Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

A ROBUST OPTIMIZATION APPROACH FOR PRODUCTION PLANNING UNDER EXOGENOUS PLANNED LEAD TIMES

Erinç Albey

Department of Industrial Engineering
Özyeğin University
İstanbul, 34794, TURKEY

İhsan Yanıkoğlu

Department of Industrial Engineering
Özyeğin University
İstanbul, 34794, TURKEY

Reha Uzsoy

Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, NC 27695-7906, USA

ABSTRACT

Many production planning models applied in semiconductor manufacturing represent lead times as fixed exogenous parameters. However, in reality, lead times must be treated as realizations of released lots' cycle times, which are in fact random variables. In this paper, we present a distributionally robust release planning model that allows planned lead time probability estimates to vary over a specified ambiguity set. We evaluate the performance of non-robust and robust approaches using a simulation model of a scaled-down wafer fabrication facility. We examine the effect of increasing uncertainty in the estimated lead time parameters on the objective function value and compare the worst-case, average optimality, and feasibility of the two approaches. The numerical results show that the average objective function value of the robust solutions are better than that of the nominal solution by a margin of almost 20% in the scenario with the highest uncertainty level.

1 INTRODUCTION

The problem of planning releases into capacitated production facilities is central to the domain of production planning and supply chain management. The objective is to determine the timing and quantity of releases into the facility to ensure that production meets demand in some optimal manner. This, in turn, requires explicit consideration of the cycle times, the time elapsing between the release of a unit of work into the facility and its completion as finished product that can be used to meet demand. In highly capital-intensive industries such as semiconductor manufacturing, whose complex manufacturing processes require hundreds of unit operations, average cycle times of the order of 10-12 weeks are common. Hence any effective release planning mechanism must take cycle times into account in some manner.

The problem of modeling cycle times in release planning is further complicated by the nonlinear, stochastic behavior of production systems. Both queuing models (Curry and Feldman 2010) and simulation analysis (Mönch et al. 2013) have shown that the cycle time of jobs moving through a production unit is a random variable whose probability distribution is affected by the level of resource utilization as well as, potentially, the entire history of the stochastic processes governing the arrival and processing of work at the machines making up the system. Since these factors are determined by the release decisions made by the release planning model itself, cycle times should be treated as endogenous to the release planning problem.

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

However, explicitly addressing this endogeneity of cycle times results in extremely complex optimization models. Throughout this paper, the term "cycle time" will refer to the random variable representing the time between the release of a unit of work and its completion, while the term "lead time" will refer to a parameter representing the cycle time in a release planning model. We shall refer to the basic unit of work released into the production system as a job, and the quantity of work released into the system in a given period as a production lot.

The vast majority of release planning models in both the academic literature and industrial practice represent cycle times as exogenous parameters independent of system workload that are unaffected by the decisions made by the release planning model in which they are used. These models include the widely used Material Requirements Planning (MRP) procedure (Vollmann et al. 2005) as well as the linear and integer programming models proposed in the production planning domain (Hackman and Leachman 1989; Voß and Woodruff 2006; Missbauer and Uzsoy 2011). The prevalence of these tools raises the interesting question of how they can be modified to accommodate the stochastic nature of cycle times to some degree.

One approach that has been proposed for modeling cycle time distributions in release planning is the use of load factors e_{ipt}, which specify the fraction of material of product i released in period p that will emerge as output in period t. These can be viewed as specifying a deterministic output distribution over time for the material released in a given period, and constitute a generalization of several other ways of representing exogenous lead times. In this paper we seek to incorporate stochastic cycle times into release planning models by treating the load factors as random variables that can vary over a specified uncertainty set. We test the model on a data set derived from a scaled-down semiconductor manufacturing facility (Kayton et al. 1997), and find that the use of a distributionally robust formulation can yield significant improvements over purely deterministic approaches.

2 PREVIOUS RELATED WORK

The literature on representing cycle times in production planning problems has generally taken three different approaches: the use of exogenous fixed lead times, iterative multi-model methods and the use of nonlinear clearing functions. We now discuss each of these in the following subsections. Throughout the paper we shall assume a release planning model in which a finite planning horizon is divided into T periods of equal length, as is common practice in mathematical programming models used for release planning.

The vast majority of the release planning procedures in the literature model cycle times using fixed, exogenous, workload-independent parameters. This approach can take several forms. The most common is to treat the cycle time as a deterministic, workload-independent delay of duration L_i between the release of the work into the system and its completion. In most models the lead time L_i is assumed to be an integer multiple of the basic planning period, and to be associated with an entire planning period, in the sense that each unit of work released in that period will take exactly L_i time units to complete its processing. Hence a uniform rate of releases over the planning period t leads to a uniform rate of output in a later period $t+L_i$. If we denote the amount of product i released into the production system in period t by R_{it} and the output of finished product by X_{it}, this relation can be expressed as

$$X_{it} = R_{i,t-L_i}. \tag{1}$$

Under this approach the output of the production resource or production system is simply the release schedule right-shifted in the amount of the lead time parameter L_i. Hackman and Leachman (1989) have shown that this approach can be extended in a straightforward manner to the case where L_i is a non-integer value, assuming that the releases R_{it} are uniformly distributed over the planning period in which they take place. Essentially the same relation is used in the backwards scheduling phase of the well-known Material Requirements Planning procedure (Vollmann et al. 2005). An interesting consequence of this modeling approach is that under integer lead times all material released in a given period is assumed to move through the production process as a unit, consuming capacity at various machines and entering finished inventory in a single period.

