INT J COMPUT COMMUN, ISSN 1841-9836
7(5):798-806, December, 2012.

Multi-period Customer Service Level Maximization under Limited Production Capacity

S. Babarogić, D. Makajić-Nikolić, D. Lečić-Cvetković, N. Atanasov

Sladjan Babarogić, Dragana Makajić-Nikolić
Danica Lečić-Cvetković, Nikola Atanasov
University of Belgrade, Faculty of Organizational Sciences
Serbia, 11000 Belgrade, Jove Ilića 154
E-mail: sladjan@fon.bg.ac.rs, gis@fon.bg.ac.rs
danica@fon.bg.ac.rs, nikola.atanasov@fon.bg.ac.rs

Abstract:
This paper will focus on a make-to-stock multi-period order fulfilment system with random orders from different classes of customers under limited production circumstances. For this purpose a heuristic algorithm has been developed aimed at maximizing the customer service level in any cycle and in the entire multi-period. In this paper, in order to validate the results obtained with this algorithm, a mixed integer programming model was developed that is based on the same assumptions as the algorithm. The model takes into account the priorities of customer groups and the balanced customer service level within the same group. The presented approaches are applied to a real example of Fast Moving Consumer Goods. Their comparison was carried out in several scenarios.
Keywords: limited production capacity, customer service level, heuristic algorithm, mixed integer programming.

1 Introduction

The distribution of available finished products among customer orders requires an efficient distribution system aimed at improving the effectiveness of the entire business. The effectiveness of business is directly related to products quantities and the profits through sales. In addition to profit-oriented decisions on the selection of orders to be met, it is necessary also to take into account the customer service level. According to [6], key performance indicators in manufacturing companies are identified in measuring the customer service level and customer satisfaction. Customers whose purchases represent a large share of the company's sales require special attention and the company should make sure that they achieve the highest possible fulfilment of each order. There are also customers that continuously increase their orders and based on that also expect a corresponding service level. Due to their large number, small customers influence the overall sales of manufacturing companies. Some of them also represent a potential for future sales growth and increase in revenues of the manufacturing company. These facts underline the importance of making the right decisions when selecting orders to be met. Therefore, a heuristic algorithm has been developed [5] that is used for decision-making concerning the customer service level in each cycle by taking into account the priorities of the customers. Traditional approaches to fulfil orders based on the make-to-stock (MTS) production system are described in [1] by taking into account the available supplies of finished products to satisfy customer orders following the principle of First Come - First Served (FCFS) without assigning priorities to customers and orders. The basic idea of the approach described in [8] is the segmentation of customers in order to increase the total revenue of the manufacturing company by accepting and delivering orders which provide maximum profit.

In this paper the customers are clustered into priority groups based on the size of their orders. Due to their potential growth, there is tendency to provide protected quantities of products to

the customers with the lowest priority. If a manufacturing company has a limited production capacity, it is clear that the company will decide to reject some of the orders received which will have a direct impact on the profitability of the company. The decision to reject orders is made based on a comparison of orders where less profitable orders are rejected. According to [2] this issue has also been defined as dynamic models for managing orders under limited production capacity based on profitability analysis.

The problem discussed in this paper refers to meeting the demand in a multi-period, where the unmet demand in one cycle is not compensated in the following cycles, i.e. there are no backorders. Demand is a weekly phenomenon which requires dynamic decision-making. The heuristics shown in [9] refer to the problem of replenishment of multiple products in order to meet the demand when the storage capacity is limited. Authors in [13] present a mixed integer programming (MIP) model that applies to small and medium-sized enterprises with limited Available to Promise (ATP) quantities and which has to decide which customers they will accept and in each cycle which part of the demand of accepted customers they are going to meet.

