Modelling Congestion for Aggregate Production Planning in Open Queuing Networks

Dang Phi Van Hai
PhD Candidate, Department of Industrial Engineering
Dalhousie University
Halifax, Nova Scotia, Canada
hai.dang@dal.ca

Uday Ventakadri
Professor, Department of Industrial Engineering
Dalhousie University
Halifax, Nova Scotia, Canada
uday.venkatadri@dal.ca

Tri Nguyen-Quang
Associate Professor, Department of Engineering, Faculty of Agriculture
Dalhousie University
Truro, Nova Scotia, Canada
tri.nguyen-quang@dal.ca

Claver Diallo
Professor, Department of Industrial Engineering
Dalhousie University
Halifax, Nova Scotia, Canada
claver.diallo@dal.ca

Young Ki Chang
Assistant Professor, Department of Engineering, Faculty of Agriculture
Dalhousie University
Truro, Nova Scotia, Canada
YoungChang@dal.ca

Abstract

The challenge in aggregate production planning for high-tech manufacturing industries such as aerospace, semiconductor manufacturing, or high precision components production is the variability in cycle times or cycle steps due to the rework required to meet very high levels (6-sigma) of quality. This variability at the lower planning level needs to be accounted for in aggregate planning level. Planning circularity, whereby cycle time depends on resource utilization while resource utilization is determined by cycle time continues to be an important problem in the aggregate planning literature. It is well known that ignoring congestion, as is the case in MRP-II based systems still widely in use, is inaccurate. In the presence of congestion, the relationship of WIP and throughput is nonlinear and bottleneck resources may shift constantly. The most common approach of addressing the nonlinear relationship between the WIP and the throughput is through the clearing function. Recent work by Omar et al. (2017) proposed a mixed-integer linear model for closed queuing production networks using fixed release planning. The challenge with this model is that it is difficult to scale up for typically sized problems scalability. This work extends the approach in Omar et al.

(2017) by proposing a non-linear model based for the open queuing production network which shows some promising results.

Keywords
Aggregate, Production Planning, Congestion, Clearing Function and Nonlinear Programming.

1. Introduction

Putting more effort on cooperation and coordination among supply chain partners is necessary for the companies to hold their competitiveness in this globalization era. Improving the effectiveness of Production Planning and Control (PPC) is a part of that effort. Effective PPC is challenging because it requires synchronization among different stakeholders inside and outside a company. The PPC process is divided into two levels, aggregate planning at the top level and detailed production planning at the shop-floor level. The main objective of PPC is not only to allocate resources efficiently and responsively but also to maximize collaboration among resources at the lower level. It is difficult in aggregate planning to estimate accurately the variability at the lower level because of the circularity, when the cycle time at the detail level depends on the resource utilization level, but the resource utilization is determined by the workload which defined by the release decision from aggregate planning. In many well-known approaches such as the MRP, it is assumed that lead time is a constant and load independent. However, this assumption leads to the overestimation of resources, and could create an infeasible aggregate production plan. As stated by Asmundsson et al. (2006), many production systems must consider lead time in their planning to match supply with demand. However, in queuing models, lead time in turn depends on resource utilization which determined by the assignment of works to resources by the planning models. Graves (1986) introduced a linear clearing function assuming that the average cycle time is kept in constant while in the MRP, a fixed lead time representing a time-shift of the input is used (Orcun et al., 2006).

The clearing function represents the non-linear relationship between WIP level and resource throughput; it has been used to model the effect of congestion at the aggregate planning phase. It helps the production planner improve the feasibility of a planning solution since it integrates the congestion effect. The CF could be estimated by using analytical or simulation approaches. In many analytical methods, the CF could be estimated at just bottleneck resource, but this approach could be challenging since bottleneck can shift as product mix changes. Moreover, the analytical approach is considered suitable solution for solving small problem instances. On the other hand, although the simulation approach can overcome the issues of analytical methods, it is time-consuming.

There have been many methods to linearize the CF as well as to incorporate into optimization models for aggregate production planning. The problem is discussed in multi-product, multi-period planning context. The approach of Missbauer (2002) improves the performance of order release problem by considering the behavior of a queuing network. This study derives the CF for a bottleneck machine while the other non-bottleneck machines are assumed to have fixed lead time. The relationship of WIP and throughput is then integrated into a linear programming model. Since the CF is analytically built from the bottleneck machine, the CFs are different for different bottle neck machines when the product mix changes. The effect of product mix in lead time and production capacity is introduced using a partitioned CF for linear programming (Asmundson et al., 2009). In this approach, the CF is formed empirically using a simulation model and the correlation between the utilization and lead time is presented as a concave curve. Recently, Omar et al. (2017) introduced a data-driven method to obtain CF for closed queuing network in which the throughput is calculated based on the WIP of an entire network. This approach can overcome the problem of shifting bottleneck stations as the product mix cases because the system throughput is estimated for discrete combinations of product mix generated beforehand. However, the CF is demonstrated for closed systems using the Mean-Value Analysis (MVA) method assuming CONWIP systems. Moreover, by integrating the CF into a MILP using binary integer variables, computational time and memory consumption can be challenging as the network size increases.