A refinement of this approach that seeks to represent the fact that releases in a given period will encounter a range of cycle times specifies a set of load factors w_{ist} associated with a planning period p, representing the fraction of work released in planning period p that will emerge as finished product in period t. The load factor e_{ipt} can be viewed as an estimate of the probability that a unit of material released in period p will emerge as finished product in period t. In this case we have the relations

$$X_{it} = \sum_{p=1}^{t} e_{ipt} R_{ip}, \tag{2}$$

$$\sum_{t} e_{ipt} = 1. \tag{3}$$

Setting one of the load factors in (2) (and also in (3)) to one and the rest to zero recovers the equation (1) above. While it is possible to make both load factors e_{ipt} and lead times L_i time-dependent, their values must be selected with care to avoid inconsistencies such as material released in earlier planning period overtaking material released later. It is also not at all evident how to select time-dependent parameters in a consistent manner. These issues are discussed by Missbauer and Uzsoy (2011); closely related issues are discussed in the context of traffic modeling by Carey (1992).

Another refinement suggested by Hung and Leachman (1996) is to associate lead time parameters with the boundaries between periods. Especially when fractional lead times are used, this allows great flexibility in modeling, such that releases made in a single planning period may emerge as output over multiple periods. This is accomplished by computing load factors based on the lead time estimates, again assuming uniform rates of releases in planning periods.

The use of exogenous lead times also raises the issue of at what point in the lead time the job consumes capacity on the machines. In this case, the exogenous lead time represents the delay between the job arriving at a particular machine and its completing its processing there. Capacity on the machine may be consumed in the last period of this lead time, i.e., the period in which the job is completed, or at any point within this lead time (Spitter et al. 2005).

A number of authors have addressed the problem of determining optimal lead times for different production and inventory systems, including Ben-Daya and Raouf (1994) for inventory systems and Gong et al. (1994) and Milne et al. (2015) for MRP systems. Albey and Uzsoy (2015) use simulation optimization to determine the cost-minimizing set of time-varying lead times for a production planning model, and find that significant improvements in plan performance can be obtained. Kacar et al. (2016) find that the use of fractional exogenous lead times can lead to significant improvements in performance over integer lead times.

In summary, a variety of models with fixed exogenous lead times of various kinds have been proposed, generally yielding computationally tractable linear optimization models (Tardif and Spearman 1997; Voß and Woodruff 2006). However, the nonlinear relation between resource utilization and the mean and variance of the cycle time is completely neglected, as is the fact that different jobs released as part of the same production lot in the same period may emerge in different planning periods due to the stochastic nature of the production process. An interesting consequence of the failure to model congestion is that these release planning models yield meaningful dual prices for resource capacity only when a resource is fully utilized, failing to capture the effect of resources that are heavily but not fully utilized and thus contribute significantly to the cycle time. While our focus in this paper is on models with fixed exogenous lead times, for the sake of completeness we now briefly review alternative approaches to modeling the dependency between release decisions and cycle times.

Multimodel iterative approaches decompose the release planning problem into two subproblems, iterating between them until a mutually consistent solution is obtained. The first subproblem, generally formulated as a linear program, computes the optimal releases for a given set of lead time estimates. The second subproblem, usually a simulation model of the production system of interest, then computes estimates of the lead times induced by these releases, and feeds them back to the first subproblem. Several algorithms of this type have been proposed (Hung and Leachman 1996; Byrne and Bakir 1999; Hung and Hou 2001; Kim and Kim 2001; Byrne and Hossain 2005; Bang and Kim 2010; Albey and Bilge 2011), which differ in how specific parameters of the release planning subproblem are updated. While this approach combines two well-understood modeling techniques - simulation and linear programming - in an intuitive manner, their convergence behavior is not well understood. Irdem et al. (2010) compared the performance of the approaches of Hung and Leachman (1996) and Kim and Kim (2001), finding marked differences in their convergence behavior. The solution reached by the Kim and Kim approach depends on the initial solution, suggesting the strong possibility of local optimality. Finally, the need to perform multiple replications of a large simulation model during the planning computations results in high computational burden for this class of methods.

Nonlinear clearing functions are essentially a metamodel for a queuing system, relating some measure of its planned workload in a planning period to its expected output in that period. The concept was originally proposed by several researchers in the late 1980's (Graves 1986; Srinivasan et al. 1988; Karmarkar 1989), and in its most common form expresses the planned output of a resource as a concave non-decreasing function of its workload, the amount of work available to the resource in the planning period. These functions can then be embedded in a mathematical programming model for release planning. The state of the art of these models at the present time is the Allocated Clearing Function model of Asmundsson et al. (Asmundsson et al. 2006; Asmundsson et al. 2009), which has been tested computationally in a wide range of environments and found to outperform models with fixed lead times when appropriately parameterized (Haeussler and Missbauer 2014; Albey et al. 2014; Kacar et al. 2016; Albey et al. 2017; Ziarnetzky et al. 2018).

Parameter estimation of clearing functions severely affects the functions' performance and a satisfactory framework for estimating these parameters for practical production systems have been lacking. Estimates in the literature are mostly obtained using conventional least-squares regression approaches. Albey et al. (2017) takes an alternative approach which seeks to explicitly account for the uncertainties in the clearing function parameters induced by estimation errors using robust optimization (RO). These authors compare the performance of a production planning model using a multi-dimensional clearing function and its robust counterpart under several experimental settings. An extensive discussion of clearing function based production planning models is given by Missbauer and Uzsoy (2011). However, since the focus of this paper is on models with fixed exogenous lead times, we will not discuss this body of work further.

