Applying Theory of Constraints to Moving Assembly Lines

Trumone Sims* and Hung-da Wan

Center for Advanced Manufacturing and Lean Systems and Mechanical Engineering Department
University of Texas at San Antonio
San Antonio, Texas 78249, U.S.A.

ABSTRACT

The first step in the Theory of Constraints (TOC) methodology is to identify the constraint. Several methods have been recommended in literature, such as looking for a backup of inventory (i.e., the operation that the inventory is waiting for is the constraint), or using linear programming or other analytical models. Yet, these methods may not be useful in a matured lean environment, which may have moving assembly lines where constraints are not obvious. This paper proposes two new methods for this purpose. The first method, Flow Constraint Analysis, takes a holistic view and evaluates whether the customer's demand is being satisfied. This evaluation is made by comparing the takt times and the cycle times of resources in the manufacturing system in order to identify the constraint(s). The second method, Effective Utilization Analysis, can be employed to pinpoint the location of the system constraint to a specific process or station. The actual production throughput is compared against the ideal capacity of the system to locate the bottleneck. This method is based on the relationship between WIP, bottleneck rate and lead time for a constant work in process (CONWIP) system. A case study of both methods applied to an actual production facility is presented.

1. Introduction

Theory of Constraints (TOC) and Lean Manufacturing are both proven methodologies to achieve excellence in manufacturing systems. The lean principles strive to improve the flow of a value stream by eliminating waste [1]. Meanwhile, "every value stream has a primary bottleneck (constraint) that limits its ability to reach its goal" [2]. In mass production environments, constraints are usually easy to find; just look for large stockpiles of WIP, backlogs, and frequent expediting [2]. But in a matured lean manufacturing environment, especially in a moving assembly line, none of these conditions should exist. Therefore a different approach has to be taken in order to identify the system constraint(s). In this paper, two new methods are presented, i.e., the Flow Constraint Analysis and the Effective Utilization Analysis.

1.1. The Use of Theory of Constraints

The TOC started out as scheduling software and has advanced into an important management philosophy [3]. The Goal provides an outline of the steps required to improve a system. The Five Focusing Steps allow the implementation of the theory of constraints concepts. The five focusing steps (5FS) are: 1. Identify the System Constraint, 2. Decide How to Exploit the Constraint, 3. Subordinate Everything Else, 4. Elevate the Constraint and 5. Go Back to Step 1, but beware of "Inertia".

The five focusing steps are constantly evolving. They have evolved into the Process Of OnGoing Improvement (POOGI), which includes the original five steps united with two prerequisites. The first prerequisite is to define the system under investigation and identify its purpose, while the second is to define measurements that align the system to its purpose [3].

The effectiveness of TOC has been reviewed extensively over time. For example, the extended literature survey by Naor et al. [4] provides a great insight into the theoretical foundation of TOC. Furthermore, the second evolution is taking place now. Pretorius [5] has identified several shortcomings with the five focusing steps. To address these shortcomings, the 5FS are transformed into a decision map that includes all five steps and the two prerequisites, but allows decision points to guide the user through the process. The answer to the first decision point, "Is the constraint physical?" is yes. Therefore to analyze the manufacturing system being studied, the first two steps of the five focusing steps do not change.

---

* Corresponding author: Tel.: (248) 252-3249; E-mail: trumone@myway.com

Constraints are often referred to as bottlenecks. By exploiting the constraint, we should wring every bit of capability out of the constraining component as it currently exists. In other words, TOC urges us to rethink what we can do to get the most out of this constraint without committing to potentially expensive changes or upgrades and implement in a short period of time [6]. Constraints can be both external and internal. External constraints are often beyond the control of management because they are market driven. External or market constraints affect demand, they influence product mix, which in turn affects resource utilization [7]. The product mix for the manufacturing system being studied in this paper is 60% of product A and 40% of product B. Internal constraints come in many forms, e.g., management philosophy, labor skills, inflexible work rules and limited capacity at various resources [7]. During this study only limited capacity will be considered.