A large number of MIP models include unlimited production capacity. The uncapacitated requirement planning model, with demand fulfilment flexibility, is shown in [7]. In each cycle separately a decision is taken regarding the launch of production and the part of demand of each customer that will be met in the respective cycle. A similar MIP model is presented in [12]. This model implies that the manufacturer may, in each cycle, decide whether to start the production, which order to fulfil and to what extent. The above mentioned papers consider the maximization of profit as the main criterion. The approach presented in this paper, however, aims at maximizing the customer service level. In [11] the orders of customers are clustered into two groups: small-size orders and large-size (divisible) orders In the presented basic MIP model, it is in each cycle determined which of the small-size orders will be fulfilled and what fraction of the large-size order will be met in order to maximize the customer service level. The multi-objective MI nonlinear mathematical model, in which the maximization of average customer service levels is one of the four objectives, is presented in [3].

The remainder of this paper is organized as follows: the second section defines the problem of allocating scarce resources in a manufacturing company. The allocation problem is related to the distribution of limited production capacity to customer orders in order to maximize the customer service level. The third section deals with the computational results and provides an analysis of these two approaches in a real example of Fast Moving Consumer Goods (FMCG). In the Conclusions section the authors lists the principal advantages of the proposed algorithm for the allocation of limited production capacity, as well as possible directions for the further development.

2 Problem definition and model formulation

The algorithm has been developed for solving the problem of FMCG industry products. By introducing minor modifications it can also be applied to the product allocation problem in other industries. The basic assumptions of the problem discussed in this paper are the following:

- It studies a multi-period and a set of customers that place order in all or almost all of the cycles. Demand is uneven and is known only for one cycle in advance;
- The production capacity is limited and constant in the entire period;
- If the incoming customer orders in a single cycle do not exceed the available stock of finished goods, the allocation is complete and all customer orders are fulfilled, while any surplus products are stored for the next cycle. Inventory holding costs are neglected;

- When the total of all orders exceeds the available stock of products, it is necessary to define the distribution of products i.e. rules based on which the allocation of products will be done according to the received customer orders. The allocation has to maximize the customer service level.

- Considering the type of products that are being studied, orders that have not been fully met in the reporting cycle shall not be compensated in the subsequent cycles;

- Order of customer priority is known. These are clustered into priority groups in which they have the same order of priority as the other group members;

- The product unit price is the same for all customers.

Based on the assumptions of problem, in previous research effort we have developed heuristic algorithm that introduces the concepts of partitions and tokens. The algorithm aims to maximize the cumulative customer service level with a balanced customer service level within the same group. The detailed explanation of the proposed algorithm is given in [5]. The important features of the algorithm are:

- Classification in groups, provides the order of allocation with primarily focus to satisfy customers that are important for the company.

- Application of mechanism of Partitions, ensures certain groups of lower priority within protected partitions to be involved in allocation so that the low-ranked customers would be at least partially satisfied.

- Application of mechanism of Group Memory Token, allows all customers within a marked group to receive the unsatisfied demand from the previous cycle, with the extended delivery lead time, with which they improve the overall customer service.

2.1 MIP customer service level maximization model

A mixed integer customer service level maximization model (CSL model) model was developed in order to validate the results obtained with this algorithm. The model is based on the same assumption as the algorithm. However, here the total demand of all customers in all cycles is known in advance. Given the purpose of the model, these are the assumptions on which the model is based:

- An r number of cycles is observed and the demand of any of n customers is known in each of those cycles, $OC_{il},\ i = 1, \ldots, n,\ l = 1, \ldots, r$. The demand might not be fulfilled and the demand that was not fulfilled is not compensated for in the next cycle;

- Production is limited and constant in all cycles and it equals $PT_l,\ l = 1, \ldots, r$. If the production exceeds the total demand of the cycles, the surplus products are stored for the following cycle, so as that the total demand ST_l in any cycles equals $PT_l + max\{0, PT_{l-1} - OT_{l-1}\}$. Inventory holding costs are neglected;

- Customers are grouped according to priority. Lowest-priority customers are a protected group to which the amount of protected products $AP_{il},\ i = 1, \ldots, n,\ l = 1, \ldots, r$ is allocated in each cycle.

Model variables:

- z_{il} - customer service level, defined as the fraction of customer order demands OC_{il} delivered on time [10],
- AC_{il} - allocated quantity of the customer i in the cycle l.

