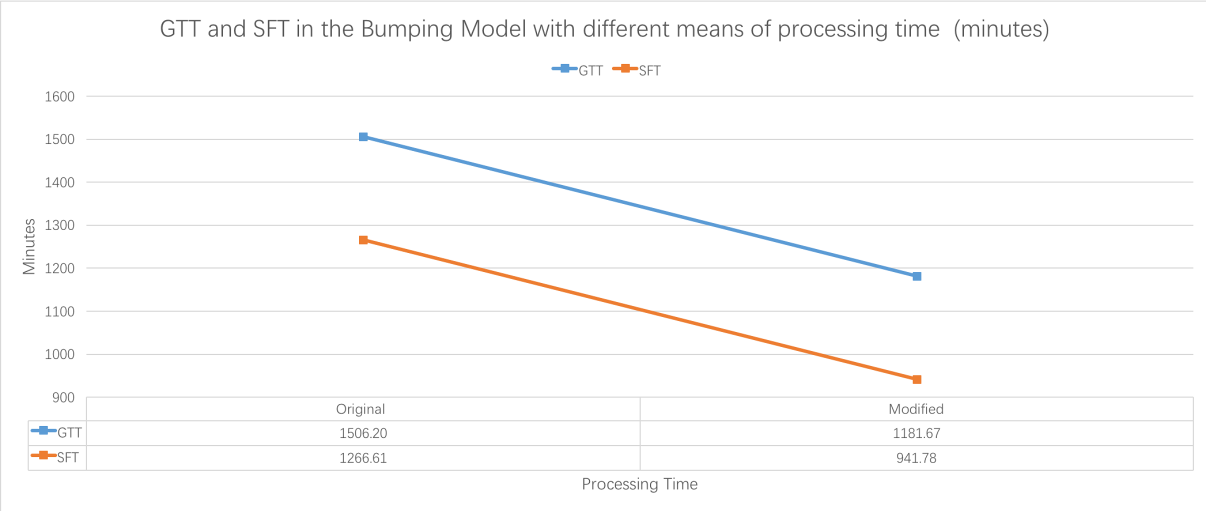


Figure 0.6: GTT and SFT of the flow shop in the Bumping Model with different processing time

Limitations and future researches

The Bumping Model gives a new method of capacity adjusting in the flow shop. And the conclusion is that the new model improves the performances of the flow shop. At the same time, the thesis studies on the impacts of some factors on performances of the flow shop in order to provide an improved Bumping Model. However, there are some limitations in the thesis: (a), the results got from the simulation depend on the assumptions made for all the parameters of the production process such as the arrival rate distribution, processing time distribution and the contractual due dates policy. (b), only one type of shop configuration (pure flow shop) is considered in the thesis, all the orders in the model are processed in the same route. (c), due to the constraint of time, the thesis only considers one level of workforce flexibility with four levels of workforce efficiency; (d), a modified attempt on processing time is done in the thesis, but it is not systematic and rigorous research, so there should be more scientific research on this topic; (e), the thesis only combines the periodical release (PR) with the Bumping workforce. Based on these limitations, future researches on the Bumping Model includes: (a), the Bumping models with higher levels of worker flexibility; (b), the Bumping models with different shop configurations; (c), the Bumping models with different configurations of processing time and (d), combining other order release methods with the Bumping Model.

1 INTRODUCTION

1.1 Background

Make to order (MTO) is a manufacturing management system where the production of a product only starts when the order is received by the company. Usually, the production system of MTO company faces the interference of many uncertain factors such as the changing needs of external customers, the shortage of internal resources, the imbalance of workload, the bottleneck etc. These factors make the production system very dynamic, versatile and complex. Therefore, the effective management and control of the production process in MTO company are critical in order to ensure a smooth, balanced and efficient operation of the manufacturing system.

In face with the challenge, the MTO companies usually prefer a job shop configuration, which allows for several types of product requiring different stations with different routings. However, in order to satisfy the need of customers, configuration and optimization the use of resources, production management modes and methods are constantly updated and transformed: from MRP to ERP, from JIT to Lean Production and from Optimized Production Technology(OPT) to Theory of Constraints(TOC). Nowadays, the increasing trend to implement lean manufacturing recently has brought many companies to lean streamline the process, moving toward flow shop configurations. These are characterized by products being processed by the same stations in the same sequence.

Lean manufacturing is a hot topic that always discussed by many scholars and practitioners in manufacturing industries. In the book *The lean practitioner's filed book: Proven, Practical, Profitable and Powerful Techniques for Making Lean Really Work*, we witness the improvement of the production system from job shop to flow shop. Besides, there is also a deeper analysis of the flow shop, station balancing as a very traditional flow balancing approach is discussed. In a flow shop considering station balancing, when receiving a job, managers prefer to arrange the station with the same cycle time (CT), in order to satisfy the demand, the cycle time has to be lower than or equal to the Takt time (TT). Through balancing the workload in each station, it seems it is sufficient to achieve a one-piece flow and the aim of a shorter lead time. However, it is obvious that there are two main factors can affect the balance of the flow: (a), the factors of workers, such as worker efficiency, it would vary every time for different reasons; (b), the factors of orders, for instance, types of order, MTO companies are featured with a highly personalized order, which means there will different workload in each station for different orders. In this case, it seems that station balancing is not a proper method for the flow shop of MTO company.

Fortunately, the author of that book, Protzman (2016), proposed a new technique called Baton Zone Bumping to ensure pull production and "one-piece flow"[2]. Both "pull production" and "one-piece flow" are the most important principles in lean manufacturing. The Baton Zone

Balancing describes a line where the workers are assigned along the line according to an increasing worker efficiencies and the workers from downstream can catch the tasks from upstream, hence the idle time and WIPs can be reduced. The catching or bumping process will happen only when the downstream workers finished their jobs or when there is interruption from the downstream. This method is proved to improve the utilization of workers and reduce the WIPs effectively compared with the traditional station balancing in the flow shop. The bumping process gives the inspiration to introduce a new method into capacity adjustment of the flow shop in MTO: The Bumping Model.

This thesis introduces the relevant concepts within the capacity adjustment and compares performances of the flow shop in Bumping Model with that in the traditional Static Model. Then, some factors are tested to find their impacts on the Bumping Model. Finally, based on some assumptions, an improved baton zone bumping model is proposed.

1.2 Research questions

As aforementioned, the Bumping Model is a new algorithm that will apply to control the outflow of MTO production system. Following the aim of the thesis, two research questions were considered in the thesis:

Question 1: What are the contributions of the Bumping Model to performances of the flow shop?

Question 2: How do factors, release period, worker efficiency, transfer time and the processing time, affect performances of the flow shop in the Bumping Model?

1.3 Research methodology

A simulation model has been used to test the Bumping Model and measure flow shop performances under every environmental condition and according to all the possible managerial decisions taken into consideration in this study. The model simulates a flow shop with five stations, each station consists of one machine and one worker. The approach of releasing orders is periodic release (PR). After arrival, orders are stored in a Pre-shop pool before being released to the shop floor. The orders are release one time every day, but the release period is also a factor that is tested in the thesis. Worker performs according to the algorithm of the Bumping Model. There is a bumping process if the worker moves to another station.

Stochastic orders' arrival rate and processing time at each station are considered in order to create more realistic experimental configurations. The production system of MTO, actually, is usually characterized by different job types requiring different processing, routing within the shop floor and production time. Stochastic arrivals rate and processing times allow to include these properties in the model.

With the reason that delivering on time is one of the most important requirement to MTO company, four system performances are used to evaluate and compare. Two are concerned with orders flow time. The first is the Gross Throughput Time (GTT) which is the average time between order's arrival in the Pre-shop pool and the final departure of the finished product; the second is Shop Floor Throughput Time (SFT) which represents the time spent in the shop after order's release. The other two performances are related to the ability of the system to respect due dates and deliver on time: average tardiness and the average number of tardy orders per day. Besides, considering the Bumping Model derives from lean manufacturing, another three important indicators are measured to find the reasons why the four performances proposed above change. They are workload in the system (WIP), idleness rate of each worker (Utilization) and percentage of time worker spend on the bumping work (Bumping Frequency).

1.4 Thesis outlines

The remainder of this thesis is organized as follows:

Chapter 2: A thorough literature review is presented. The first main topic is the concept of workload control (WLC) and the main methods concerning capacity adjustment proposed in the literature. A short description of labor flexibility is then introduced, mainly about "where rules", "when rules" and "who rules" in a dual resources constrained (DRC) system. The final paragraph of chapter 2 is entirely dedicated to the Baton Zone Balancing

Chapter3: Introducing the Bumping Model and the research questions that are addressed by this study. Then a detailed description of the simulation model developed and used for the aim of the research is made followed by the configuration of experiments carried out with all the parameters and variables considered.

Chapter4: Presenting the results obtained through simulations, answering the research questions presented above.

Chapter5: Results are summarized and the practical implication is discussed. Then, a brief introduction of limitations in the thesis and the further possible study on the Bumping Model.

2 LITERATURE REVIEW

The articles cited in the following literature review have been collected by using www.scopus.com and www.cnki.net and searching “workload control”, “labor flexibility” and “baton zone balance” as keywords. Relevant papers were selected and analyzed. The references of these articles were useful in order to find more concerned material for the literature review.

2.1 Workload control and capacity adjustment

The workload control was derived from the I/O concept proposed by Wight in 1970[3], the task waiting for delivery is first placed in the task pool at the front of a shop floor. According to a certain task delivery mechanism, the tasks in the task pool are gradually and orderly placed into the shop floor to alleviate the phenomenon of congestion and work-in-process accumulation in the production process of the shop floor [错误！未找到引用源。]. Simplifying the production process at the shop floor into a series of short, easy-to-manage processing queues. In general, the job’s input to the shop floor is controlled according to the shop’s work center capability to adjust and maintain WIP at a stable level. The theoretical basis is “funnel model”[3] showed in *Figure 2.1*. Shimoyashiro (1984)[5] and other researchers studied the production of the job shop by using the simulation methods and revealed the relationship between WIP, output rate and average flow time. The results showed that when WIP increases to one point, the output rate growth slows down, while the average flow time continues to grow. Therefore, it is obvious that there will an optimum value of WIP that can improve the performance of a production system by balancing the workload of each machine, which means that it can achieve the aim of higher output rate and short shop floor time in the same WIP. It is precisely that the idea of WLC is to create a pre-controllable shop floor environment by controlling the amount of work in the shop floor to achieve the aim of higher utilization/output rate, lower inventory, and shorter production cycle.

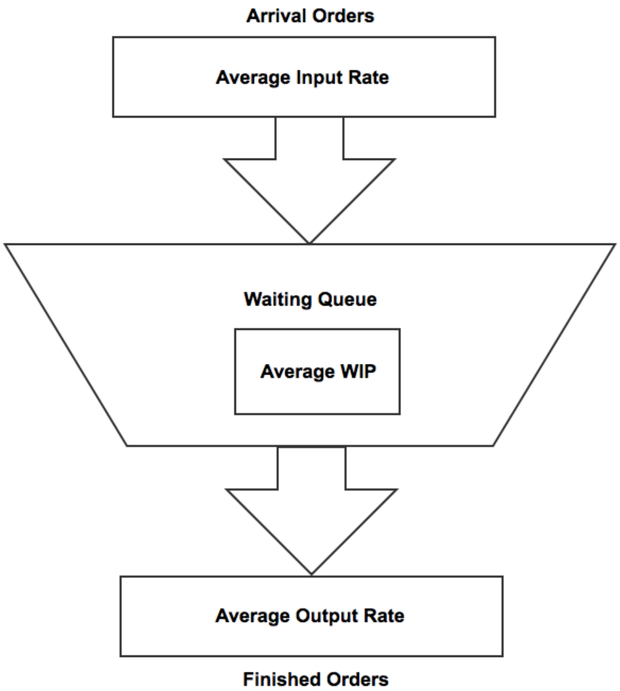


Figure 2.1: The funnel model

In fact, the WLC links the production plan and control system to the manufacturing layer by deciding when to release the order to the shop. Bertrand and Wortmann (1981)[6], Bechte (1988)[7], Kingsman et al. (1989)[8] respectively proposed three WLC concepts that are roughly comparable. These three concepts all propose to divide the WLC into three layers of control structures according to different stages of the order flow. These three layers are the order entry layer, order release layer, and the priority dispatch layer, as is seen in *Figure 2.2*.

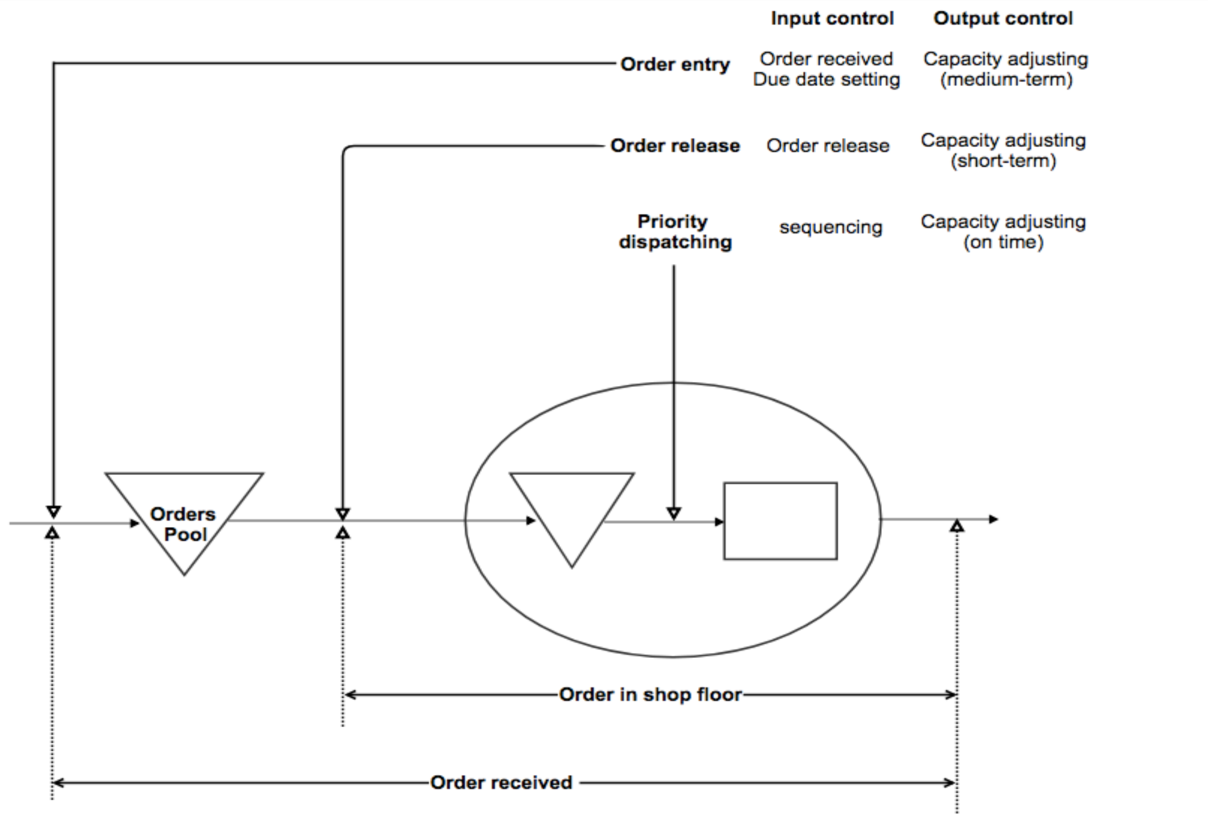


Figure 2.2: The three levels of WLC

Order entry: Deciding if an order can accept. The input control of the decision point includes the delivery date/price coordination decision and the delivery date verification problem; the output control adjusts the medium-term production capacity according to the order quantity and the load status of the system (monthly production plan). The decision cycle of this level is from weeks to months depending on the changing conditions of the production environment.

Order release: Deciding which strategy to use to release the order to the shop floor for processing. The input control of the decision point includes the time to release order and the selection strategy of the order to be placed. By calculating the total amount of materials required for the order delivery and the equipment occupation time, the order processing can be known. The release plan at this level have to fully consider the impact of the orders released on the machine load, control the load of the equipment within the preset standard range, and maintain the workload balance between different equipment. Output control is to adjust short-term production capacity (weekly production capacity)

Priority dispatching: After the order is released, it enters the production queue, the input control at this stage is to select the appropriate dispatch rules to guide the processing order of the order in the queue, the processing order is the order queue before the shop floor. The goal of processing sequencing is to achieve a predetermined lead time and prevent useless waiting on the next process equipment. This process is simple to describe, but it is very difficult in practice. To solve this problem, there are many rules about the ordering priority. Output control is daily production capacity adjustment (overtime)

All in all, through controlling the three levels, the aim of WLC is as following:(a), reduce and balance WIP inventory to keep it at as low as possible; (b), shorten the shop floor time and achieve fast and punctual flow of orders in the workshop; (c), improve equipment utilization; (d), on-time delivery.

As proposed above, WLC integrates two control mechanisms: (a), input control, to regulate the inflow of work to the system; and (b), output control, which uses capacity adjustments to regulate the outflow of work from the system. Much WLC research has focused on input control, while output control has been largely neglected. Normally, in the order entry phase, companies prefer to accept all the orders arriving in a period and still deliver them within due dates issued. The reason is that they can use some methods to adjust their production capacity for a period of time (Kingsman, 2000)[13] or they will try to extend capacity or at most to delay due date. In WLC literature, we find many researchers are focusing on accounting the workload while the description of capacity is simplified to standard output rates (Yuan 2017)[16]. However, real production capacity may be very different from the standard one while it depends on the resources in the production system. And it is well-known that there are three resources in a production system:(a), Machine; (b), Workforce and (c), Subcontracting.

In other words, it means that companies can easily adjust their production capacity by changing these resources, as follows:

- Working overtime: it is normal in many manufacturing company, there are usually more than 1 shift in their production plan, which making the station operated for more hours.
- Relocating workers: there will be several machines in a shop floor, but not all the machine is operated, workers can go to the machine overload to help.
- Subcontracting: some of the orders received can be done by a sub-contractor

From the literatures founded in website www.scopu.com, only these following researchers have done deep exploration on capacity adjustment.

In 2000, Kingsman, started to give some discussion on the output control, he said the output control to the ability to reduce the lead times is same to the input control to the ability to keep a balanced amount of work in the production system. Then, he went deep on this topic with Hendry (2002)[19], a new model called Lancaster University Management School (LUMS) is described by them. This model focuses on keeping a preset maximum time limit Backlog Length which is defined by the value of the total amount of planned workload divided by the planned capacity). Under this model, if the BL limit is overcome by a release of an order, the method of output control will be applied to the system to introduce additional capacity. There are three kinds of output control:(a), worker reallocation (worker flexibility); (b), overtime;(c), worker reallocation and overtime, are tested. Two conclusions they have gotten in this research: one is that the model provides an extra capacity to the bottleneck process; other is that the bottleneck is "wandering" in production system of MTO companies for the variance of orders. Finally, they conclude that input control is to reduce lead times by reducing the accepted workload, while output control makes lead time lower for the same workloads compared to input control alone.

In the next several years, another model called Input- Output Control (PIOC) was proposed (Moreira & Alves, 2006)[21] to control the input and output. The main idea of this model is setting an upper limit on the workload of the shop floor by calculating the workload in the pre-shop pool. The decision on release order in the PIOC depends on the information about orders, the shop floor, and its capacity. Sometimes increasing short-term capacity (like hiring temporary workers, working for more shift or overtime) is necessary for the reason of overload of pre-shop pool. The results in this research show that PIOC improves performances especially when due dates are not strictly defined. The most visible measure improvement is mean tardiness, the percentage of tardy, and the mean time in shop floor.

However, both Kingsman and Moreira focused on controlling output with the change of the total workload of the shop floor and their studies are only applied at the order entry stage. They are not so practical for the company prefers to increasing the capacity or negotiating with the customer to rejecting orders. So further researches have been done by some scholars. Land (2015) [24]proposed a model to adjust capacity when the planned workload of a certain station is overcome. The planned workload is the corrected aggregate load because it gives the best representation of the future expected direct load of a station. In the meantime, the shop load and released workload can also be used as an indicator to trigger the release of orders. In other words, this is a model integrating the input control and output control.

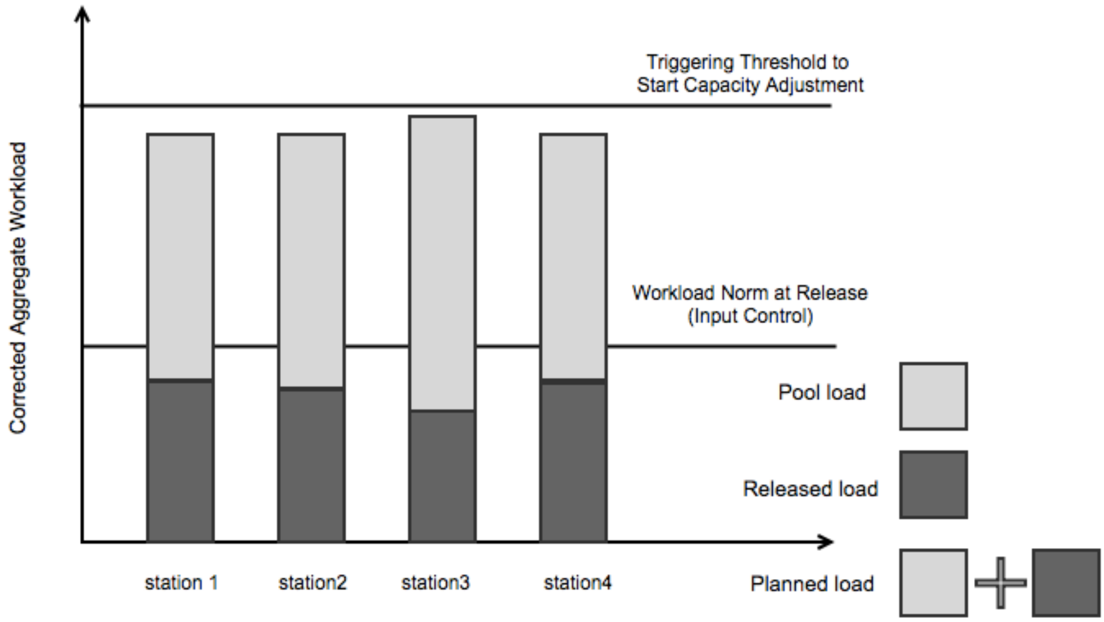


Figure 2.3: Land framework for output control (Thürer 2016)[28]

Thürer et al, as a famous researcher in the topic of workload control, did a lot of worker on the output control in the base of previous researches. He refined LUMS and created LUMS COR (Lancaster University Management School Corrected Order Release). With introducing a new load calculating approach corrected aggregate load approach instead of the classic one in periodic release mechanism and a real pull approach (once a direct load of any work center falls to zero, the worker forward from the pool) in continuous release and comparing with other releasing methodologies, he proposed that LUMS COR is the best model to shorten the throughput time and reducing the WIP and percentage of order tardy.

All in all, results of these researches show that input control has a stronger effect on the lead times and percentage tardy, while output control mostly impacts mean tardiness. Therefore, the two approaches shall be combined [错误！未找到引用源。]. But it is obvious researchers spend more time studying the reallocation of the workforce, which is related to the concept of worker flexibility. Literature of worker flexibility which is the premise of worker relocation will be summarized in the next part of this chapter.

2.2 Worker flexibility

From the literature on workload control, especial in the capacity adjustment. The workforce seems to play an important role. Actually, in a real production system, the configurations of the production system are more complicated than simulation one, the workers and machines will bring an incredible impact on the performance of the company. Many researchers assume station is the minimum unit of a system which is a unique entity with a fixed capacity and it can work for all the time. However, as is known to us, there are machines and operators within stations. Machines need workers to run, while workers need machines and tooling in order to handle jobs and to perform tasks. Compared with operators, machines are more specialized, thus, they are usually idle or under-loaded for the dynamic demand of the marketing. As a result, the number of operators is frequently lower than the number of machine and station. In this case, some operators would work on different stations, this kind of ability is called worker flexibility which gives the production system a flexible workforce. Usually, a flexible workforce allows the manager of the shop floor to move around to respond to temporarily overload department. Actually, worker flexibility is the most relative topic with the production control for it can give more approaches to adjust capacity.

However, it is obvious that if a company want to get a flexible workforce, it should invest in the cross-training program to make the workers be multi-skilled (Hopp & Van Oyen, 2004)[30]. Because of the requires for investment and ongoing expense, many companies can't implement this way quickly when facing the period of high load of work, but in the view of long-term, for companies less willing to discard cost-based criteria or companies under less pressure to improve, increasing labor flexibility without changes in staffing levels may represent the best choice (Felan 1993). Besides, there are others opinions to support the implementation of worker flexibility, one is proposed by the authors in Malachowski (2015)[31], worker flexibility is considered as a competitive advantage while machines can be bought by company provided the availability of enough capital. Others consider more on the economic view, they stated workers are less expensive than machines and they easily allow switching from the production of one product to another one.

Form the ideas above, the machines and workers are considered as two constraints of a production system. So, a concept called Dual-Resource-Constrained is proposed by scholars. After studying on the topic for several years, the research can be summarized in the following three rules:

- Where' rules or assignment rules: they define the policy by which a worker is assigned to a particular station.

- 'When' rules or control rules: they define when a worker is available for transfer to another department.
- 'Who' rules: they define which worker has to be transferred to another department.

2.2.1 Where rules

"Where" rules have shown to significantly impact on system performances. Their purpose is to define where the workers can be transferred, or in other words, which stations the worker can work on. There are some examples:

- Random (RND) - free choice, the selection of the next station is completely random.
- First Come, First Served (FCFS) – workers are transferred to the station with first come job in the queue.
- First in System, First Served (FIFS): workers are transferred to the station with first in the system, first served job in the queue.
- Shortest Processing Time (SPT): workers are transferred to the workstation whose first job has the shortest operation time.
- Longest Number of jobs Queuing (LNQ): workers are transferred to the workstation with the highest number of jobs pooling.
- Largest Total Processing Time (LTPT): workers are transferred to the workstation with the largest total processing time of jobs queuing.
- Worker efficiency (WE): workers are transferred to the workstation where they are more

efficient.

- Upstream (UST): workers are transferred to the most upstream loaded station.
- Downstream (DNS): workers are transferred to the most downstream loaded station.
- Closest (CLS): workers are transferred to the closest loaded station.
- Max-time (MCT): workers are transferred to the station with the longest cycle time.
- Min-time (MCT): workers are reallocated to the machine with shortest cycle time.

From the examples presented above, it is clear to us that the load of a station and the efficiency of a worker on the station are the two main criteria.

2.2.2 When rules

These rules define when the workers can leave the current working stations. In DRC systems authors usually adopt two types of "when" rules:

- Centralized control – workers can be transferred to other workstations after the completion of the current job. (Darwin, 2009; Sammarco, 2014)[32]
- Decentralized control: workers can be transferred to other workstations as soon as they are idle, which means the queue upstream their home department has been depleted.

Much literature gives the result that "Decentralized control" is a more effective rule than "Centralized control". However, both two rules perform worse if there are too many transfer time, since workers have to spend the time to transfer from one station to another station.

2.2.3 Who rules

After determining the time (when) and station (where), a worker should be selected to transfer to the station that has already defined. "Who rules" give us a guidance to select the worker. From the literature, we know that the "who rules" mainly depends on cross-train activities. And the results get by the previous scholars shows that usually the workers should be trained to own the ability to work on at least two machines. But the higher level of flexibility does not significantly improve the performances. Besides, making few workers more specialized on certain stations is better than training all workers homogeneously.

2.3 Baton Zone Balancing

In this thesis, a new method of capacity adjustment called Bumping Model is proposed. Bumping Model derives from the Baton Zone Balancing. Therefore, it is necessary to give some introduction of Baton Zone Balancing.

As we all know that manufacturing is experiencing a dramatic change nowadays. The market is not like the last century; it is gradually dominated by customers. In this background, many manufacturing companies are trying to change the production mode and introducing more suitable managerial philosophy. "Lean" as a famous and proved effective philosophy get more and more attention from companies and the concept called "lean manufacturing" is becoming popular in manufacturing industries.

When talking about lean manufacturing, four principles should be mentioned: (a) flow, (b) pace, (c) pull and (d) zero defects. Thanks to these principles, more and more companies are optimizing their production system in the following five steps:

- Identify value: value is defined by the customer, all that customer is willing to pay
- Map the value stream: The Value Stream is the set of activities that are made on the inputs to give to the customer a product/service.
- Create flow: Anything that stops the flow is waste and identify it and remove it
- Establish pull: In lean manufacturing, forecasts are wrong, instead of trying to make a better forecast, we should focus on having a better response capacity
- Seek perfection: seeking perfect means a high level of quality, low cost and low waste, this action is always on the journey and not on the destination

Besides, in order to test whether a production system is lean or not, two items on your lines or in the office process are examining. These are excessing inventory and idle time. Thanks to the book "The Lean Practitioner's Field Book: Proven, Practical, Profitable, and Powerful Techniques for making Lean Really Work", the thesis introduces the Baton Zone Balance with a simple game written in the book and results that authors have attained.

This game is a very simple exercise asking for people to cut out about 50 strips of paper about 2 in. × 6 in. and write first and last name on these strips. If three people are participating, run the exercise for 4–5 minutes. Once the time is up, recording on a flip chart when the first piece was completed, how many strips of paper were completed, how much work in process (WIP) was in the system, and if there were any defects, that is, names unreadable or spelled wrong.

