Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

STICK TO THE PLAN OR ADJUST DYNAMICALLY? COMBINING ORDER RELEASE AND OVERTIME PLANNING FOR VARYING DEMAND AND PROCESS UNCERTAINTY

Julian Fodor
Stefan Haeussler

Department of Information Systems, Production and Logistics Management
University of Innsbruck
Universitaetsstraße 15
Innsbruck, 6020, AUSTRIA

ABSTRACT

Within the area of manufacturing planning and control, there is a long ongoing debate on when and if decisions should be integrated into a centralized model or split into separate planning levels. While a centralized monolithic model is capable of solving separate decisions simultaneously, a hierarchical approach offers more degrees of freedom since a local planner always has more accurate information. The focus of this paper is on the design and mathematical assumptions of optimization models for overtime and order release decisions in order to cope with different degrees of demand and process uncertainty. We execute the optimal decisions within a simulation model of a multi-stage, multi-product stylized flow shop. Our results show that a fully centralized is outperformed by a hierarchical design and that planning order release quantities centrally, in combination with flexible overtime planning, yields the lowest costs for high process uncertainty on the shop floor.

1 PROBLEM DESCRIPTION

The manufacturing planning and control field has had a continuous discussion about whether certain decisions should be integrated into a single *centralized* decision model or separated into distinct planning levels (Gelders and Van Wassenhove 1982; de Kok and Fransoo 2003; Vogel et al. 2017; Wang et al. 2021; Ghadimi et al. 2022; Kasper et al. 2023). One argument for a centralized model is that it can handle multiple decisions simultaneously, while a *hierarchical* approach offers more flexibility since a local planner at the shop floor (i.e., the bottom level) always has more information and a better representation of the actual processes. The focus of this paper is on the design and mathematical assumptions of optimization models that make *order release* and/or *overtime* decisions in order to cope with different degree of demand and process uncertainty. While order release decides on when to release which order from an pre-shop order pool (Bertrand and Wortmann 1981; Bergamaschi et al. 1997; Haeussler et al. 2019), short-term capacity decisions (i.e, overtime) try to match resources, work centers and jobs based on the specific job's requirements (Chen et al. 2009). There are many methods of order release and the research subject itself spans a long tradition since the early 1950s (Adam and Surkis 1977; Bertrand and Wortmann 1981; Cigolini et al. 1998; Thuerer et al. 2012; Haeussler et al. 2020; Thürer et al. 2023). In this paper, we either use order release as part of an optimization model or use a simple order release rule, called backward infinite loading (BIL), where the lead time is predetermined and constant over time.

Capacity adjustment includes many research strands within production research, mainly being the worker allocation problem (Costa et al. 2019), outsourcing as a form of capacity increase (Qi 2011) and the application of overtime (Özdamar and Yazgaç 1997). Overtime can be used for various methods, be it negating all tardy orders (Shi et al. 2023), simply generating more possible capacity (Abdel-Aal 2019)

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

or minimizing costs because the scheduling of overtime is possible (Jaramillo and Erkoc 2017). Again, we either include the overtime decision in a central model (i.e., an optimization model) or, on the lower decision level, adjust the capacity by using a heuristic rule that decides on overtime using simple workload thresholds. The main goal of these two planning decisions is threefold: (i) achieve short flow times, i.e. the duration of orders from release to completion, (ii) maintain a high machine output and (iii) reach a good due-date performance of the orders. Thus, the main challenge for order release and capacity planning is to balance overtime, holding (WIP and finished goods inventory) and backorder costs.

In this paper, we compare different optimization models that either simultaneously or successively - in combination with a rule or a heuristic - decide on overtime and order release. The successive planning models consist of two planning hierarchies where we vary the decisions to be made on the central (top) planning stage: On the one hand we use an optimization model for order release, and on the other, we decide on overtime centrally. The top level decision is then combined with an overtime or order release decision on the lower planning level. The decision whether to centralize or decentralize the one or the other decision is not straightforward and largely depends on the degree of variability and uncertainty in demand and the underlying production process: While an integrated centralized model can simultaneously decide on both decisions and thus should be superior in coordinating release and capacity adjustments, it has less information and a more inaccurate representation of the shop floor. On the contrary, the hierarchical approach is more flexible to react to the changing environment in the short-term and thus should be superior for high process uncertainty. Therefore, our research question for this paper is as follows:

To what extent should overtime and order release decisions be integrated on a central planning level under different demand variability and process uncertainties?