CSL model:

$$\max \sum_{l=1}^{r} \sum_{i=1}^{n} w_i z_{il} \tag{1}$$

s.t.

$$AC_{il} - z_{il} \cdot OC_{il} = 0,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{2}$$

$$\sum_{i=1}^{n} AC_{il} \leq ST_l,\ l \in \{1, \ldots, r\} \tag{3}$$

$$AC_{il} \geq AP_{il},\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{4}$$

$$z_{il} \leq 1,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{5}$$

$$AC_{il} \in \mathbb{Z}^+,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{6}$$

The objective function (1) represents the maximization of the total customer service level, where customers are differentiated by using weights. The w_i parameter, which is the weight coefficient, is used to cluster the customers into priority groups. The optimum solution of the CSL model is extremely sensitive to the given values of this parameter, in particular when production is significantly less than the total demand. The first constraint (2) refers to the percentage and absolute satisfaction of demand. The second constraint (3) models the total demand and the total fulfilment of demand in each of the cycles. The total demand in a cycle is made up of the production in the given cycle and the eventual surplus from the previous cycle. The third constraint (4) allows the creation of a protected partition. The AP_{il} parameter, which is the minimum quantity of products to be delivered to the customer i in the cycle l, represents the reserved quantity for the customer i which is in the protected partition. The value of this parameter for customers outside the protected partition is 0. The value of this parameter has a direct impact on the service level of customers from the protected partition and indirect impact on the service level of customers outside this partition. The last constraint (5) represents the upper bound of the fractional customer service level for the customer i in the cycle l.

3 Computational results and discussion

In order to analyze the impact of the production capacity to the customer service level, the algorithm and the CLS model were tested in three scenarios. All the parameters are the same, except for the production level which equals 1000, 1300 and 1400 FMCG units respectively. For the scenario when the production level reaches 1000 units, the supply is significantly lower than the total demand, for the 1300 units scenario the supply almost meets the minimum total demand. In case of the 1400 units scenario, there are unallocated products only in a few cycles.

Table 1 shows the demand of nine customers over a period of nine weeks. The customers are clustered into three groups. Customers A1 and A2 belong to the first group, while customers B1-B4 make part of the second one. The C-customers are in the third group. The first and the second group of customers are in the first partition, while the third one is in the second, protected partition.

Table 1: Input parameters

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3	Demand
W1	330	575	280	110	40	121	100	52	25	1633
W2	360	393	110	170	135	157	74	40	0	1439
W3	220	700	60	100	160	130	100	65	40	1575
W4	230	480	120	140	80	146	74	94	20	1384
W5	270	650	390	110	100	241	140	83	0	1984
W6	381	751	89	140	260	95	110	48	30	1904
W7	320	615	20	120	90	100	46	27	20	1358
W8	390	1.055	120	120	190	130	110	75	0	2190
W9	305	780	290	90	60	110	30	92	11	1768
TOTAL	2806	5999	1479	1100	1115	1230	784	576	146	

The parameter value KP (protective percentage quota) for all three scenarios is 0.95 for the first partition and 0.05 for the second one. These values are based on the decision of the company management which is founded on the realistic assumption that it is necessary to satisfy even the small customers in order to keep them in the system and take advantage of their potential growth. In this way the dependence on major customers would also be reduced. The given KP parameter values are used to determine the AP (amount of allocated products) parameter value . For each customer, the value of AP parameter is set to 5

For the purpose of benchmarking the presented algorithm it was necessary to set the weights in the CSL model that will provide the best balance of the customer service level for the given data. The customers from the protected partition are not included into this sensitivity analysis, as they belong to the lowest-priority group and in any weight distribution their weights are 1. The AP parameter values have a much greater influence on the fulfilment of their demands. The GNU Linear Programming Kit [4] was used to solve the model. The programming kit includes the branch-and-cut algorithm for solving the MIP problem. Figure 1 shows five value variants of weight coefficients. The first value in brackets represents the weight coefficient of the customers from the first group, while the second value is the weight coefficient of the second group. Given the fact that the first group is a higher priority group, the weight coefficients of the customers from this group have to be greater than the weight coefficients of the customers from the second group. The figure shows that the significant difference between weight coefficients has a negative effect on the balancing of the second group. However, small differences have a negative effect on the balancing of the first group. Using the Bisection method has shown that for the given data it is necessary to use the weight 65 for the first customer group and 10 for the second one.