The contribution made by this paper is to demonstrate how the congestion effect can be incorporated for aggregate planning in the open queuing network case. This paper is organized in four more sections as follows: We present the literature relating to this problem in Section 2, the reviewed work summarizes the main theoretical background for this research. Section 3 explains the methodology used. In Section 4, a case study is presented to demonstrate the viability of the concept. Section 5 concludes the paper and offers a few directions for future research.

2. Literature Review

The first mathematical representation of manufacturing resource in literature to describe the relationship of enterprise's resources with customer demand anticipation was of Modigliani and Hohn (1955). In 1957, Manne solved this model by using LP formula for transportation model. In this formula, lead time is the factor to connect top-level planning with shop floor planning. Therefore, in study of Hackman and Leachman (1989), they also proposed a production planning model in which based on linear programming and stated that lead-time is an exogenous parameter to the model.

The clearing function concept is understood as the planning process to coordinate material flow in whole supply chain in such an effective way is a complicate work. In the simple context of a single product single machine M/M/1 queue, the CF is formulated easily. However, the problem becomes more complicated when both the number of products and machines increase. The first CF approach is the study of Graves (1986), where it is stated that the speed of production system is manageable by increasing or decreasing the production process to keep the average lead time constant. In the linear CF of Graves, is considered as managerial policy, the lead time is assumed as constant value which means that it does not depend on the WIP level at period t. This leads to the production system is implied to have unlimited resources. However, in the most common popular tool MRP, the fixed lead-time is implied as a time-shift of the input. By using a concave nonlinear non-decreasing function, Karmarkar (1989) proposed another method to define CF by analyzing the congestion effect at shopfloor for a single station a generalized formula. Those first approaches of CF have shown the importance and effectiveness of this function in aggregate planning. This leads to many consecutive approaches in literature later. CF could be derived for closed production system with high processing time and failure rate (Spearman, 1991). In other research, a set of CFs is obtained to find relationship of WIP level with the practical worse-case analysis of throughput (Spearman & Hopp, 1996). Missbauer (2002) introduced an order release model which trying to optimize the balance between order and capacity. In his study, it was assumed that the bottleneck station is represented by the M/G/1 steady-state system and then the CF for this station is presented as the relationship between the expected workload and the expected output. Although the CF is derived by steady state and Little's Law, there are other approaches (Selçuk et al. , 2008; Missbauer, 2009;). In Kacar et al. (2011), the CF is derived from the steady state of a queuing network from the transient state of production resources or from estimation of production system using simulation. When the queuing system is in steady state, the CF is measured by relationship between WIP and throughput by Little's Law. In the study of Kacar et al. (2016), a comparison is made among integer and non-integer lead-time estimation model of Hackman and Leachman (1989) with the Aggregate Clearing Function (ACF) model of Asmundsson et al. (2009). In their conclusion, they confirmed that non-integer lead-time provides a better estimate of throughput compared to an integer lead-time and can provide results comparable to the ACF. However, non-integer lead-time is more favorable for a system with low variability to reduce complexity of achieving and fitting the clear function (CF).

In the context of multi-product systems, Asmundsson et al. (2009) introduced the ACF to distribute a partial resource capacity to each type of product according to customer demand. In this approach, the authors developed a production planning model for multi-product, single-stage manufacturing process that can capture the nonlinear behavior between workload and lead times. By using outer linearization of the non-linear model, they obtained LP formulation for multi-stage system. CFs are derived for each station and included it in LP independently although there are correlation flows between stations in a queuing network. This also happens in other studies when simulation is used for open queuing network to derive the piecewise linear CF at every station in ACF model (Kacar, 2009; Kacar et al., 2011; Kacar & Uzsoy, 2014)

Orcun et al. (2009) tried to include the safety stock into the ACF to optimal the planning process while considering the stochastic demand for a single product production plant. This can be done by empirically deriving the CF using simulation model. From known values of WIP, the CF is formed by estimating simulated throughput and then release schedules for two different products are established for each period. Finally, the derived CF is integrated to the ACF production planning model.

By comparing the production planning model of Karmarkar (1989) with classical linear programing model, Kefeli et al. (2011) concluded that the linear programming model using dual price underestimates the capacity constraint because it does not express the congestion effect adequately. The CF model is more realistic than the classical model since it holds finished good inventory at utilization levels of less than one. However, the dual price approach is hard to apply for more complex production-inventory system with multiple stages, multiple products, and multiple interaction between WIP and demand. To integrate the production planning and engineering process improvement for single-source re-entrance system in Kim and Uzsoy (2013), the CF concept was employed to assign resource capacity in many production stages and improvement stages. They applied Karush-Kuhn-Tucker (KKT) optimality conditions to analyze marginal cost to focus inside the effectiveness of resource management.