The comparison is done by using a multi-model approach where, for the top level, the decisions of a mixed integer or linear program are executed over a certain planning horizon in a simulation model of a multi-stage, multi-product stylized flow shop. For the decentralized decisions we use either an overtime heuristic or static release, in our case backward infinite loading. In our experimental setup we test several demand patterns (i.e., constant and uncorrelated seasonal demand) and different sources of process uncertainty like machine failures and exponentially distributed processing times. Performance is measured by using cost-based measures consisting of overtime, inventory and backorder costs as well as the service level.

2 MATHEMATICAL MODELS

Our approaches combine optimization methods on a planning level with reactive short time heuristics or static rules on the shop floor to test their performance on varying levels of uncertainty. There are three models that are analyzed within this paper. The three models are similar in their foundation but have a distinct difference in what they control. Our first model is centralized and integrates all decisions and generates a completely static plan that determines release quantities as well as overtime usage (OverTime and RELease planning - OTREL). The optimal decisions are rigidly executed by the simulation. The second model determines release quantities centrally and uses an overtime heuristic to schedule overtime within the simulation (RELease planning and Heuristic OverTime - RELHOT) and the third model plans overtime centrally and release quantities by using a rule based approached with fixed lead times (i.e., BIL) inside the simulation (OverTime planning and static release with BIL - OTBIL). The central models are all represented by optimization models and all three have a similar basis. One will be explained in detail and then the adjusted logic for the other two will be presented shortly. We assume that the full day can be used for production (equal to 1440 minutes), where 960 minutes is the normal work time and an additional 120 (U_t^n), 240 (V_t^n) or 480 (Y_t^n) minutes can be gained by assigning overtime. We assume that the scheduling of overtime is always possible, but only in steps of either 120, 240 or 480 minutes. The optimization models themselves are based on the overtime model proposed by Özdamar and Yazgaç (1997) and adapted to fit with our three approaches.

2.1 Overtime Planning and Static Release with BIL (OTBIL)

The first optimization model is called OTBIL as it combines overtime planning from an optimization point of view with a static application of backward infinite loading. It optimizes the number of overtime hours in each planning period at each work center and it can map the actual release quantities by using the same lead time as BIL on the shop floor τ (see the variable $d_{j,t+\tau}$ in equation 2). The objective is to minimize the total costs. Put simple, the decision variable is the amount of overtime and the release is governed within the simulation by BIL. The assumed lead time for BIL is an experimental factor which we will vary over scenarios, but within each scenario we use the same lead times for both levels.

Table 1: Notation of our optimization models.

Indices	
n	work centers $(n = 1,...,N)$
j	products $(j = 1,...,J)$
t	periods $(t = 1,...,T)$
Variables	
$W^n_{j,t}$	WIP of product j in front of work center n at the end of period t
$F_{j,t}$	Finished Goods Inventory (FGI) of product j at the end of period t
$B_{j,t}$	Backorders of product j at the end of period t
U^n_t, V^n_t, Y^n_t	Binary variable indicating 2, 4, 8 hours of overtime
$X^n_{j,t}$	Output of product j at work center n in time period t
$R_{j,t}$	Release amount of orders j in period t
Parameters	
ω	holding costs for the WIP (per order and period)
ϕ	holding costs for finished goods (per order and period)
χ	backorder costs (per order and period)
θ	overtime costs per hour
$d_{j,t}$	demand of product j in period t
τ	periods of lead time used in BIL
ξ^n_j	operation time of product j at work center n
c^n	capacity of work center n
c^{MF}_{red}	factor for capacity reduction based on probabilities of machine failure (MF) [0...1]
c^{HOT}_{red}	factor for anticipated overtime on the shop floor [0...1]

$$min \sum_{n=1}^{N}\sum_{j=1}^{J}\sum_{t=1}^{T}(\omega \cdot W^n_{j,t}) + \sum_{j=1}^{J}\sum_{t=1}^{T}(\phi \cdot F_{j,t} + \chi \cdot B_{j,t}) + \sum_{n=1}^{N}\sum_{t=1}^{T}(2\cdot\theta\cdot U^n_t + 4\cdot\theta\cdot V^n_t + 8\cdot\theta\cdot Y^n_t) \tag{1}$$