The vast majority of the release planning models used in the literature have taken the form of deterministic mathematical programs (Pochet and Wolsey 2006; Voß and Woodruff 2006). While the load factor approach does indeed allow material released in a single period to emerge over a number of periods, representing a distribution of the cycle times, this distribution is completely deterministic and exogenous to the release planning model. Given the interdependency between the release decisions the planning model seeks to make and the cycle times that will result, as well as the inherently stochastic nature of the cycle times in any realistic production system, a tractable solution to the general problem appears difficult. Hence in this paper we shall focus on a simpler problem, in which demand is assumed to be known with certainty for the duration of the planning horizon and the load factors relating the releases to the output are subject to uncertainty. Even this more limited problem is challenging. The use of scenario-based stochastic programming (Birge and Louveaux 2011) would result in a scenario tree that grows exponentially in the number of planning periods, resulting in very large formulations. In addition, the interrelation between the releases and the cycle times raises the question of how one would generate meaningful scenarios for the load factors, which we would expect to be highly correlated across periods. The robust optimization approach of Bertsimas and Sim (2004), particularly its adaptation to inventory planning by Bertsimas and Thiele (2006), offer a promising approach to the problem we address in this paper. In particular, the robust formulation yields a well-structured, tractable linear program with considerable computational advantages. We now describe the proposed model in the following section.

3 RELEASE PLANNING MODEL

For simplicity of exposition we assume that all products follow the same sequence of machines during the production process, and that unsatisfied demand in a period can be backordered at a cost while deriving the robust formulation. The first of these assumptions is relaxed in the numerical experiments. The deterministic formulation of the *release planning problem* (RPP) is then as follows.

(RPP):

$$\min \quad \sum_{i\in I}\sum_{t\in T}\left(\sum_{p=1}^{t} c_i R_{ip} e^0_{i|K|pt} + \max\{h_{it}I_{it}, \beta_{it}B_{it}\}\right)$$

$$\text{s.t.} \quad \sum_{p=1}^{t}\sum_{\tau=p}^{t} R_{ip}e^0_{i|K|p\tau} - \sum_{\tau=1}^{t} d_{i\tau} \le I_{it} \quad \forall i\in I, \forall t\in T \tag{4}$$

$$\sum_{\tau=1}^{t} d_{i\tau} - \sum_{p=1}^{t}\sum_{\tau=p}^{t} R_{ip}e^0_{i|K|p\tau} \le B_{it} \quad \forall i\in I, \forall t\in T \tag{5}$$

$$\sum_{i\in I} a_{ik}\sum_{p=1}^{t} R_{ip}e^0_{ikpt} \le Q_{kt} \quad \forall k\in K, \forall t\in T \tag{6}$$

$$B_{it}, I_{it}, R_{it} \ge 0 \quad \forall i\in I, \forall t\in T, \tag{7}$$

where e^0_{ikpt} is the load factor representing the *nominal* (or expected) probability that product type $i\in I$ that is released at period $p\in T$ will be completed on machine $k\in K$ in period $t\in T$; R_{ip} denotes the release amount of product $i\in I$ released in period $p\in T$; I_{it} and B_{it} denote the inventory and backorder levels of product type $i\in I$ at the end of period $t\in T$, respectively; d_{it} is the demand for product type i for period $t\in T$; a_{ik} denotes the resource usage of unit product type $i\in I$ on machine $k\in N$; and Q_{kt} is the available capacity of machine $k\in K$ at period $t\in T$. Constraints (4) and (5) enforce material balance for the finished goods inventory, while (6) ensures that the resource capacity Q_{kt} is not exceeded for each period $t\in T$ and machine $k\in K$. (7) are the non-negativity constraints. Notice that $|K|$ denotes the last machine in the routing and is used only once. Ultimately, the objective of RPP is to minimize the sum of expected release, inventory holding/backorder costs.

It is important to point out that the load factor probabilities ($\boldsymbol{e}$) are assumed to be known in RPP. However, in practice, they follow an ambiguous distribution due to errors in estimation; possible sources of such uncertainties are discussed in Stinstra and Den Hertog (2008) and Yanıkoğlu et al. (2016). To yield distributionally robust solutions for such errors, we shall use the robust optimization paradigm. First, we introduce the ambiguity set U that shall be used to model the uncertain load factor probability vector as

$$U_{ikp} = \left\{\boldsymbol{e}_{ikp}\in[0,1]^{|T|}: \quad \sum_{t\in T} e_{ikpt} = \Gamma_{ip}; \quad \ell_{ikpt} \le e_{ikpt} \le u_{ikpt} \quad \forall t\in T\right\}.$$

The ambiguity set ensures that the elements of the probability vector sum to $\Gamma_{ip}\in[0,1]$ (notice that $\Gamma_{ip}=0$ denotes item i released in period p cannot leave the system in the planning horizon $\{p,\dots,|T|\}$), and reside between lower and upper bounds determined by vectors $\boldsymbol{\ell}\in[0,1]^{|T|}$ and $\boldsymbol{u}\in[0,1]^{|T|}$. Following this approach, the *semi-infinite release planning problem* (SRPP) yielding a distributionally robust solution against the uncertainties in the load factor probabilities is given as follows.