Batching model

In order to simulate the batch production process. First, have each person print their first and last names on the strips of paper in 10-piece batches. After the first 10 are completed, they are passed to the next person and then the first person starts on the next 10. Continue the game until the time is up. The batching model is simplified as *Figure 2.4*.

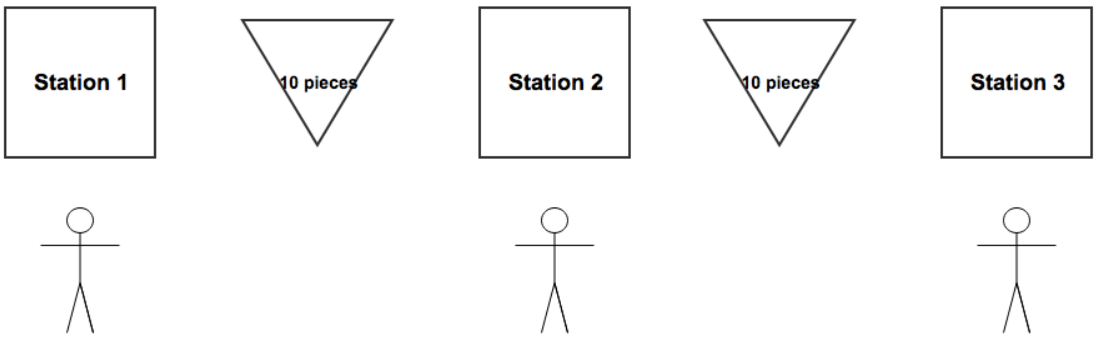


Figure 2.4: Batching model

Results: We can find that between persons or stations there are always 10 pieces WIP and the persons usually become idle.

Station balancing model

Secondly, run the game again implementing one-piece flow. Each person now fills in their name on a strip and immediately passes it to the next person. In some cases, WIP will build up between participants. This is normal. Compare the same results information at the end. The process is can be seen in *Figure 2.5*.

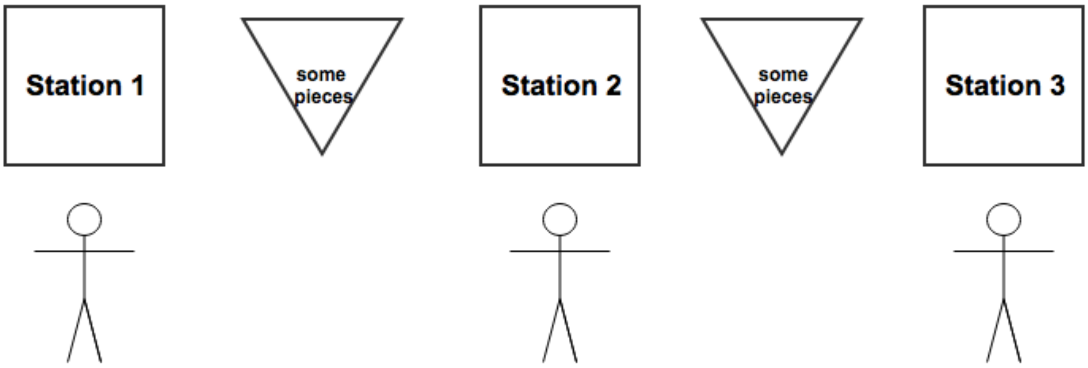


Figure 2.5: Station balancing model

Results: The results show that there is a significant improvement in output but probably found there was still some pieces between the players even if there was much less WIP and some people had some idle time. The WIP would have built up between someone with a short name vs. someone with a longer name. The idle time would have been witnessed with the person after the longer name that had a shorter name. This is the problem with station balancing. One cannot capture the time lost between stations with different labor times.

Actually, the station balancing is very common in the real production system, managers are spending much time dividing the whole time by the stations to get the same cycle time (CT) of each station to keep the station balanced. And considered Takt Time (TT), they change the number of stations or increasing an operator in a station to meet the requirement: $CT \leq TT$.

However, as is proposed in the results of the second exercise, sometimes, because of the labor efficiency or product types, the stations are impossible to be perfectly balanced. There are the following kind of problems in station balancing:

- The major problem with this approach is that people are not robots and we all work at different rates, sometimes a worker will in a low spirit or a worker will become more efficient. Both of the two phenomena will bring an imbalanced line
- If the line needs to run more than one model or type of product, it would also be difficult to keep the labor balanced. It is very common in current marketing for the products are becoming more and more customized. May one type of product requires 1 minute in station 1 while other one asks for 2 minutes.
- Because station balancing was designed based on a set number of operators, if an operator is missing, it creates problems.
- Most station balanced lines are sit-down lines. The problem with sit-down lines is that the product flow is normally lost.
- Sit-down stations are 10%–30% less efficient than stand-up and walking lines.
- The next problem with station balanced system is when starting a production flow. In this case, each operator has to sit and wait until enough inventory is in the line to start working. If there is a large amount of labor in a unit, two hours, for example, some operators near the end of the line may sit idle for hours.

- Sometimes, if one operator is sick or on vacation. Since the operators are not cross-trained, so there will be a huge WIP upstream the station that the missing operator works on.
- If we have a fast worker in a station balanced line, he/she will finish the work and sit idle, but the we still have to pay the same salary
- Another problem we confront is where a line is balanced to Takt time (customer demand) instead of cycle time. For example, let’s say the Takt time is 3 minutes, but we have only 2 minutes and 40 seconds worth of TLT. 160 seconds divided by 3 = 53.3 seconds. But our Takt time is 60 seconds. So we will have to put 60 seconds worth of work at stations 1 and 2, and 40 seconds at station 3. So every cycle, station 3 is idle for 20 seconds.
- Another issue is when the total labor changes because of a different product or even sometimes a slower operator. We find engineer, or team leader will either continue to run it the same way creating more variation and WIP in the process or they may try to rebalance the line. However, with an aim to rebalance the line, they have to completely move tools, material, and equipment to different stations. In many cases, the process flow (PFA) is lost, and the product will start to travel backward which is an obvious violation of Lean flow. First in, first out (FIFO) is also lost.

Because of these problems presented above, we end up with either idle time, WIP inventory between stations, or both. Definitely, extra WIP or idle time is always the signal of a problem. Totally, we can found that station balance is mainly influenced by the workers’ efficiency and the types of product.

Baton Zone Balancing model

Finally, run the exercise and put the slowest person first, the second slowest person second, and the fastest person last. As soon as the fastest person is done filling in their name, they pull from number two regardless of how much of their name number two has completed. Number two then pulls from number one. Number three then finishes number two's work and adds their work and sets it down and pulls the next one from number two, again regardless of how much number two has done. Sometimes number two may not have anything done and sometimes number one's work might not be done in which case number three finishes number one's work, number two's work, and their own work and sets it down. Number two may have to finish number one's work after they pull. This means the last person will normally end up writing their own name and then part or all of the second person's name and the second person will end up writing part of the first person's name.

Results: After compared with the station balance model, the Baton Zone Balancing brings an increase with the bumping process, normally 2 to 10 additional strips depending on the variation of the size of the names. This normally represents 5 to 20% increases in productivity. One can now see with bumping our output increases and the WIP is limited to 2 or 3 pieces in the flow if done properly.

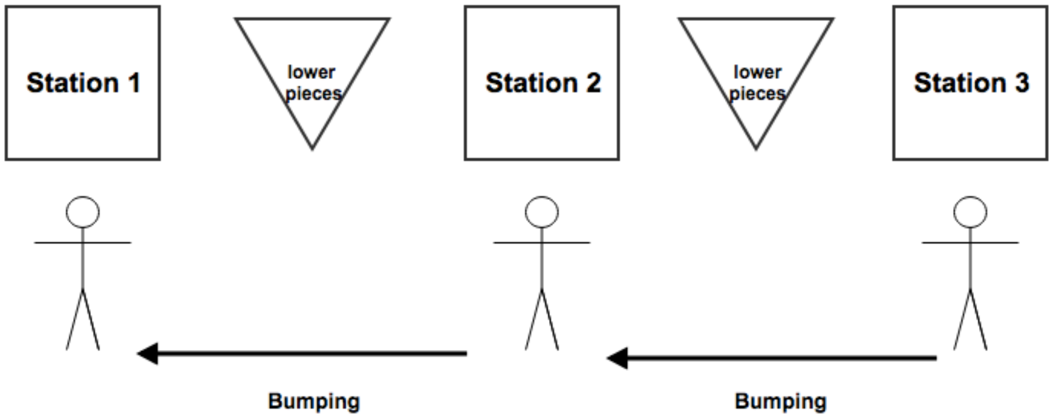


Figure 2.6: The Baton Zone Balancing model

From the game, we can see the change in production mode clearly and the experiment done by authors of the book gives us the results explained above. But we still don't know why we called this kind way of work balancing "Baton Zone Balance". Actually, lean practitioners did many kinds of research on it and give a more detail description to the "Baton Zone Balance". Compared with the station balancing, they summarized the Baton Zone Balancing in following words, also with the help of *Figure 2.7*.

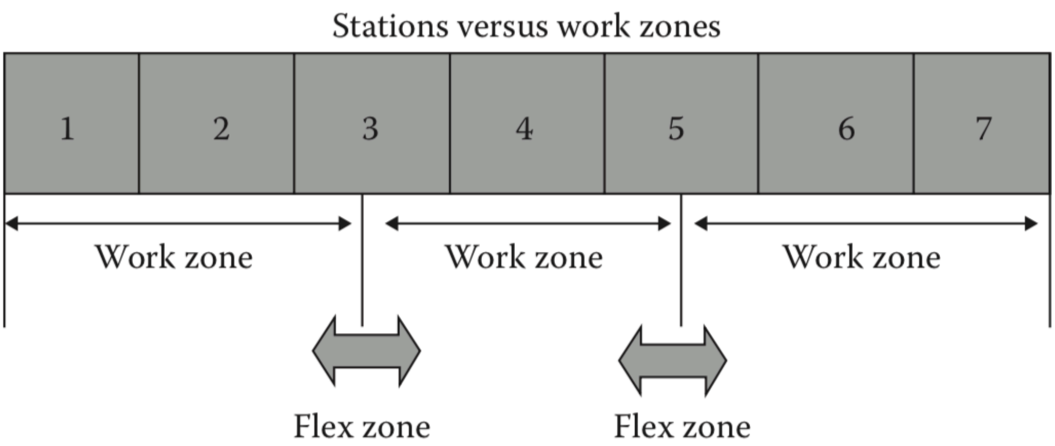


Figure 2.7: Station Balancing vs. Baton Zone Balancing

To facilitate work balancing, we use flex zones referred to as Baton zones, which are areas where handoffs occur between operators. The layout must be designed with short, easily shared steps around the zone. Long operations should be split into smaller steps.

In this system, the operators or workers are spread out on the line to work just like runners in a relay race. Like the baton handoff in the relay race, the operators hand off the product to the operator after them and then bump them to the operator before them until the first operator is reached who goes to the beginning of the line and starts a new part. The main difference from station balancing is all of the operators in this type of line are not waiting for the person before them to hand them the product like in a push-type station balanced system. Instead, they are literally pulling the part from that operator even if they are not finished their station's work.

The entire baton handoff cycle is triggered by the completion of a product at the end of the line. Therefore, it is a true pull system. When the operator at the end of the line (let's call them $operator\ 3$) completes the product and places it in a bin, tray, or hopefully the shipping box, then $operator\ 3$ moves back down the line to the operator immediately prior to him ($operator\ 2$) and takes his product from him regardless of where it is in the process. It is very difficult to get operators to hand off their parts initially. They want to finish their work at the station first, which will cause $operator\ 3$ to be idle until they ($operator\ 2$) hands it off. When the product is given to $operator\ 3$, $operator\ 2$ then moves down to take the part from $operator\ 1$, this is considered a handoff or what we call bumping. This bumping down the line causes a baton handoff and bump by each subsequent operator on the assembly line until the operator closest to the beginning of the assembly line no longer has a product to work on. This first operator ($operator\ 1$) will then start another product at the beginning of the line. The rule is that the operators on the line must continue to work on the product in their hands until they are bumped or complete a product.[2]

And thanks to the concept of Baton Zone Balancing, combined with the properties of MTO companies, the uncertain process time in each station, there is a reasonable decision to give us to introduce this kind of rules into the MTO companies. To explain the decision more detailed, baton zone bumping model definitely brings a light to the flow shop of the MTO company.

3 OBJECTIVES AND RESEARCH METHODOLOGY

3.1 Research gap

Most capacity research deals with worker reallocation. For what regards operator reallocation or operator flexibility, studies are more focus on the workload of each station as the trigger to relocate the workers along the production line. They proposed many kinds of “where”, “when” and “who” rules in the flow shop and the results showed that a proper capacity adjusting can bring a benefit to the performance of production system. Baton Zone Balancing, as a new concept in the line balancing system, give us a new option to think about a new method in the MTO companies. we try to introduce the algorithm (bumping) originating from the Baton Zone Balance to adjust the capacity of the flow shop. The trigger is always the completion of a job and interrupted by a worker. We call it “Bumping Model”. This thesis will compare “Bumping Model” with Static Model and studies on the impact of different factors on the Bumping Model.

3.2 Research questions

Question 1: What are the contributions of the Bumping Model to performances of the flow shop?

Question 2: How do factors, release period, worker efficiency, transfer time and the processing time affect performances of the flow shop in the Bumping Model?

3.3 Research methodology

In order to investigate the issues presented in the research questions and fill the research gap in the particular field, a simulation model was developed. Python as one of very popular programming language was used to write the model, and thanks to the external libraries SimPy which provide a simple way to create the simulation environment.

Here, I want to thanks to the work done by the previous scholars, their studies provide me with most of the parameters adopted for the simulation. All of them should be equal to the ones presented in this research to ensure that we can make an exact comparison.

Then, a more detail description of the shop configuration, the parameters adopted, the experiment design and the implementation is provided to give the reader the opportunity to check the model and the simulation results.

3.3.1 Simulation model

This model simulates a shop floor made of 5 stations, each station includes one machine and one worker. The flow of the shop is directed which means that the job should be processed in fixed routing namely from station 1 to station 5. One pool is allocated to each station, which plays the role to decouples the production flow from the station upstream to the one downstream. In addition, there is a Pre-shop pool, it performs as a buffer to decouples orders arrival in the system to orders release to the shop floor. It enables to release the order to shop floor in a periodic way and do a late modification on orders without incurring in extra costs like a reworking of semi-finished products or scraps. Totally, there are one PSP, five machines, five pools and five workers in the shop, we can see it clearly in *Figure 3.1*. In order to simulate the shop, there are seven types of entities in our simulation model, the relations and attributions of these entities can be seen in *Figure 3.2*.

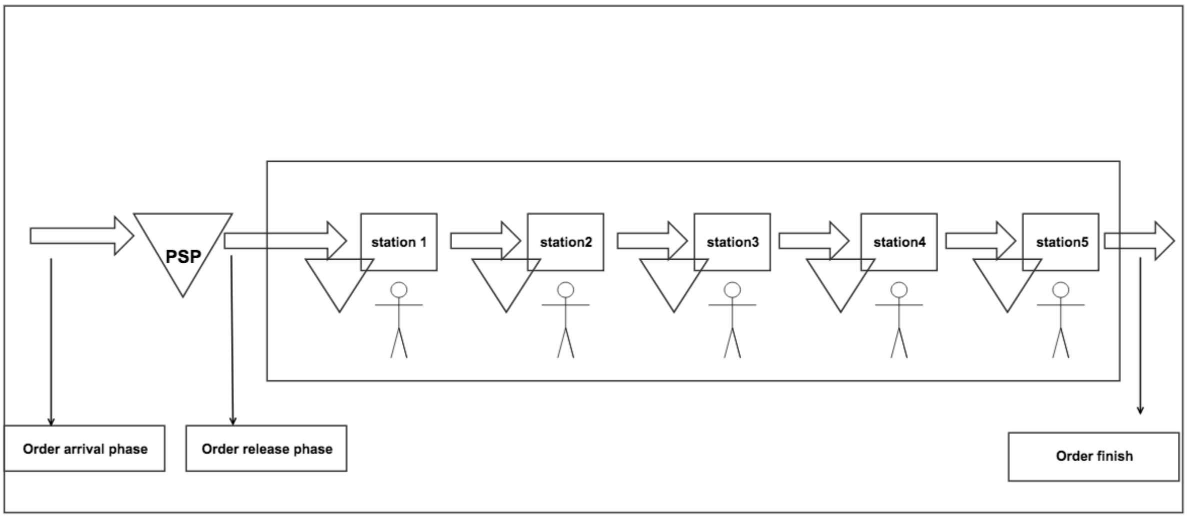


Figure 3.1: Shop configuration of the simulation model in the thesis

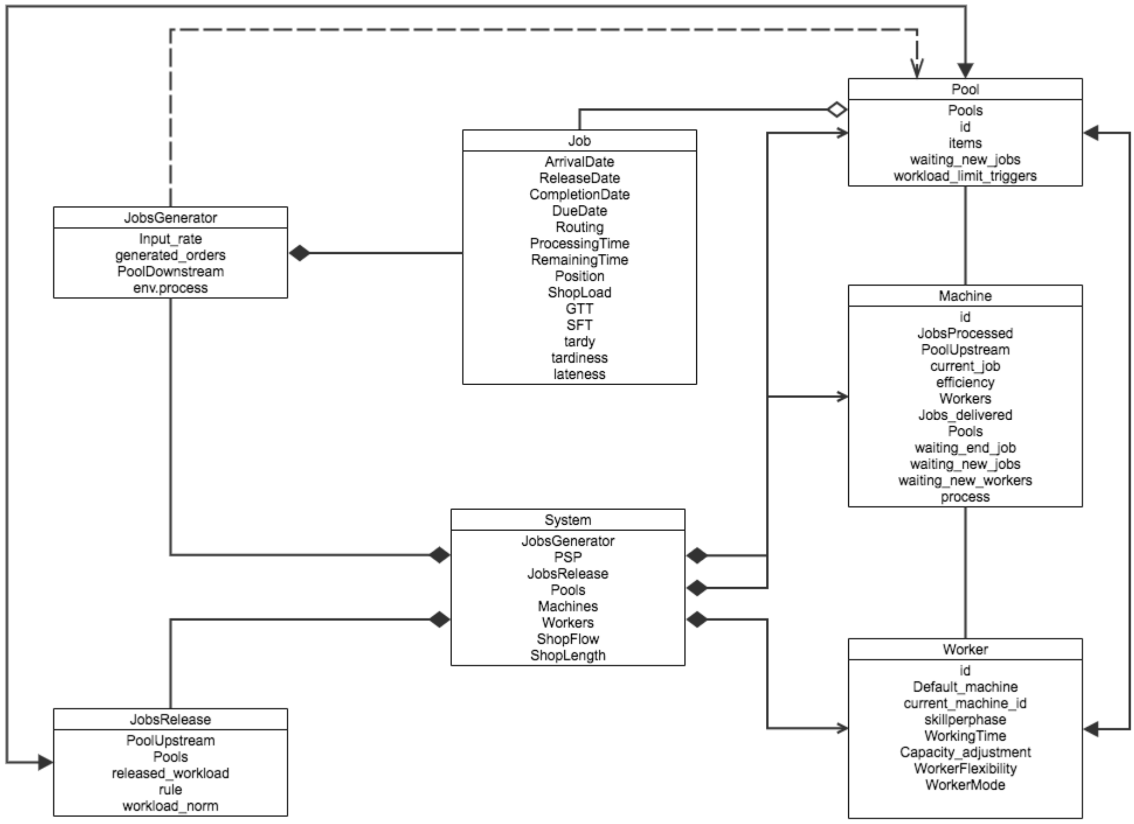


Figure 3.2: E-R graph of the simulation model

3.3.2 Types of workforce

There are five workers on the shop floor, according to the concept of Baton Zone Balancing, some of them can move to another station to catch the job. In order to simulate the bumping process, when the worker downstream finish his current job and there is no job in the upstream pool, he will go to the station upstream to catch the current job of the previous worker, the model replace the movement of worker with the mandatory completion of job on the previous station whenever is finished or not. Maybe some people think the workforce is static to a station, but actually, it is not. In order to recognize this type of workforce from the tradition one that we will compare with, we called this type of workforce "Bumping". In this case, there are only two kinds of the workforce in our simulation model:

- Static – workers are constrained at their main station and they cannot provide extra capacity to the others.
- Bumping – workers are assigned to a main station in the simulation model, but they can move to the station upstream to catch the job in a practical scene. Here, in order to simulate the algorithm easily, the movement of job replaces that of worker

According to the work flexibility literature and our experiences, the main station is undoubted to become the one where the worker has the highest level of proficiency. So the worker always has the priority to work on the main station. Most researches on worker flexibility are substantially studying on the question: when the workers and which of them(who) will go to which station(where) in order to improve the capacity of a production system, and the researchers have proposed many algorithms to the questions. In fact, the Bumping Model provides an answer to the question. The answer is showed as following:

When the **worker** $\boldsymbol{i}\,(1 < i \leq 5)$ in the shop floor finishes his current job and there is no workload in the pool of the main station and the **worker** $(\boldsymbol{i-1})$ is processing a job in **station** $(\boldsymbol{i-1})$ **(when)**, **worker** $\boldsymbol{i}$ **(who)** will go to catch the job that **worker** $(\boldsymbol{i-1})$ is doing on **station** $(\boldsymbol{i-1})$ **(where)**, otherwise waiting. Here we use a flowchart *Figure3.3* to show the process.

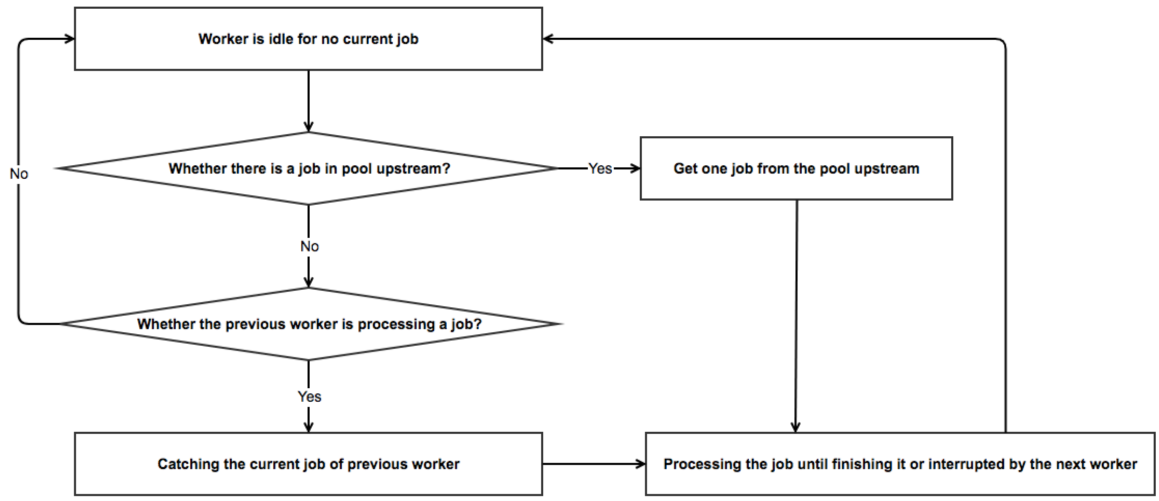


Figure 3.3: The bumping workforce

Furthermore, there are different levels of worker flexibility and efficiency in different research as is shown in the literature review about worker flexibility. However, in this thesis the level of worker flexibility is tested in the Bumping Model is $level\ 1$. Thus, we consider the circumstance that worker can only catch the job when the previous worker is doing the work of the current job that should be done on the main station of the previous worker. Here we can give an example to describe the Bumping Model in $level$ 1 worker flexibility:

Example: the worker can only catch the job from the upstream station, for instance, the $worker\ 4$ can only do the works should be done on $station\ 4$ or $stations\ 3$, if the $worker$ 3 is doing the work should be finished on $station$ 2, the $worker$ 4 can't be able to catch the job from the $worker$ 3.

Fortunately, the different level of worker efficiency has been tested in the thesis. Four values of efficiency have been tested: 100%, 85%, 70%, and 55%. The underlying assumption is that workers can process jobs at their main station with the maximum efficiency (100%), while their proficiency at the station upstream will reduce. *Tables 3.1, 3.2, 3.3, and 3.4* show that workers' skills on each station in different levels of worker efficiency (Notes: the worker efficiency refers to the ability to do the work of a job in different phases, not really go to work on the station):

	Station1	Station2	Station3	Station4	Station5
Worker1	100%				
Worker2	100%	100%			
Worker3		100%	100%		
Worker4			100%	100%	
Worker5				100%	100%

Table 3.1: Skills distribution in 100% worker efficiency

	Station1	Station2	Station3	Station4	Station5
Worker1	100%				
Worker2	85%	100%			
Worker3		85%	100%		
Worker4			85%	100%	
Worker5				85%	100%

Table 3.2: Skills distribution in 85% worker efficiency

	Station1	Station2	Station3	Station4	Station5
Worker1	100%				
Worker2	70%	100%			
Worker3		70%	100%		
Worker4			70%	100%	
Worker5				70%	100%

Table 3.3: Skills distribution in 70% worker efficiency

	Station1	Station2	Station3	Station4	Station5
Worker1	100%				
Worker2	55%	100%			
Worker3		55%	100%		
Worker4			55%	100%	
Worker5				55%	100%

Table 3.4: Skills distribution in 55% worker efficiency

3.3.3 Order release method

In the simulation model of the thesis, orders are generated stochastically. And all the orders will be received and put on a pre-shop pool until the release to the shop floor. And the order release method is periodic release for it is the most common way in a realistic production system

Periodic release (PR) is a similar method to the immediate release, periodical release performs almost no control since it only postpones the release of orders. As mentioned in Bergamaschi (1997), a practical reason why this method is more used than IM is that usually production orders are not directly released to the shop floor. Instead, once they are confirmed, they are placed in the in-tray of the production manager who evaluate their release periodically (e.g. every hour, day or week). The release period which indicates the time of delaying the order release is the main parameter in PR. It may have some impact on the performances of the Bumping Model, so the thesis test two types of Periodic release: release one time a day (Period = 480) and release two times a day (Period = 240).

Combined the order release method with the different types of the workforce. There are two models compared in the thesis: (a), the Static Model; (b), the Bumping Model. *Table 3.5* shows the two models and *Table 3.6* summarizes some parameters of the system.

	Workforce algorithm	
	Static	Bumping
PR	PR + Static	PR + Bumping

Table 3.5: Two models tested

Shop length	5
Shop flow	Directed
Number of machines	5

Number of workers	5
Release period	240minutes, 480minutes
Dispatching rule	FCFS
Capacity of each station	480minutes/day
Working days' length	500days
Arrival rate distribution	Exponential
Process time distribution	Truncated Log-normal Mean 30, variance 900 Minimum 0, maximum 360 And an modified one for Bumping Model
Jobs contractual due date	Uniform [2418, 3600] + arrival date
Workforce flexibility	Bumping (level 1)
Workforce efficiency	100%, 85%, 70%, and 55%
Transfer time	5, 10, 15, 30

Table 3.6: Simulation parameters

There are some system parameters should be explained more detailed:

Arrival rate

Much evidence from the researches shows that jobs arrival follows an exponential distribution for it can simulate the actual arrival well. Therefore, new orders are generated continuously and the time between two consecutive arrivals is not fixed. How to define the value of the Inter-arrival time becomes a problem. Thanks to the work have been done by scholars, it can be calculated in the following formula. As in article of Portioli (2012), the utilization of the system has been set 93.3% for most of the experiments, so the arrival rate only depends on shop configurations considered in our simulation model, actually after calculating, it is equal to 15 jobs/day

$$Interarrivaltime = \frac{Workinghours * TargetUtilization}{AverageProcessingTime * AverageNumberofMachines}$$

Due date

When talking about the due date, many studies don't want to spend much time on it, because, the companies usually don't have a strong bargain power. Normally, the due dates are proposed by customers. However, it is obvious that jobs should be set a due date in the thesis in order to analyze the performances. Fortunately, some approaches have been proposed by some researchers, the first one is setting the value of due date through a constant allowance, then adding a constant to the entry date of jobs in the system; the second one is considering both a constant allowance and a random variables used to generate variability and the third one is taking a constant allowance and a factor depends on variability and the number of stations in the system into account. In this thesis, the value of the due date is calculated according to the concept of approach 3. The values of due dates are equal to the arrival date plus an allowance and the allowance is picked for a uniform distribution between α and β, where: α is obtained through the sum of the average number of machines in the shop configurations times the 95% of the processing time, plus a constant allowance (2000 minutes) (Bertolini, 2015; Thürer, 2016). β is set to the value which makes the percentage of tardy jobs in case of immediate release of orders is equal to 20%. And thanks to the work done by predecessors who have already done in this way and got the results: $\alpha = 2418$ and $\beta = 3600$ for five stations directed pure flow shop. So we can set the due date equal to the arrival date plus an allowance picked from the $Uniform\ [2418, 3600]$.

3.3.4 Factors tested in the Bumping Model

Release period

As aforementioned, periodic release is used in this thesis, and considering the release period is a very common and significant parameter in the periodic release, the thesis test two values of release period: 240 minutes and 480 minutes. They correspond releasing two times a day and one time a day, they are common in the real world. *Table 3.7* shows the experiments on release period.