$$W^n_{j,t} = W^n_{j,t-1} + d_{j,t+\tau} - X^n_{j,t} \quad \forall j,\ n = 1, t \in \{1\dots T-\tau\} \tag{2}$$

$$W^n_{j,t} = W^n_{j,t-1} - X^n_{j,t} \quad \forall j,\ n = 1, t \in \{T-\tau+1\dots T\} \tag{3}$$

$$W^n_{j,t} = W^n_{j,t-1} + X^{n-1}_{j,t} - X^n_{j,t} \quad \forall\ n > 1, j, t \tag{4}$$

$$F_{j,t} - B_{j,t} = F_{j,t-1} - B_{j,t-1} + X^N_{j,t} - d_{j,t} \quad \forall\ j, t, n = N \tag{5}$$

$$\sum_{j=1}^{J}(\xi_j^n \cdot X_{j,t}^n) \leq (c^n + 120 \cdot U_t^n + 240 \cdot V_t^n + 480 \cdot Y_t^n) \cdot c_{red}^{MF} \quad \forall n,t \tag{6}$$

$$U_t^n + V_t^n + Y_t^n \leq 1 \quad \forall n,t \tag{7}$$

$$W_{j,t}^n, X_{j,t}^n, F_{j,t}, B_{j,t} \geq 0 \tag{8}$$

$$U_t^n, V_t^n, Y_t^n \in \{0,1\} \tag{9}$$

The objective is to minimize the costs, consisting of work in progress (WIP), finished goods inventory (FGI) holding, backorder (BO) and overtime costs (OT), as seen in (1). The costs are calculated for each period, so if products have to wait in FGI or are multiple periods late, they will generate costs for each period that they are early/late. (2) - (5) represent the material flow within the production and (6) displays the capacity for the work centers with the possibility of overtime. c_{red}^{MF} represents the capacity reduction if machine failures are introduced into the system and becomes 0.9, i.e., a 10% reduction of capacity, in scenarios where we test machine failures. This optimization model governs overtime setting. This is done by setting the three binary Variables U_t^n, V_t^n, Y_t^n either one or zero (9), whereas the sum of all three has to be ≤ 1 for each work center in each period, ensuring that only one overtime mode is active for each work center during each period (7). Constraint (8) ensures that all variables are non negative.

Backward Infinite Loading (BIL) Backward infinite loading has several variations, but all seek to deduct a certain lead time from each order's due date to determine its' release date. If the release date is on or before the current date, the job is released to the shop regardless of current shop loads. Otherwise, the job stays in the pre-shop pool until the aforementioned is true. There are multiple BIL calculation techniques, where many utilize the following method to calculate the release date according to Wisner (1995), which we adapted as can be seen in the following Algorithm 1.

Algorithm 1 Backward Infinite Loading.

```
if due_date - τ ≥ current_date then
    do nothing                              ▷ Product stays in waiting
else if due_date - τ < current_date then
    release product immediately             ▷ Product starts production independent of shop load
end if
```

BIL, in our case, therefore checks if there is at least τ periods left until the products' due date, and once this threshold is undercut, the product is released. As mentioned above, BIL uses the same lead time τ as fed into the optimization model. Therefore the optimization model has the perfect anticipation of release quantities. Please note that this model is the only model that uses fixed lead times for order release, as the other models have no fixed lead time and calculate release times individually for each order.

2.2 Release Planning and Heuristic Overtime (RELHOT)

This optimization model governs release decision and uses a simple heuristic to schedule overtime on the shop floor. The optimization model needs to be adjusted as follows: Constraint (2) has to be replaced by a constraint that allows the model to govern releases through $R_{j,t}$ and remove overtime application (see constraint 10). Described in short, this model plans release times and quantities through $R_{j,t}$ and relies on the overtime heuristic (HOT), described below, to schedule overtime during execution of the plan in the simulation. The mathematical model for this optimization method is given by the following equations:

$$min \sum_{n=1}^{N}\sum_{j=1}^{J}\sum_{t=1}^{T}(\omega \cdot W_{j,t}^n) + \sum_{j=1}^{J}\sum_{t=1}^{T}(\phi \cdot F_{j,t} + \chi \cdot B_{j,t})$$