Two major benefits are achieved by following the TOC methodology: (1) realizing the maximum system improvement from the least investment in resources and (2) we learn exactly how much effect improving a specific system component has on overall system performance [6].

1.2. TWO NEW METHODS

In this paper, two new methods are proposed to pinpoint constraints in mature lean systems, such as moving assembly line. The first method, named Flow Constraint Analysis, is a holistic approach that evaluates whether the customer's demand is being satisfied. This evaluation is made by comparing the takt times and the cycle times of resources in the manufacturing system. Cycle times will come from one of two sources, i.e., a moving assembly line or an automatic station. Resources with cycle times higher than a calculated target value are likely to be the constraints.

The second proposed method is the Effective Utilization Analysis. It can be employed to pinpoint the location of the system constraint to a specific process or station. The actual production throughput is compared against the ideal capacity of the system. This method is based on the relationship between WIP, bottleneck rate and lead time for a constant work in process (CONWIP) system. Resources with the low effective utilization are likely to be the constraints.

Both methods can be used to identify the constraint where resource time is constant or random. The two methods can be used independently or in conjunction depending on the situation. Both methods provide valuable insight into the performance of the enterprise under analysis.

1.3. CASE DESCRIPTION

A case study of both methods applied to an actual production facility is presented in the paper. The value stream under consideration consists of tandem assembly lines separated by a work in process (WIP) buffer. The buffer holds enough WIP to allow the assembly lines to temporarily run at different speeds without affecting (blocking or starving) one another.

This facility produces vehicles. There are two models, and they are transported down the assembly line on a carrier. There is a main assembly line which is fed by a sub-assembly line. The main line is made up of five moving assembly lines. The names of the lines are Frame 1, Frame 2/Final 1, Chassis 2/Final 2, Final 3/Chassis 3 and Inspection. The name of the sub-assembly lines are Trim 1, Trim 2 and Chassis 1. Due to page limit, the complete picture is not included in this paper. Figure 4 in Section 3 shows a segment of the assembly line.

2. METHOD #1: FLOW CONSTRAINT ANALYSIS

The Flow Constraint Analysis method involves a two-step process. The first step of the analysis is to determine if a true constraint is located in the manufacturing system. The second step is to identify secondary/tertiary constraints. In the first step, the existence of a constraint is determined by calculating and comparing the takt times and the cycle times. If the takt time is greater than the associated cycle time for each resource in the manufacturing system, the system is capable of meeting customer demand. The system constraint is then defined as being external. If demand exceeds the capacity of any of the resources of the manufacturing system a true bottleneck is said to exist [7], which means the constraint is internal. Another method of defining an internal constraint resource is through spare capacity. Spare capacity is the difference between cycle time and the takt time [2]. The resource with the least amount of spare capacity is the primary bottleneck for manufacturing systems with resources that have varying cycle times.

The manufacturing facility under analysis has an Andon system which collects the time and duration of events that occur. The Andon system also states a takt time for each assembly line in the manufacturing facility. This given takt time will be used during the analysis.

Another type of constraint resource also exists. These constraint resources have sufficient capacity when managed and scheduled carefully, but they could adversely affect the system's performance when managed inappropriately [7]. That is the purpose of the second step in the flow constraint method of analysis. The second step is to identify secondary/tertiary constraints by comparing individual times against each other.

When applying the flow constraint method, the process time for resources falls into one of four categories based on two characteristics. The first of the two characteristics is if the process time is dependent on the model mix, while the second is if the process time remains constant or varies from job to job.

2.1. MOVING ASSEMBLY LINES

2.1.1 EVALUATION

Statistical fluctuations apply to the performance of all resources [8]. One of the most prevalent sources of fluctuations is "natural" variability. Natural variability includes minor fluctuation in process time due to differences in operators, machines, and material [9]. Natural variability is ignored in the flow constraint method. Also, there are no effects to the assembly lines due to model complexity. The assembly lines are assumed to operate at a deterministic cycle. Usually, the cycle time can be obtained from the control panel. For this analysis the average cycle time was calculated for the resources and assumed constant. Continuous moving assembly lines process times can be categorized as non-model dependent and constant.