(SRPP):

$$\min \quad \sum_{i\in I}\sum_{t\in T}\left(\left(\sum_{p=1}^{t} c_i R_{ip} e^0_{i|K|pt}\right) + \max\{h_{it}I_{it}, \beta_{it}B_{it}\}\right)$$

$$\text{s.t.} \quad \sum_{p=1}^{t}\sum_{\tau=p}^{t} R_{ip}e_{i|K|p\tau} - \sum_{\tau=1}^{t} d_{i\tau} \le I_{it} \quad \forall i \in I, \forall t \in T, \forall \boldsymbol{e} \in U \tag{8}$$

$$\sum_{\tau=1}^{t} d_{i\tau} - \sum_{p=1}^{t}\sum_{\tau=p}^{t} R_{ip}e_{i|K|p\tau} \le B_{it} \quad \forall i \in I, \forall t \in T, \forall \boldsymbol{e} \in U \tag{9}$$

$$\sum_{i\in I} a_{ik} \sum_{p=1}^{t} R_{ip}e_{ikpt} \le Q_{kt} \quad \forall k \in K, \forall t \in T, \forall \boldsymbol{e} \in U \tag{10}$$

$$B_{it}, I_{it}, R_{it} \ge 0 \quad \forall i \in I, \forall t \in T.$$

SRPP is intractable because it contains infinitely many constraints as expressed by the universal quantifier in the inventory balance constraints (8) and (9). To obtain the tractable reformulation of the problem, we shall apply the three step procedure of the RO paradigm per Gorissen et al. (2015), §2). For the sake of brevity, we shall present the paradigm over constraint (9). Notice that RO is a constraint-wise approach and the same procedure is also adapted to derive the tractable robust reformulations of the objective function, and constraints (9) and (10).

The first step is to maximize the left-hand side of (8) over the uncertainty set for each $i \in I$ and $t \in T$ as follows.

$$\max_{\boldsymbol{e}\in \boldsymbol{U}} \left\{ \sum_{p=1}^{t}\sum_{\tau=p}^{t} R_{ip}e_{i|K|p\tau} \right\} \le I_{it} + \sum_{\tau=1}^{t} d_{i\tau} \qquad \forall i \in I, \forall t \in T, \tag{11}$$

where $[\boldsymbol{e} \in \boldsymbol{U}]$ denotes $[\sum_{t\in T} e_{ikpt} = \Gamma_{ip} \quad \forall i \in I, k \in K, p \in T; \quad \ell_{ikpt} \le e_{ikpt} \le u_{ikpt} \quad \forall i \in I, k \in K, (p,t) \in T]$ The practical implication of the first step is that if constraint (8) is satisfied with respect to all realizations of the uncertainty for the robust solution at hand $(\boldsymbol{R}^*, \boldsymbol{I}^*)$, it must also be satisfied for the worst-case realization of the uncertainty. Notice that when the decision variables are fixed to the solution at hand, the problem is linear in the uncertainty parameter $\boldsymbol{e}$.

In the second step, we take the dual of the maximization problems for each $i \in I$ and $t \in T$, obtaining the dual problem shown in (12) and (13).

$$\min_{\boldsymbol{v},\boldsymbol{R}} \sum_{p=1}^{t} \left(\Gamma_{ip}v_{ipt}^3 + \sum_{\tau=p}^{|T|} (\ell_{i|K|p\tau}\ v_{ip\tau t}^1 + u_{i|K|p\tau}\ v_{ip\tau t}^2)\right) \ \left(\le I_{it} + \sum_{\tau=1}^{t} d_{i\tau}\right) \tag{12}$$

$$\text{s.t.} \quad v_{ip\tau t}^1 + v_{ip\tau t}^2 + v_{ipt}^3 \ge \beta_{ip\tau}^t R_{ip} \quad \forall i \in I, \forall t \in T, \forall p \in T : p \le t, \forall \tau \in T : p \le \tau \le t, \tag{13}$$

where $(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{v}^3)$ denote the dual variables for the lower bound, upper bound, and equality constraints determined by the uncertainty set $(\boldsymbol{U})$; respectively. Notice that the minimization problem yields an upper bound for the problem in Step 1 by weak duality, i.e., we may remove the minimization in the LHS of the constraint in the final step (shown in (14) and (15)), obtaining the robust counterpart (RC) for constraint (8).

$$\sum_{p=1}^{t} \left(\Gamma_{ip}v_{ipt}^3 + \sum_{\tau=p}^{|T|} (\ell_{i|K|p\tau}\ v_{ip\tau t}^1 + u_{i|K|p\tau}\ v_{ip\tau t}^2)\right) \le I_{it} + \sum_{\tau=1}^{t} d_{i\tau} \quad \forall i \in I, t \in T \tag{14}$$

$$v_{ip\tau t}^1 + v_{ip\tau t}^2 + v_{ipt}^3 \ge \beta_{ip\tau}^t R_{ip} \quad \forall i \in I, \forall t \in T, \forall p \in T : p \le t, \forall \tau \in T : p \le \tau \le t. \tag{15}$$

The same RO procedure is also adapted in the other constraints and the objective function, and the resulting *RC of the release planning problem* (RRPP) is given as follows; for further details on RO we refer the reader to Gorissen et al. (2015):