$$W^n_{j,t} = W^n_{j,t-1} + R_{j,t} - X^n_{j,t} \quad \forall j,t,\ n = 1 \tag{10}$$

$$W^n_{j,t} = W^n_{j,t-1} + X^{n-1}_{j,t} - X^n_{j,t} \quad \forall\ n > 1, j, t$$

$$F_{j,t} - B_{j,t} = F_{j,t-1} - B_{j,t-1} + X^N_{j,t} - d_{j,t} \quad \forall n = N, j, t$$

$$\sum_{j=1}^{J} (\xi^n_j \cdot X^n_{j,t}) \leq c^n \cdot c^{MF}_{red} \cdot c^{HOT}_{red} \quad \forall n, t$$

$$W^n_{j,t}, X^n_{j,t}, F_{j,t}, B_{j,t} \geq 0$$

c^{HOT}_{red} determines the anticipated amount of overtime the heuristic plans during execution. Dependent on the scenario and initial test runs, this reduction is between 2-20% and changes depending on the scenario. This is introduced to link the model and the overtime heuristic and is based on preliminary simulation runs.

Heuristic Overtime (HOT) This model employs a simple heuristic for overtime usage on the shop floor. The heuristic takes the load in the queue of each work center into consideration and if there is enough load for a certain mode of overtime, it applies it. The 1.125, 1.25 and 1.5 multipliers are the factors of load that guarantee that there is at least 16 + 2, 16 + 4 or 16 + 8 hours of work available. Algorithm 2 describes the simple heuristic in detail.

Algorithm 2 Simple Overtime Heuristic.

```
if work center_load < period_length then
    Do nothing
else if work center_load > period_length·1.5 then
    Apply eight hours of overtime
else if work center_load > period_length·1.25 then
    Apply four hours of overtime
else if work center_load > period_length·1.125 then
    Apply two hours of overtime
end if
```

2.3 Overtime and Release Planning (OTREL)

This model is purely static and reliant on the plan, there are no adjustments to be made during execution. Therefore it needs to be able to calculate both overtime usages and release. It is a combination of both previous ones using only the optimization part and not the reactive shop floor part. Described in short, this model plans both release and overtime in a plan and that plan is then executed by the simulation without any changes. For the model, see Section 2.1 and instead of equation (2) we use equation (10) repeated here for convenience:

$$W^n_{j,t} = W^n_{j,t-1} + R_{j,t} - X^n_{j,t} \quad \forall j,\ n = 1, t$$

3 SIMULATION MODEL AND EXPERIMENTAL DESIGN

The simulation was created using Python 3.11 with the simpy 4.0.1 library for the simulation and the Pyomo 6.5.0 library for the mathematical optimization environment using the gurobi solver. Each iteration starts with the optimization model calculating a plan, and depending on which model is tested, feeding different information into the simulation that it has to abide by. After the simulation is done, costs as well as service level information gets collected and then averaged over the amount of iterations. The simulation model calculates costs based on each period, for each product that has a certain state of either work in progress (WIP), waiting in finished goods inventory (FGI) or being late (BO) it applies a cost of 1$, 4$ or 16$ respectively per period. If overtime is scheduled, each 60 minutes of overtime cost 2$, so a maximum of 16$ cost per work center per period is possible if scheduling eight hours of overtime. The costs were chosen this way as there is an increase in value once there are finished products or late (Haeussler et al. 2022). We use a processing time of 160 minutes per step and a total period length of 960 minutes (expandable to a maximum of 1440 with eight hours of overtime). The simulation also uses two forms of demand generation and machine failures which are shortly described in their own paragraphs below. As we calculate plans (with the optimization models) and execute them on the shop floor, there is no "warm-up" period, although we added eight periods of no demand to the beginning and end of each plan (eight being the used due date slack used within the simulation), to avoid end-of-horizon effects, guarantee that the optimization models have full information and can schedule products everywhere within their planning horizon.