Another new way to model the CF was also introduced by Kacar and Uzsoy (2014) using linear regression. In their approach, after doing comparison, they concluded that the load-based CF performs better than product-based CF in most cases. In the later publish, Kacar and Uzsoy (2015) improved the way to fit CF to empirical data by using simulation to optimize decision variables related to CF parameters.

Albey et al. (2014) built a multiple dimension CF model using simulation to show the impact of the product mix change on aggregate planning. The total workload is disaggregated into many individual products and creating set of different combinations of product ratio. These combinations are used to understand the relationship between resource utilization and lead time. The ACF model of Asmundsson et al. (2009) is used by Kefeli and Ozsoy (2016) to extend their previous work (Kefeli et al., 2011) to define the bottleneck station for multi-product, multi-machine system presented in their case study. In the publications by Asmundsson et al. (2009) and Albey et al. (2014), the importance of integrating the product mix in top level planning models was emphasized. In both approaches, the CF was derived by running simulation of production facility, this leads to two disadvantages. The first drawback is that to simulate a real production system in detail requires much effort and information about the system performance which is difficult to obtain. Moreover, computing simulation is always time-consuming task.

3. Methodology

To directly integrate the clearing function into the aggregate planning based on the congestion concept for the open queuing network, the first step is to formulate the relationship between the total WIP level and the throughput of the system, and the second step is to integrate the formula of WIP-throughput into the aggregate planning model.

In the first step, the multi-product, multi entry production system in which raw material arrival follows the Poisson process is considered as an open queuing network. Therefore, it has all the normal properties of a Jackson open queuing network such as many workstations, many products, the part can visit a workstation multiple times, and the queue in front of any workstation has infinite waiting space. By using the formula of queuing phenomenon in an open queuing network, the relationship between the total WIP of the whole system and the throughput is captured.

The formula for calculating the total WIP of the system is used as a constraint for the aggregate planning model in the second step. The objective function is to minimize the total cost in which including raw material release, WIP release, inventory, and backorder cost. The model has five constraints. The first and the second constraints are the restriction related to flow conservation for WIP and finished good inventory. In the third constraint, the throughput of system of product p at period t must be less than or equal to the number of external arrivals to the system. The fourth constraint ensures that the resource utility at each resource m is always less than or equal to 1 to avoid the overload for the system. The fifth constraint is the calculation of exact value of average level of WIP of the whole system.

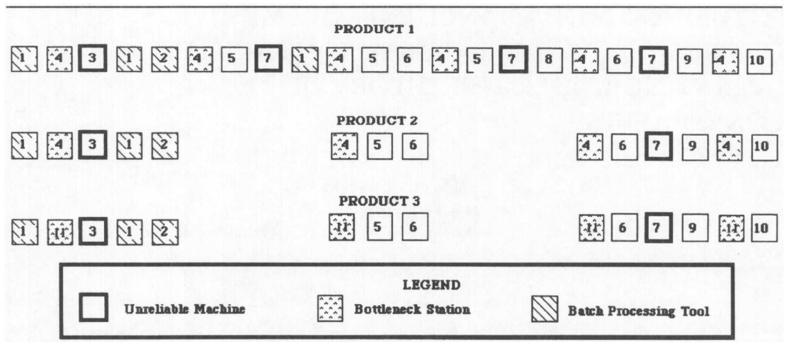
The aggregate planning model is the nonlinear programming (NLP) model since it has the nonlinear constraint representing the relationship between the WIP and the throughput. This NLP model is solved by the Knitro solver 9.0 in MPL 5.0. The result from the NLP model is used to compare with the discrete simulation model of the queuing network which is run in Arena Simulation to make the comparison.

4. Case study and Results

The case study from the wafer fabrication of Kayton et al. (1997) is used to illustrate the viability of the proposed model. However, there are some adjustments in this example to make it more suitable with this model. Firstly, the machine reliability and batching in workstations are ignored. Secondly, the machine processing time at workstation is assumed to be exponential distribution. Thirdly, since this model is applied for open queuing network, there is external arrival to the first workstation under the Poisson process.

A simplified facility is represented as in table 1 to show the basic routing for three different products. There are 11 different working stations that the three products will go through. For example, the product 2 first goes to Station 1 then to Station 4, then to Station 3 and so on. In the original paper, Station 4 and 11 are considered as bottleneck points. Station 3, 7 are unreliable and Station 1 could process in batch. However, since this example is used to illustrate the analyze of output measures for queuing network, machine reliability could be removed for simplification.

Table 1. Representation of production flow (Kayton et al., 1997)



Each product might return to a station many times and has different routines, each station serves products with the same processing time when the products visit. The batching station 1 in original paper operating three batches at the same time could be simplified by dividing the service time at this station to 3. Table 2 and Table 3 show average processing time for eleven stations and the number of visits to each station of three products.