(RRPP):

$$\min \quad \sum_{i\in I}\sum_{t\in T} \left(\sum_{p=1}^{t} c_i R_{ip} e_{i|K|pt}^0 \right) + Z_{it}$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{p=1}^{t} \left(\Gamma_{ip} v_{ipt}^{3} + \sum_{\tau=p}^{|T|} (\ell_{i|K|p\tau}\ v_{ip\tau t}^{1} + u_{i|K|p\tau}\ v_{ip\tau t}^{2}) \right) - \sum_{\tau=1}^{t} d_{i\tau} \le I_{it} \quad \forall i \in I, \forall t \in T \\
& v_{ip\tau t}^{1} + v_{ip\tau t}^{2} + v_{ipt}^{3} \ge \beta_{ip\tau}^{t} R_{ip} \quad \forall i \in I, \forall (p,t) \in T : p \le t, p \le \tau \le t \\
& \sum_{\tau=1}^{t} d_{i\tau} - \sum_{p=1}^{t} \left(\Gamma_{ip} w_{ipt}^{3} + \sum_{\tau=p}^{|T|} (\ell_{i|K|p\tau}\ w_{ip\tau t}^{1} + u_{i|K|p\tau}\ w_{ip\tau t}^{2}) \right) \le B_{it} \quad \forall i \in I, \forall t \in T \\
& w_{ip\tau t}^{1} + w_{ip\tau t}^{2} + w_{ipt}^{3} \le \beta_{ip\tau}^{t} R_{ip} \quad \forall i \in I, \forall (p,t) \in T : p \le t, p \le \tau \le t \\
& \sum_{i \in I} a_{ik} \sum_{p=1}^{t} R_{ip} u_{ikpt} \le Q_{kt} \quad \forall k \in K, \forall t \in T \\
& h_{it} I_{it} \le Z_{it} \quad \forall i \in I, \forall t \in T \\
& \theta_{it} B_{it} \le Z_{it} \quad \forall i \in I, \forall t \in T \\
& B_{it}, I_{it}, R_{it}, Z_{it} \ge 0 \quad \forall i \in I, \forall t \in T \\
& v_{ip\tau t}^{1},\ w_{ip\tau t}^{2} \le 0 \quad \forall i \in I, \forall p \in T, \forall \tau, t \in T \\
& v_{ip\tau t}^{2},\ w_{ip\tau t}^{1} \ge 0 \quad \forall i \in I, \forall p \in T, \forall \tau, t \in T \\
& v_{ipt}^{3}, w_{ipt}^{3} \quad (\text{urs}) \quad \forall i \in I, \forall p, t \in T,
\end{aligned}
$$

where $(\boldsymbol{w}^1, \boldsymbol{w}^2, \boldsymbol{w}^3)$ denote the dual variables when the RO paradigm is adapted in constraint (9), and Z_{it} the auxiliary variable used to reformulate the inner maximization related to the backorder cost.

The resulting robust counterpart of SRPP, denoted by RRPP, is a tractable linear programming problem that can be solved efficiently by commercial solvers. In the next section, we present the production environment used in the numerical analysis and details of the simulation study conducted.

4 NUMERICAL EXPERIMENTS

We evaluate the performance of our proposed release planning models using a multistage multi-product production system representing a scaled-down semiconductor wafer fabrication facility that has been previously studied by several authors (Kayton et al. 1997; Kacar et al. 2012). The data set reflects the major characteristics of wafer fabrication, including a re-entrant bottleneck process, batch processing machines and multiple products with varying process routings.

There are three products (Products 1, 2 and 3) flowing through the facility shown in Figure 1. Each row of nodes in the figure represents the routing of each product. Material flows from left to right, i.e. all products start the process at Station 1 and complete their processing at Station 10. Product 1 has 22 process steps including 6 visits to machine 4, which is the bottleneck station for a system producing only product 1. On the other hand, Products 2 and Product 3 have 14 process steps.

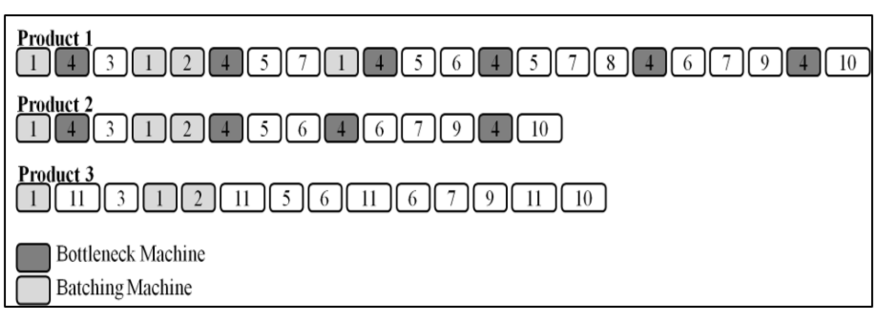

Figure 1: Product flow routes over the machines.

Table 1: Processing time distributions and batch sizes.

Machine #	Mean	Std. Dev.	Batch Size
1	80	7	4
2	220	16	4
3	45	4	1
4	40	4	1
5	25	2	1
6	22	2,4	1
7	20	2	1
8	100	12	1
9	50	4	1
10	50	5	1
11	70	2,5	1

Processing time parameters and batch sizes are listed in Table 1. All processing times follow a log-normal distribution and are given in minutes. We assume instantaneous material transfer between consecutive operations on a routing. We use backorder cost of 50, release cost of 3, inventory holding cost of 15 for all products. We use a demand scenario similar to that studied in Kacar et al. (2012) and assume a 3:1:1 product mix, i.e., 3 units of Product 1 are produces for every unit of Products 2 and 3, based on the industrial facility motivating the data set. The mean and standard deviation of the demand are $\mu_i = 60, 20, 20$ and $\sigma_i = 6, 2, 2$ for product types 1, 2 and 3 respectively. A planning horizon of 12 weeks is considered.