	Release period	
Model	240 minutes	480 minutes
Static	Static + Period 240	Static + Period 480
Bumping	Bumping + Period 240	Bumping + Period 480

Table 3.7: Experiments on release period

Worker efficiency

The workforce is usually a very important factor in a production system, as it proposed in the previous part, the type of workforce in this thesis is positioning on Bumping $level\ 1$ in worker flexibility. However, as we all know, different level of worker efficiencies means different output rate and also it will give impact on the cost to train the workers. In this sense, it is very meaningful to test the impact of workforce efficiency on the performance of shop floor in Bumping Model and try to find the best value of factor to improving the performance of a production system with a low cost. The thesis has tested four values of workforce efficiency (100%, 85%, 70%, and 55%). *Table 3.8* shows that there are another three experiments.

	Worker efficiency		
Model	85%	70%	55%
Bumping	Bumping + Effi. 85%	Bumping + Effi. 70%	Bumping + Effi. 55%

Table 3.8: Three experiments on worker efficiency

Transfer time

In order to simulate a more practical circumstance, the thesis introduce the factor transfer time. As we all know, in Bumping Model, if the worker wants to catch a job from the previous worker, he has to take time to go to the previous station and do it. So it is very necessary to consider the transfer time in the thesis, which may be a significant factor. Because it can affect the idle rate of a worker. In this thesis, four values of transfer time are set and tested: 5minutes, 10minutes, 15minutes, and 30minutes. *Table 3.9* gives a short view of another four experiments.

	Transfer time			
Model	5 minutes	10 minutes	15 minutes	30 minutes
Bumping	Bumping + TT-5	Bumping + TT-10	Bumping + TT-15	Bumping + TT-30

Table 3.9: Experiments on transfer time

Processing time

At first, how to give a process time in each station to a job is a tough problem to the research. Fortunately, Portioli (2012) proposed that the processing times of jobs should be computed according to the same stochastic distribution and with the same parameters for all stations for sake of simplicity and in order to avoid the creation explicit bottlenecks in the system. The distribution adopted is a log-normal with mean 30 minutes and variance 900 $minutes^2$, which observations can range between 0 and 360 minutes. However, in order to simulate a more realistic order. The distribution has been truncated to make sure the values of process times in each station are not in a very big gap. In the meantime, the range of possible observation is wide and it guarantees a sufficient variability of processing times. In fact, the maximum observable value (i.e. 360 minutes) is 12 times the average value (i.e. 30 minutes). *Figure 3.4* shows a graphical representation of processing time density function based on the sample used during the simulation

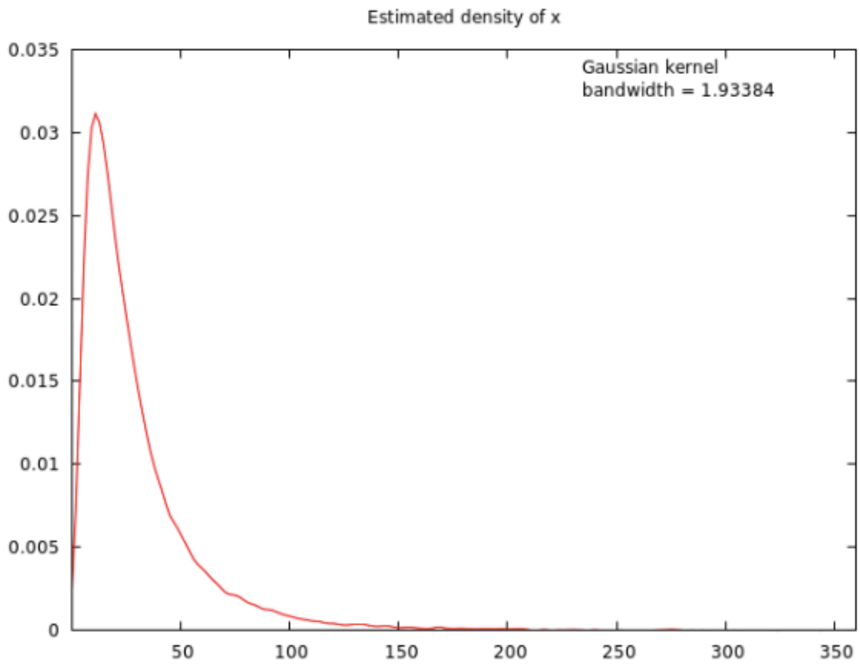


Figure 3.4: The density function of processing time

In addition, as is introduced in the part of Baton Zone Balancing, we should arrange the workers along the production line according to increasing worker efficiency for it can boost the bumping process. Besides, it is obvious that if we apply the Bumping Model to the flow shop, the $worker\ 1$ and $worker\ 5$ are different with another three workers. So, in order to get some deeper conclusion, the process times of jobs in five stations are modified before the jobs are released to the shop floor, but only the mean processing times in each station are changed. The sum of mean processing time and the variance are not changed. What is most important is that the modification algorithm depends on the results of the Bumping Model in original processing time setting. Aim at testing whether the Bumping Model can benefit from a modified processing time. So another two runs have been done in the thesis (as is seen in *Table 3.10*)

	Processing time	
Models	Original	Modified
Bumping	Bumping+ OR	Bumping + MO

Table 3.10: Two experiments on processing time

3.3.5 Software description

The codes of this thesis are written in the Python language, which including objects(classes), functions and loops etc., to simulate the Bumping Model. There are main four common properties of the object in the model:

Autonomy – Objects decisions are taken considering both their state (e.g. idle, processing, waiting) and the behavior of the others.

Intelligence – Objects have some kind of internal intelligence and they do not require that an external system tell them what to do, but rather they take decision individually.

Interaction – Objects are able to interact with the system and with other objects.

Adaption – Objects can change their behaviors according to their programmed intelligence

Actually, because of these properties, there are many processes during the simulations, they run synchronously and interact with each other to achieve the common goal, the processing and completion of jobs. This coding approach makes the research easily to carry on, even if some researchers are not proficient at programming, and permit scalability and flexibility of the software. This part is written with the aim to show the reader the functioning of the simulator so to provide a clear introduction of the assumptions behind the results get from the simulation model and attributions of objects. There are six types of object in the simulation model: job, jobs generator, pool, worker, machine, and order releaser, but for the reason of the difference between stations in Bumping Model, $station\ 1$ and $station\ 5$ are different with $station\ 2$, $station\ 3$ and $station\ 4$, so machine in our model are divided into five classes.

Jobs generator object

The order generator plays the role of the customers in the thesis, which can provide the function to continuously provide the order to the system. As aforementioned, the studies show that exponential distribution can simulate the process well. Meanwhile, in the introduction of arrival rate, there is a specific elaboration about the order generating.

Job object

Job object which represents the customer order is the one that most easily to understand but the most important one in the system. The reason is it stores the main information (arrival date, release date, and completion date etc.) about a customer order which will be used to calculate the performance indicators (GTT, SFT, Lateness, Tardiness and Tardy etc.) of the production system. the formula to calculate these performance indicators are shown as following:

- Gross throughput time(GTT), also known as manufacturing lead time:

$$GTT(i) = Completion\ date(i) - Arrival\ date(i)$$

- Shop floor throughput time(SFT), the time job stays in the shop floor:

$$SFT(i) = Completion\ date(i) - Release\ date(i)$$

- Lateness:

$$Lateness(i) = Completion\ date(i) - Due\ date(i)$$

- Tardiness:

$$Tardiness(i) = \max(0; Completion\ date(i) - Due\ date(i))$$

- Tardy:

$$Tardy(i) = \begin{cases} 1 \ \ when\ Due\ date(i) < Completion\ date(i) \\ 0\ when\ Due\ date(i) > Completion\ date(i) \end{cases}$$

The job's process routing in this thesis is fixed, orders are always processed from $station\,1$ to $station\,5$. The attributions of jobs should be focused on are the processing time and remaining time. These two attributions of jobs are stored in two vectors respectively. The value of remaining time is decremented while jobs are progressively processed through a station, and then they move downstream to the next station.

Pool object

Pool objects play the role to store order, or semi-product in other words, between stations, every pool is assigned to one machine. It is very important to remember that a waiting station will be triggered if the worker finds there is an order in the pool. In this thesis, a job whether it is finished or not in one station can be put into the pool downstream in order to simulate the catching process of a worker.

Worker objects

The most important attribution of worker objects is the level of flexibility and efficiency, a vector called *"skillperphase"* is used to summarize the skills. The first position represents the skills at the first station, the second position represents the skills at the second station and so on. The value in each position represents the worker's efficiency on the station. It can range from 100% to 0%. Workers cannot work at the station where their efficiency is 0% since it means that they are not skilled at that station.

For example, the *"skillperphase"* of $worker\ 3$ in the model can be represented by the vector $[0\%, 85\%, 100\%, 0\%, 0\%]$, we can explain it in the following way:

- the $worker\ 3$ has maximum efficiency at the $station\ 3$, therefore, if the processing time of a job is 30 minutes, the $worker\ 3$ will process the job on $station\ 3$ for 30 minutes.

- The $worker\ 3$ has 85% efficiency at $station\ 2$. Then, if process time of a job is 30 minutes by $worker\ 2$, the $worker\ 3$ will spend 35.3 minutes on processing it.

- The $worker\ 3$ can't do the work that should be finished by $worker\ 1$, $worker\ 4$ and $worker\ 5$.

There are two types of the workforce in this study:

- Static: workers are fixed to their main station, thus they cannot be reallocated

- Bumping: workers can go to catch a job from the previous worker.

Actually, for the reason of simplicity and easily to calculate the workload in each station, the Bumping workforce is simulated simply similar with the static one but add a function to check whether there is a job in the pool upstream and whether the previous worker is doing the work of job should be done on the previous station. So we can describe the Bumping worker object in our simulation model in the following way:

1. Assign workers to their main stations

2. Get the skills in the five stations by setting the $skillperphase$

3. If the machine objects are waiting for a new worker, trigger it

4. If the machine objects are waiting for a new job, trigger it if there are jobs in pool upstream or the previous worker is doing the work of job should be done on the previous station.

The reason why we don't simulate Bumping workforce in the worker objects is that it more easily for us to simulate the bumping process in the machine objects.

Machine objects

Machine objects are the core component of the simulation model in our thesis, they work in synergy with workers. Every machine has a pool upstream that is used to store the job from the previous stations. In this thesis, we assume only one worker can work on one machine at the same time. Once completed or caught by the worker downstream, they are moved to the pool upstream of next station. The bumping process will show in the machine objects. Typically, there are three kinds of machine object:

- Interrupted: $machine1$, the $machine1$ and $worker1$ belong to $station\ 1$, they can only be interrupted by the worker form $station$ 2. The working process can be shown as *Figure 3.5.*

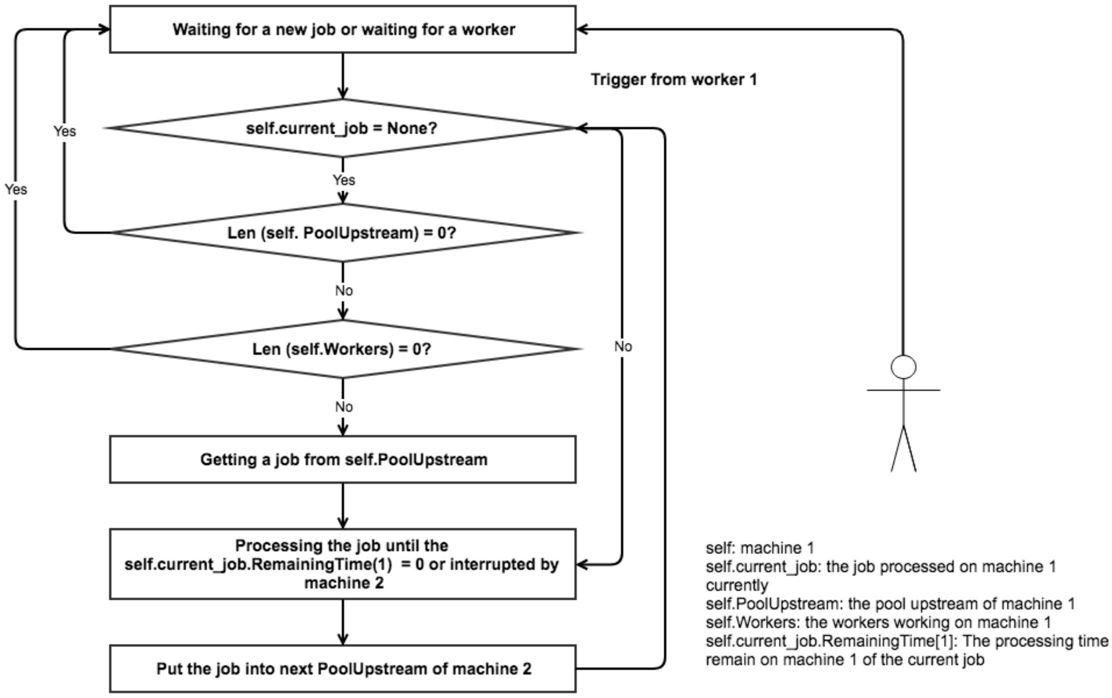


Figure 3.5: The working process of machine 1

- Interrupted and interrupting machine: $machine\ 2$, $machine3$ and $machine4$, they can be interrupted by the $worker3$, $worker4$ and $worker5$, meanwhile $worker2$, $worker3$ and $worker4$ can also interrupting the process of $machine\ 1$, $machine2$ and $machine\ 3$. Their working process can be seen in *Figure 3.6.*

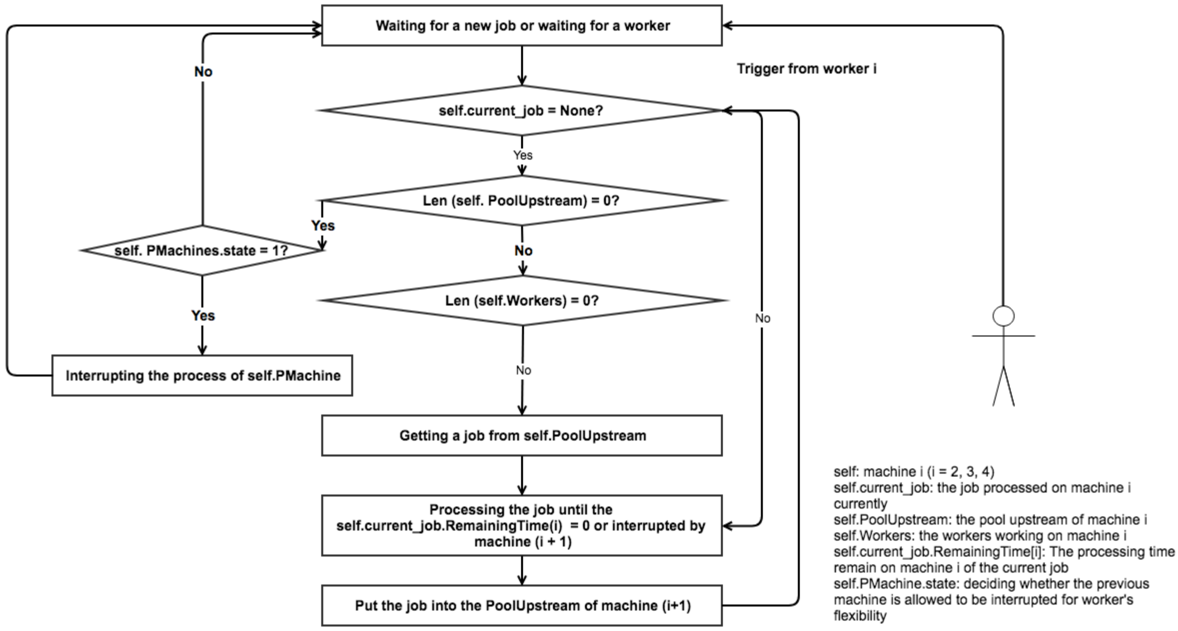


Figure 3.6: The working process of machine 2, machine 3 and machine 4

- Interrupting: $machine\ 5$, it is special as $machine1$, it can only interrupt the $machine\ 4$ by sending the $worker\ 5$ to catch the job from the $machine\ 4$. Its process can be seen in *Figure 3.7*.

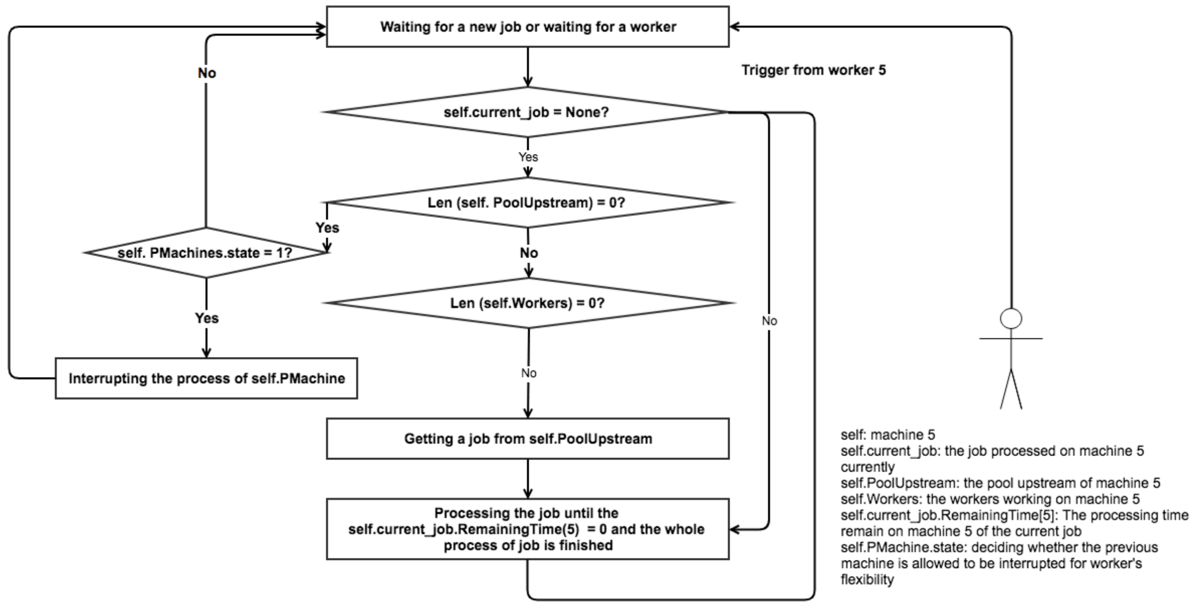


Figure 3.7: The working process of machine 5

As proposed in the worker objects, the Bumping Model affect the actual process time by workers, there is detail explanation about the actual process time in each machine (or worker).

$\boldsymbol{For\ \ machine\ 1:}$

$$Actual\ processing\ time = \frac{Actual\ time}{Worker\ efficiency}$$

$where$:

$Actual\ time$: $the\ time\ that\ job\ stays\ in\ machine\ 1$

$Worker\ efficiency$: $the\ efficiency\ the\ worker\ 1\ working\ on\ the\ machine 1$

$\boldsymbol{For\ machine\ i\ (1\ < i\ \leq 4):}$

$$Actual\ processing\ time = \frac{RemainningTime(i-1)}{Worker\ efficiency(i, i-1)} + \frac{Actual\ time(i)}{worker\ efficiency(i, i)}$$

$where$:

$RemainningTime(i-1)$: $the\ remaining\ time\ of\ job\ in\ machine(i-1)$

$Worker\ efficiency(i, i-1)$: $the\ efficiency\ of\ worker(i)\ working\ on\ the\ machine(i - 1)$

$Actual\ time(i)$: $real\ time\ worker(i)\ working\ on\ the\ job's\ work\ should\ be\ processed\ in\ machine\ i$

$Worker\ efficiency(i, i)$: $the\ efficiency\ of\ worker(i)\ working\ on\ the\ machine$

$\boldsymbol{For\ \ machine\ 5}$

$$Actual\ process\ time\ = \frac{RemainingTime(4)}{Worker\ efficiency(5,4)} + \frac{Processing\ time(5)}{Worker\ efficiency(5,5)}$$

$where$:

$Remaining\ Time(4)$: $the\ remaining\ time\ of\ job\ on\ machine\ 4$

$Worker\ efficiency(5,4)$: $the\ efficiency\ of\ worker5\ working\ on\ the\ machine4$

$Processing\ time(5)$: $the\ time\ of\ job\ should\ be\ processed\ in\ machine\ 5$

$Worker\ efficiency(5,5)$: $the\ efficiency\ of\ worker\ 5\ working\ on\ the\ machine5$

3.3.6 Warm-up period, number of runs and length of the simulation

With the aim to obtain more believable results, it is necessary to define a warm-up period, the length of the simulation and the number of runs in a simulation. Besides, in order to reduce the variance between experiments and to focus only on the influence due to the parameters of the simulation, the common random technique has been used. Therefore, for each replication of the experiment different sample of jobs is used, while the same run of different experiments shares the same sample of customer orders.

In this simulating study, the length of the simulation time is 500 days. Each simulation is replicated in 50 runs (Land, 2006). This with the objective of reducing the experimental error and the variance of the parameters. Both parameters (simulation time length and number of runs) are calculated following the procedure presented in the research by Mosco, et al. (1982).

It is obvious that some simulation time is necessary to make the model to reach a steady state for the production system is empty at the beginning of the simulation. During this period, the data are not accurate and representative of the potential performance of the system, so they can't be considered. The concept of "Warm-up" period is proposed in this situation, and then how to determine the length of the Warm-up period. As we find in literature, different methods to determine the Warm-up period's length exist and were used. In this thesis, we adopt the graphical method proposed by Welch (1987). An experimental run should be made for Static and Bumping Model, and we chose the workload process daily as the main criteria for it represents the stability of the system. The graph in *Figure 3.8* represents the workload processed daily distribution over 500 simulation days of the system in the Static Model. Data reach the steady state after the 80th day, while in the initial period the system does not have consistent reliable performance. *Figure 3.9* shows the workload processed daily distribution over 500 simulation days of the system in the Bumping Model. As we can see, it is very quick to be stable thanks to the bumping process, only after 50 days the workload processed daily becomes stable. Also, there are other graphs showed in *Figure 3.10, 3.11, 3.12, and 3.13* about the workload processed daily distribution over the 500 simulating days in different experiment configurations. All in all, for the result of the simulation we can set the length of Warm-up day equal to 100 days which is definitely enough to get reliable results.

In the end, the Warm-up period is set 100 simulating days, the total length of the simulation is set to 500days and the number of runs of each experiment is set 20. All statics are collected after the Warm-up period. Besides what is most important information we find here is that the Bumping Model can promote the production system to be stable.

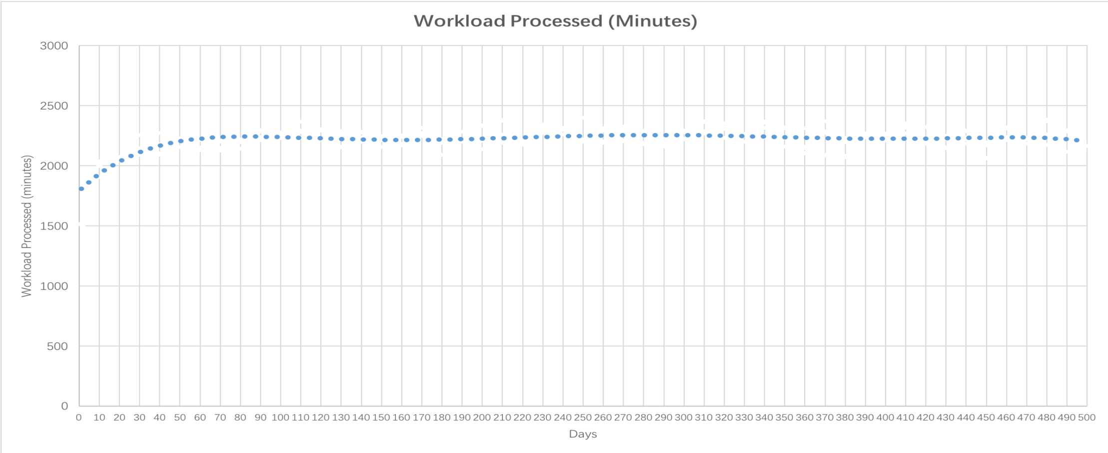


Figure 3.8: Workload Processed daily in Static Model

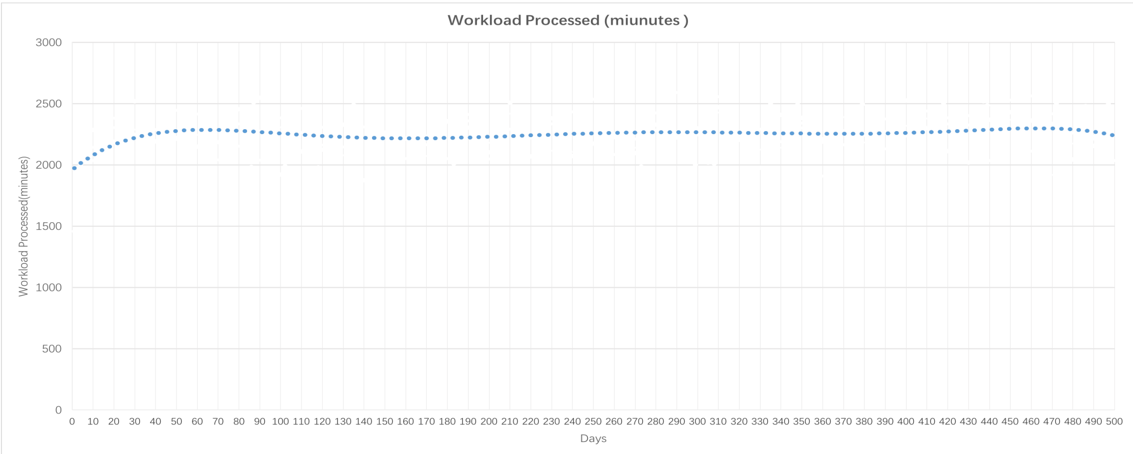


Figure 3.9: Workload Processed daily in Bumping Model

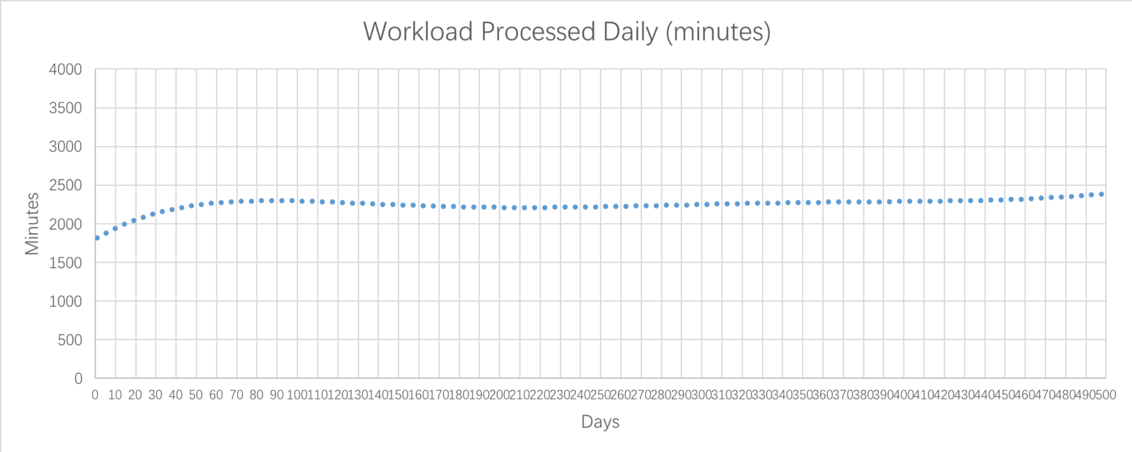


Figure 3.10: Workload processed daily in the Bumping Model (Period = 240)

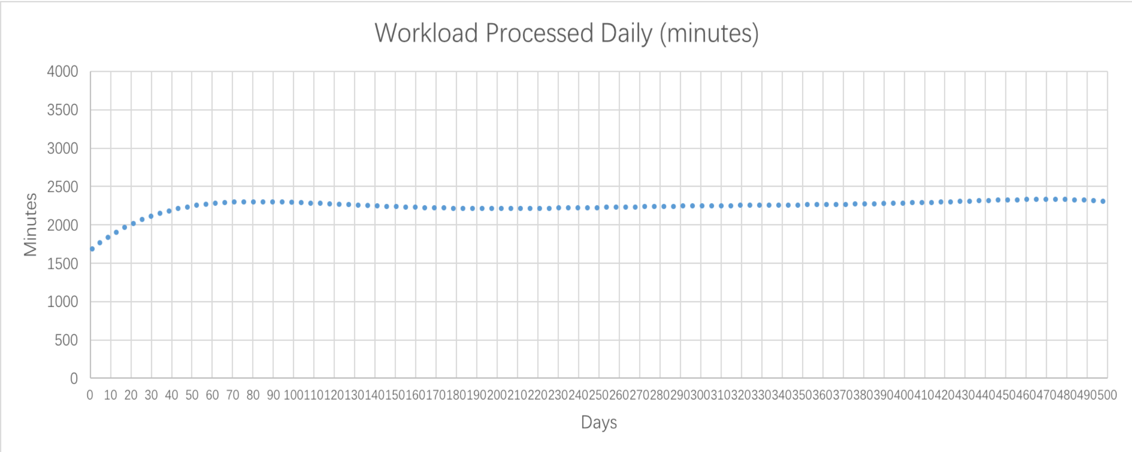


Figure 3.11: Workload processed daily in the Bumping Model (Effi. 70%)

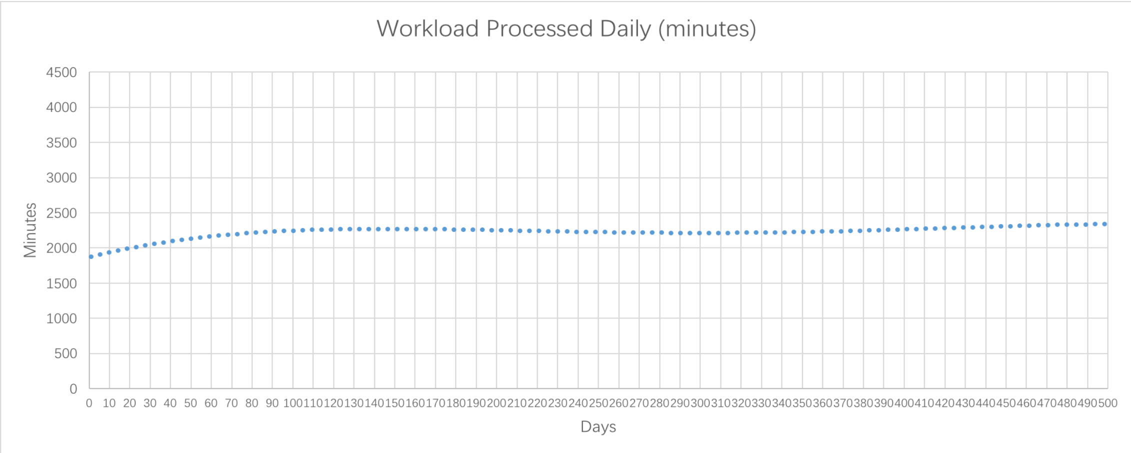


Figure 3.12: Workload processed daily in the Bumping Model (Transfer time = 15minutes)

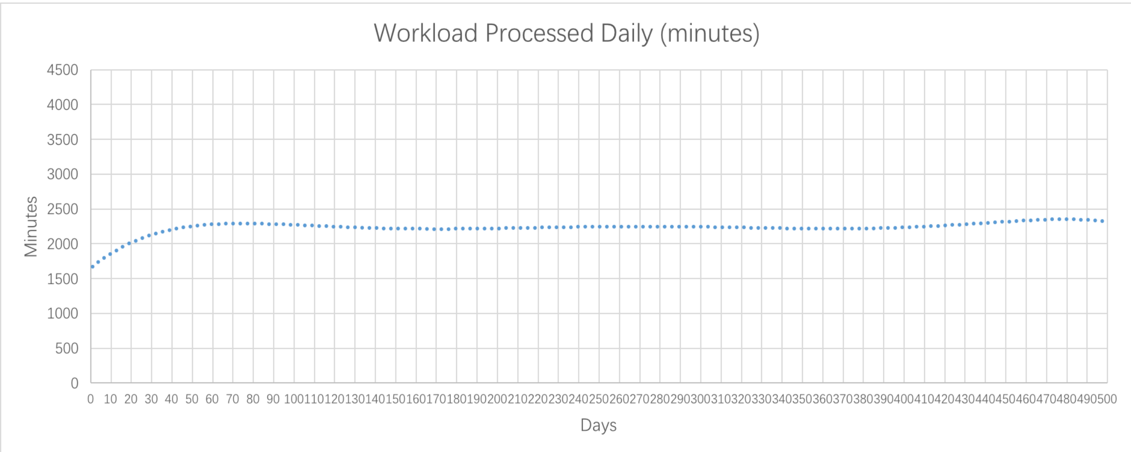


Figure 3.13: Workload processed daily in the Bumping Model (modified processing time)

4. ANALYSIS OF RESULTS

4.1 Research question 1

What are the contributions of the Bumping Model to performances of the flow shop?