Table 2. Processing time at stations

	Machine time										
Station	1	2	3	4	5	6	7	8	9	10	11
Processing time (min)	80	220	45	40	25	22	20	100	50	50	70

Table 3. Number of visits to stations

	Product	Station											Total
		1	2	3	4	5	6	7	8	9	10	11	
Number of visits	1	3	1	1	6	3	2	3	1	1	1	0	22
	2	2	1	1	4	1	2	1	0	1	1	0	14
	3	2	1	1	0	1	2	1	0	1	1	4	14

The proposed aggregate model is applied with the parameters are:

- Each planning period is a week, the production time is seven days per week, three shifts per day, eight hours per shift. So, the total production time is 168 hours.
- Raw material cost, good-finished cost, WIP cost, and backorder cost are 6, 3, 1, and 15 correspondingly.
- Demand for the three products over 20 periods is shown in Table 4:

Table 4. Demand for the three products over 20 periods

Period 1	Product 1	Product 2	Product 3
1	817	520	535
2	1497	879	1352
3	1963	1179	1658
4	869	609	754
5	1423	852	1331
6	1812	1032	1382

7	1049	672	1080
8	865	594	696
9	1289	740	1146
10	1848	1174	1545
11	803	284	450
12	365	274	384
13	62	76	209
14	941	622	1001
15	1985	1255	1697
16	805	512	533
17	1867	1178	1648
18	1346	758	1304
19	259	182	295
20	846	573	683

Results from NLP model include the:

- Arrival rates of the three products for 20 periods
- Quantities of raw material input, throughput, WIP, backorder and finished-goods inventory for 20 periods.

To estimate the accuracy of the WIP and throughput derived from the NLP model, the arrival rates of three products are put into a simulation model to check for the WIP and the throughput. Those WIP and throughput calculating from NLP are compared to the corresponding WIP and throughput getting from the simulation.

The Simulation model of the case study in Kayton et al. (1997) is created using Arena Simulation. The model is graphically designed as shown in Figure 1. The open queuing network in this simulation model also contents eleven working stations and produces three different products as the same to the model in Kayton et al (1997). There are twenty scenarios representing for twenty different periods, the combination of arrival rate at every period for three products is input to the simulation model in each scenario.

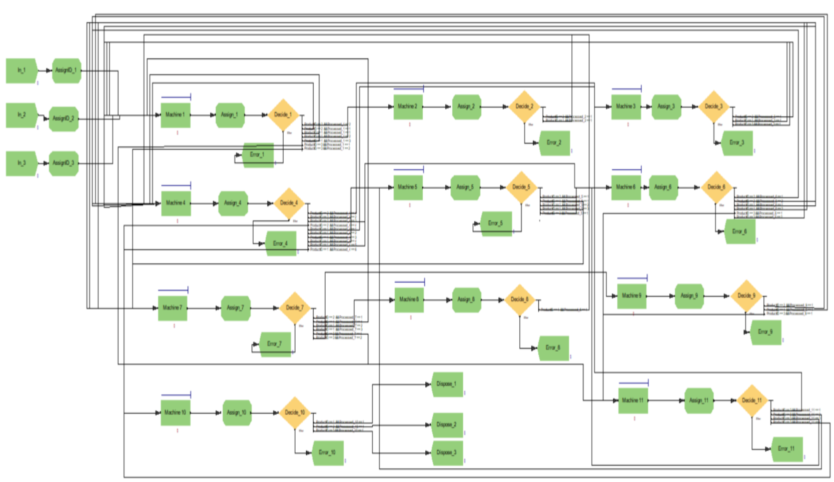


Figure 1. Simulation model of the case study in Kayton et al., (1997)

By running the Knitro solver in MPL, the results of NLP model for three products are the arrival rates, WIP, and throughput over 20 periods. Table 5 shows the arrival rates of three products:

Table 5. Arrival rate of three products over 20 periods

Period	Arrival rate of product 1	Arrival rate of product 2	Arrival rate of product 3
1	0.3782	0.2407	0.2477
2	0.6931	0.4069	0.6259
3	0.9088	0.5458	0.7676
4	0.4023	0.2819	0.3491
5	0.6588	0.3944	0.6162
6	0.8389	0.4778	0.6398
7	0.4856	0.3111	0.5000
8	0.4005	0.2750	0.3222
9	0.5968	0.3426	0.5306
10	0.8556	0.5435	0.7153
11	0.3718	0.1315	0.2083
12	0.1690	0.1269	0.1778
13	0.0287	0.0352	0.0968
14	0.4356	0.2880	0.4634
15	0.9190	0.5810	0.7856
16	0.3727	0.2370	0.2468
17	0.8644	0.5454	0.7630
18	0.6231	0.3509	0.6037
19	0.1199	0.0843	0.1366
20	0.3917	0.2653	0.3162

Those 20 sets of arrival rates of three products are used to generate 20 scenarios for the Simulation model as shown in Table 6

Table 6. WIP and Throughput from Simulation model.