3.1 Stylized Flow Shop Design

Nowadays stylized flow shop is a popular manufacturing system that embraces different industrial areas including electronics (Aurich et al. 2016), metal industry (Semini et al. 2006), semiconductor manufacturing (Lin and Chen 2015) as well as many more. For this paper, the stylized flow shop design depicted in Figure 1 has been chosen because of its applicability to many industries. This makes it possible to have simple as well as very complex scenarios within one system if paired with all the stochastic influences described previously. The stylized flow shop has three stages with two work centers (WC) each, and each possible combination of work centers describes one product type, resulting in a total of eight different products (abbreviated as P1 to P8 in Figure 1). The processing times are equal for all work centers.The results were obtained by simulating various scenarios with (i) three different combinations of optimization and flexibility on the shop floor, (ii) different processing time distributions (i.e., exponentially (exp) and uniformly (uniform) distributed) , (iii) machine failures on all or no work centers, (iv) seasonal or constant demand levels and (v) three levels of machine utilization based on a shop floor without any overtime and machine failures, this results in 72 total scenarios per planning horizon. We also tested a planning horizon of 52, 26 and 13 periods.

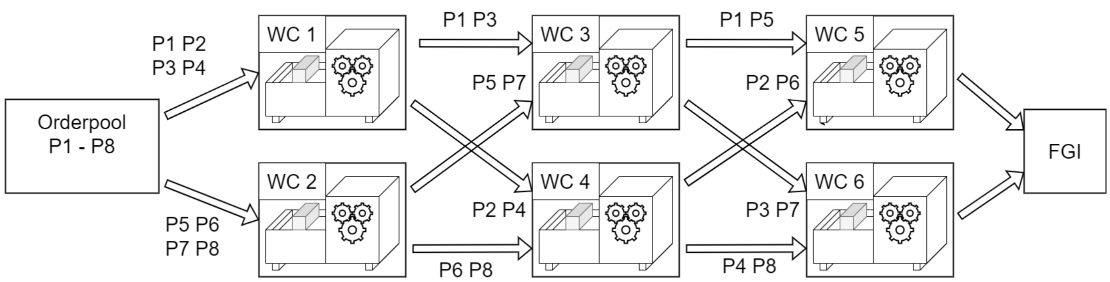

Figure 1: Our stylized flow shop design.

For OPT_OT_BIL, lead times (τ) from two to eight are tested and will be denoted as BIL2 to BIL8.

For each result, an optimization gap of 0.1% was allowed for the optimization models and the number of iterations per scenario was determined to be 30 by using Welch's method (Welch 1967). We applied Welch's Method to a basis of 50 iterations and the amount of deviation from the result was allowed to be within 2.5% which resulted in 30 iterations. All the displayed results are tested to be significant using Wilcoxon's significance test (Wilcoxon 1945), the significance tests are also available in our Data Repository. The experimental design can be seen in Table 2.

Table 2: Experimental design.

	Parameter	**Values**
Planning Characteristics	Optimization Models	OTREL, OTBIL and RELHOT
	Planning Horizon	52, 26 and 13 periods
	Lead time Variation	only for OTBIL (τ between two and eight)
Shop Characteristics	Demand	Constant or Seasonal (two Sine with a oscillation of 13 periods that have been shifted by half their oscillation to average out the same as constant demand)
	Machine Failures	Either all or no work centers, gamma distributed mean time to failure = 2760 minutes with σ = 26, mean time to repair = 345 minutes with σ= 7
Job Characteristics	Processing Times	160 minutes for each stage (exp or uniform distributed)
	Inter Arrival Times	79.3 (85%), 85.3 (80%), 91.8 (75%) minutes for 75% and 80% utilization respectively; product types uniformly distributed
	Due Date Setting	Fixed to entry period + eight periods

3.2 Demand Generation

For our experiments, we compare a static (Constant) to a seasonal demand (Sine) with a cycle length of 13 periods (as depicted in Figure 2). Both of these demands are calculated based on the inter arrival time of products, where for the constant demand, the amount of products is randomly distributed among all product types, resulting in a uniform distribution over the planning horizon (Puergstaller and Missbauer 2012). For the seasonal demand, we follow the procedure given by Enns and Suwanruji (2004). We generate two sine curves that are offset by a half season and assign products that are processed on either work center three (WC 3) or four (WC 4), so there is a bottleneck at the second production stage, which shifts over time between these two work centers.