2.1.2 RESULTS ANALYSIS

The point chart below, Figure 1, shows the takt time and cycle time for each assembly line. The first three lines are the sub-assembly lines, while the remaining lines are part of the main line. The cycle time for the Frame 2/Final 1 assembly line is greater than the takt time for the line. Therefore the system is not able to meet customer demand and a true bottleneck exists, which is the Frame 2/Final 1 assembly line.

Protective capacity is the capacity required to correct for late starts in the assembly of products [8]. Chassis 2/Final 2 and Final 3/Chassis 3 assembly lines cycle times are the same which means the system has potential dual constraints. Since both assembly lines have the same cycle time there is not a sufficient amount of protective capacity.

The resources upstream of the inspection assembly line have higher average cycle times. The inspection assembly line therefore is not the designed constraint. This fact is being mentioned because this is a plant management philosophy.

The sub-assembly products are assembled to the main line products on the Frame 2/Final 1 assembly line. Comparing the cycle time for the sub-assembly lines with the Frame 2/Final 1 cycle time shows that two thirds of the sub-assembly lines have a higher cycle time. The sub-assembly lines could potentially starve the main assembly line of parts.

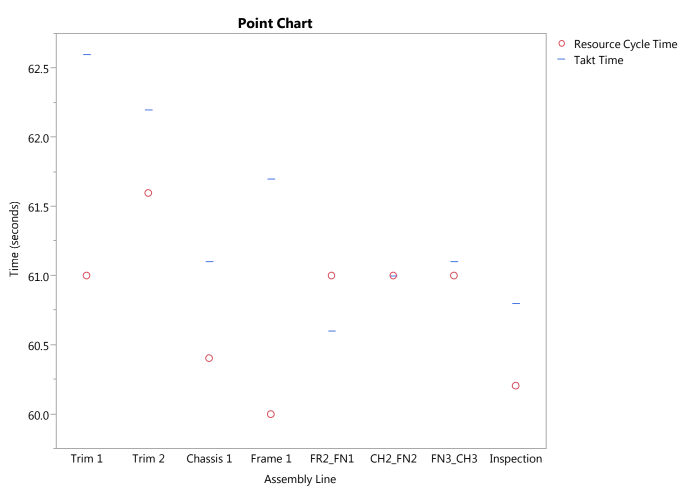

Figure 1. Takt Time vs. Cycle Time.

2.2. INDIVIDUAL STATIONS

2.2.1 EVALUATION

Additional resources are located between Frame 1 and Frame 2 assembly lines. These resources will be added to the discussion to demonstrate the use of the flow constraint method to automatic stations. There is a turnover, two transfers, two alignment/adjust work stations and a belt elevator.

To aid in understanding the function of the resources, the sequence of operations will be described. The product is removed from its skillet (pallet) and rotated 180 degrees. The first transfer removes the product from the turnover equipment and places it into one of two alignment/adjust stations. The alignment/adjust stations are dedicated resources, which means only one type of product is processed by each. Next, transfer #2 removes the product from one of the alignment stations and places it one the belt elevator. The elevator lowers the product onto its skillet where it travels through the rest of the manufacturing system. The definition of station cycle time presented by Hopp and Spearman [9] will be used to determine the cycle time for the automatic stations.

$$Cycle\ time = move\ time + queue\ time + setup\ time + process\ time + wait - to - batch\ time$$
$$+wait - in - batch\ time + wait - to - match\ time \quad (1)$$

The system follows single piece flow and first in first out rules. The alignment/adjust station that services Part B has an automatic changeover process which starts prior to the part's arrival, therefore the setup time will be assumed to zero. Queue time will not be considered in the analysis using the Flow Constraint Method. Move time will be called process time for the transfers since this is their only job function; equation (1) reduces to,

$$Cycle\ time = process\ time \quad (2)$$

The turnover and belt lift fall into the independent and constant category. The transfers process time (move time) fall into the dependent and constant category. While the alignment/adjust stations have independent and variable process times. As before with assembly lines, the first step of the analysis is to calculate and compare the takt times and cycle times. For stations with random cycle times the maximum cycle time values are compared against the takt time. If only one station has maximum cycle time values that exceeds the takt time, then that station is the constraint. If multiple stations have maximum cycle time values that exceeds the takt time, a different comparison method is required. A comparison of the probability of the cycle time exceeding the takt time should be performed.