	Scenario Properties		Controls			Responses			Responses		
	S	Name	in_1_time	in_2_time	in_3_time	Entity 1.Num	Entity 2.Num	Entity 3.Num	Entity 1.WIP	Entity 2.WIP	Entity 3.WIP
1		Scenario 1	2.6441	4.1545	4.0371	788.000	519.000	475.000	0.6349	0.2762	0.3004
2		Scenario 10	1.4428	2.4576	1.5977	1410.000	811.000	1311.000	1.3655	0.5426	0.9351
3		Scenario 11	1.1004	1.8322	1.3028	1908.000	1128.000	1588.000	2.0139	0.8407	1.2844
4		Scenario 12	2.4857	3.5474	2.8645	863.000	599.000	802.000	0.6907	0.3317	0.4418
5		Scenario 13	1.5179	2.5355	1.6228	1417.000	821.000	1326.000	1.2746	0.5230	0.9072
6		Scenario 14	1.1920	2.0929	1.5630	1737.000	1033.000	1348.000	1.7698	0.6902	0.9856
7		Scenario 15	2.0593	3.2144	2.0000	1027.000	682.000	1046.000	0.8669	0.3777	0.6831
8		Scenario 16	2.4969	3.6364	3.1037	844.000	558.000	660.000	0.6843	0.3221	0.4053
9		Scenario 17	1.6756	2.9189	1.8847	1216.000	745.000	1130.000	1.1075	0.4363	0.7405
10		Scenario 18	1.1688	1.8399	1.3980	1835.000	1130.000	1522.000	1.8518	0.8105	1.1358
11		Scenario 19	2.6896	7.6046	4.8008	796.000	263.000	411.000	0.6077	0.1469	0.2442
12		Scenario 2	5.9172	7.8802	5.6243	368.000	256.000	385.000	0.2604	0.1338	0.2030
13		Scenario 20	34.8432	28.4091	10.3306	70.000	71.000	198.000	0.0423	0.0357	0.1076
14		Scenario 3	2.2957	3.4722	2.1580	940.000	624.000	923.000	0.7558	0.3396	0.6114
15		Scenario 4	1.0881	1.7212	1.2729	1908.000	1252.000	1683.000	2.0983	0.9124	1.3311
16		Scenario 5	2.6831	4.2194	4.0519	775.000	499.000	541.000	0.6195	0.2721	0.3007
17		Scenario 6	1.1569	1.8335	1.3106	1826.000	1146.000	1521.000	1.9033	0.8206	1.2575
18		Scenario 7	1.6049	2.8498	1.6565	1285.000	734.000	1244.000	1.1791	0.4564	0.8707
19		Scenario 8	8.3403	11.8624	7.3206	264.000	193.000	274.000	0.1809	0.0857	0.1551
20		Scenario 9	2.5530	3.7693	3.1626	817.000	563.000	652.000	0.6641	0.3060	0.3895

The WIP and Throughput from NLP model and Simulation model are put into Table 7 for comparison.

Table 7. WIP and Throughput from NLP and Simulation over 20 periods

WIP-NLP			WIP-Sim			Throughput-NLP			Throughput-Sim		
P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
0.6326	0.2742	0.3	0.6349	0.2762	0.3004	817	520	535	788	519	475
1.3609	0.5468	0.9276	1.3655	0.5426	0.9351	1497	879	1352	1410	811	1311
2.0392	0.8409	1.2825	2.0139	0.8407	1.2844	1963	1179	1658	1908	1128	1588
0.6877	0.3286	0.4408	0.6907	0.3317	0.4418	869	609	754	863	599	802
1.274	0.5218	0.9037	1.2746	0.523	0.9072	1423	852	1331	1417	821	1326
1.7678	0.6903	0.9831	1.7698	0.6902	0.9856	1812	1032	1382	1737	1033	1348

0.8654	0.3784	0.6774	0.8669	0.3777	0.6831	1049	672	1080	1027	682	1046
0.6819	0.3192	0.4027	0.6843	0.3221	0.4053	865	594	696	844	558	660
1.1093	0.4351	0.7399	1.1075	0.4363	0.7405	1289	740	1146	1216	745	1130
1.8642	0.813	1.1519	1.8518	0.8105	1.1358	1848	1174	1545	1835	1130	1522
0.6071	0.1459	0.2469	0.6077	0.1469	0.2442	803	284	450	796	263	411
0.262	0.1335	0.2045	0.2604	0.1338	0.203	365	274	384	368	256	385
0.0423	0.0351	0.1065	0.0423	0.0357	0.1076	62	76	209	70	71	198
0.7595	0.3424	0.614	0.7558	0.3396	0.6114	941	622	1001	940	624	923
2.0976	0.9112	1.3341	2.0983	0.9124	1.3311	1985	1255	1697	1908	1252	1683
0.6219	0.2694	0.2985	0.6195	0.2721	0.3007	805	512	533	775	499	541
1.903	0.8243	1.2612	1.9033	0.8206	1.2575	1867	1178	1648	1826	1146	1521
1.1784	0.4535	0.8727	1.1791	0.4564	0.8707	1346	758	1304	1285	734	1244
0.1821	0.0868	0.1539	0.1809	0.0857	0.1551	259	182	295	264	193	274
0.6638	0.3064	0.3936	0.6641	0.306	0.3895	846	573	683	817	563	652

To show whether the CFs getting from NLP model and Simulation model follow the form of theoretical clearing function, the data from the Table 8 are graphically presented as clearing function of each product Figure 2.