To answer this question, we have to compare many performance indicators of the productions system in the Bumping model with those in the Static model. Here, the level of workforce efficiency is 100% in order to reduce the impact of work efficiency on the results, as the same another factors are neglected such as Transfer time and release period etc.

At first, we compare the differences in average Gross Throughput Time (GTT), which show the lead time of a manufacturing company and the average SFT, which represent the time of job stay in the shop floor. As is illustrated in *Figure 4.1*, the value of GTT decrease by nearly 40% (from 2509 minutes to 1506 minutes), and that of SFT of order reduce 44% compared with Static Model. While the Bumping Model improves the two indicators, we can also conclude that it brings higher improvement on the SFT. As for tardy and tardiness, the value of both indicators drop around 95%, which is a good signal in satisfying the customer's order respect to the due date. For the value of lateness, it reduces dramatically from (-499.07) minutes to (-1503.3) minutes, about 200% reduction, which means the order can be completed in advance for nearly 1503 minute in Bumping Model.

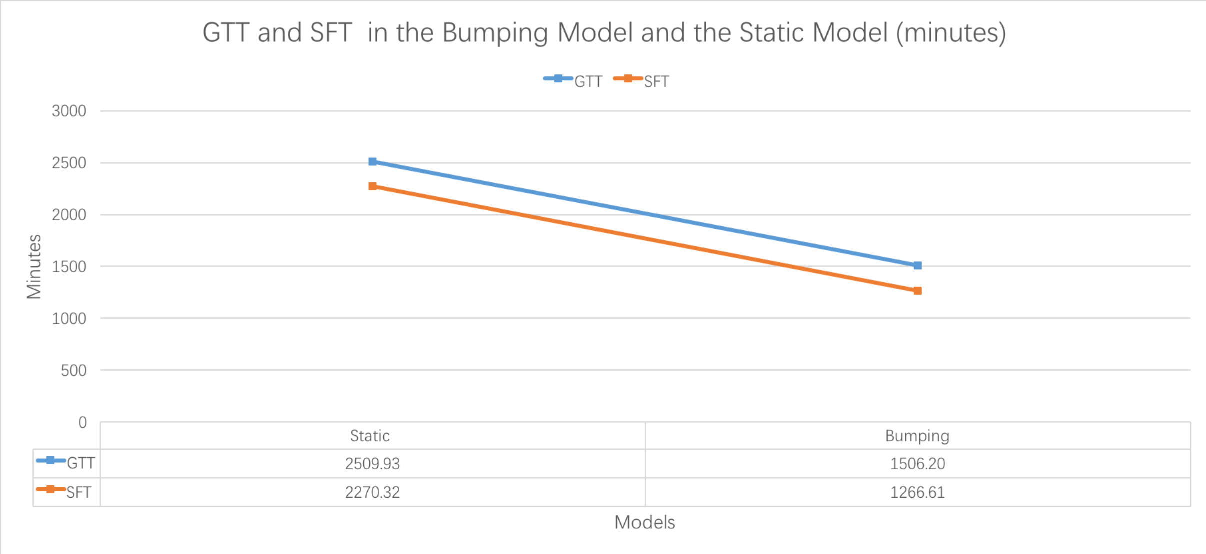


Figure 4.1: Average GTT and SFT in the Static Model and the Bumping Model

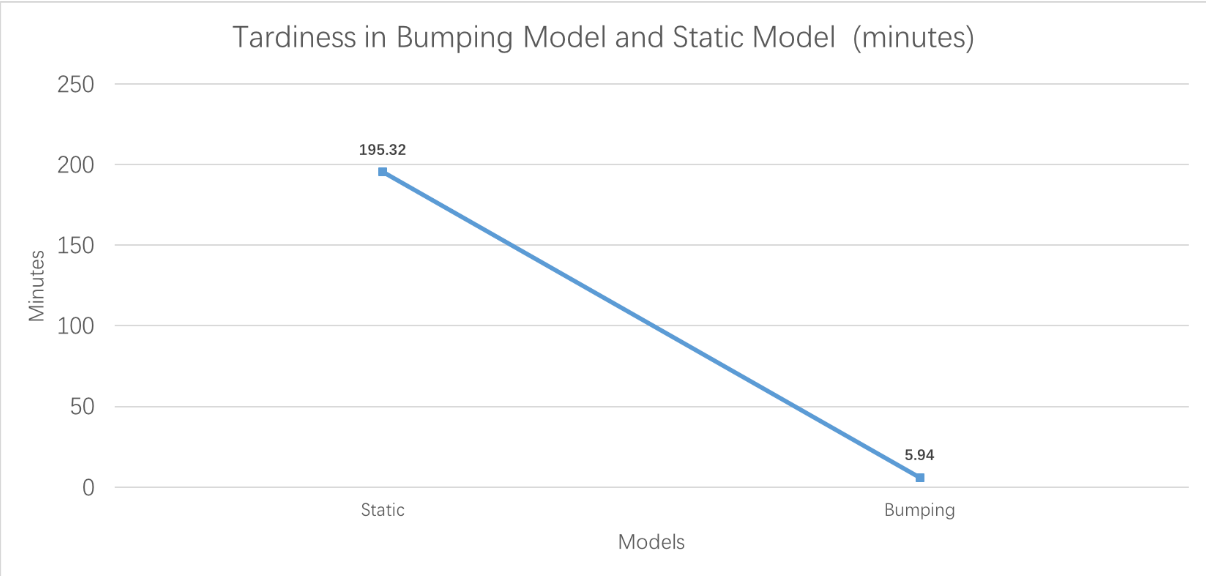


Figure 4.2: Average tardiness in the Static Model and the Bumping Model

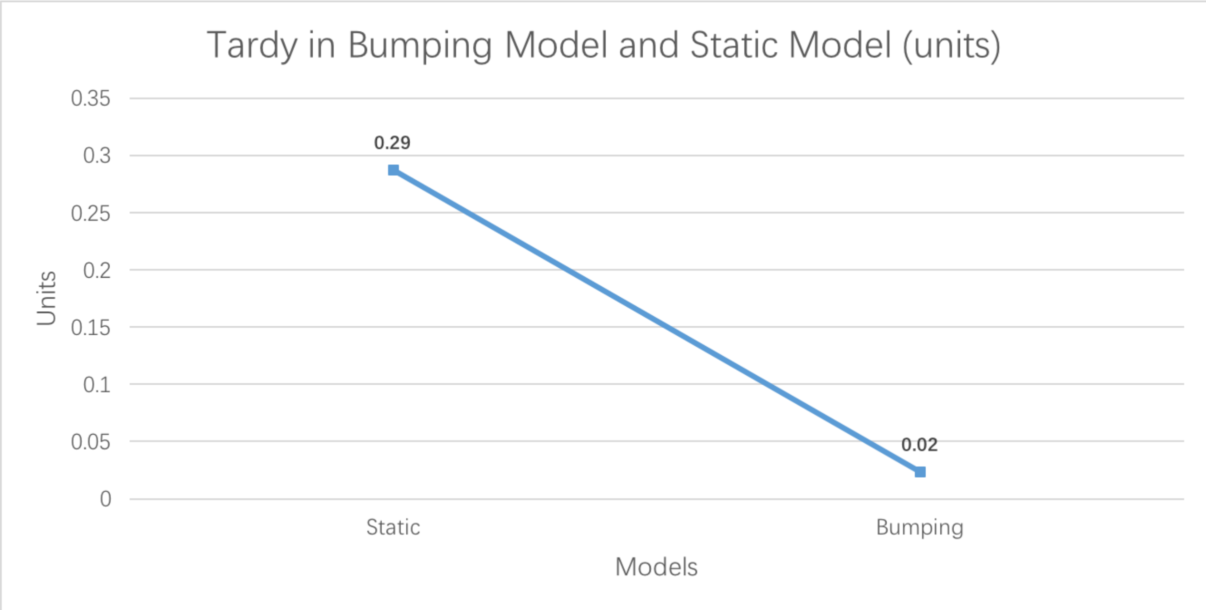


Figure 4.3: Average Tardy in the Static Model and the Bumping model

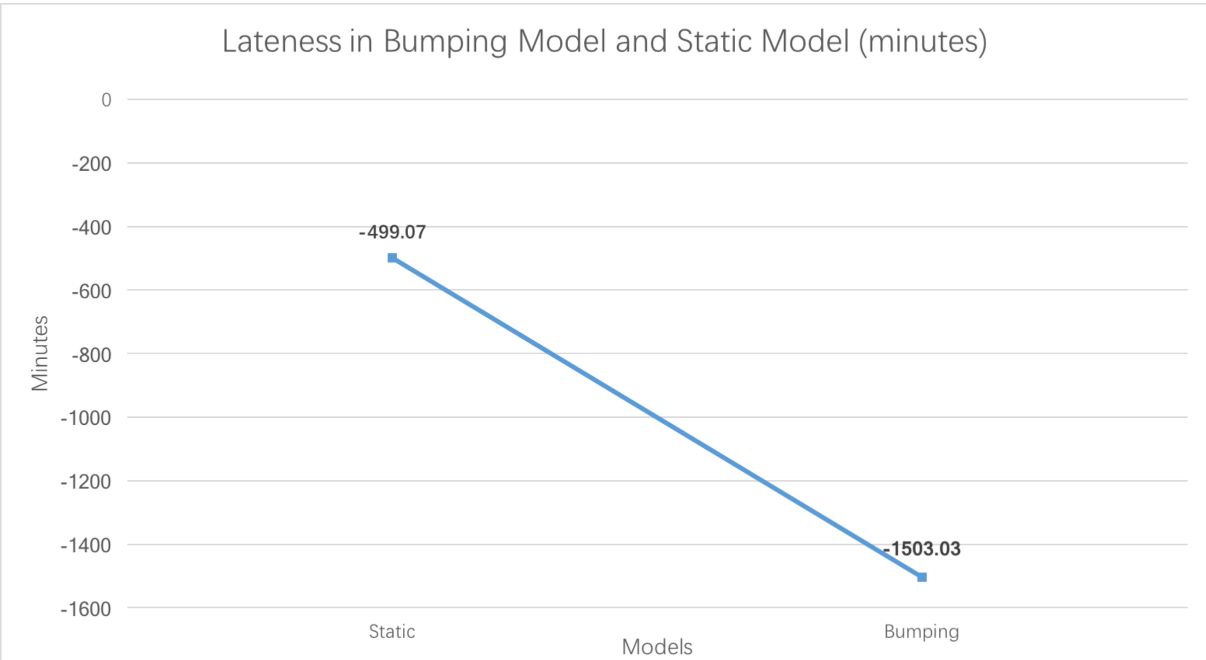


Figure 4.4: Average Lateness in the Static Model and the Bumping Model

In order to find the reasons of the improved performances, other two main indicators should be calculated:(a), idleness rate of each worker, which represents the utilization of resources; (b), average workload in the system, which refers to the WIPs in the system.

The idleness rate of each worker (utilization)

The idleness of worker is a very important factor in every kind of production system, by reducing the idleness rate of workers, the performance of a production system can improve a lot without spending extra expense. Thus, usually, the managers of companies do their best to exploit the human resource. *Figure 4.5* shows the difference of idleness rate of workers between Bumping model and Static Model, as we can see that in Bumping Model, except the $worker\ 1$, all the idleness rates of other workers are lower than those in Static Model.

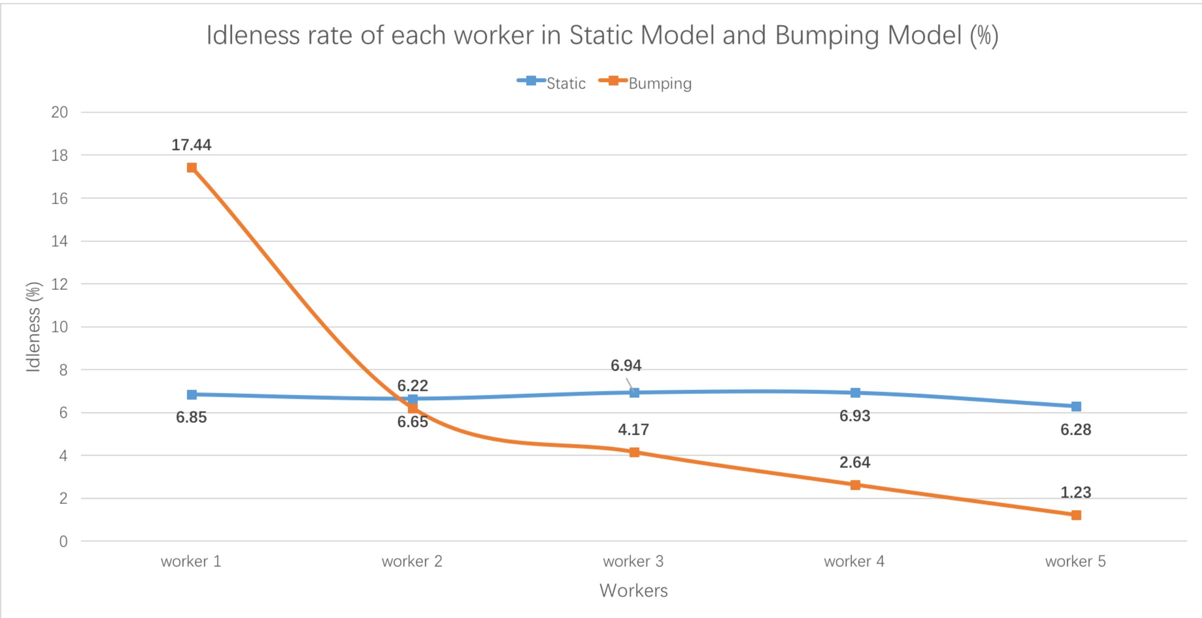


Figure 4.5: Idleness rate of each worker in the Static Model and the Bumping Model

However, the average idleness rates of workers in the Static model and the Bumping Model are 6.7% and 6.3% respectively, which is not a noticeable improvement. In order to find a more persuasive reason, we go to analyze the workload in the system.

The workload in the system (WIP)

As aforementioned, the value of SFT decreases by almost 44%, but the average idleness rate of workers drops only from 6.7% to 6.3%, so there should be more crucial reason to facilitate the improvement. In a manufacturing company, another very significant indicator is WIP, which is represented by the Shop load of each station in the thesis. The value is the sum of the processing time of jobs in the pools, it also considers the contribution of processing times already processed of jobs. *Figure 4.6* shows the difference in the average shop load of each machines between the two models. What is clear to us is that the average shop load of each station in the Bumping Model is approximately half of that in the Static Model, which means jobs processed in the Bumping Model flow quickly in the production system. It is a very specific phenomenon to decrease the SFT.

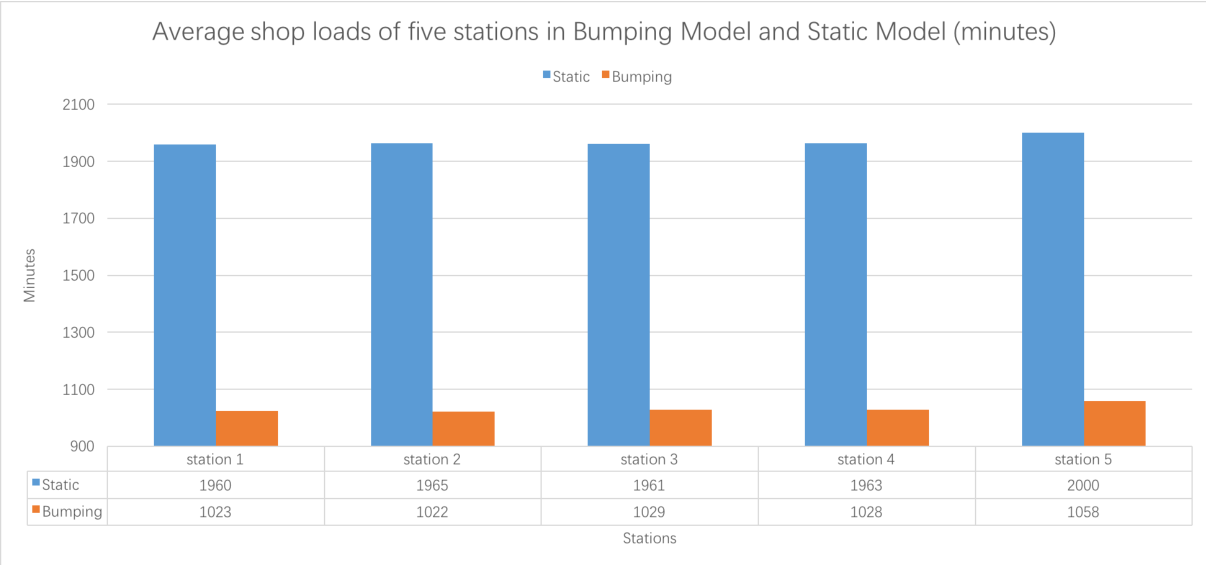


Figure 4.6: Average shop loads of five machines in the Static Model and the Bumping Model

Furthermore, the thesis also considered the fluctuation of the total workload in the system during the whole simulation process period. It surprises us that the total workload of the system in the Bumping Model shows a more volatile state than that in the Static Model, even if the average workload presents oppositely. The standard deviation of total workload of the system in the Bumping Model is 1044.24 minutes, while it is only 452.35 in the Static Model. For the result, the thesis didn't go on with the phenomenon, but we think it is because of the interruption happening during the bumping process. This process makes there are fewer orders in the system which can amplify the influence of variance of jobs.

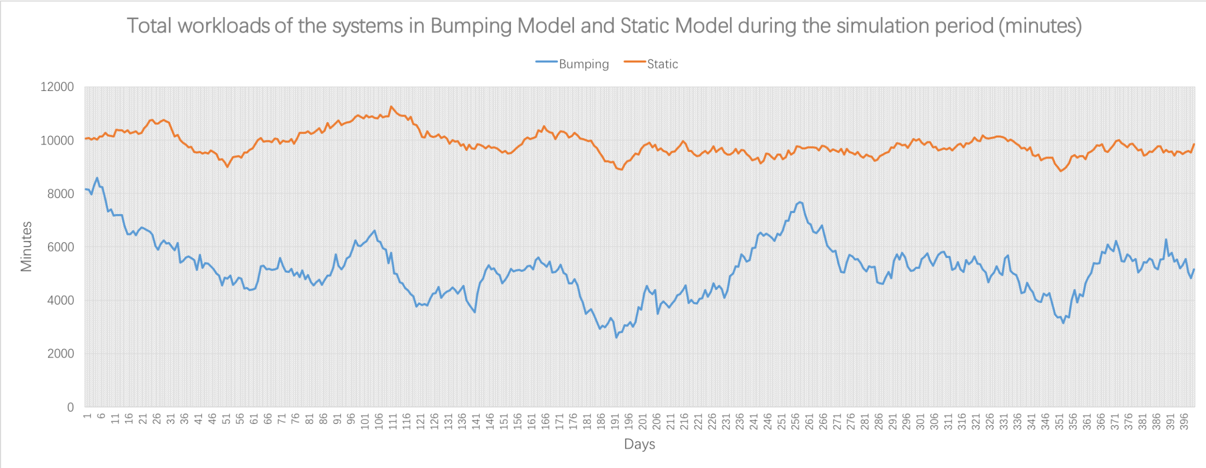


Figure 4.7: Total workloads of the systems in the Bumping Model and the Static Model along the whole simulation period

The percentage of time worker spend on the bumping work (bumping frequency)

It is obvious for us that workers are constrained to the main station in the Static Model, so they can't go to catch a job to work. Thus, these percentages of workers equal to zero. In contrast, in the Bumping Model, some of the workers are allowed to leave the main station to catch a job to process for the worker flexibility. As we can see in *Figure 4.8*, $worker\ 1$ never spend time on bumping work for he can't catch a job in our model. The percentage decreases from $worker\ 2$ to $worker\ 5$. This trend can be explained for the different types of machines as is mentioned in chapter 3.3.5. $Worker\ 5$ can't be interrupted by other workers, which means there is no worker can help him, but when he is idle, he has to go to help $worker\ 4$. With this reason, worker 5 is the worker who are the busiest one and has less time to spend on bumping work. For $worker\ 4$, $worker\ 5$ can spend some time helping him, so he is not as busy as $worker\ 5$. In this case, he can go to spend more time than $worker\ 5$ on bumping work. Then, the $worker\ 3$ works in the same way, he can spend more time than $worker\ 4$ helping $worker\ 2$. Finally, the $worker\ 2$ will be the people who can spend most of the time on bumping worker, which makes $worker\ 1$ be the idlest worker as is shown.

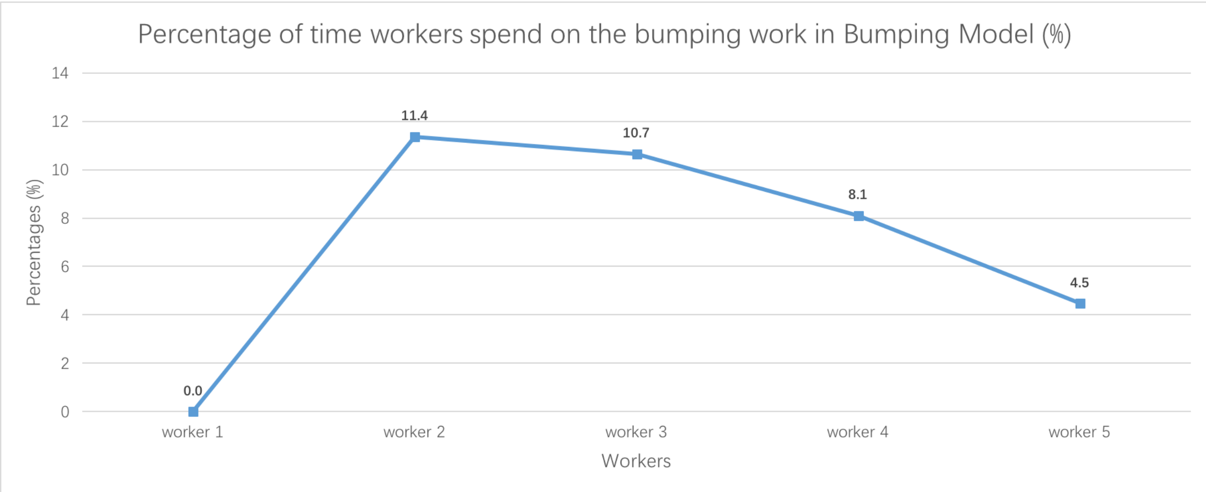


Figure 4.8: Percentage of time workers spend on bumping work in the Bumping Model

4.2 Research question 2

How do factors, release period, worker efficiency, the transfer time and the processing time affect performances of the flow shop in the Bumping Model?

4.2.1 The impact of release period

As is mentioned in the thesis, the periodic release is chosen as the order release method in the simulation model. The values of release period are set to 240 minutes (two times a day) and 480 (one time a day) to test the influence. *Figure 4.9* illustrates that Period 240 performs better in both two models (Static and Bumping). Besides, after calculating, we find that decreasing rates of GTT in Period 240 and Period 480 situation are 45.8% and 40.0% and those of SFT are 48.3% and 44.2% when adopting the Bumping Model. Combined the results found above, it gives a very valuable evidence to prove that the Bumping Model benefits from the shorter release period in GTT and SFT ($P_{period\ 240} > P_{period\ 480}$). As for the Tardiness and Tardy, the results are same, Tardiness reduces around 97.5% in Period 240, 96.9% in Period 480 and Tardy drops about 94.06% in Period 240, 92% in Period 240. Besides, when compared the Bumping Model in Period 480 with that in Period 240, the GTT increases by 17.15% while the SFT rises 10.97%. The conclusion that longer release period brings more harm to the GTT is gained.