Scenario 1

Table 2 presents the results obtained through the application of the algorithm, while Table 3 contains results of the CLS model optimization. The given production capacity in both cases was 1000 FMCG units. Both tables show the customer service level for every customer over a period of nine weeks. The last row provides the average customer service level for every customer.

By using the CLS model, the objective function value, which represents the weighted customer service level, is 1213.77. By weighting the results of the algorithm, the objective function becomes the value of 1092.77 (90.03% of the optimum value). This is an expected result because the CLS model maximizes the weighted customer service level. However, from the point of view of the company management it is more important that the customer service level for the customers in the same group is balanced.

Based on the results for customers in the first group (A1 and A2), it can be concluded that

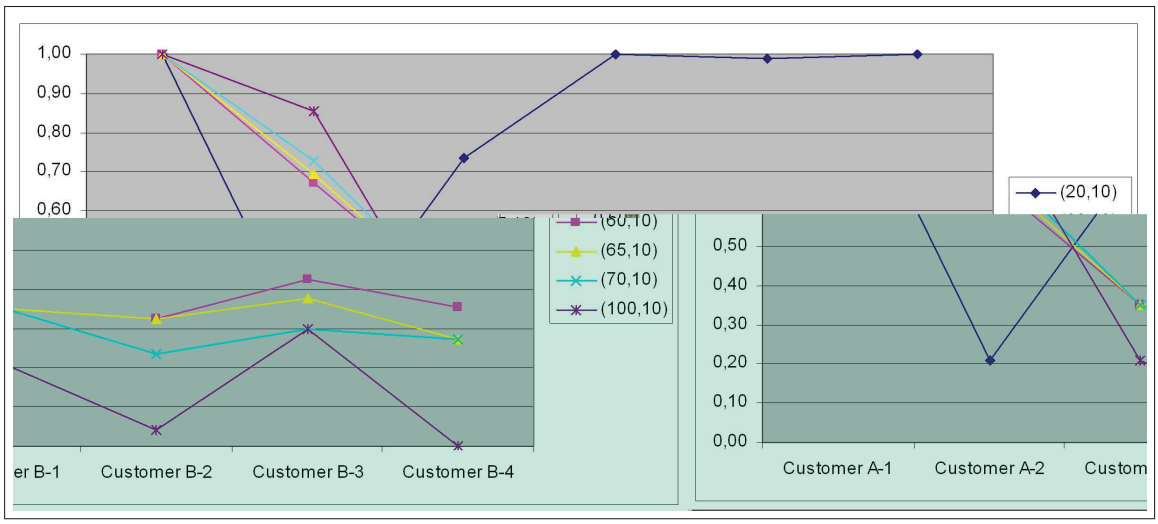

Figure 1: CLS sensitivity to weights

Table 2: Algorithm results for 1000 units

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3
W1	1.000	1.000	0.082	0.082	0.075	0.083	0.280	0.288	0.280
W2	0.786	0.784	1.000	0.594	0.274	0.707	0.432	0.450	-
W3	1.000	1.000	0.067	0.070	0.063	0.069	0.240	0.246	0.250
W4	0.843	0.846	0.467	0.664	1.000	0.829	0.270	0.266	0.250
W5	1.000	1.000	0.036	0.036	0.030	0.037	0.221	0.229	-
W6	0.496	0.498	1.000	0.757	0.373	1.000	0.264	0.271	0.267
W7	1.000	1.000	0.050	0.042	0.044	0.050	0.543	0.519	0.550
W8	0.438	0.440	0.158	0.958	0.453	0.731	0.273	0.267	-
W9	0.846	0.887	0.000	0.000	0.000	0.000	0.367	0.380	0.364
AVG	0.797	0.797	0.214	0.400	0.287	0.370	0.293	0.304	0.308