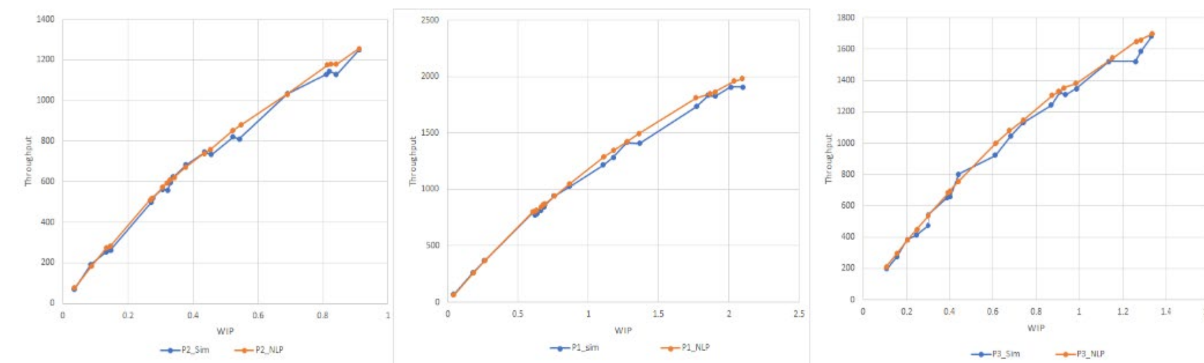


Figure 2. Clearing function of product 1, 2 and 3 from NLP and Simulation

To graphically present the clearing function across all three products, the processing time of each product is used as weight of each product. Based on the processing time at each machine and the number of visits of each product through those machines, the total processing time to complete a product is calculated in Table 8. As a result, the weight of each product is also calculated corresponding.

Table 8. Processing time and weight of each product

Machine	1	2	3	4	5	6	7	8	9	10	11	Total processing time	Weight
Processing time (mins)	80	220	45	40	25	22	20	100	50	50	70		
Number of visits of P1	3	1	1	6	3	2	3	1	1	1	0	944	**0.398**
Number of visits of P2	2	1	1	4	1	2	1	0	1	1	0	654	**0.276**
Number of visits of P3	2	1	1	0	1	2	1	0	1	1	4	774	**0.326**

After using the weights calculated in Table 8. the total WIP and Throughput for all product getting from NLP and Simulation model are in Table 9.

Table 9. Total WIP and Throughput for all product after weighted

Total WIP from Simulation	Total Throughput from Simulation	Total WIP from NLP	Total Throughput from NLP
0.06178811	112.043013	0.06126375	113.827167
0.14623302	247.687163	0.14662231	249.516843
0.20676421	342.666915	0.20780771	346.109586
0.36203701	523.414644	0.36240354	544.716506
0.41959135	601.010008	0.41918207	635.459415
0.42694851	633.234298	0.42525332	643.093477
0.47576184	693.127227	0.47709059	717.541218
0.49339565	705.10445	0.49079253	735.13481
0.51049978	770.305985	0.50812483	759.788275
0.59392785	847.326209	0.59702077	872.624697
0.67204402	938.077468	0.66978011	955.170208
0.80268408	1058.07489	0.80287379	1090.96947
0.87920625	1119.70135	0.8787807	1170.17351
0.94748591	1222.97873	0.94577419	1235.54535
0.99816995	1212.54112	0.99504997	1279.29238
1.21624598	1415.96181	1.21466183	1456.62873
1.33106008	1538.48542	1.34193781	1563.29572
1.39405214	1538.98881	1.39616024	1605.56975
1.45238786	1588.5225	1.46189181	1647.31424
1.52098448	1653.71057	1.52135396	1689.75017

The clearing function for all products after weighted are presented in Figure 3.

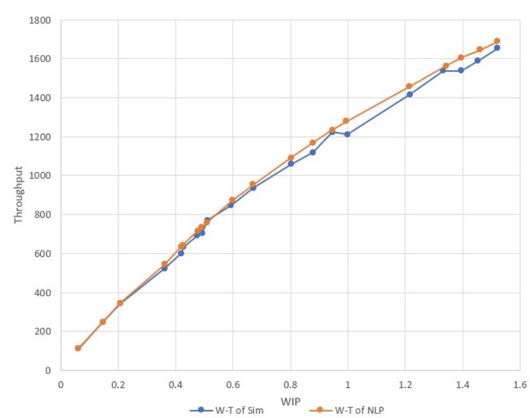


Figure 3. Clearing function of total product from NLP and Simulation

From the graphs getting presented in Figure 2 and 3, we can conclude that the clearing functions obtained from NLP proposed model in this research and from the simulation model are very similar.