2.2.2 RESULTS ANALYSIS

There is not a takt time given for the automatic stations from the Andon system. For demonstration purposes, the cycle time for the upstream assembly line, Frame 1, will be used as the reference time value the equipment should be operating below. This value, 60 seconds, has been selected to reduce the effects of blocking the Frame 1 assembly line.

The turnover and belt lift follow the same type of analysis as the moving conveyors. After the data collection process, the average cycle time for the turnover station was 52 seconds and the maximum cycle time value observed on the belt lift was 39.8 seconds. Therefore these stations were eliminated from consideration as a possible constraint.

The next resources, the transfers are analysed by considering the effects of model mix. The processing time for transfer 1 is the longest for Part A, with a value of 33 seconds. While processing time for transfer 2 is longest for Part B, with a value of 24 seconds. These stations are also eliminated as a possible constraint.

The cycle time data for the remaining two pieces of automatic equipment is random and therefore a histogram has been created for the sets of data. The Part 1 alignment/adjust operating range, see Figure 2 is well below the target value and is not a system constraint. The Part 2 alignment/adjust histogram shows that the cycle time data is bimodal, see Figure 3. Since we are concerned with constraint identification, the first mode is not considered during the analysis. The second mode, is also well below the target value. None of the automatic equipment is a system constraint.

If both of the histograms had contained the target cycle time value, a probability distribution would have been fitted to the data. Next, the probability of the resource running at a cycle time greater than the target value would have been calculated and the resource with the highest probability would be the constraint resource.

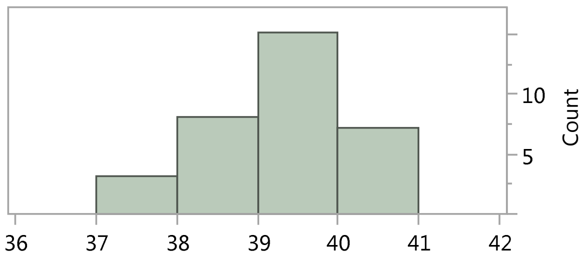

Figure 2. Part 1
Alignment/Adjust Cycle Time Histogram.

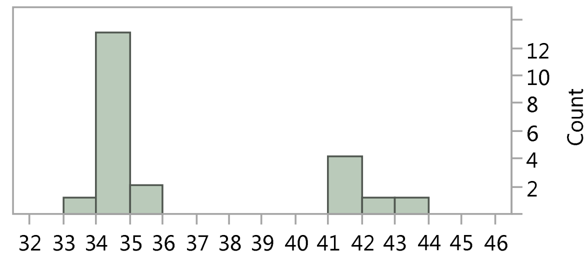

Figure 3. Part 2
Alignment/Adjust Cycle Time Histogram.

3. Method #2: Effective Utilization Analysis

Before expounding on the effective utilization method a brief definition of some of the method's terminology is required. There is a design reference line that is theoretically located through the center of the front axes of the vehicle. This line is known as the L10 line. The distance between the L10s is constant and is known as the assembly line pitch.

The work performed by an individual operator is known as a process. The process should begin and end within a designated area on the plant floor. This area is known as the process pitch. All the processes don't have the same work content. Because of the differences in work content, the process pitch is not the same for all the processes. The start of the process pitches could also be different. In some rare instances the process pitch is greater than the assembly line pitch.