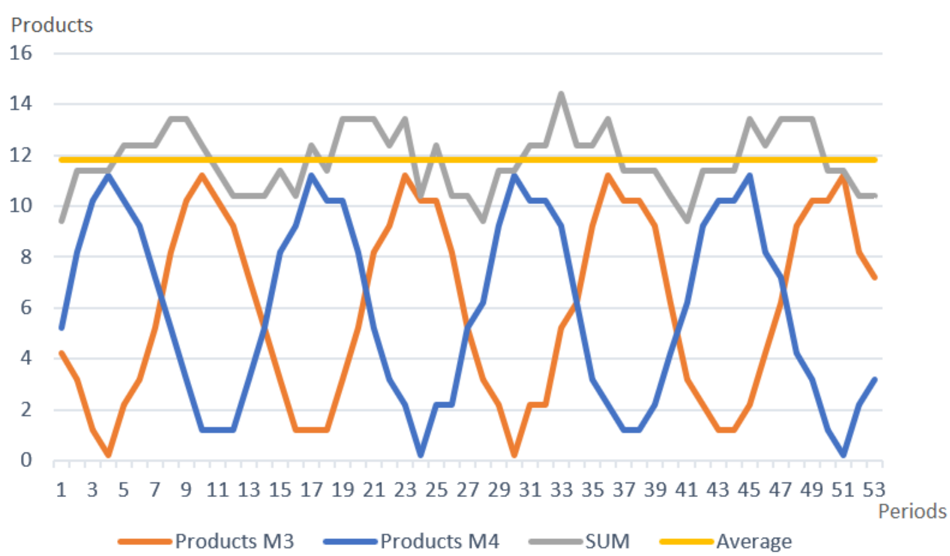

Figure 2: Example sine demand for inter arrival time = 79.3 for 53 periods.

3.3 Machine Failures

Machine failures have been included as another source of process uncertainty. Machine failures and their repair times are based on a certain distribution given in Table 2 which are taken from earlier literature (Kacar and Uzsoy 2010). Further, we assume that either no or all machines break down which results in a loss of 10% capacity for each work center.

4 RESULTS

For brevity, we only present the results for the best performing scenarios in terms of total cost measures but we provide the results for all scenarios in our Data Repository. The presented results are based on a planning horizon of 13 periods and can be seen in Table 3.We structured the results section based on the model categories where we start with OTREL, thereafter discuss the performance of OTBIL and end this section with RELHOT.

4.1 Centralized Monolithic Model: OTREL

The centralized monolithic model OTREL should have the advantage of planning overtime and order release quantities simultaneously. Thus, on the one hand it should be best in scenarios with seasonal demand since it is known by the model (i.e., deterministic) and, on the other, it should get worse when process uncertainty increases. This means that we conjecture OTREL to perform best in cases of seasonal demand and low process uncertainty which means in cases of no machine failures and uniform processing times. Therefore, the following Figures 3 and 4 show all scenarios with seasonal demand and varying process uncertainty. This means that we do not depict scenarios with high process uncertainty (i.e, scenarios with machine failures and exponential processing times) which will be discussed below.

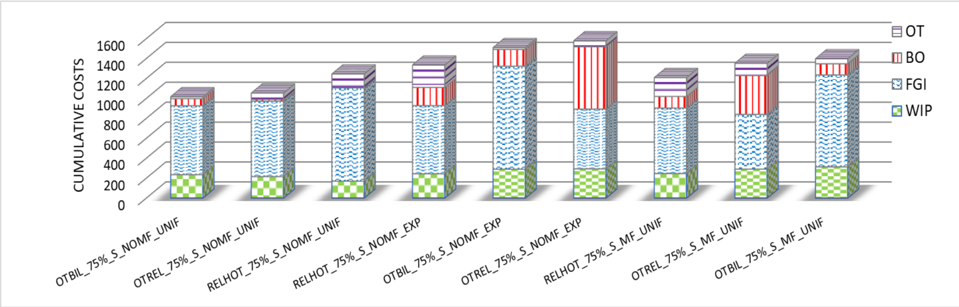

Figure 3: Results for seasonal demand and 75% machine utilization with varying process uncertainty.

Surprisingly, OTREL does not perform best in any of the cases with 75% utilization and is even slightly outperformed (3.66% more total costs) by OTBIL in the scenario with the lowest process uncertainty (*75%_S_NOMF_UNIF*).