Moreover, to show the accuracy of the WIP and Throughput calculated from proposed NLP model and simulation model, the Mean Absolute Percentage Error (MAPE) is used to calculate the percentage error between the WIP and throughput of the three products calculated from NLP model and derived from discrete simulation. We take the WIP of the 1st product from NLP and simulation as an example of how to calculate the MAPE; the MAPEs of other products are calculated similarly. Table 10 shows the WIP calculated from NLP model and simulation.

Table 10. WIP value derived from NLP model and Simulation

WIP1 – NLP (W1)	WIP1 – Simulation (W1s)
1.1779	1.1823
1.3336	1.3256
1.5794	1.5851
1.6145	1.6024
2.8079	2.8012
2.8085	2.8368
2.8322	2.8431
2.9164	2.9239
2.9419	2.9333
2.9636	2.9717
2.9913	3.0127
3.0864	3.1101
3.7169	3.6437
3.7513	3.7422
4.1002	4.1729
5.6076	5.5931
6.1655	6.0498
6.9878	6.9476
7.2234	7.2151
8.5666	8.7569

The MAPE of the WIP of product 1 is calculated as:

$$MAPE_{W1} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|W1_i - W1s_i|}{W1_i}$$

Using the data of the Table 9, the MAPE of the WIP of product 1 is:

$$MAPE_{W1} = 0.75264\%$$

We can calculate the other values of MAPE by doing the similar process for throughput and WIP of the second and the third product. As a result, the MAPE are stated in Table 11, Table 12, and Table 13.

Table 11. MAPE of WIP of each three products

MAPE of WIP of product 1	MAPE of WIP of product 2	MAPE of WIP of product 2
0.3143%	0.5721%	0.5756%

Table 12. MAPE of Throughput of each three products

MAPE of Throughout of product 1	MAPE of Throughput of product 2	MAPE of Throughput of product 2
3.135%	3.3458%	4.3575%

Table 13. MAPE of WIP and Throughput of total products

MAPE of total WIP	0.3344%
MAPE of total Throughput	2.8745%

From the results of MAPR, we can see that the differences between values from the NLP model and simulation are acceptable.

6. Conclusions and Future Study

In this paper, a NLP model was developed for aggregate production planning by directly inserting the formula for system throughput under the Jackson open-queuing network assumption. Numerical results show promise. In future work, several cases from the literature need to be evaluated for accuracy, scalability, computational feasibility with previous approaches in the literature.

References

Albey, E., Bilge, Ü., & Uzsoy, R. . An exploratory study of disaggregated clearing functions for production systems with multiple products. *International Journal of Production Research*, *52*(18), 5301-5322. 2014

Asmundsson, J., Rardin, R. L., & Uzsoy, R, Tractable nonlinear production planning models for semiconductor wafer fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing*, 19(1), 95-111. 2006

Asmundsson, J., Rardin, R. L., Turkseven, C. H., & Uzsoy, R, Production planning with resources subject to congestion. *Naval Research Logistics (NRL)*, *56*(2), 142-157. 2009

Graves, S. C. ,A tactical planning model for a job shop. *Operations Research*, *34*(4), 522-533. 1986

Hackman, S. T., & Leachman, R. C, A general framework for modeling production. *Management Science*, 35(4), 478-495. 1989

Kacar, N. B., & Uzsoy, R. , A comparison of multiple linear regression approaches for fitting clearing functions to empirical data. *International Journal of Production Research*, 52(11), 3164-3184. 2014

Kacar, N. B., Irdem, D. F., & Uzsoy, R, An experimental comparison of production planning using clearing functions and iterative linear programming-simulation algorithms. *IEEE Transactions on Semiconductor Manufacturing*, 25(1), 104-117. 2011

Kacar, N. B., Mönch, L., & Uzsoy, R.,. Modeling cycle times in production planning models for wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 29(2), 153-167, 2016.

Karmarkar, U. S, Capacity loading and release planning with work-in-progress (WIP) and leadtimes. *Journal of Manufacturing and Operations Management*, 2, 105-123, 1989.

Kayton, D., Teyner, T., Schwartz, C., & Uzsoy, R. , Focusing maintenance improvement efforts in a wafer fabrication facility operating under the theory of constraints. *Production and Inventory Management Journal*, 38(4), 51. 1997

Kefeli, A., & Uzsoy, R. , Identifying potential bottlenecks in production systems using dual prices from a mathematical programming model. *International Journal of Production Research*, *54*(7), 2000-2018, .2016.