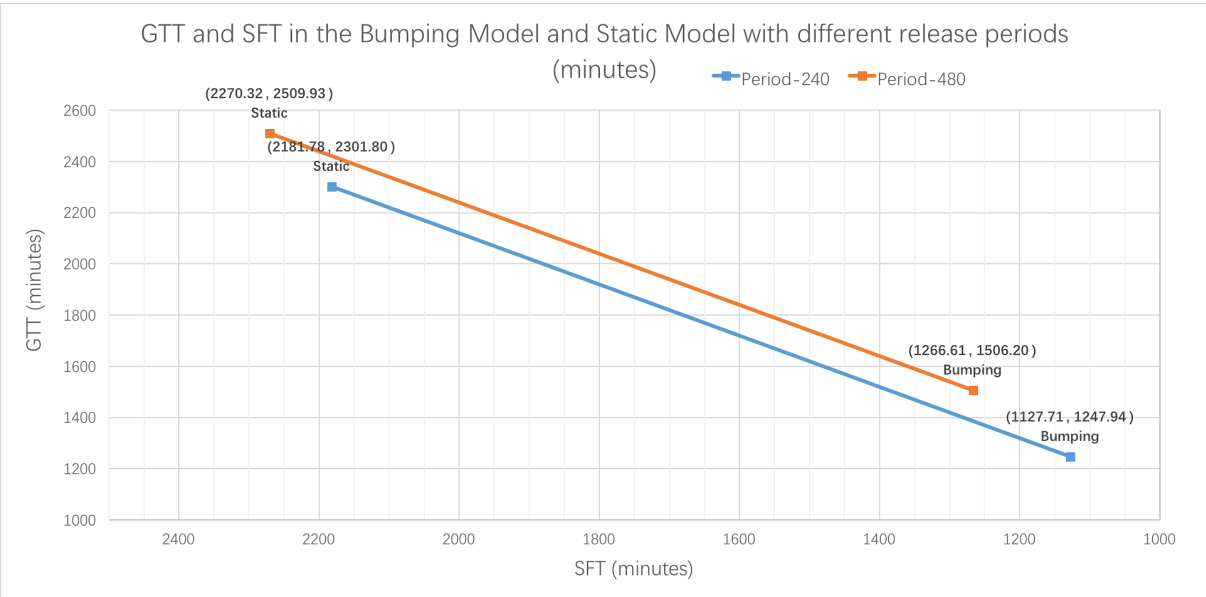


Figure 4.9: GTT, SFT in the Bumping Model and the Static Model with Period = 240 and Period =480

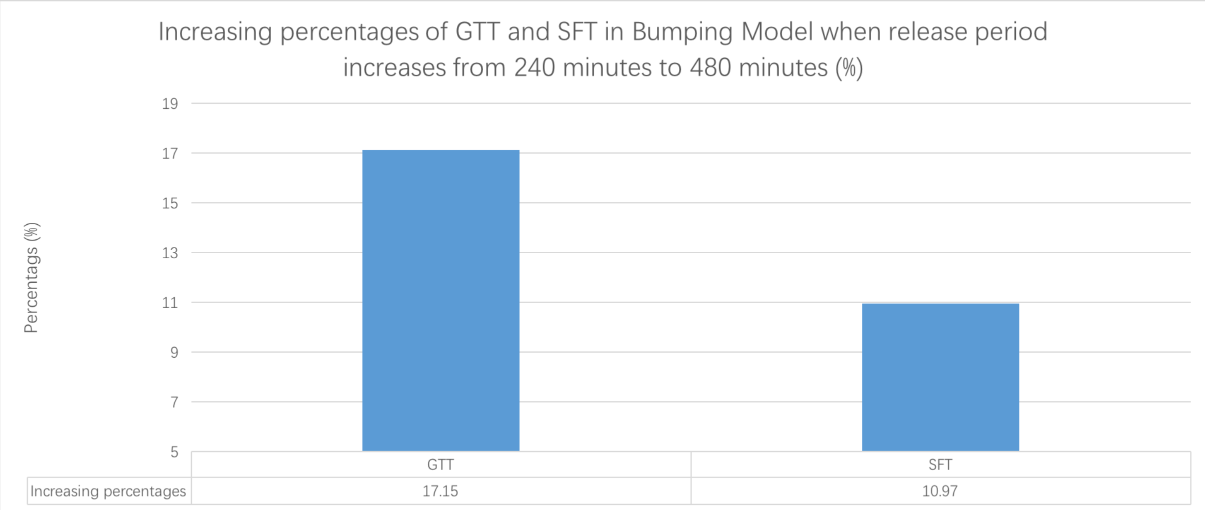


Figure 4.10: Increasing percentage of GTT and SFT in the Bumping Model when release period increases from 240 minutes to 480 minutes

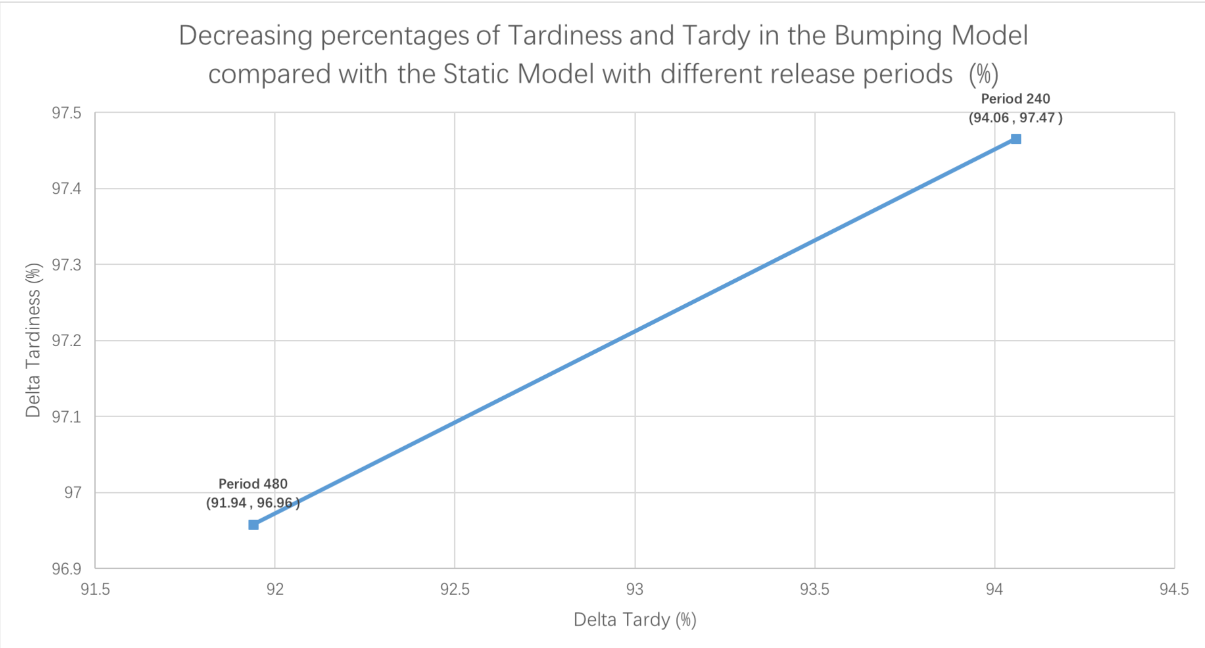
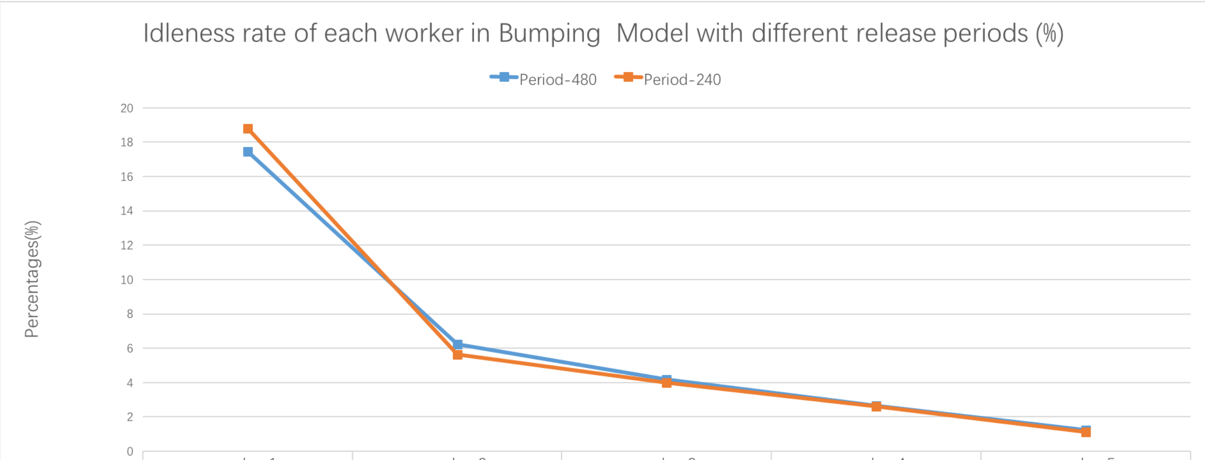


Figure 4.11: Decreasing percentages of Tardiness and Tardy in Bumping Model compared with Static Model with different release periods

In conclusion, the results show that the release two times a day performs better than release one times a day. Here, we consider the three main factors: idleness rate of each worker, workload in the system and percentage of time worker spend on the bumping work, and try to find the reason for this phenomenon.

The idleness rate of each worker (utilization)

As we can see in *Figure 4.12 and Figure 4.13*, compared with Period 240, Period 480 makes the idleness rate of the first worker increase by 7.71%, while idleness rates of others workers drop. However, what surprises us is that the average idleness rate increases approximately by 1.35% (from 6.34% to 6.43%). In this case, the idleness rate may not the reason why the Bumping Model benefits the shorter release period in periodic order release.



	worker 1	worker 2	worker 3	worker 4	worker 5
Period-480	17.4	6.2	4.2	2.6	1.2
Period-240	18.8	5.6	4.0	2.6	1.1

Figure 4.12: Idleness rate of each worker in Bumping Model with different release periods

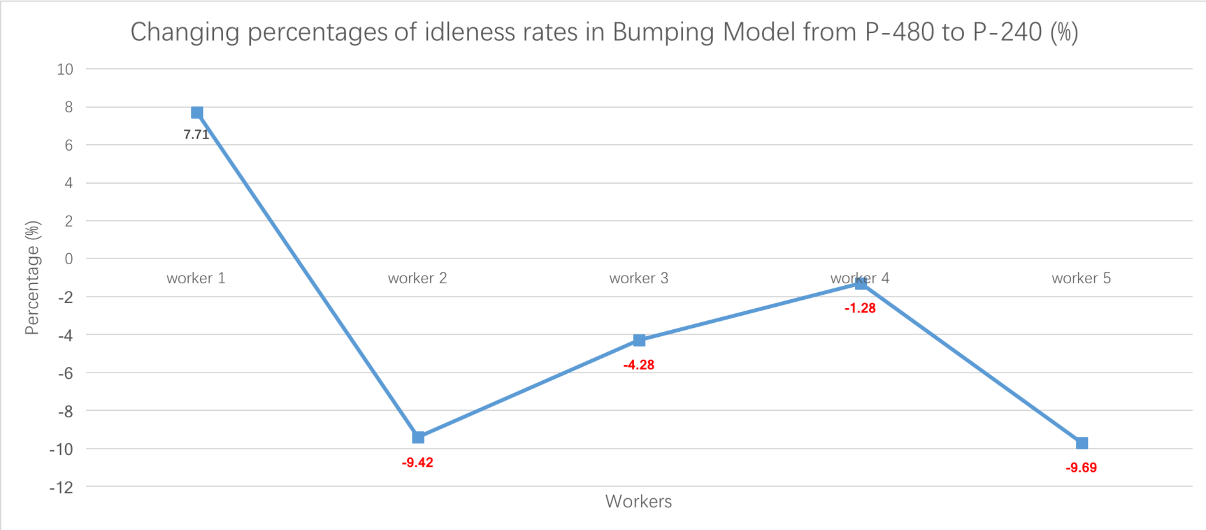


Figure 4.13: Percentage of idleness rate's change of five workers from P-480 to P-240 in Bumping Model

The workload in the system (WIP)

The idleness rate of each worker can't provide a reason why the Bumping Model benefit from the shorter release period, and we try to consider other common indicators, the workload in the system. As is shown in *Figure 4.14*, compared with the Period 480, the average shop loads in Period 240 decrease by around 21% in the five stations. Also, the standard deviation of total workload in the system becomes smaller (from 1044 minutes to 486 minutes), we are very confident that the lower shop load of each machine make the performance of Period 240 better than that of Period 480. However, it is not the root reason for the results, we go deep and study on the question: why the shop load reduces? Based on the core idea of the Bumping Model, the bumping process, the percentage of time worker spend on the bumping work gradually steps into our eyes.

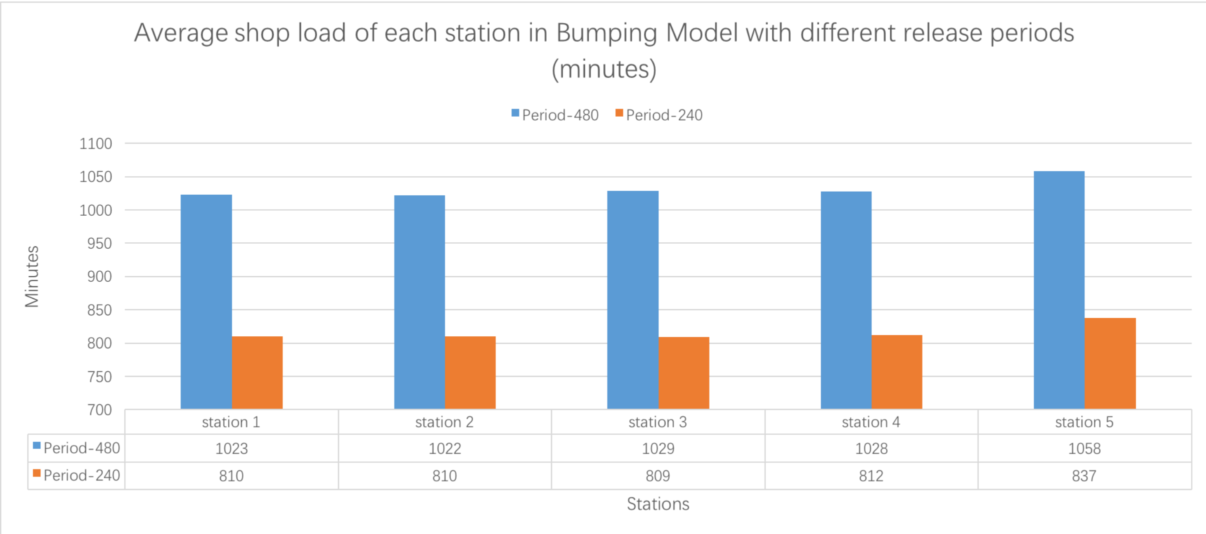


Figure 4.14: Average shop load of each station in Bumping Model with P-480 and P-240

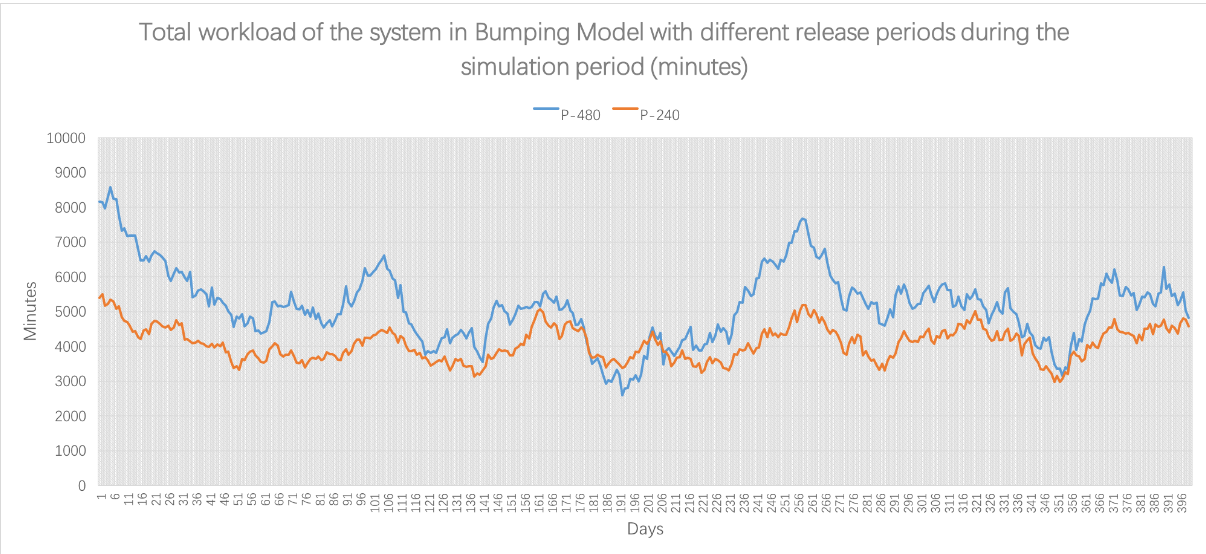
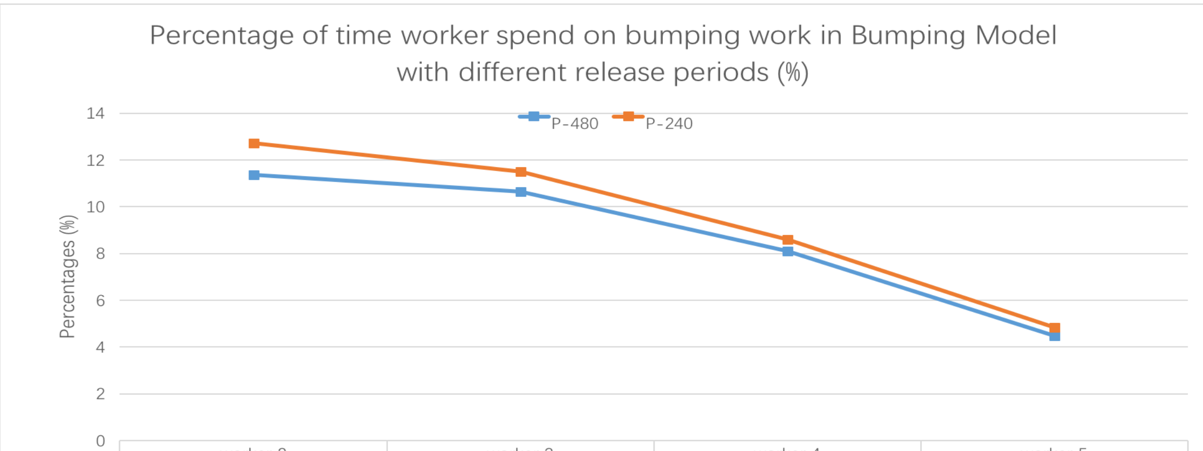


Figure 4.15: Total workload of the system in the Bumping Model with P-480 and P-240 along whole simulation period

The percentage of time worker spend on the bumping work (bumping frequency)

The main difference between the Bumping Model and Static Model is the worker can go to the upstream station to catch the job of the previous worker. It is easy for us to know that the first worker never works on bumping work for the algorithm of Bumping Model, all the remaining workers can catch a job from upstream station. The following *Figure 4.17* shows the increasing percentage of the percentage of time worker spend on the bumping work in the Bumping Model with release period = 240 compared with release period = 480: $worker\ 2(11.92\%)$, $worker\ 3(7.93\%)$, $worker\ 4(6.12\%)$ and $worker\ 5\ (7.95\%)$ respectively. After comparing this percentage, the thesis finds that the reason of lower shop load in Bumping Model with release period = 240 is the increasing bumping process, which provide the evidence again to the conclusion: the Bumping Model reduces the WIP in production system. Besides, give the conclusion that shorter release period boosts the performances of the Bumping Model.



	worker 2	worker 3	worker 4	worker 5
P-480	11.4	10.7	8.1	4.5
P-240	12.7	11.5	8.6	4.8

Figure 4.16: Percentage of time worker spend on the bumping work in P-480 and P-240

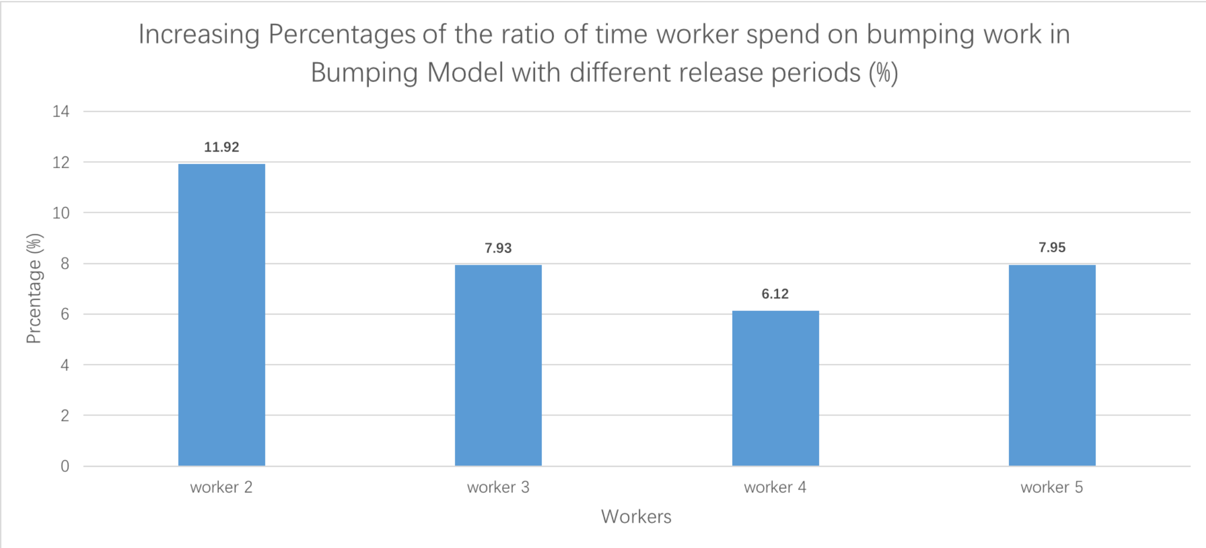


Figure 4.17: The increasing percentage of the ratio of time worker spend on bumping work in Bumping Model when release period decreases from 480 minutes to 240 minutes

4.2.2 The impact of worker efficiency on the bumping work

Worker efficiency is a very import part of worker flexibility, there are three levels of worker efficiency are tested: 85%, 70%, and 55%. As is usual, at first, discussing on the change of GTT and SFT, *Figure 4.18* shows the increasing percentage of GTT and SFT, compared with the 100% efficiency, the GTT increases by around 4.55% in Effi. 85%, 14.84% in Effi.70%, and 33.8% in Effi. 55%. The SFT increases by about 5.38% in Effi.85%, 17.63% in Effi. 70%, and 40.2% in Effi. 55%. It is very clear to us that the change of SFT is bigger than that of GTT. Besides, as we can see that, from the Effi. 100% to Effi.85%, the GTT and SFT grow 4.55% and 5.38% respectively, the GTT and SFT separately rise 9.84% and 11.69% from Effi 85% to Effi. 70%, and from Effi. 70% to Effi. 55%, the GTT and SFT increase by 16.5% and 19.2%. The results give us a signal that with the decreasing of worker efficiency, the GTT and SFT will drop faster and faster and it is more harmful to SFT.

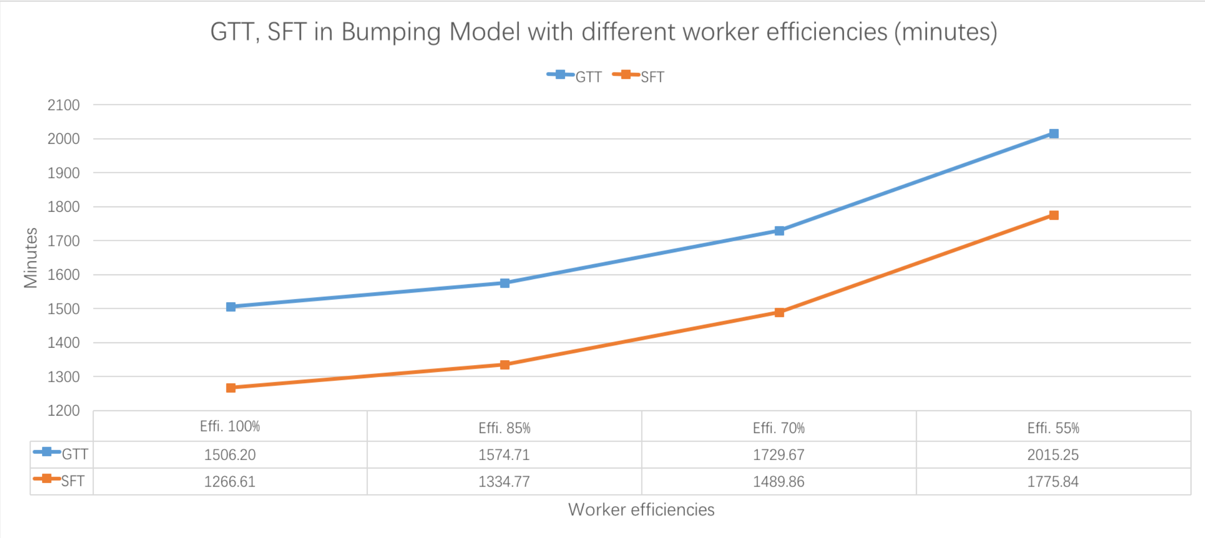


Figure 4.18: GTT and SFT in the Bumping Model with Effi. 100%, Effi. 85%, Effi. 70%, and Effi.55%

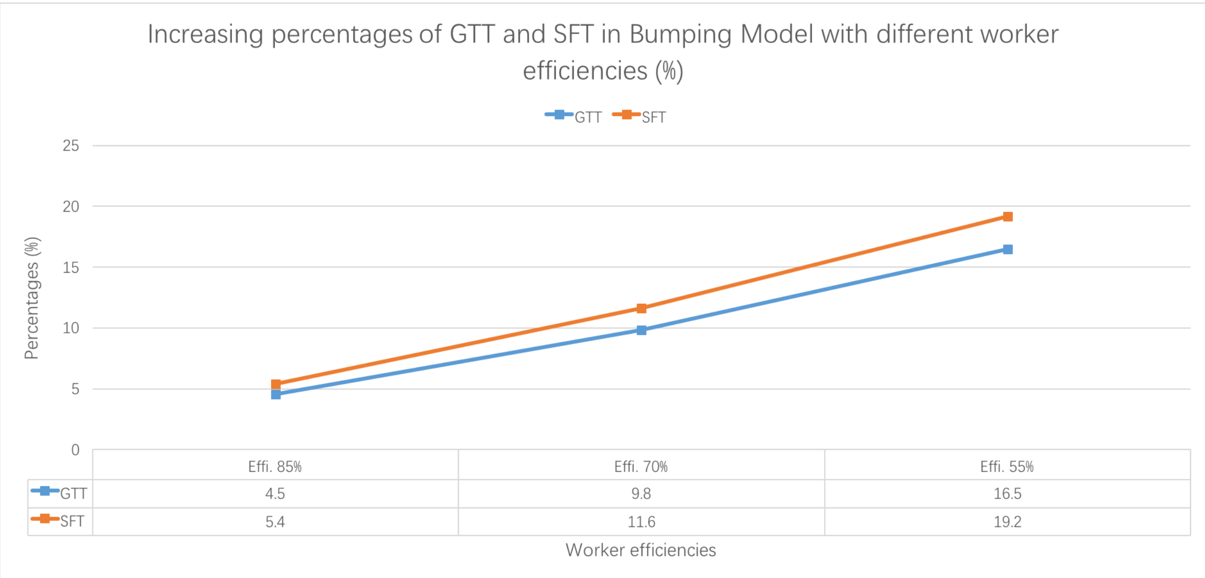


Figure 4.19: Increasing percentages of average GTT and SFT in each change of worker efficiencies in the Bumping Model

About Tardiness and Tardy, *Figures 4.20* shows that they are increasing with a stable speed firstly and then the speed become higher and higher. The speed of increasing is keeping more or less in a stable value when the worker efficiency changes from Effi. 100% to Effi.85% and then to Effi. 70%, the tardiness and tardy increase 114% and 63.01%; from Effi.85% to 70%, the two indicators increase 79.18% and 60.65%, while from Effi. 70% t0 Effi. 55%, they increase by 131.3% and 107.4%, which is a higher speed.

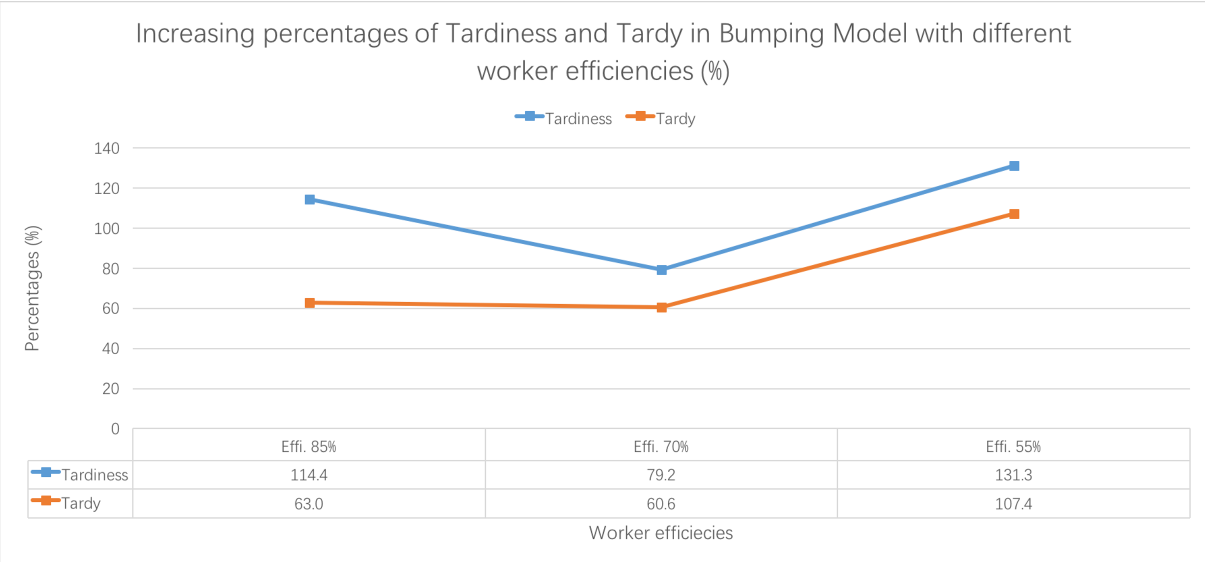


Figure 4.20: The increasing percentages of Tardiness and Tardy in the Bumping Model with each change of work efficiency

Considered the same way in research question 1 and the part of release period, the idleness rate of each worker, workload in the system, and the percentage of time worker spend on the bumping work have to be analyzed.