The numerical results in Table 2 show that as the level of uncertainty γ (i.e., the percentage of dispersion from the nominal data $[(1-\gamma)\boldsymbol{e}^0, (1+\gamma)\boldsymbol{e}^0]$) increases, the objective function value deteriorates. Namely, when we increase the uncertainty from 3% to 50%, the objective function value increases by 45% (compare the optimal function values of the RRPP when the level of uncertainty is 3%, 5%, and 10%).

Table 2: Objective function values of RRPP model under uncertainty levels 3, 5, 10, 20, and 50%.

Uncertainty	3%	5%	10%	20%	50%
Total cost	58018	59399	62746	69050	84535

The optimal objective function value of the nominal problem (RPP), i.e., 55899, outperforms that of the robust reformulation (RRPP). The difference between the two objectives is around 7% when the level of uncertainty is 3% and gets as high as 52% when the uncertainty level increases to 50%. Therefore, as the level of uncertainty increase, the objective function value becomes more conservative as anticipated. It is important to point out that a direct comparison of the robust and nominal solutions may be misleading because both approaches perform the best in their own domains, and it is known in advance that robust solution yields conservative objectives. Therefore, the nominal solution outperforms the robust solution when the nominal (or expected) data for the load factor probabilities is realized, as seen in the numerical results in Table 3. Similarly, the robust optimal solution outperforms the nominal optimal when the worst-case is realized as seen in Table 4.

Table 3: Objective function values of RPP model, when release plans from RRPP-3%, RRPP-5%, RRPP-10%, RRPP-20% and RRPP-50% are used.

Release Plan	RRPP-3%	RRPP-5%	RRPP-10%	RRPP-20%	RRPP-50%
RPP-Total cost	57838	59103	62152	67950	82258

In the remainder of the numerical results, we shall compare the worst-case and average optimality and feasibility performances of the two approaches. Note that the nominal solution may violate the capacity constraint when load factor probabilities differ from the nominal probability vector. Note that the robust optimal solution is always feasible since it is designed to be feasible even for the worst-case realization of the probabilities. Table 4 presents the percentages of infeasibility of the capacity constraint when the worst-case probabilities are realized. When the level of uncertainty is 3%, the capacity constraint is violated by 1% (on average) and 40% of the capacity constraints are violated for the nominal solution. When the level of uncertainty is 50%, the associated violation increases up to 9%. It is important to point out that to compare the optimality performances of two approaches, we define a cost for each unit of extra capacity as the dual price of the bottleneck machine. The dual price information is retrieved from the solution of the nominal model and added into the objective function with the assumption that all extra capacity requirements can be met from a third party supplier at an outsourcing cost, which is assumed to be equal to the dual price of the bottleneck machine. When this cost is included in the objective, the difference between robust and nominal solutions gets as high as 52% in favor of the robust approach, when the worst-case uncertainty for the 50% case is realized (compare, the second and the third columns in Table 4).

Table 4: Comparison of RPP and RRPP when worst-case scenarios are realized for RPP solutions.

Uncertainty	3%	5%	10%	20%	50%
RRPP-Total cost	58018	59399	62746	69050	84535
RPP-Total cost + outsourcing cost	60151	62989	70083	84271	128849
Extra capacity requirement	1%	1%	2%	3%	9%
% of violated capacity constraints	40%	40%	40%	40%	46%

We use a Monte Carlo simulation to compare the average performances of the two approaches with respect to the sampled load factor probability distributions. Probability vectors are sampled from the ambiguity set U by using Algorithm 1.

Algorithm 1: Sampling algorithm.

Inputs: the nominal probability vector $\boldsymbol{e}^0 \in [0,1]^S$,
the scaling vector $\boldsymbol{\gamma} \in [0,1]^S$, *sample size*

Output: *sample*

While # of sampled vectors $\leq$ *sample size*
  for $i = 1:1:|S|-1$
    sample $p_i \sim$ Uniform$\left[(1-\gamma_i)e_i^0, (1+\gamma_i)e_i^0\right]$
  end
  calculate $e_{|S|} = 1 - \sum_{i=1}^{|S|-1} e_i$
  If $(1-\gamma_i)e_{|S|}^0 \leq p_{|S|} \leq (1+\gamma_i)e_{|S|}^0$
    add $\boldsymbol{e}^0$ to *sample*; # of sampled vectors ++
  end
end

For each sampled distribution, we implement the nominal and robust solutions to the release planning problem and optimize the problem over B and I, and report the optimal objective function values. The average optimal objective function values of the two approaches over 100 sampled distributions are presented below in Table 5.

The numerical results show that the average objective function performance of the robust solutions are slightly better than that of the nominal solution when the level of uncertainty is lower than 10%.

Table 5: Comparison of RPP and RRPP models under 3%, 5%, 10%, 20% and 50% uncertainty levels.

Uncertainty	3%	5%	10%	20%	50%
RRPP - Avg. Total Cost	57839	59104	62160	67968	82321
RPP - Avg. Total Cost	59547	60557	66586	80128	98118

Nevertheless, the associated difference becomes more significant as the level of uncertainty increases and the associated difference gets as high as 19% when the level of uncertainty increases to 50%.