Table 3: Results of CLS model optimization for 1000 units

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3
W1	1.000	0.963	0.050	0.055	1.000	0.058	0.170	0.308	0.640
W2	1.000	1.000	1.000	0.053	0.644	0.051	0.230	0.400	-
W3	1.000	0.794	1.000	1.000	0.050	0.054	0.170	0.246	0.400
W4	1.000	1.000	1.000	0.236	1.000	0.055	0.230	0.170	0.800
W5	1.000	1.000	0.051	0.055	0.080	0.054	0.121	0.193	-
W6	1.000	0.487	1.000	0.050	0.050	1.000	0.155	0.333	0.533
W7	1.000	0.829	1.000	0.050	1.000	0.050	0.370	0.593	0.800
W8	1.000	0.187	1.000	1.000	0.053	1.000	0.155	0.213	-
W9	1.000	0.482	0.052	1.000	1.000	1.000	0.567	0.174	1.000
AVG	1.000	0.680	0.384	0.343	0.355	0.311	0.195	0.250	0.623

the use of algorithm keeps the customer service level completely balanced. By using the CLS models, customer A1, whose demand is much lower than the one of the customer A2, has a much higher customer service level. This is because the costs of high customer satisfaction within the same group are minimal when demand is the lowest, because the lowest demand causes the highest increase of the objective function. In the second group, when the CLS model was used, the customer service level for all nine weeks was more balanced than with the application of the algorithm. However, analyzing customer service level by week, it becomes clear that every week the algorithm assigns a certain amount of products to every customer from the second group and that the balance is very good every odd week. This is primarily due to use of tokens. On the other hand, based on the results of the CLS model, it is evident that every week at least one customer from the second group receives protected amount of products. In other words, looking at it by week, the balancing is inadequate. The third group belongs to a protected partition and always gets the guaranteed amount of products. The algorithm provides perfect balancing. However, based on the results of the CLS model the last customer in the third group has the highest customer service level due to its low demand which is often less than the guarantied amount of products. This happens due to the maximization of the satisfaction fractions and not the absolute amount of assigned products.

Scenario 2

The main feature of a scenario with 1300 FMCG units is that the production of each week still does not meet the total demand, but the lack of finished products is lesser than in the first scenario. Table 4 provides the average customer service level for each customer over a period of all nine weeks based on the results of the algorithm and the optimization of the CLS model.

Table 4: Results of the algorithm and the CLS model for 1300 units

	A1	A2	B1	B2	B3	B4	C1	C2	C3
Algorithm	0.937	0.929	0.539	0.706	0.556	0.675	0.434	0.452	0.429
CLS model	1.000	0.870	0.412	0.792	0.497	0.891	0.305	0.328	0.808

The weighted overall customer service level based on the results of the CLS model equals 1416.17. The results of the algorithm make this value reach 1361.42 (96.13 % of the optimum value). The balancing obtained by applying the algorithm is better than the balancing obtained through the model and it is even more prominent than in scenario 1.

Scenario 3

When the production reaches 1400 FMCG units, in two weeks (W4 and W7) the supply exceeds the demand and the surplus products are stored for the next week. The average customer service levels are shown in 5. Based on the results, the weighted overall customer service level for the CLS model equals 1463.77 and 1419.76 for the algorithm (96.99% of the optimum value).

Table 5: Results of the algorithm and the CLS model for 1400 units

	A1	A2	B1	B2	B3	B4	C1	C2	C3
Algorithm	0.971	0.968	0.627	0.733	0.575	0.765	0.458	0.542	0.548
CLS model	1.000	0.924	0.477	0.879	0.581	0.982	0.383	0.521	0.836

Looking at the data from the three scenarios above, it can be concluded that by increasing the production the total value of customer service level obtained through the algorithm, is nearing the optimum value. If observed from the balancing point of view, the advantage of the algorithm is increasingly more prominent compared to the CLS model.