The idleness rate of each worker (utilization)

As is shown in *Figure 4.21*, the idleness rates of workers are all decreasing as the worker efficiency drops from 100% to 85%, to 70%, and then to 55%. Besides, we can see in *Table 4.1*, the idles rates of all the workers expect $worker\ 5$ is increasing more and more quickly, while that of $worker\ 5$ increases in a slow-down speed.

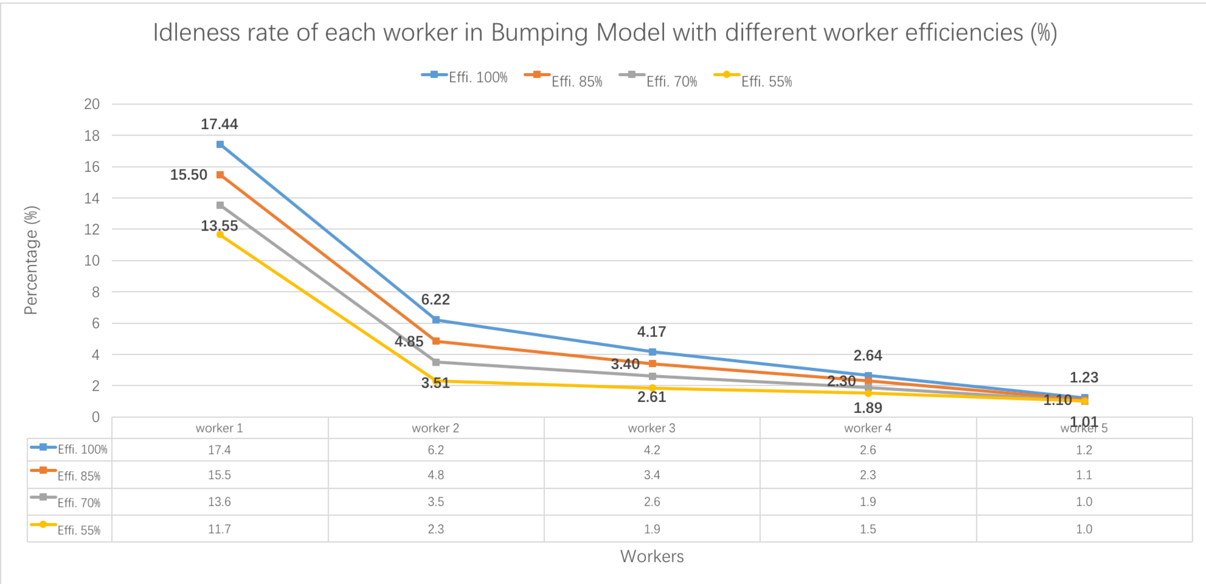


Figure 4.21: The idleness rates of workers in Effi.100%, Effi.85% and Effi.75%

	Worker 1	Worker 2	Worker 3	Worker 4	Worker 5
Effi. 100% - Effi. 85%	-11.13%	-22.05%	-18.37%	-12.79%	-10.67%
Effi. 85% - Effi. 70%	-12.55%	-27.6%	-23.25%	-17.9%	-8.12%
Effi. 70% -Effi. 55%	-13.93%	-34.69%	-29.06%	-19.17%	-0.10%

Table 4.1: Increasing percentages of idleness rates of workers in each change of worker efficiency

The workload in the system (WIP)

As another indicator, the average shop loads of five stations increase step by step with the worker efficiency drops from 100% to 85%, then to 70% and finally to 55%, which gives us the answer why the SFT and GTT increase with lower worker efficiency in the Bumping Model. The reason is the lower efficiency affects the speed of removing the overload by workers and also make the frequency of the bumping process reduce. Besides after calculating, we gain that the standard deviations of the total workload in the system are 1044 minutes, 528 minutes, 554 minutes, and 1232.9 minutes respectively when the efficiencies are 100%, 85%, 70% and 55%. It seems that the Bumping Model with worker efficiency equals 85% is the better choice if we want to reduce the variance of the total workload in the system without sacrificing the GTT and SFT too much.

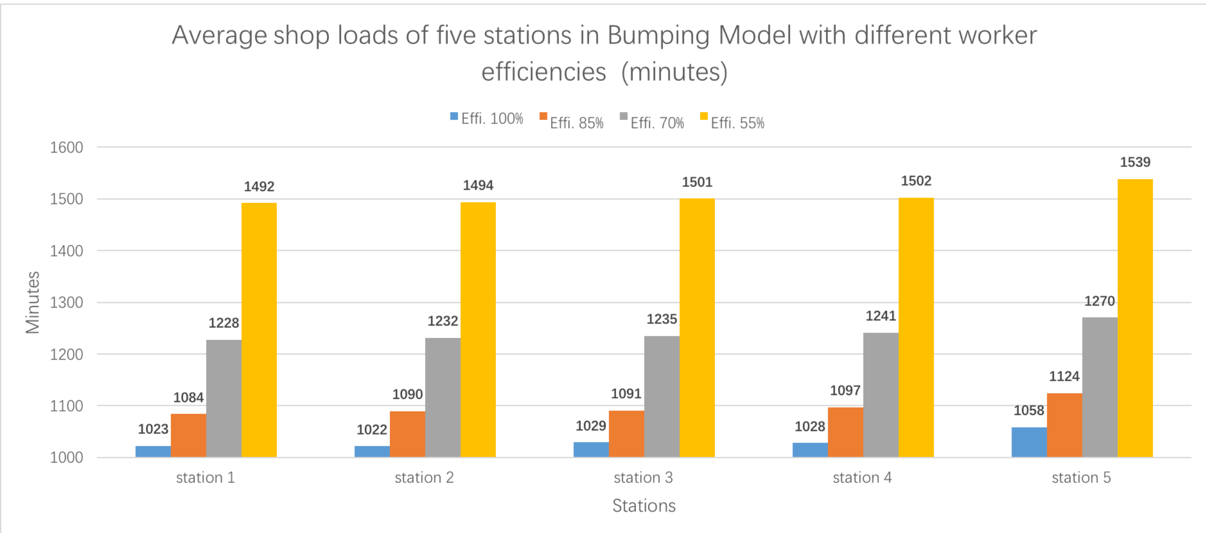


Figure 4.22: Average shop loads of five stations in the Bumping Model with Effi. 100%, Effi. 85%, Effi. 70%, and Effi. 55%

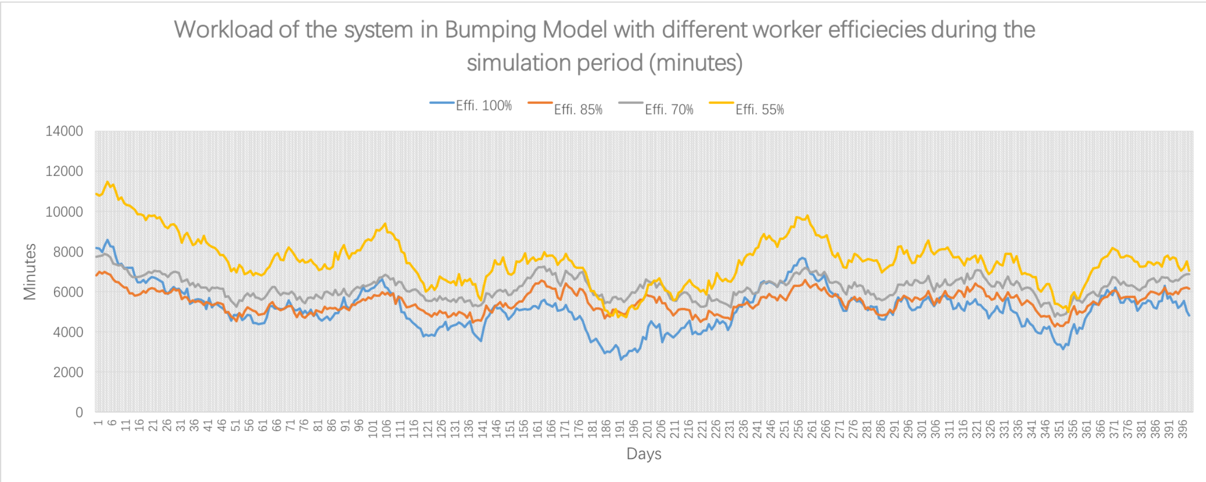
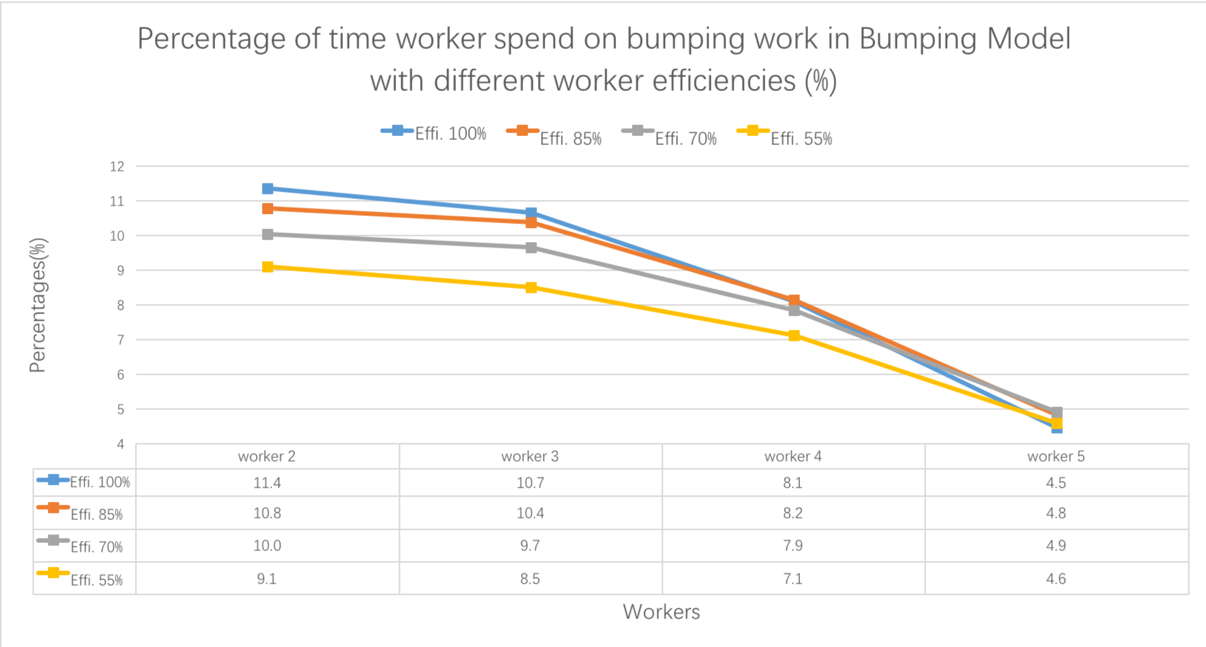


Figure 4.23: Total Workload of the system in the Bumping Model with Effi.100%, Effi. 85%, Effi. 70%, and Effi. 55% along the whole simulation period

The percentage of time worker spend on the bumping work (bumping frequency)

With the lower and lower worker efficiency, *Figure 4.24* shows that the percentages of time $worker\ 2$ and $worker\ 3$ spend on bumping work decreases, while that of $worker\ 4$ and $worker\ 5$ increase firstly and then decrease. *Figure 4.25* gives us an interesting phenomenon that the dropping rate always increases while the rising rate always decreases. Finally, we can conclude that with the worker efficiency decrease, the percentage of time all workers spend on bumping work will drop till they are equal to zero.



	worker 2	worker 3	worker 4	worker 5
Effi. 100%	11.4	10.7	8.1	4.5
Effi. 85%	10.8	10.4	8.2	4.8
Effi. 70%	10.0	9.7	7.9	4.9
Effi. 55%	9.1	8.5	7.1	4.6

Figure 4.24: Percentage of time worker spend on the bumping work in the Bumping Model with Effi. 100%, Effi. 85%, Effi. 70%, and Effi. 55%

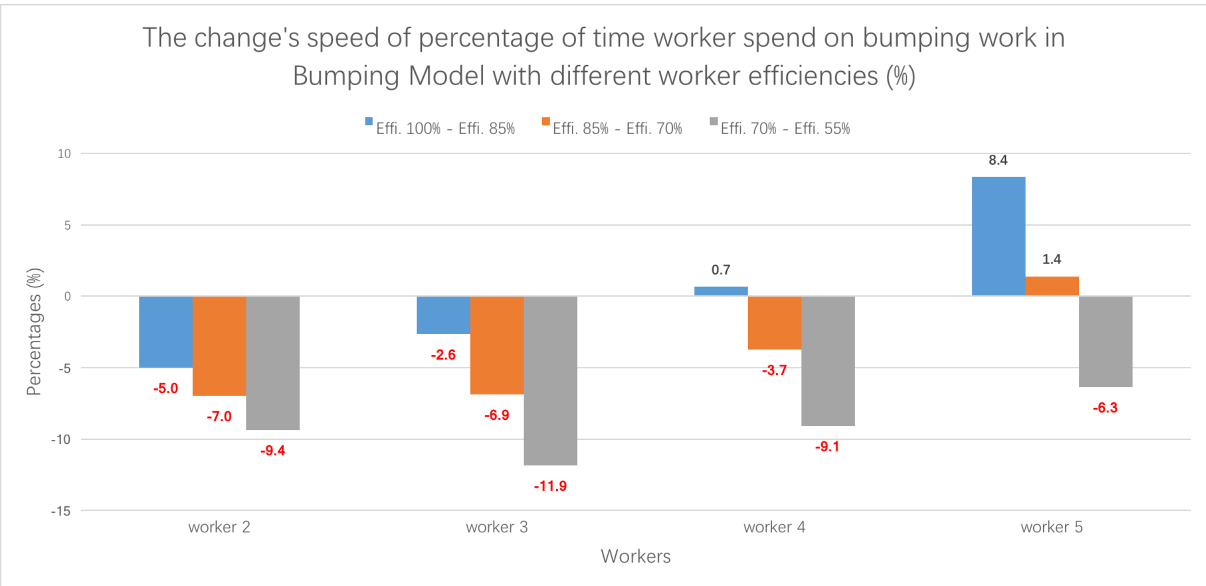


Figure 4.25: The change's speed of percentage of time worker spend on the bumping work in the Bumping Model with Effi. 100%, Effi. 85%, Effi.70% and Effi. 55%

4.2.3 The impact of transfer time

As we all know that if the worker wants to catch the job from the previous worker, he has to spend the time to go to the upstream station to do it in a real production scene. In this case, it is necessary for us to study on the influence on production performances of this kind of time. In our simulation model, it is called transfer time. Base on the configuration of our simulating production system, the thesis set the transfer time equal to 5 minutes, 10 minutes and 15 minutes to test. *Figure 4.26* shows that GTT and SFT of orders go in the same trend with the change of transfer time. And we also try to find the relationship between the values of GTT, SFT and that of transfer time. Interestingly, the values of GTT, SFT have linear relationships with the value of transfer time. And the relationship functions can be seen in the *Figure 4.26* and *Table 12*. With this assumption, we set two equations in order to calculate the value of transfer time which can destroy all the advantages of Bumping Model. The result shows that if the value transfer time is bigger than 70 minutes, the benefits brought by Bumping Model will be removed or be extremely discounted. In order to prove the assumption, the thesis test the Bumping Model in Transfer Time equal to 70%. The *Figure 4.27* gives that the GTT and SFT only decrease around 8% and 9%, so with the transfer time longer than 70 minutes, Bumping Model may fail.

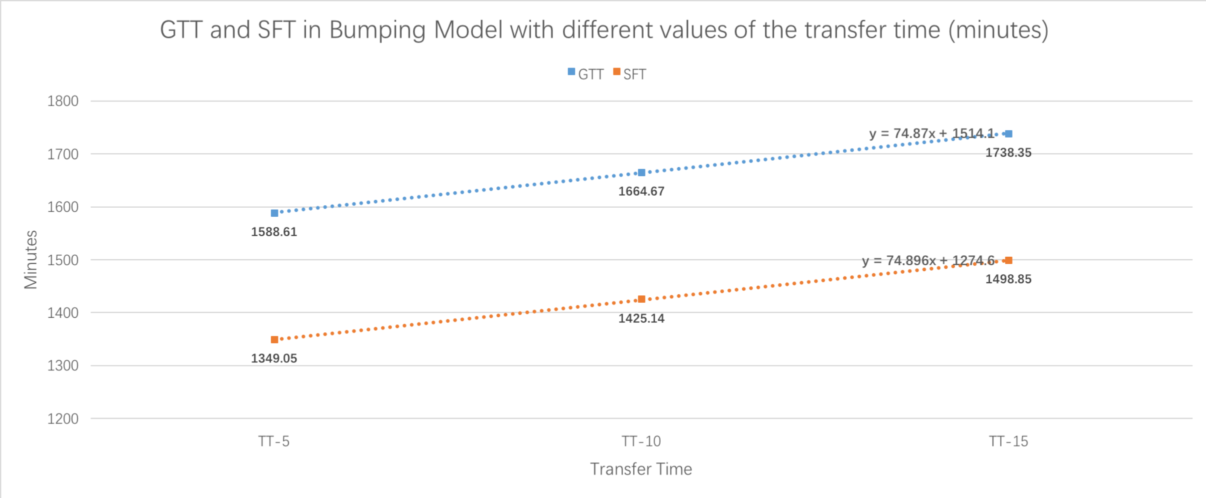


Figure 4.26: Average GTT and SFT in the Bumping Model with TT-5, TT-10 and TT-15

	GTT	Delta GTT	SFT	Delta SFT
TT-0	1506.20	0.00	1266.61	0.00
TT-5	1588.61	82.41	1349.05	82.45
TT-10	1664.67	76.05	1425.14	76.08
TT-15	1738.35	73.69	1498.85	73.71

Table 4.2: The Delta GTT and Delta SFT in TT-5, TT-10 and TT-15

$$\begin{cases} 2510 = 74.87x_1 + 1514.1 \\ 2270.32 = 74.896x_2 + 1274.6 \end{cases} \rightarrow \begin{cases} x_1 = 69.8 \\ x_2 = 69.8 \end{cases}$$

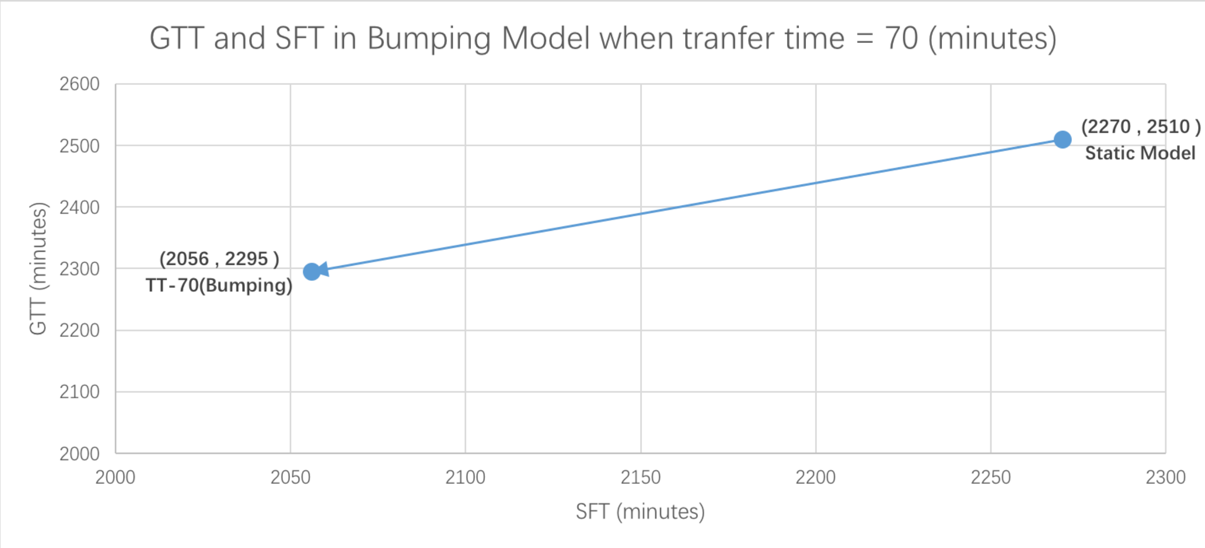


Figure 4.27: GTT and SFT in Bumping Model with TT-70

Besides, as for Tardiness and Tardy, *Figure 4.28* presents the same results as GTT and SFT to us with introducing the factor of transfer time.

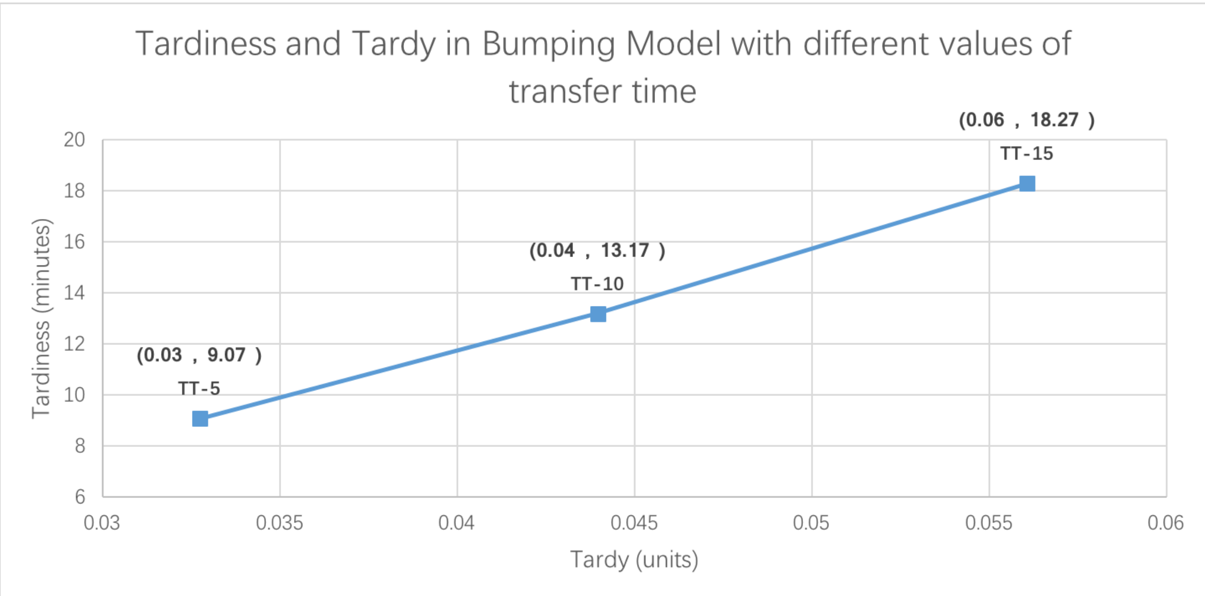


Figure 4.28: Tardiness and Tardy in Bumping Model with TT-5, TT-10 and TT-15

The idleness rate of each worker (utilization)

In order to get the root cause, there is still an analysis on the idleness rates of workers. It is clear to see in *Figure 4.29* only the idleness rate of $worker\ 1$ decreases while those of other workers increase. In the meantime, the speed of decrement or increment is slowing down. Obviously, it is easy for us to understand the results, the $worker\ 1$ spend more time on working for decreased times of bumping, and the other workers spend much more time on the transfer process.

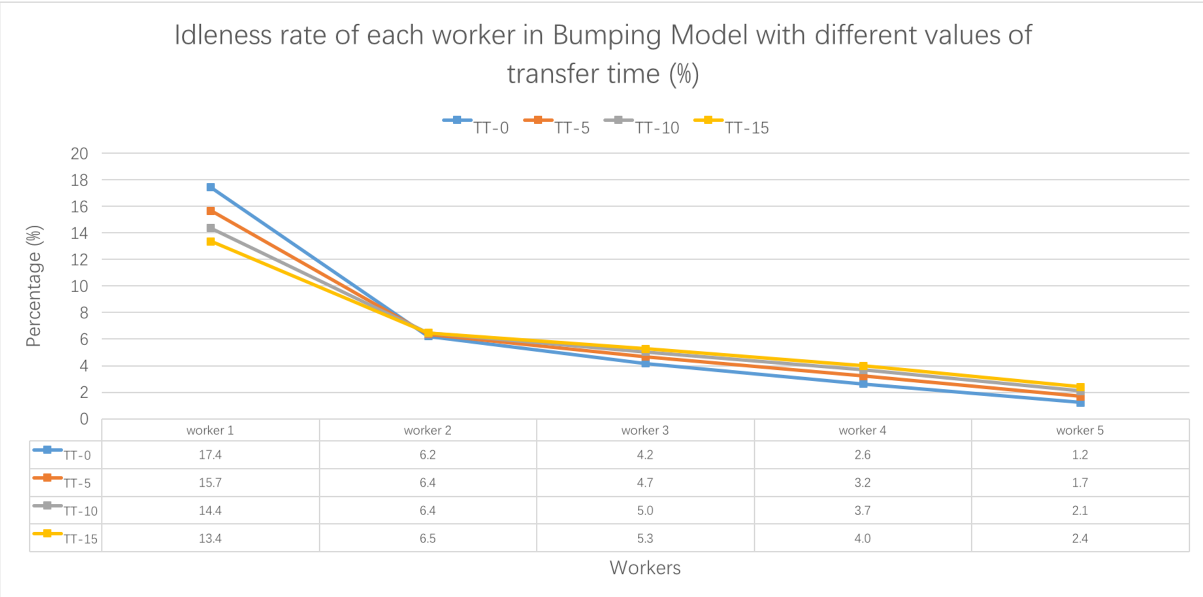


Figure 4.29: Idleness rates of workers in the Bumping Model with TT-5, TT-10 and TT-15

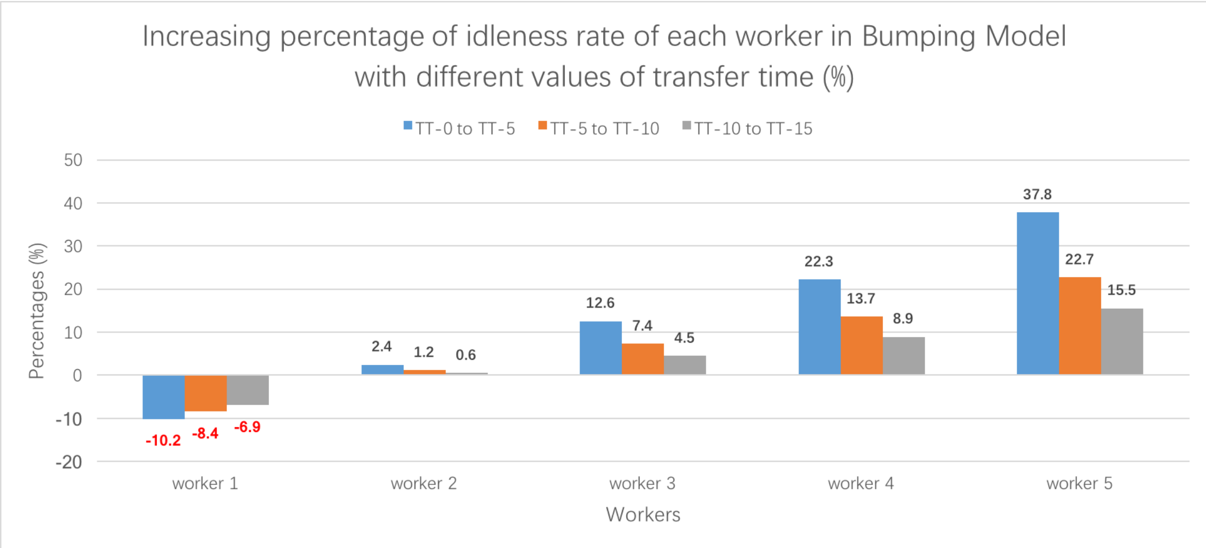


Figure 4.30: Increasing percentage of idleness rate of each worker in the Bumping Model with TT-5, TT-10 and TT-15

The workload in the system (WIP)

With the analysis above, so it is also easy for us to understand the growths of the average shop load of each station. However, the standard deviation of the total workload in the system shows an unaffected state, they are 1044 minutes, 1053 minutes, 1073 minutes and 1101 respectively, even if the transfer time is increasing.