5 CONCLUSIONS

In multimodel iterative approaches, point estimates for lead times are obtained using iterations between subproblems. This paper takes an alternative approach which seeks to explicitly account for the uncertainties in the lead times induced by estimation errors using robust optimization. We compare the performance of a production planning model using a fixed exogenous set of lead times, which are obtained through a multimodel iterative approach and its robust counterpart under several experimental settings. We use a Monte Carlo simulation setting to estimate the affect of uncertainty in the lead time parameters. As expected, as the level of uncertainty in the function parameters in the robust counterpart is increased, the resulting production plan deviates from the optimal solution of the deterministic model. However, the production plans found by the robust counterpart are proven to be less vulnerable to estimation errors.

A direction for future research is to compare the realized performance of the production system under different planning models by simulating the execution of the proposed release schedules. It is quite possible that although the objective function value of the deterministic model may be better than that of the robust models, the robust model may result in significantly better cost performance. It would also be interesting to test the performance of the proposed robust model in settings, where there are parallel machines and in the cases when it is possible to produce the same product through alternative (flexible) routings. In the presence of manufacturing flexibility, the uncertainty in the load factors should not only be defined over time domain but also over alternative resources.

REFERENCES

Albey, E., and Ü. Bilge. 2011. "A Hierarchical Approach to FMS Planning and Control with Simulation-based Capacity Anticipation". *International Journal of Production Research* 49(11):3319–3342.

Albey, E., Ü. Bilge, and R. Uzsoy. 2014. "An Exploratory Study of Disaggregated Clearing Functions for Production Systems with Multiple Products". *International Journal of Production Research* 52(18):5301–5322.

Albey, E., Ü. Bilge, and R. Uzsoy. 2017. "Multi-Dimensional Clearing Functions for Aggregate Capacity Modelling in Multi-Stage Production Systems". *International Journal of Production Research* 55(14):4164–4179.

Albey, E., and R. Uzsoy. 2015. "Lead Time Modeling in Production Planning". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1996–2007. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Albey, E., İ. Yanıkoğlu, and R. Uzsoy. 2017. "Load Dependent Lead Time Modelling: A Robust Optimization Approach". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 3531–3542. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Asmundsson, J., R. L. Rardin, C. H. Turkseven, and R. Uzsoy. 2009. "Production Planning with Resources Subject to Congestion". *Naval Research Logistics* 56(2):142–157.

Asmundsson, J., R. L. Rardin, and R. Uzsoy. 2006. "Tractable Nonlinear Production Planning Models for Semiconductor Wafer Fabrication Facilities". *IEEE Transactions on Semiconductor Manufacturing* 19(1):95–111.

Bang, J.-Y., and Y.-D. Kim. 2010. "Hierarchical Production Planning for Semiconductor Wafer Fabrication Based on Linear Programming and Discrete-event Simulation". *IEEE Transactions on Automation Science and Engineering* 7(2):326–336.

Ben-Daya, M. a., and A. Raouf. 1994. "Inventory Models Involving Lead Time as a Decision Variable". *Journal of the Operational Research Society* 45(5):579–582.

Bertsimas, D., and M. Sim. 2004. "The Price of Robustness". *Operations Research* 52(1):35–53.

Bertsimas, D., and A. Thiele. 2006. "A robust Optimization Approach to Inventory Theory". *Operations Research* 54(1):150–168.