4 Conclusions and Future Works

This paper presents benchmarking of a heuristic algorithm for the dynamic solving of the problem of allocating limited supplies of products to received customer orders with the aim of maximizing the customer service level. In order to validate the algorithm an MIP model was developed. The model is used for the distribution of products based on demands that are known in advance for the entire period. Computational results have indicated that the proposed algorithm, with the increase in the production capacity, ensures a value of the total customer service level that is closer to the optimum values obtained using the CLS model. In addition, in all three scenarios the balancing results of service level for customers from the same group achieved with the algorithm were better than the balancing obtained using the CLS model.

Further analysis of customer demand by week and the obtained customer service levels have shown that the further research might have to be directed towards the analysis of the correlation between the customer service level and the fluctuations in demand. Besides, the algorithm could be modified by introducing the assumption that the selling price depends on the customer affiliation to a certain group or on the quantity of products ordered. The growth rate of a customer's demand is one of the essential elements used by the management in manufacturing companies in planning the sales. The inclusion of this parameter into the problem would require the modification of the algorithm making it a more useful tool in the decision-making process.

Acknowledgement

This research was partially supported by the Ministry of Education and Science, Republic of Serbia, Project number: TR35045.

Bibliography

[1] Cederborg O., Rudberg M., Customer Segmentation and Capable-to-Promise in a Capacity Constrained Manufacturing Environment, *16th Int. Annual EurOMA Conference*, Goteborg, Sweden, 2009, http://www.iei.liu.se/prodek/forskning/iscaps/filarkiv/1.120209/CederborgandRUdbergEurOMA2009.pdf, accessed 12 January 2010.

[2] Chan F.T., Chung S.H., A Modified Multi-Criterion Genetic Algorithm for Order Fulfillment in Manufacturing Network, *Proceedings of the 9th Asia Pacific Industrial Engineering & Management System Conference*, APIEMS, Indonesia, 2221-2226, 2008.

[3] Chen C., Lee W., Multi-objective optimization of multiechelon supply chain networks with uncertain product demands and prices, *COMPUT. CHEM. ENG.*, ISSN 0098-1354, No 28: 1131-1144, 2004.

[4] GLPK - GNU Linear Programming Kit. http://www.gnu.org/software/glpk, accessed 25 December 2011.

[5] Lecic-Cvetkovic D., Atanasov N., Babarogic S., An Algorithm for Customer Order Fulfillment in a Make-to-Stock Manufacturing System, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(5): 983-791, 2010.

[6] Lin J., Chen J.H., Enhance Order Promising with ATP Allocation Planning Considering Material and Capacity Constraints, *JCIIE*, ISSN 2151-7606, 22(4): 282-292, 2005.

[7] Merzifonluoglu Y., Geunes J., Uncapacitated production and location planning models with demand fulfilment flexibility, *INT J PROD ECON*, ISSN 0925-5273, 102: 199-216, 2006.

[8] Meyr H., Customer Segmentation, Allocation Planning and Order Promising in Make-to-Stock Production, *OR SPECTRUM*, ISSN 01716468, 31(1): 229-256, 2009.

[9] Minner S., A comparison of simple heuristics for multi-product dynamic demand lot-sizing with limited warehouse capacity, *INT J PROD ECON*, ISSN 0925-5273, 118: 305-310, 2009.

[10] Pochet Y.,Wolsey L.A., *Production Planning by Mixed Integer Programming*, Springer, 2010.

[11] Sawik T., Integer programming approach to reactive scheduling in make-to-order manufacturing, *MATH COMPUT MODEL*, ISSN 0895-7177, 46(11-12): 1373-1387, 2007.

[12] Xiao Y., Taaffe K., Satisfying market demands with delivery obligations or delivery charges, *COMPUT OPER RES*, ISSN 0305-0548, 37(2): 396-405, 2010.

[13] Xiong M.H. et al, A DSS approach to managing customer enquiries for SMEs at the customer enquiry stage, *INT J PROD ECON*, ISSN 0925-5273, 103(1): 332-346, 2006.