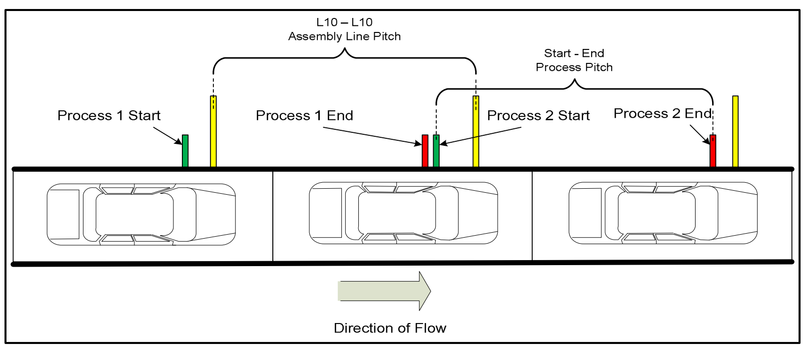

Figure 4. Shows the relationship between assembly line and process pitch.

3.1. Moving Assembly Lines

3.1.1 Evaluation

For assembly lines analyzed using the effective utilization method, model complexity and downtime effect the determination of the constraint resource. The effect of model complexity is captured in a variable named average cycle time. The processes performed by an operator are product model dependent. Since there is process time variability, the time weighted average is calculated for each process pitch and recorded as the average cycle time. Downtime durations are recorded in a variable named average downtime. First, the total downtime is calculated by summing up the downtime across all shifts over several days for each pitch. Next, the total downtime is divided by the number of products produced over the same time period. This calculation is performed for each process pitch. The average cycle time and average downtime values are summed. The summation is called process cycle time. This value will typically not be analyzed by itself. The reason is because of manpower allocation. There are typically two operators, one on each side of the product on the moving assembly line.

A comparison has to be made among all the operators working in a process pitch to ensure that the maximum time seen by the product in the process pitch is recorded. Logic is used in the next step of the method to determine the correct value to be recorded. The name of the recorded value is *pitch cycle time*.

Pitch cycle time can be assigned one of three values. The first logical decision is to determine if there are any processes in the pitch. If there are processes in the pitch, then the maximum *process cycle time* is compared against the inverse of the bottleneck rate for the assembly line. If the *process cycle time* is greater than the inverse of the bottleneck rate, then the maximum *process cycle time* is recorded as the *pitch cycle time.*

The next scenario occurs is if there are no processes in the pitch. In this case, the summation of the average downtime(s) for the pitch is compared against the inverse of the bottleneck rate. If the summation of the average downtime(s) is less than 1% of the inverse of the bottleneck rate, then a value of 0 is recorded as the *pitch cycle time.* The 1% value is based off of historical system performance data for the manufacturing system being studied. In theory is there are no processes taking place in a pitch, then that those pitch should not produce any downtime. Even though the product still has to physically travel through the pitch, a value of 0 is recorded because the only other acceptable value would be the inverse of the bottleneck rate. Recording the inverse of the bottleneck for pitches tends to improve the utilization value for the assembly line, therefore 0 is recorded instead.

Let us now discuss the third and final value that the *pitch cycle time* can assume. For the third possible value to be valid, two events can occur and produce the same result. The first event is; yes, there are processes in the pitch but no, the maximum *process cycle time* is not greater than the inverse of the bottleneck rate. The second event is; no, there are no processes in the pitch and no, the summation of the average downtime(s) is not less than 1% of the inverse of the bottleneck rate. If either of these events occurs, the *pitch cycle time* is calculated by summing the inverse of the bottleneck rate and the maximum average downtime value for the pitch.

3.1.2 RESULTS ANALYSIS

The analysis begins with the calculation of the practical lead time for the assembly line. This is accomplished by summing the *pitch cycle time* values. The assembly lines maintain a constant work in process (WIP). The actual WIP value has to be reduced to account for pitches with no processes and very little downtime. Therefore, only non-zero *pitch cycle time* values are counted to determine the WIP value for the assembly line. With the practical lead time and WIP calculated for the assembly line, the practical production rate can be determined from equation 3.