Kefeli, A., Uzsoy, R., Fathi, Y., & Kay, M, Using a mathematical programming model to examine the marginal price of capacitated resources. *International Journal of Production Economics*, *131*(1), 383-391. 2011

Kim, S., & Uzsoy, R, Modeling and analysis of integrated planning of production and engineering process improvement. *IEEE Transactions on Semiconductor Manufacturing*, 26(3), 414-422. 2013

Manne, A. S.,A note on the Modigliani-Hohn production smoothing model. *Management Science*, 3(4), 371-379. 1957

Missbauer, H. Aggregate order release planning for time-varying demand. *International Journal of Production Research*, *40*(3), 699-718. 2002

Missbauer, H. ,. Models of the transient behaviour of production units to optimize the aggregate material flow. *International Journal of Production Economics*, 118(2), 387-397. 2009

Modigliani, F., & Hohn, F. E, Production planning over time and the nature of the expectation and planning horizon. Econometrica, *Journal of the Econometric Society*, 46-66. 1955

Omar, R. S. M., Venkatadri, U., Diallo, C., & Mrishih, S. , A data-driven approach to multi-product production network planning. *International Journal of Production Research*, *55*(23), 7110-7134. 2017

Orcun, S., Uzsoy, R., & Kempf, K. (December). Using system dynamics simulations to compare capacity models for production planning. In *Proceedings of the 2006 winter simulation conference* (pp. 1855-1862). 2006,

Selçuk, B., Fransoo, J. C., & De Kok, A. G. (). Work-in-process clearing in supply chain operations planning. *Iie Transactions*, 40(3), 206-220. 2008

Spearman, M. L. , An analytic congestion model for closed production systems with IFR processing times. *Management Science*, 37(8), 1015-1029. 1991

Spearman, M. L., & Hopp, W. J, Factory Physics: Foundations of Manufacturing Management. *Irwin, Chicago, IL,* 439. 1996

Biographies

Dang Phi Van Hai is a Ph.D. candidate in the Department of Industrial Engineering, Dalhousie University. Ms. Hai holds a B.Eng. in Automation from University of Technology – Viet Nam National University HCMC, Viet Nam and an M.Eng. in Industrial Engineering from Asian Institute of Technology, Bangkok, Thailand. She has been working as a lecturer in the Department of Industrial Systems Engineering, University of Technology – Viet Nam National University HCMC, Viet Nam from 2010 to 2017.

Uday Venkatadri, Ph.D., P.Eng., is a Professor in the Department of Industrial Engineering at Dalhousie University in Halifax, Nova Scotia. Dr. Venkatadri holds a B.Tech. in Mechanical Engineering from Indian Institute of Technology (BHU) Varanasi, an MS in Industrial Engineering from Clemson University, and a Ph.D. in Industrial Engineering from Purdue University. He has taught at Dalhousie University since July 2001. Before joining Dalhousie, he was a Lead Architect for supply chain planning products at Baan. He has also worked as a Research Associate at Université Laval in Québec City. Dr. Venkatadri's research interests are in facilities planning and supply chain management. He teaches courses in these areas as well as in operations research, modeling of industrial systems, information systems, algorithms, and quality control.

Tri Nguyen-Quang, Ph.D., is an Associate Professor in the Department of Engineering, Faculty of Agriculture at Dalhousie University in Truro, Nova Scotia. Dr. Nguyen-Quang holds a B.Sc. in Engineering from Institut National Polytechnique de Grenoble, an M.Sc. in Soil Structure from Institut National Polytechnique de Grenoble, a Ph.D. in Biofluid Mechanics, from Mechanical Institute of Marseille, and a Ph.D. in Mechanical Engineering from École Polytechnique, University of Montreal. Dr. Tri Nguyen-Quang is Director of Biofluids and Biosystems Modeling Lab (BBML) at Dalhousie University, Canada. His threefold research program which combines theorical aspects, fieldwork/experiments, and mathematical models, mainly focusses on 1) The monitoring and modeling of freshwater Harmful algal Blooms); 2) Computational fluid dynamics applied to coastal erosion and flooding area; 3) Coupling biological and physical models; 4) Supply chain and optimization.

Claver Diallo, Ph.D., is a Professor in the Department of Industrial Engineering at Dalhousie University in Halifax, Nova Scotia. Dr. Diallo holds a B.Eng., M.Sc., and Ph.D. from the Université Laval. His research interests are in performance optimization for production and service systems, specifically maintenance engineering and management, reliability and availability engineering, design and optimization of closed loop supply chains, design for sustainability, and remanufacturing.

Young Ki Chang, Ph.D., is an Assistant Professor in the Department of Engineering, Faculty of Agriculture at Dalhousie University in Truro. Dr. Chang holds a BSc, M.Sc., and Ph.D. from Seoul National University, South Korea. He has worked at Samsung Electronics and was a Postdoctoral Research Fellow at University of California, Davis, Bio-image and Sensing Center at North Dakota State University, and Precision Agriculture Research Program at Nova Scotia Agricultural College, Canada. He founded the bio-systems automation & robotics lab in Dalhousie University. His research interests are in real-time modelling, machine vision and AI applications, and sensor/data fusion.