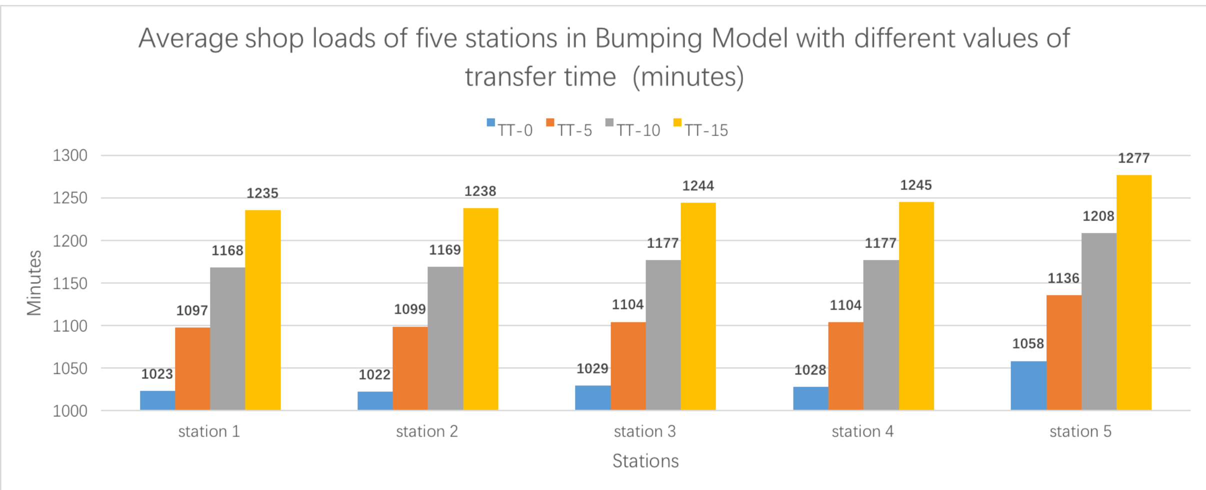


Figure 4.31: Average shop loads of five stations in Bumping Model with TT-5, TT-10 and TT-15

	station 1	%	station 2	%	station 3	%	station 4	%	station 5	%
TT-0	1022.8		1022.0		1029.2		1028.1		1058.3	
TT-5	1097.5	7.3	1098.9	7.5	1104.4	7.3	1104.3	7.4	1135.7	7.3
TT-10	1168.4	6.5	1169.1	6.4	1176.7	6.5	1176.6	6.6	1208.3	6.4
TT-15	1235.4	5.7	1237.8	5.9	1244.0	5.7	1244.9	5.8	1276.7	5.7

Table 4.3: Increasing percentage of average shop loads in the Bumping Model with TT-5, TT-10 and TT-15

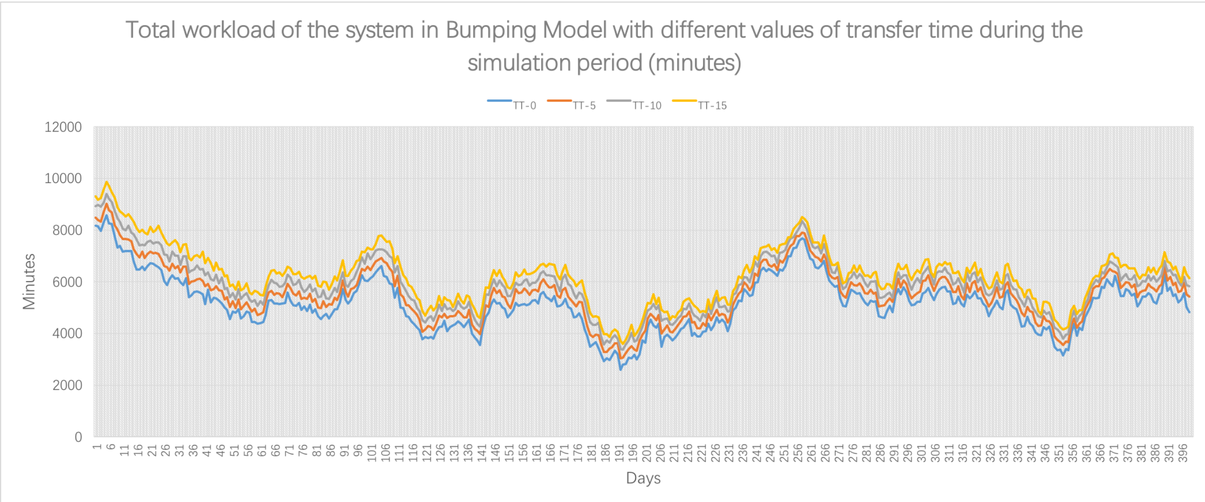


Figure 4.32: Total workload of the system in the Bumping Model with TT-0, TT-5, TT-10 and TT-15 along the whole simulation period

The percentage of time worker spend on bumping work (bumping frequency)

Figure 4.33 and *Figure 4.34* tell us the percentage of time worker spend on the bumping work in the Bumping Model is decreasing with the increasing transfer time, and the speed of decreasing becomes lower and lower from worker 2 to worker 5 and from TT-5 to TT-15.

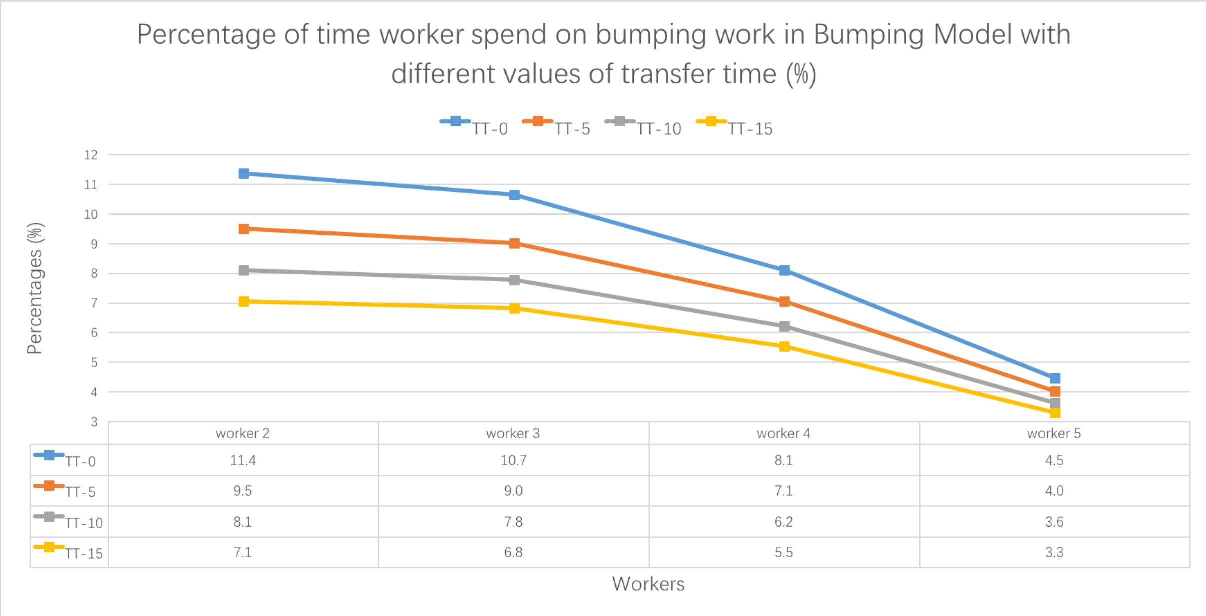


Figure 4.33: The percentage of time worker spend on the bumping work in the Bumping Model with TT-0, TT-5, TT-10 and TT-15

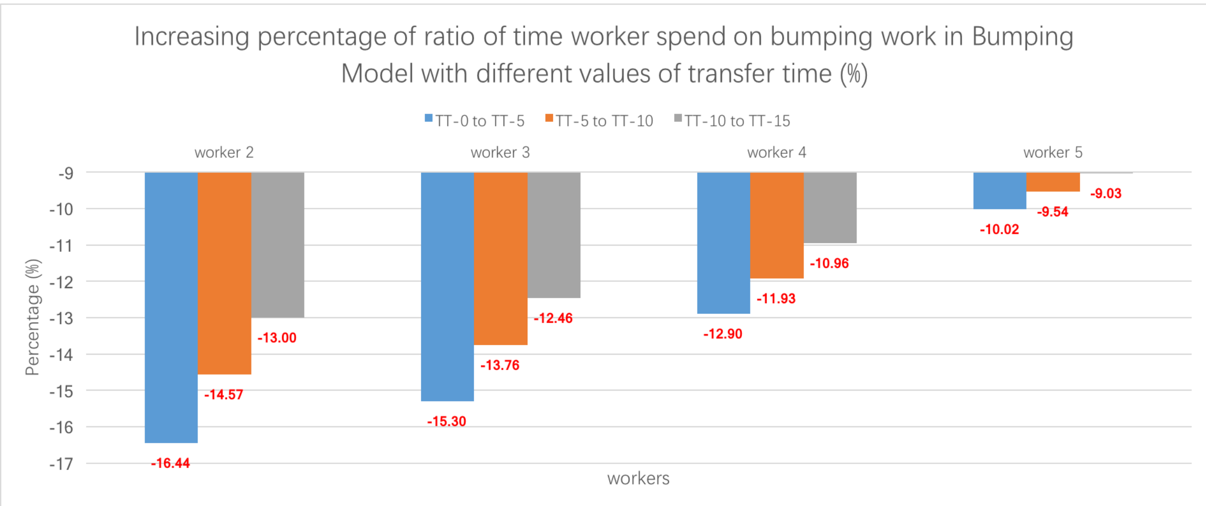


Figure 4.34: Increasing percentage of ratio of time worker spend on bumping work in Bumping Model with TT-0, TT-5, TT-10 and TT-15

4.2.4 Impact of combining the three factors

After a long analysis of the single factors, here we try to combine all the factors. The table below gives all the combinations and all experiment configurations that have been tested in the thesis. The results are shown in *Figure 4.35*.

Models	Period	Level of efficiency	Transfer time (minutes)	GTT(minutes)	SFT(minutes)
Static	P-240	/	/	2301.80	2181.78
Static	P-480	/	/	2509.93	2270.32
Bumping	P-240	100%	0	1247.94	1127.71
Bumping	P-240	100%	15	1529.54	1409.27
Bumping	P-240	100%	30	1719.46	1599.13
Bumping	P-240	85%	0	1418.25	1297.90
Bumping	P-240	85%	15	1632.29	1512.16
Bumping	P-240	85%	30	1804.48	1684.11
Bumping	P-240	70%	0	1583.03	1462.77
Bumping	P-240	70%	15	1764.07	1643.77
Bumping	P-240	70%	30	1918.19	1797.91
Bumping	P-480	100%	0	1506.20	1266.61
Bumping	P-480	100%	15	1738.35	1498.85
Bumping	P-480	100%	30	1933.18	1693.87

Bumping	P-480	85%	0	1574.71	1334.77
Bumping	P-480	85%	15	1838.81	1599.32
Bumping	P-480	85%	30	2014.68	1775.27
Bumping	P-480	70%	0	1729.67	1489.86
Bumping	P-480	70%	15	1979.96	1740.55
Bumping	P-480	70%	30	2133.20	1893.81

Table 4.4: Results of some experiment configurations (combining the three factors)

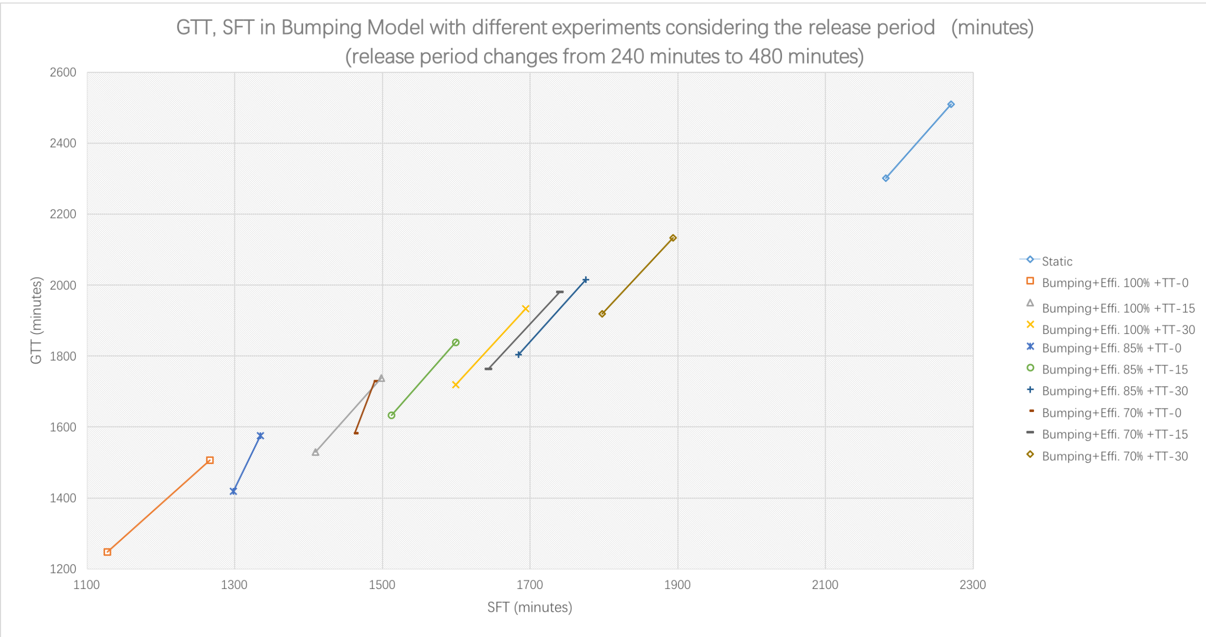


Figure 4.35: GTT and SFT in Bumping Model with different experiments considering release period

As we can see in *Figure 4.35*, nearly all the changes of GTT and SFT from P-240 to P-480 are in the same trend, but there are two special one, there are the combinations: Bumping+Effi.70% + TT-0 and Bumping + Effi. 85% + TT-0. They give us this information that in low efficiency will release the impact of release period on GTT and SFT when the value of transfer time is 0. But the transfer time doesn't have the same impact on the release period.

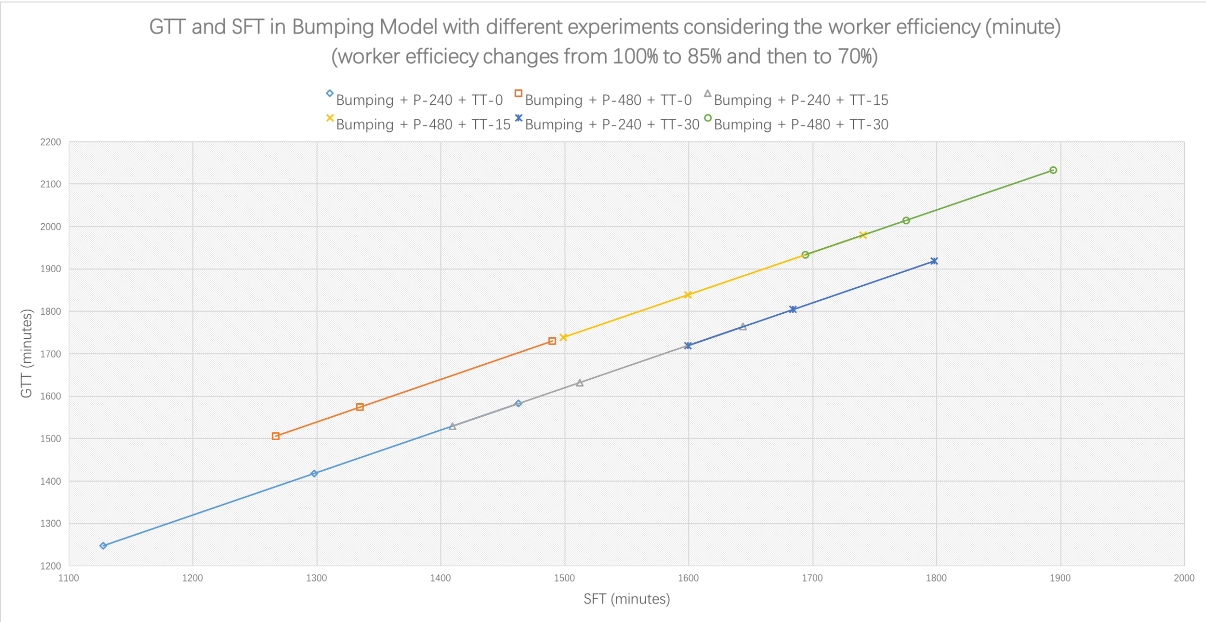


Figure 4.36: The average GTT and SFT in the Bumping Model with different experiments considering the worker efficiency

Figure 4.36 shows us that transfer time and release period will release the impact of worker efficiency on the performances of the system in the Bumping Model. It is very obvious when worker efficiency changes from 100% to 85% and the value of transfer time is equal to zero, the GTT and SFT increase a lot in Period 240 while only a little in Period 480.

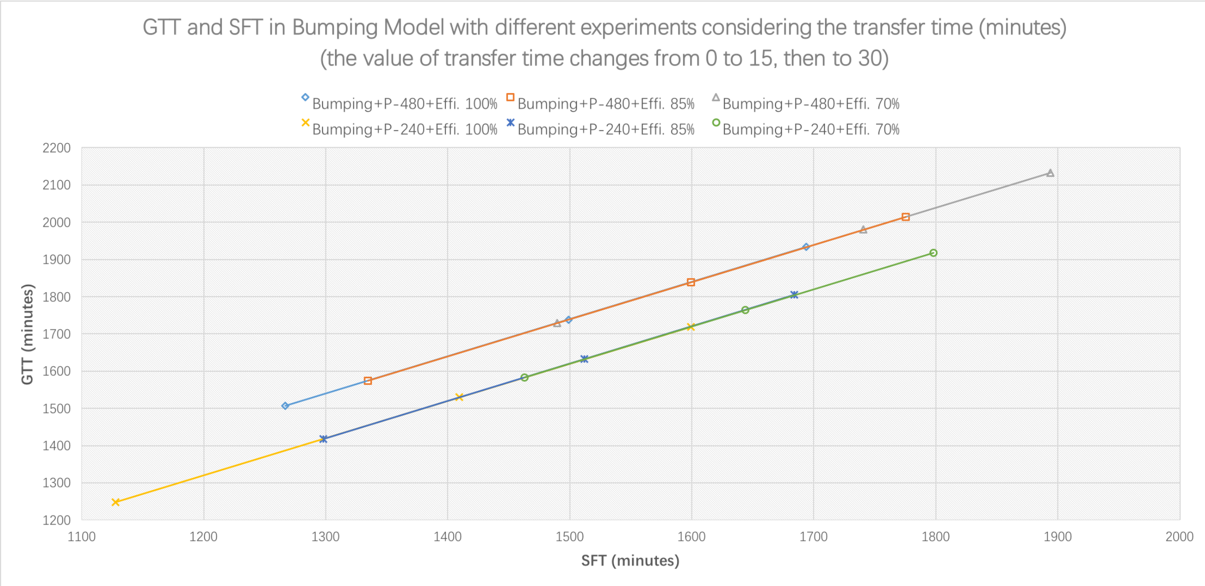


Figure 4.37: The GTT and SFT in Bumping Model with different experiments considering the transfer time

Figure 4.37 illustrates that both the release period and worker efficiency will relieve the impact of transfer time on performances of the system in the Bumping Model.

4.2.5 Deeper analysis on the processing time

Actually, this part is the only an attempt to study the impact of processing time on the Bumping Model. There are two reasons that why we do this attempt: (a), one is that the original of Bumping Model, the Baton Zone Balancing, benefits from a decreasing processing time configuration; (b), another one is that the huge idleness rate of $worker\ 1$ in Bumping Model shown in research question one. The huge idleness rate of $worker\ 1$ gives us the motivation to assign more worker to $worker\ 1$ so that the simulation system can utilize the human resources sufficiently. In order to do this, firstly, we should propose some assumptions to modify the processing time in the five stations. With the start point that this part is only an attempt, in other words, it is only an exploration, so we assume that: (a), the total of mean processing time is a constant which equals to 150 minutes (30 minutes * 5); (b) the variance of processing time on each station keep immobile. (c) the modifications of the processing times are based on the idleness rate of worker of the previous experiment; (d) the start point is the original one that the means of processing time in five stations are same (30minutes).

As we can know in the research question 1, the idleness rates of workers are showing in the following table.

	Worker 1	Worker 2	Worker 3	Worker 4	Worker 5
Idleness rate	17.44%	6.22%	4.17%	2.64%	1.23%

Table 4.5: Idleness rates of workers in Bumping Model (P-480 + Effi.100% + TT-0)

The modifying steps

1. Calculating the average idleness rate
2. Calculating the differences between the idleness rates of five workers and average idleness rate respectively
3. Modifying the mean of processing time in following formula

$$Modified\ mean_i = 30 + 30 * Delta\ idleness\ rates_i$$

$where$:

$Modified\ mean_i$: $the\ modified\ mean\ of\ processing\ time\ of\ station\ i$

$$Delta\ idleness\ rates_{(i)}: the\ difference\ between\ idleness\ rate\ of\ worker\ i\ and\ the\ average\ idleness\ rate$$

Here, we test the modified means of the processing time of five stations.

The modified one: $JOBS\_MEAN = [33.3, 30.0, 29.3, 28.9, 28.5]$, the sum of means is same with the original one, also the variances.

The results show in *Figure 4.38* that both GTT and SFT decrease after the mean of process time in five stations are modified. They decrease by around 21.55% and 25.65% respectively, which is actually a huge reduction, and the improvement of SFT is slightly better than that of GTT. As for Tardiness and Tardy, they are nearly equal to zero after this modification (see in *Figure 4.39 and Figure 4.40*). In order to find the reason, the thesis analyzes the idleness rate of each worker, workload in the system and the percentage of time worker spend on bumping work.

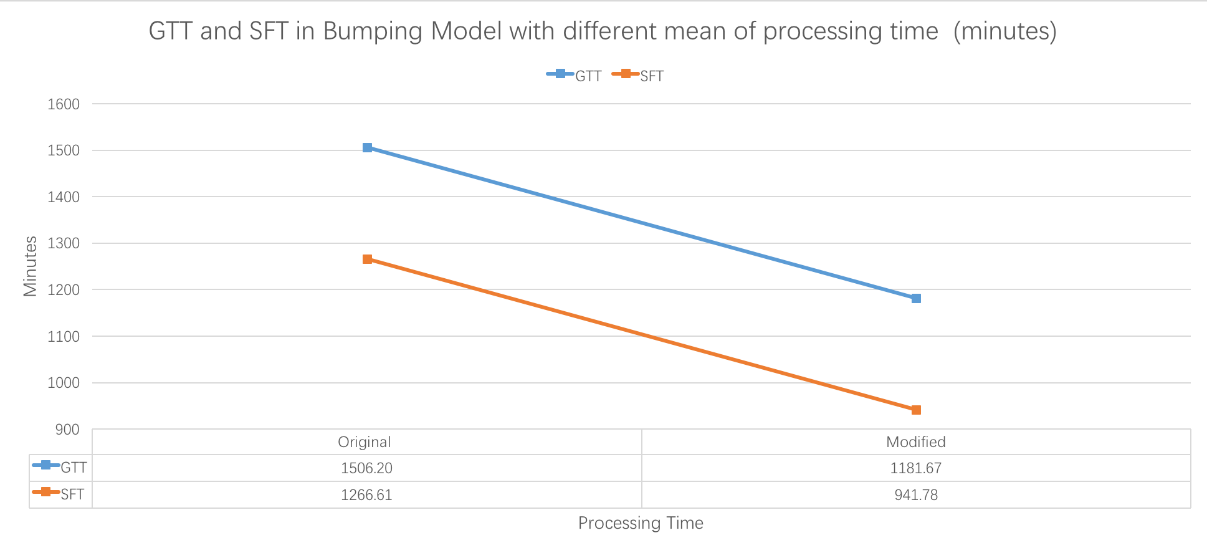


Figure 4.38: GTT and SFT in the Bumping Model with Original and Modified Processing Time

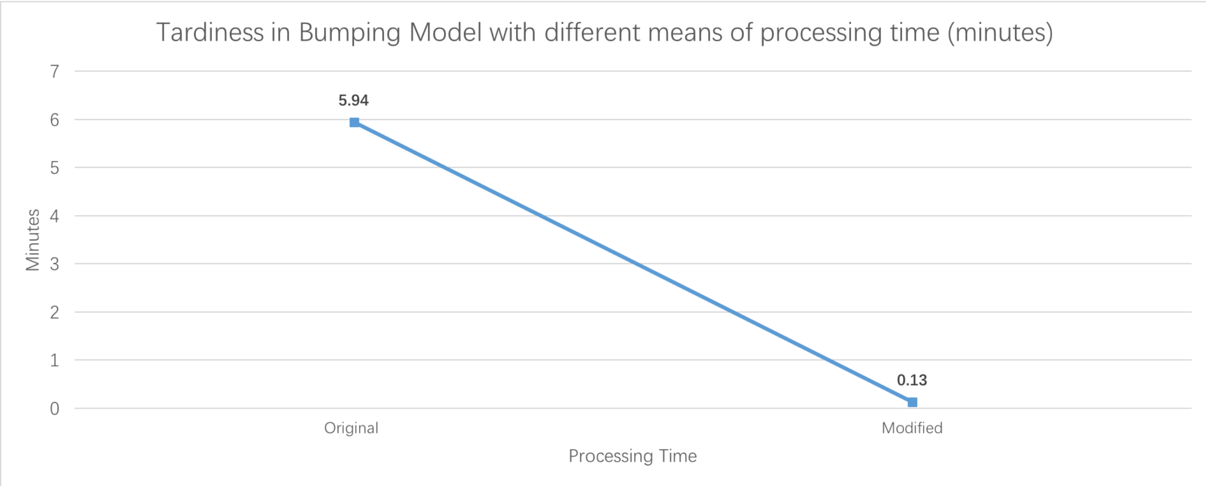


Figure 4.39: Tardiness in the Bumping Model with Original and Modified processing time

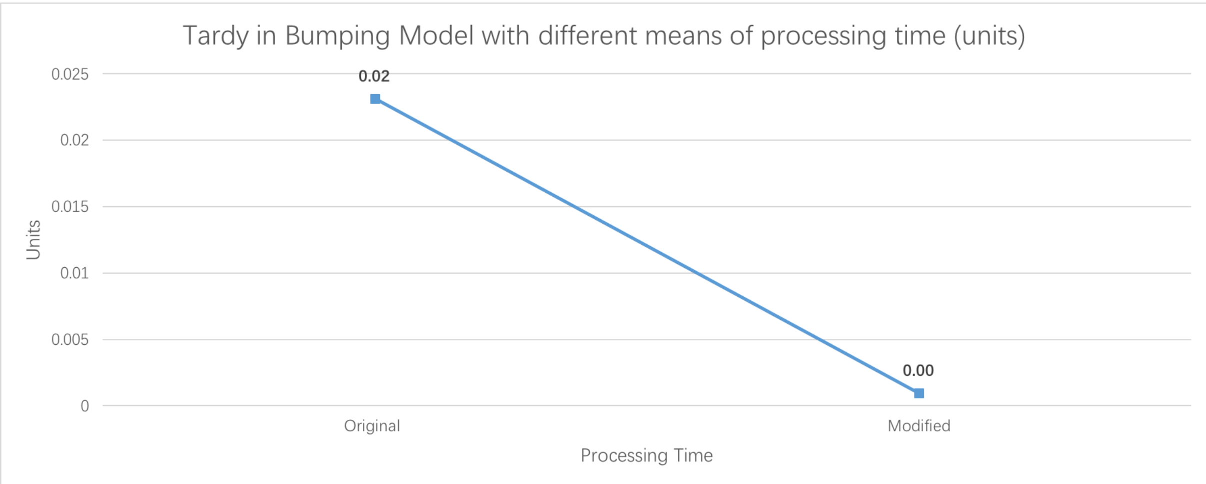


Figure 4.40: Tardy in the Bumping Model with Original and Modified Processing Time

The idleness rate of each worker (utilization)

As we can see in the *Figure 4.41*, only the $worker\ 1$ and $worker\ 2$ become busier and other workers attain more time to be idle. Especially, the idleness rate of $worker\ 1$ reduce a lot, around 28.8%. However, the average idleness rate of workers doesn't change.

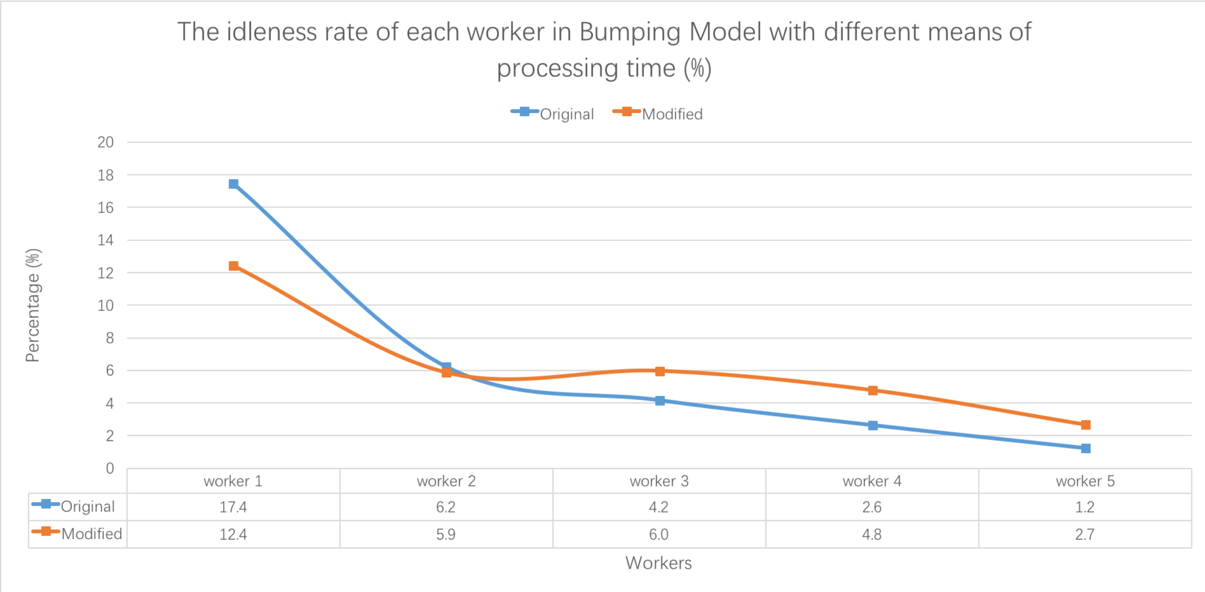


Figure 4.41: Idleness rate of each worker in Bumping Model with Original and Modified processing time

Workload in the system (WIP)

It is meaningful to study on the shop load for it is a good representation of WIP. *Figure 4.42* illustrates a decreasing trend of average shop loads of five stations in the simulation model. From the original processing time to the modified one, the average shop loads of five stations decrease 21.46%, 29.26%, 31.21%, and 32.35% respectively, which are noticeable reductions. Besides, about the fluctuation of total workload of the system during the whole simulating period, the standard deviation of the total workload in the system changes from 1044 minutes to 790 minutes, which is also an improvement.