$$r_b^P = \frac{W}{T_0^P} \tag{3}$$

The utilization for the assembly line is determined by evaluating the ratio of the practical production rate with respect to the assembly line bottleneck rate, see equation 4. The bottleneck rate will be larger than the production rate because the bottleneck rate doesn't assume any losses due model complexity or downtime. This property can be used to verify calculations while performing the analysis.

$$U = \frac{r_b^P}{r_b} \tag{4}$$

The flow chart shown in figure 5 will aid in following the steps required to perform the various calculations. Unlike traditional thinking where high utilization rates are associated with constraint resources, the opposite is true using this methodology. High utilization rate are desired, low utilization rates represent assembly lines that are constantly stopping. Among the sub-assembly lines Trim 1 has the lowest value, while Frame 1 has the lowest value for the main assembly lines, see table 1.

Table 1. Assembly Line Utilization Rates.

Assembly Line	Trim 1	Trim 2	Chassis 1	Frame 1	Frame 2 / Final 1	Chassis 2 / Final 2	Final 3 / Chassis 3	Inspection
Utilization	93%	97.5%	97.6%	96.2%	96.7%	97.8%	98.1%	99%

3.2. INDIVIDUAL STATIONS

3.2.1 EVALUATION

Now the utilization method will be used to determine the constraint for individual automatic stations. The method of calculating a utilization value for independent stations presented here is borrowed from Hopp and Spearman [9]. They did an excellent job of explaining the concept and developing the applicable equations, see equation 5. This section of the paper will apply those concepts and equations to the manufacturing system under study.

Instead of analyzing all the automatic stations, only the alignment/adjust stations will be reviewed. These stations were selected because of their more frequent stoppages.

$$Utilization = \frac{Arrival\ rate}{Effective\ production\ rate} \quad (5)$$

Model complexity and downtime effects will be considered in this analysis method. The effective production rate, in the denominator, will capture downtime effects.