Birge, J. R., and F. Louveaux. 2011. *Introduction to Stochastic Programming*. 2nd ed. New York: Springer.
Byrne, M., and M. A. Bakir. 1999. "Production Planning Using a Hybrid Simulation - Analytical Approach". *International Journal of Production Economics* 59(1-3):305–311.
Byrne, M., and M. Hossain. 2005. "Production Planning: An Improved Hybrid Approach". *International Journal of Production Economics* 93:225–229.
Carey, M. 1992. "Nonconvexity of the Dynamic Traffic Assignment Problem". *Transportation Research B* 26B(2):127–133.
Curry, G. L., and R. M. Feldman. 2010. *Manufacturing Systems Modeling and Analysis*. 2nd ed. New York: Springer.
Gong, L., T. de Kok, and J. Ding. 1994. "Optimal Leadtimes Planning in a Serial Production System". *Management Science* 40(5):629–632.
Gorissen, B. L., İ. Yanıkoğlu, and D. den Hertog. 2015. "A Practical Guide to Robust Optimization". *Omega* 53:124–137.
Graves, S. C. 1986. "A Tactical Planning Model for a Job Shop". *Operations Research* 34(4):522–533.
Hackman, S. T., and R. C. Leachman. 1989. "A General Framework for Modeling Production". *Management Science* 35(4):478–495.
Haeussler, S., and H. Missbauer. 2014. "Empirical Validation of Meta-Models of Work Centres in Order Release Planning". *International Journal of Production Economics* 149:102–116.
Hung, Y.-F., and M.-C. Hou. 2001. "A Production Planning Approach Based on Iterations of Linear Programming Optimization and Flow Time Prediction". *Journal of the Chinese Institute of Industrial Engineers* 18(3):55–67.
Hung, Y.-F., and R. C. Leachman. 1996. "A Production Planning Methodology for Semiconductor Manufacturing Based on Iterative Simulation and Linear Programming Calculations". *IEEE Transactions on Semiconductor Manufacturing* 9(2):257–269.
Irdem, D. F., N. B. Kacar, and R. Uzsoy. 2010. "An Exploratory Analysis of Two Iterative Linear Programming Simulation Approaches for Production Planning". *IEEE Transactions on Semiconductor Manufacturing* 23(3):442–455.
Kacar, N. B., D. F. Irdem, and R. Uzsoy. 2012. "An Experimental Comparison of Production Planning Using Clearing Functions and Iterative Linear Programming-Simulation Algorithms". *IEEE Transactions on Semiconductor Manufacturing* 25(1):104–117.
Kacar, N. B., L. Mönch, and R. Uzsoy. 2016. "Modeling Cycle Times in Production Planning Models for Wafer Fabrication". *IEEE Transactions on Semiconductor Manufacturing* 29(2):153–167.
Karmarkar, U. S. 1989. "Capacity Loading and Release Planning with Work-in-Progress (WIP) and Leadtimes". *Journal of Manufacturing and Operations Management* 2:105–123.
Kayton, D., T. Teyner, C. Schwartz, and R. Uzsoy. 1997. "Focusing Maintenance Improvement Efforts in Wafer Fabrication Facility Operating Under the Theory of Constraints". *Production and Inventory Management Journal* 38(4):51.
Kim, B., and S. Kim. 2001. "Extended Model for a Hybrid Production Planning Approach". *International Journal of Production Economics* 73(2):165–173.
Milne, R. J., S. Mahapatra, and C.-T. Wang. 2015. "Optimizing Planned Lead Times for Enhancing Performance of MRP Systems". *International Journal of Production Economics* 167:220–231.
Missbauer, H., and R. Uzsoy. 2011. "Optimization Models of Production Planning Problems". In *Planning Production and Inventories in the Extended Enterprise*, edited by K. G. Kempf, P. Keskinocak, and R. Uzsoy, 437–507. New York: Springer.
Mönch, L., J. W. Fowler, and S. J. Mason. 2013. *Production Planning and Control for Semiconductor Wafer Fabrication Facilities*. 2nd ed. New York: Springer.
Pochet, Y., and L. A. Wolsey. 2006. *Production Planning by Mixed Integer Programming*. New York: Springer.
Spitter, J., C. A. Hurkens, A. De Kok, J. K. Lenstra, and E. G. Negenman. 2005. "Linear Programming Models with Planned Lead Times for Supply Chain Operations Planning". *European Journal of operational research* 163(3):706–720.
Srinivasan, A., M. Carey, T. E. Morton et al. 1988. "Resource Pricing and Aggregate Scheduling in Manufacturing systems". Technical report, Carnegie Mellon University, Tepper School of Business, Pittsburgh, Pennsylvania.
Stinstra, E., and D. Den Hertog. 2008. "Robust Optimization Using Computer Experiments". *European Journal of Operational Research* 191(3):816–837.
Tardif, V., and M. L. Spearman. 1997. "Diagnostic Scheduling in Finite Capacity Production Environments". *Computers & Industrial Engineering* 32(4):867–878.
Vollmann, T. E., W. Berry, D. Whybark, and F. Jacobs. 2005. *Manufacturing Planning and Control Systems for Supply Chain*. 2nd ed. New York: McGraw Hill, Inc.
Voß, S., and D. L. Woodruff. 2006. *Introduction to Computational Optimization Models for Production Planning in a Supply Chain*, Volume 240. New York: Springer.
Yanıkoğlu, İ., D. den Hertog, and J. P. Kleijnen. 2016. "Robust Dual-Response Optimization". *IIE Transactions* 48(3):298–312.
Ziarnetzky, T., L. Mönch, and R. Uzsoy. 2018. "Rolling Horizon, Multi-product Production Planning with Chance Constraints and Forecast Evolution for Wafer Fabs". *International Journal of Production Research* 56(18):6112–6134.

AUTHOR BIOGRAPHIES

ERİNÇ ALBEY is an assistant professor in the Industrial Engineering Department at Özyeğin University. He received his B.Sc., M.Sc. and Ph.D. degrees in Industrial Engineering from Boğaziçi University, Istanbul, Turkey. He worked as a researcher at Bogazici University Flexible Automation and Intelligent Manufacturing Laboratory and as a postdoctoral research associate in the Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University. His e-mail address is erinc.albey@ozyegin.edu.tr.

İHSAN YANIKOĞLU is an assistant professor in the Industrial Engineering Department at Özyeğin University. He received his B.Sc. and M.Sc. degrees in Industrial Engineering from Bilkent University, and his M.Phil. and Ph.D. degrees in Operations Research and Econometrics from Tilburg University. His research interests are mainly based on robust and stochastic optimization. In the recent years, his research focuses on developing practicable robust optimization methodologies and their applications. His email address is ihsan.yanikoglu@ozyegin.edu.tr.

REHA UZSOY is Clifton A. Anderson Distinguished Professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He holds B.Sc. degrees in Industrial Engineering and Mathematics and an MS in Industrial Engineering from Boğaziçi University, Istanbul, Turkey. He received his Ph.D. in Industrial and Systems Engineering in 1990 from the University of Florida. His teaching and research interests are in production planning, scheduling, and supply chain management. He was named a Fellow of the Institute of Industrial Engineers in 2005, Outstanding Young Industrial Engineer in Education in 1997, and has received awards for both undergraduate and graduate teaching. His email address is ruzsoy@ncsu.edu.