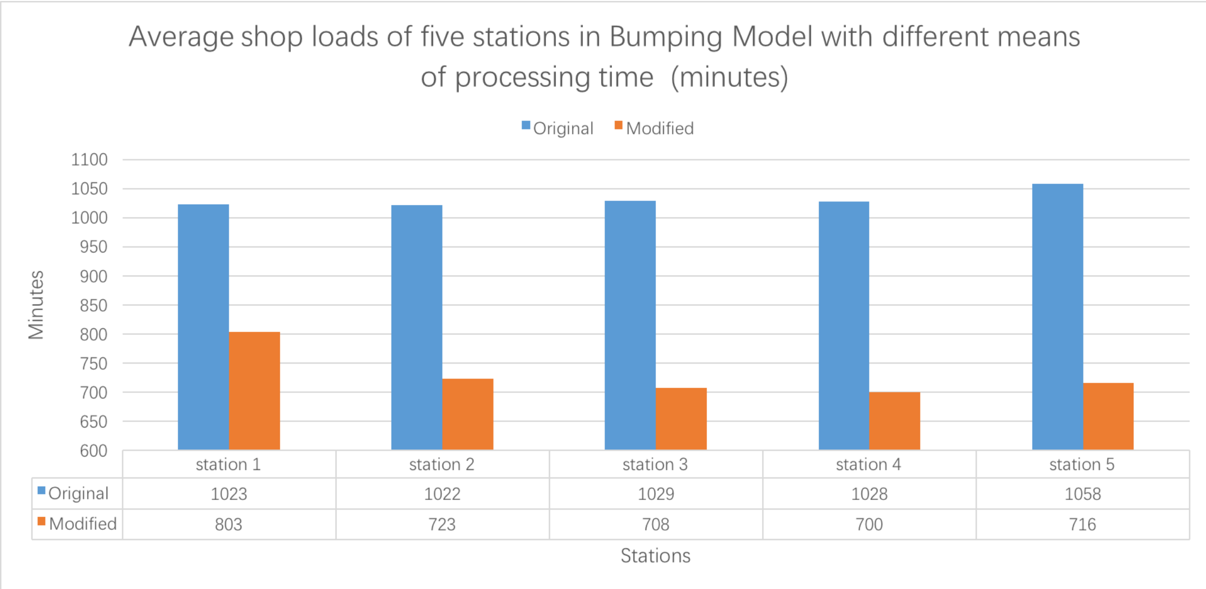


Figure 4.42: Average shop loads of five stations in Bumping Model with Original and Modified processing time

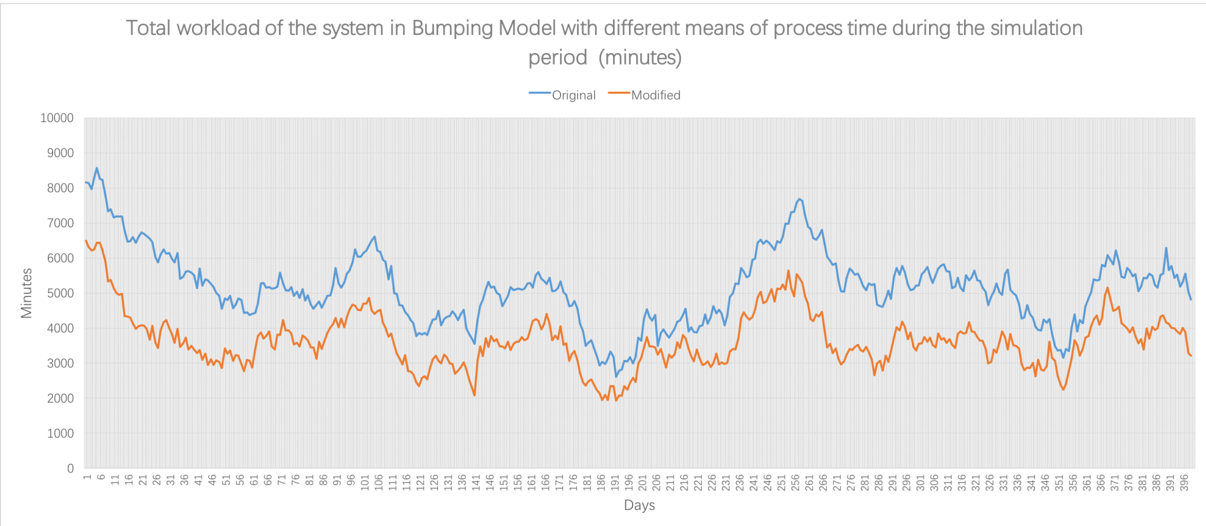


Figure 4.43: Total workload in the system along the whole simulation period in different processing time

The percentage of time worker spend on the bumping work (bumping frequency)

The percentage of time worker spend on bumping work increase a lot after modifying the mean of processing time on each station. The increasing percentage of the four workers are 53.34%, 54.51%, 65.14% and 81.60% respectively. Which tells us that there is more bumping process after modifying the means.

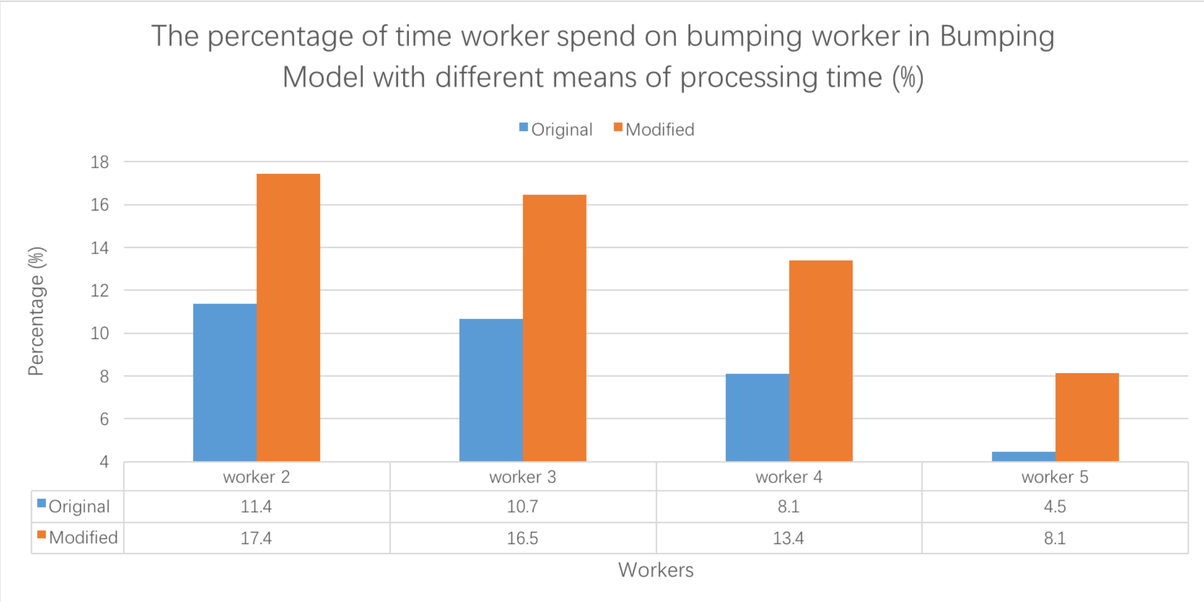


Figure 4.44: The percentage of time worker spend on the bumping work in the Bumping Model with original and modified processing time

5. DISCUSSION AND CONCLUSIONS

This chapter of the thesis is composed of a summary of the findings while answering the research questions. The first part is the answer to the research questions and some managerial implications. The second part discusses the limitations and future researches.

5.1 Answers to the research questions

1. What are the contributions of the Bumping Model to performances of the flow shop?

The first research question aimed at assessing the effect of Bumping Model have on the shop floor system's simulation model developed and used. In the model, the order release mechanism is simplified to periodical release and set the value of period equal to 480 (one time a day). The first reason is the order release method is common in real life, the second reason is to ensure enough order at the beginning of the day. Within the shop floor, a Bumping workforce is adopted to compare with the Static one. As aforementioned, Bumping Model is an innovative way of capacity adjusting, which is grounded from the concept of Baton Zone Balancing. From the literature, we found that the performance of Baton Zone Balancing is better than that of Station Balance. Thus, the thesis has applied the algorithm in Baton Zone Balance on the output control.

After the analysis in chapter 4.1, the results confirm the beneficial effects of Bumping Model on each system performance evaluated in the study (GTT, SFT, average tardiness, lateness and tardy etc.). The GTT decreases by 40% and SFT decreases by 44% compared with the Static Model. The average lateness, tardiness and tardy reduce more dramatically around by 200%, 96%, and 91% respectively. And we found in the thesis the reason for this huge improvement is the increments of workers' utilization and around 50% decrement of workload in the system during the simulation process. Thus, there are many reasons implement the Bumping Model in a production system to improve its performances. This methods of capacity adjustment can easily reduce the WIP and improve the utilization of human resources even if it is not so much. Besides, from the results of the simulation, the idleness rate of $worker\ 1$ is extremely higher than those of other workers, this phenomenon is corresponding to the attribute of a Baton Zone Balancing line which always put the worker in low efficiency in the start of the line. And based on this attribution, the thesis also goes deeper to try to decrease the idleness rate of $worker\ 1$ by modifying the values of the mean of processing time in the five stations of our simulation model. The research's results can be seen in chapter 4.2.5 and it will be discussed later. Totally, the thesis gives us a satisfactory answer to question 1 that the Bumping Model brings a distinct improvement of the production system. Therefore, it is meaningful for the manager to implement this kind of output control in real life.

Finally, even if the Bumping Model can improve the performance of the manufacturing system, attaining a flexible workforce means more spend of money and time in the practical world. How to trade off the improvement of performance and the cost spent to training is more realistic to managers. Also, from the results got from the chapter 4.1, although the workload of the system in the Bumping Model reduce, the variance increases a lot, which will make it tough to predict the WIP during the production process. Another problem presented in the literature review is that the Bumping Model requires more detailed process steps which can let the worker easily take over the work from the previous worker, it can bring some hard technical work to engineers.

2. ***How do factors release period, worker efficiency, transfer time and the processing time, affect performances of the flow shop in the Bumping Model?***

Although the Bumping Model is an innovative method to relocate the labor resources, it is same to other methods, in which the performance of the workforce will be influenced by the level of labor flexibility. Actually, it is easy for us to infer that system with a higher level of worker efficiency will perform better, but the start point of the thesis to study on this topic is trading off the cost of training and system's performance. As proposed above, even if the Bumping Model get advantages over the Static one, economical factor can't be neglected in a practical environment. Different level of work efficiency means different investment in a training program for workers, in this case, it is meaningful to discuss the variation of performance for the different level of labor flexibility. As aforementioned, the worker flexibility is composed of the three questions: where, when and who, but actually, the more substantial problems are about the level of worker flexibility and level of worker efficiency. In this thesis, for the time constraint, only the level of worker efficiency is studied on. With the information drawn from the simulation results, we can give following conclusions:(a), the lower worker efficiency brings increasing GTT and SFT; (b), the worker efficiency has a huger influence on SFT than GTT. (c), with the stable decreasing of worker efficiency, the GTT and SFT increase in a higher and higher speed. (d), the Tardiness, Lateness and Tardy increase in the same speed with that of worker efficiency's decreasing. For the conclusion (a), it is obvious that the reason is the lower efficiency means spending more time. Conclusion (b) give us the information that Bumping Model is definitely an output control approach which plays a main role within the shop floor. About conclusion (c), it is a useful signal to us that if a manager wants to apply Bumping Model in shop floor, it is better to make the worker be more skilled on the previous stations. When we go deeper to find the reason of these conclusions, we find that the idleness rates of workers decrease and the workload in the system increase, while it is interesting that the standard deviation of workload in the system decreases when worker efficiency changes from 100% to 85%, but increase when worker efficiency changes from 85% to 70% and then to 55%. It means that the WIP become more stable when worker efficiency is equal to 85% compared other values of worker efficiency, which gives the manager a signal to lower the efficiency properly to keep a stable WIP when implementing the Bumping Model. Besides, when we study on the percentage of time worker spend on bumping work, there is a situation that all the workers' percentages are decreasing except $worker\ 4\ and\ \ worker\ 5$, but finally, both of percentages of them decrease. After comparing the percentage of time worker spend on bumping work at different levels of worker efficiency. We find the lower efficiency try to balance the percentage of time worker spend on bumping work. Finally, considering the cost and performances of the system, we propose that the most suitable value of worker efficiency in Bumping Model is 85%.

Secondly, based on the order release method applied in the thesis, periodical release (PR) is more common but easily one for it doesn't involve with any workload controlling during the input control. But the release period as the important parameter in PR is tested in the thesis. The results illustrate that GTT and SFT decrease by 45.8% and 48.3% respectively when orders release two times a day (P-240) in the Bumping Model, they decrease by separately 40% and 44.2% when orders release one time a day (P-480) in the Bumping Model compared with the Static Model. So, the conclusion is the Bumping Model benefits from the shorter release period. And when calculating the decreasing rates of GTT and SFT in Bumping Model from P-480 to P-240, we find the decreasing rates of GTT and SFT are 17.15% and 10.97%. The results let us believe that shorter release period gives a more improvement to GTT which represents the lead time of an order. The shorter release period means order stays a shorter time in the Pre-shop pool, which means the gap between GTT and SFT is shortened, so the results are reliable. However, when we are finding the reasons for this improvement, it surprises us that the average idleness rate of workers doesn't change when releasing orders two times a day, even if GTT and SFT are decreasing. So the utilization of workers is not the reason, actually, the lower average workload in the system gives us the answer. Compared with releasing one time a day in the Bumping Model, the average workload in the system decreases by around 20%. Besides, the workload load in the system is more stable during the simulation process and its standard deviation decrease by about 53% (from 1044 minutes to 487 minutes). Also, the percentage of time worker spend on the bumping work increase. These phenomena show the bumping process is good for reducing the WIPs and facilitating the flow of orders in the production system. Finally, it is clear that a shorter release period in Bumping Model brings higher improvement.

Thirdly, as is known in the practical world, the time has to be taken when a worker goes to the upstream station to catch a job from the previous worker. It will give an influence on the utilization of workers, as a result, the performances measured (GTT, SPT, Lateness, Tardiness, and Tardy etc.) will change accordingly. It will be a very important research if managers can redesign the layout of production system every time, so they can easily control the transfer time to make the whole performance of the system to be the most optimized. In this thesis, we set the value of transfer time equal to 5 minutes, 10 minutes and 15 minutes. From the chapter 4.2.3, the thesis actually proposed these opinions:(a), GTT and SFT increase when introducing the concept of transfer time; (b), the increments of GTT and SFT are very stable when rising the value of transfer time from 5 minutes to 10 minutes and then 15 minutes; (c) an increasing transfer times give the system an increasing WIPs and a decreasing percentage of time worker spend on the bumping work; (d). Only the utilization of $worker\ 1$ increases while those of all of the other workers decrease. These conclusions actually are easier for us to understand. After introducing transfer time, the $worker\ 2$, $worker\ 3$, $worker\ 3$ and $worker\ 4$ should spend much time transferring, which giving themselves low utilizations and the system low times of bumping. As for $worker\ 1$, because he can't catch a job from other stations, at the same time, the frequency interrupted by $worker\ 2$ decreases, so his utilization increases. In addition, based on the linear relationship between GTT, SFT and transfer time, we get the maximum value of transfer time which will remove the benefit of Bumping Model. It is around 70 minutes. However, when we give this value to transfer time in the Bumping Model and simulating, there are still decrements of GTT and SFT compared with the Static Model, about 8% and 9% respectively. We believe that even if the long transfer time, the variance of the processing time of a job is enough bigger to get the benefit from the bumping process. In other words, there is more unpredictable order in real life, therefore, a bumping process definitely can give an improvement to the production system. The proper transfer time just can release the impact but can't remove it. Finally, about the fluctuation of total workload of the system in the Bumping Model with different values of transfer time, it doesn't experience any change.

After analyzing the impact of these three factors: worker efficiency, release period and transfer time, the thesis also combine these factors and do the experiments in chapter 4.2.4. After plotting all the data in graphs, these conclusions following are attained:(a), the Static Model is always worse than the Bumping Model; (b), any two factors will relieve the impact of another factor on the performances of the system in the Bumping Model. For these conclusions, we can explain in this way, at first, the release period is the most direct factor to affect the frequency of bumping, lower release period means more bumping processing. The transfer time and the worker efficiency are not the direct reasons to influence the times of bumping process, in contrast, they just influence the time worker spending on the bumping process and then influence the frequency of bumping process. So, we can say that the release period is the most important factors among the three factors, and it will influence the impact of other factors on the performance of Bumping Model.

Finally, much research in literature shows that a worker with higher efficiency should be assigned to the latter station when implementing a Baton Zone Balancing. The aim is to keep the one-piece flow in the line. However, in our model it is impossible to keep a one-piece flow, we can consider it in another concept: WIP. So one-piece flow means WIP as lower as possible. And from the discussion above every time's improvement on GTT and SFT are resulting from lower average workload in the simulation model. Based on this result, Bumping Model as an effective output control method mainly for its function to decrease the average WIP during the production system. It is same like applying Baton Zone Balancing to ensure one-piece flow. Increasing the frequency of bumping will be good for performances of a production system with this theory. However, it is incorrect to put a random and huge decreasing trend on the means of processing time in this thesis for the low level of worker flexibility. In fact, another phenomenon gives us motivation and inspiration for modifying the processing time. It is the imbalance idleness rates of workers in the Bumping Model, especially the $worker\ 1$ who are always be idle. So maybe we can assign more worker to $station\ 1$ to utilizing the resources of $worker\ 1$. In chapter 4.2.5, the thesis proposed a modified approach which is just an assumption and an attempt. However, the results give us anticipating results, the GTT and SFT decrease because of the low WIPs. So, we can conclude that implementing the Bumping Model give us more room to improve the performances of a production system if we can rearrange the production flow. And thanks to the analysis in chapter 4.2.5, we can say that the Bumping Model benefit from the decreasing trend of processing time along the production flow.

In conclusion, the thesis proves that (a), a system in the Bumping Model definitely surpasses that in the Static Model; and (b), an improved system in the Bumping Model should be in a shorter release period, proper worker efficiency (85%), lower transfer time and a decreasing trend of processing time along the production flow.

5.2 Limitations and future researches

This thesis proposes the new model, the Bumping Model, to give a new method of capacity adjusting in the shop floor. And the conclusion is that the new model benefits the performances of the production system in the MTO company. At the same time, the thesis studies on the impacts of some factors on performances of the system in order to provide an improved Bumping Model. However, there are some limitations that can't be denied in the thesis:

Firstly, the results get from the simulation depend on the assumptions made for all the parameters of the production process such as the arrival rate distribution, processing time distribution and the contractual due dates policy. So if one parameter changes, there may be another result.

Secondly, only one type of shop configuration is considered in the thesis, all the orders in the model are processed on the same route. Undoubtedly, there are many different shop configurations in a practical company and also the route of the order varies. So in order to verify the worth of the Bumping Model for MTO company, further researches are needed.

Thirdly, because of the constraint of time, the thesis only considers one level of workforce flexibility with four levels of workforce efficiency. Thus, the worker can only catch the job from the worker in the previous station. There is a large research gap to think about a Bumping Model with a higher level of worker flexibility.

Fourthly, the configuration of the processing time of jobs is considered as a factor to affect the performance of Bumping Model, because the previous results showed the Baton Zone Balancing exploit assigning workers along the line according to the working efficiency. So a modified attempt is done in the thesis. But it is not systematic and rigorous research, so there should be more scientific research on this topic, especial in the Bumping Model with a high level of worker flexibility.

Finally, with the request of enough job in the first pool at the beginning of the day and the time constraint, the thesis only combine the periodical release (PR) with the Bumping workforce. However, there are many order release methods proposed in literature such as immediately release (IM), workload limited release and workload balancing release etc. They may affect the performance of the Bumping Model.

Considering the five limitations presented above, future researches on the Bumping Model including: (a), the Bumping models with higher levels of worker flexibility; (b), the Bumping models with different shop configurations; (c), the Bumping models with different configurations of processing time and (d), combining the other order release methods with the Bumping Model should be implemented.

REFERENCES

[1] Thürer, M., Stevenson, M., Land, M.J. On the integration of input and output control: Workload Control order release[J], International Journal of Production Economics, 2016, 174, 43-53

[2] Protzman, Charles,The Lean Practitioner's Field Book:Proven, Practical, Profitable and Powerful Techniques for Making Lean Really Work, 2016

[3] Wight, O.W. InPut/OutPut control, a real handle on leadtime[J]. Production and Inventory Management, 1970(11):9-30

[4] Haoyun Yan, Hongyu LI. Production planning and control technology for MTO company[J], Modern production and management technology, 2008，25（1）：12-18

[5] S.shimoyashiro, K.Tsoda, H.Avane. Input scheduling and load balance control for a job shop[J]. International Journal of Production Research, 1984, 22(4):597-605

[6] Bertrand, J.W.M., Wortmann, J.C..Production control and information systems for component-manufacturing shops. Dissertation, Elsevier, Amsterdam, The Netherlands, 1981.

[7] Bechte, W. Theory and practice of load-oriented manufacturing control[J]. International Journal of Production Research, 1988, 26(3): 375-395.

[8] Kingsman B.G., Tatsiopoulos, I.P., Hendry, L.C.. A structural methodology for managing manufacturing lead times in make-to-order companies[J]. European Journal of Operational Research, 1989, 40(2): 196-209.

[9] Thürer, M., Stevenson, M., Qu, T. Simple subcontracting rules for make-to-order shops with limited subcontractor capacity: An assessment by simulation[J], Production Planning and Control, 2015, 13(26): 1145-1161

[10] MELNYK, S. A. and RAGATZ, G. L., Order review/release and its impact on the shop floor[J]. Production and Inventory Management, 1988, 29, 13-17.

[11] Bergamaschi, D., Cigolini, R., Perona, M., Portioli, A.. Order review and release strategies in a job shop environment: A review and a classification[J]. International Journal of Production Research, 1997, 35(2): 399-420.

[12] Portioli-Staudacher, A., Tantardini, M., A lean-based ORR system for non-repetitive manufacturing[J], International Journal of Production Research, 2012, 12, 3257-3273

[13] Kingsman, B., 2000. Modeling input–output workload control for dynamic capacity planning in production planning systems[J]. International Journal of Production Economics, 68(1), p. 73–93.

[14] B. G. Kingsman & Linda Hendry. The relative contributions of input and output controls on the performance of a workload control system in Make-To-Order companies[J], 2002

[15] Kingsman, B., Hendry, L., Mercer, A. & De Souza, A., Responding to customer enquiries in make-to-order companies: problems and solutions[J]. International Journal of Production Economics, 1996, Volume 46-47, p. 219-31.

[16] Yuan Huang, Information architecture for effective Workload Control: an insight from a successful implementation[J], Production Planning & Control, 2017, 28:5, 351-366

[17] Hendry, L. & Kingsman, B., Production planning systems and their applicability to make to order companies[J]. European Journal of Operational Research, 1989, Volume 40, p. 1- 15.

[18] Hendry, L. & Wong, S., Alternative order release mechanisms: a comparison by simulation. International Journal of Production Research[J], 1994, 32(12), p. 2827–2842.

[19] Hendry, L., Land, M., Stevensona, M. & Gaalman, G., Investigating implementation issues for workload control (WLC): A comparative case study analysis[J]. International Journal of Production Economics, 2008, Volume 112, p. 452–469.

[20] Hendry, L., Huang, Y. & Stevenson, M., Workload control Successful implementation taking a contingency-based view of production planning and control[J]. International Journal of Operations & Production Management, 2013, 33(1), p. 69 - 103.

[21] M. R. Moreira, R. Alves, A new input-output control order release mechanism: how workload control improves manufacturing operations in a job shop[J], International Journal of Production Economics, 2006.

[22] Land, M., Parameters and sensitivity in workload control[J]. International Journal of Production Economics, 2006, 104(2), p. 625–638.

[23] Land, M. & Gaalman, G., Workload control concepts in job shops: a critical assessment[J], International Journal of Production Economics, 1996, p. 535–548.

[24] Land, M.J., Stevenson, M., Thürer, M., Gaalman, G.J.C., Job shop control:in search of the key to

delivery improvements[J], International Journal of Production Economics, 2015, 168,257–266.

[25] M. Thürer, M. Stevenson, M. J. Land, Fredendall, Workload control and order release[J], Production and Operations Management 0(0), pp. 1–15, © 2012 Production and Operations Management Society

[26] Thürer, M., Stevenson, M., Silva, C., Qu, T., Drum-buffer-rope and workload control in High-variety flow and job shops with bottlenecks: An assessment by simulation[J], International Journal of Production Economics, 2017, 116-127

[27] M. Thürer, M. Stevenson, M. J. Land, Fredendall, Melnyk, Lean control for make to order companies: integrating customer enquiry management and order release[J], Production and Operations Management Society , 2014,

[28] Thürer, M., Stevenson, M., Renna, P., Workload control in dual-resource constrained high-variety shops: an assessment by simulation[J], International Journal of Production Research, 2018, 1-17

[29] Thürer, M. et al., 2015. Concerning Workload Control and Order Release: The Pre-Shop Pool Sequencing Decision[J]. Production and Operations Management, 24(7), p. 1179- 1192.

[30] Hopp J. Wallance & Oyen P. Mark Agile, Workforce evaluation: a framework for cross-training and coordination, IIE Transactions, 2004, 36:10, 919-940, DOI:10.1080/07408170490487759

[31] Małachowski Bartłomiej & Korytkowski Przemysław, Competence-based performance model of multi-skilled workers[J], Computers & Industrial Engineering 91 (2016) 165–177, DOI: 10.1016/j.cie.2015.11.018

[32] Darwin J. Davis, Hemant V. Kher, Bret J, Wagner, Influence of workload imbalances on the need for worker flexibility[J], Computers & Industrial Engineering 57 (2009) 319–329, DOI: 10.1016/j.cie.2008.11.029

[33] Sammarco M., Fruggiero F., Neumann W.P. & Lambiase A., Agent based modelling of movement rules in DRC systems for volume flexibility: human factors and technical performance[J], International Journal of Production Research, 2014, 52:3, 633-650, DOI:10.1080/00207543.2013.807952

[34] Kundu, K., Staudacher, A.P., Investigation of effectiveness of order review and release models in make to order supply chain, MATEC Web of Conferences, 2016

[35] Hutter, T., Haeussler, S., Missbauer, H., Successful implementation of an order release mechanism based on workload control: a case study of a make-to-stock manufacturer[J], International Journal of Production Research, 2018, 4(56): 1565-1580

[36] Sagawa, J.K., Land, M.J., Representing workload control of manufacturing systems as a dynamic model[J], IFAC-PapersOnLine, 2018, 2(51):825-830

[37] Fernandes, N.O., Carmo-Silva, S., Workload control under continuous order release[J], International Journal of Production Economics, 2011, 1(131):257-262

[38] Sawhney, R., Implementing labor flexibility: A missing link between acquired labor flexibility and plant performance[J], Journal of Operations Management, 2013, 98-108

[39] Davis, D.J., Kher, H.V., Wagner, B.J., Influence of workload imbalances on the need for worker flexibility[J], Computers and Industrial Engineering, 2009, 1(57):319-329

[40] Francas, D., Löhndorf, N., Minner, S., Machine and labor flexibility in manufacturing networks[J],

International Journal of Production Economics, 1(131):165-174

[41] Perona, M. & Portioli, A., The impact of parameters setting in load oriented manufacturing control[J]. International Journal of Production Economics, 1998, Volume 55, p. 133–142.

ACKNOWLEDGEMENTS

Time flies like an arrow and life is a song, the two-year studying in Italy is coming to an end. I would like to thank Polytechnic di Milano for giving me a good oversea studying experience. With the help of this educational experience, I received a job offer from Huawei company luckily. Looking back on my two-year study experience, I am full of gratitude to those who have guided me, helped me, supported me and inspired me.

At first, I would like to express my gratitude to the staffs, professors, and students of Polytechnic di Milano, they provide a good environment for me to learn scientific knowledge and exposed me to the unique culture and passion of Italy.

Secondly, it is my pleasure to acknowledge my deepest thanks and gratitude to Professor Alberto Portioli Staudacher for suggesting the topic for this thesis. It is a great honor to work under his supervision.

Thirdly, I would like to express my heartfelt gratitude to Eng. Federica Costa, for her constant encouragement and guidance. She has walked me through all the stages of the writing of this thesis. Without her consistent and illuminating instruction, this thesis could not have reached its present form.

Fourthly, my thanks would go to my beloved family and relatives for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my fellow classmates who give me their help and time to listening to me and giving me suggestions.

Finally, thanks to my beloved girlfriend for her support for my decision and attic faith in our love, which gives me filled confidence in our future.