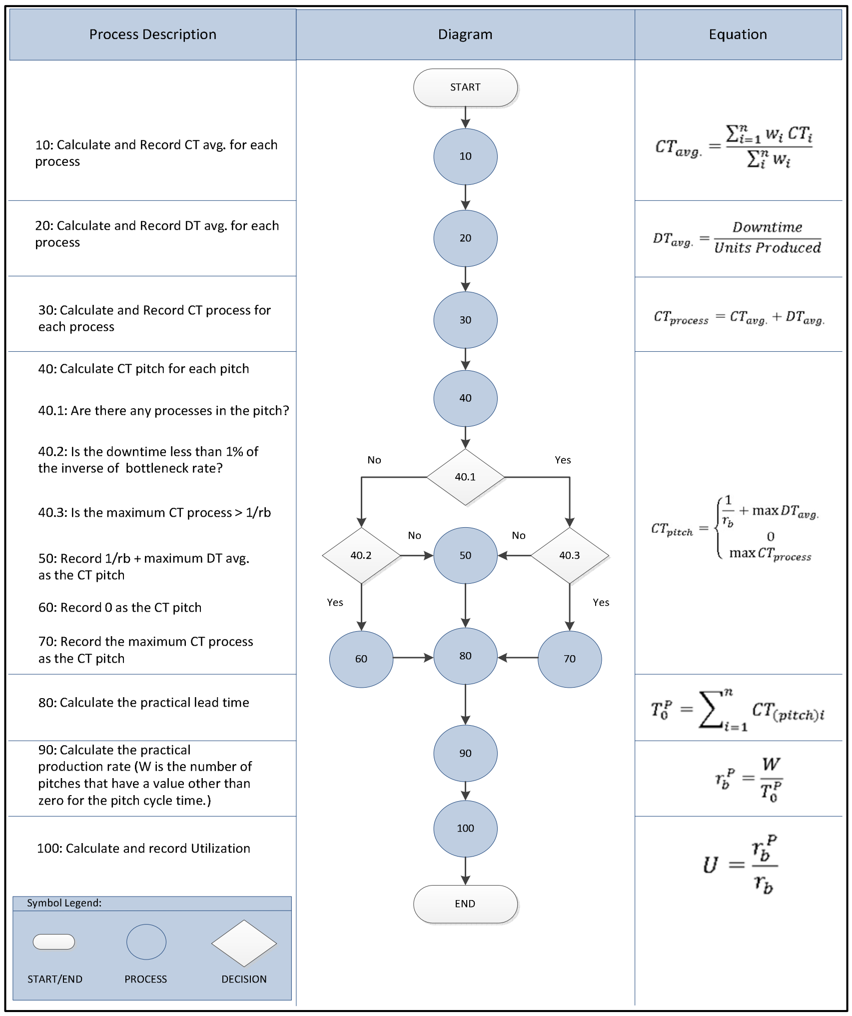

Figure 5. Flow Chart of Utilization Method (drawn by: Trumone Sims).

3.2.2 RESULTS ANALYSIS

When calculating the utilization for individual stations using this method, a lower utilization is preferred. In the case of completely reliable stations connected as a serial production line, the station with the largest cycle time (the most utilized station) is the constraint station. The station dedicated to product A had the higher utilization.

Table 2. Alignment/Adjust Stations Utilization.

Station	Utilization
Product A	55%
Product B	45%

4. Summary and Conclusion

In this paper, two methods have been proposed to locate constraints in matured lean systems, especially with moving assembly lines, which usually do not have obvious bottlenecks. A case study on a real production system shows that two methods can be used for analysis at both system level and component level. The uses of the two methods are slightly different as discussed below.

The Flow Constraint Method provides the users with the ability to determine the location of the enterprise constraint. If the there are multiple system constraints they can be quickly identified. The flow constraint method can also be used to evaluate if the design intention of the manufacturing system is being met. The Flow Constraint Method is more suitable for early planning phases, such as before the manufacturing system has been installed or when there are plans to increase the capacity of an existing facility.

The Effective Utilization Method requires more computations and data. This means the user of this method will have to spend more time and effort implementing this method. The users of this method will have the ability to narrow the focus of changes to an actual process. The users will also have deeper understandings of the manufacturing system and the effects model complexity and downtime have on the system. The Effective Utilization Method is more applicable for use as part of continuous improvement program.

The presented case study covers only a portion of a production system. For future study, the method can be applied to the other areas of the system as well as other cases. This research can be further extended to include other aspects of the TOC methodology. After the constraint is identified, appropriate decisions can be made on exploiting the constraint to further improve the system. Insights into the system constraints can also facilitate redesigning a segment of or the whole system to be compliant with the concepts of TOC.

References

[1] J. Allen, C. Robinson, and D. Stewart, Lean Manufacturing - A Plant Floor Guide Society of Manufacturing Engineers (SME), 2001.

[2] Steve Bell, Lean Enterprise Systems: Using It for Continuous Improvement. ed. by Andrew P. Sage, Wiley Series in Systems Engineering and Management, Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

[3] K. J. Watson, J. H. Blackstone, and S. C. Gardiner: "The Evolution of a Management Philosophy: The Theory of Constraints", Journal of Operations Management, Vol. 25, No.2, 387–402, 2007.

[4] M. Naor, E. S. Bernardes, and A. Coman, "Theory of Constraints: Is It a Theory and a Good One?" International Journal of Production Research, Vol.51, No.2, 542–554, 2013.

[5] P. Pretorius: "Introducing in-between Decision Points to TOC's Five Focusing Steps", International Journal of Production Research, Vol. 52 , No. 2, 496–506, 2014.

[6] H. W. Dettmer, Goldratt's Theory of Constraints - A System Approach to Continuous Improvement, ASQ Quality Press, 1997.

[7] S. E. Fawcett, and J. N. Pearson: "Understanding and Applying Constraint Management in Today's Manufacturing Environments", Production and Inventory Management Journal, Vol. 32, No. 3, pp. 46–55, 1991.

[8] R. E. Stein and M. Dekker, 'Identifying the Constraint', in Re-Engineering the Manufacturing System, CRC Press, 2003.

[9] W. Hopp and M. Spearman, Factory Physics, 3rd ed., Waveland Press, 2011.