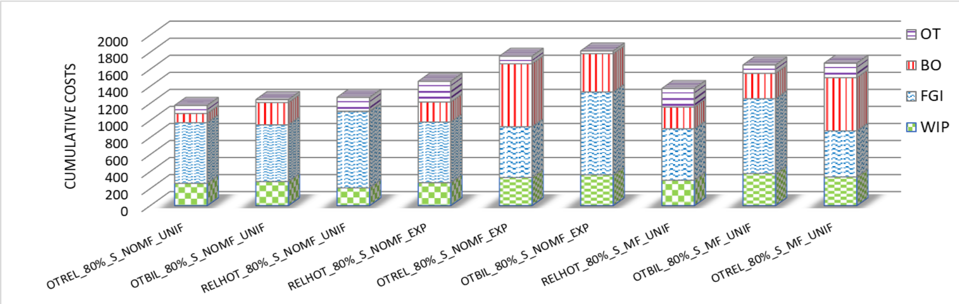

Figure 4: Results for seasonal demand and 80% machine utilization with varying process uncertainty.

As can be seen in Figure 4, for 80% utilization OTREL only performs best in case of low process utilization (*80%_S_NOMF_UNIF*), but cannot decrease costs in case of machine failures or exponential processing times. For the four remaining scenarios, RELHOT is always the best performing. This means that as soon as seasonal demand is paired with some process uncertainty it is beneficial to plan order releases centrally and schedule overtime more flexibly (on the lower level).

4.2 Centralized Overtime and Static Order Release: OTBIL

Different to above, we would expect the combination of centralized overtime planning and static order release (OTBIL) to perform best when the demand is constant. Thus, we show all scenarios with constant demand in Figures 5 and 6 except the case with the high process uncertainty.

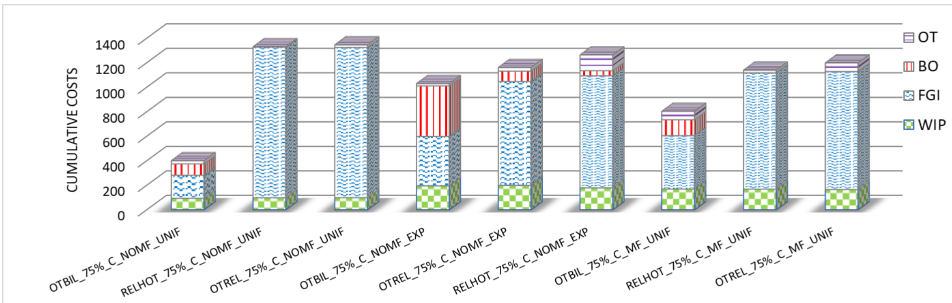

Figure 5: Results for constant demand and 75% machine utilization with varying process uncertainty.

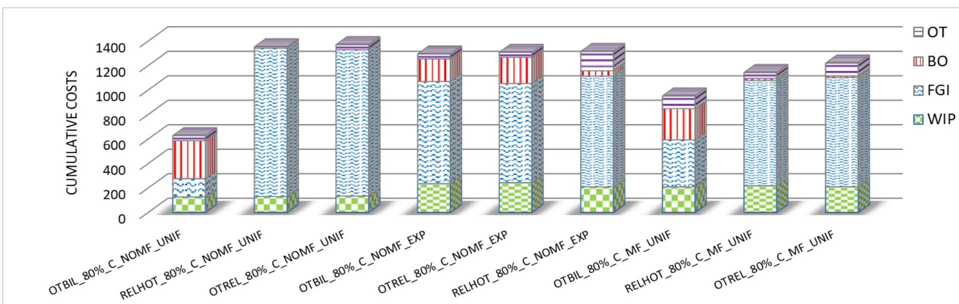

Figure 6: Results for constant demand and 80% machine utilization with varying process uncertainty.

As expected, OTBIL always yields the lowest costs in all of the depicted cases of constant demand except in one case (*80%_C_NOMF_EXP*) where there is no significant difference between the three tested models. The major advantage of OTBIL in these cases is the balancing of FGI and backorder costs which is also reflected in a relatively high service level of always more than 85% (see Table 3).

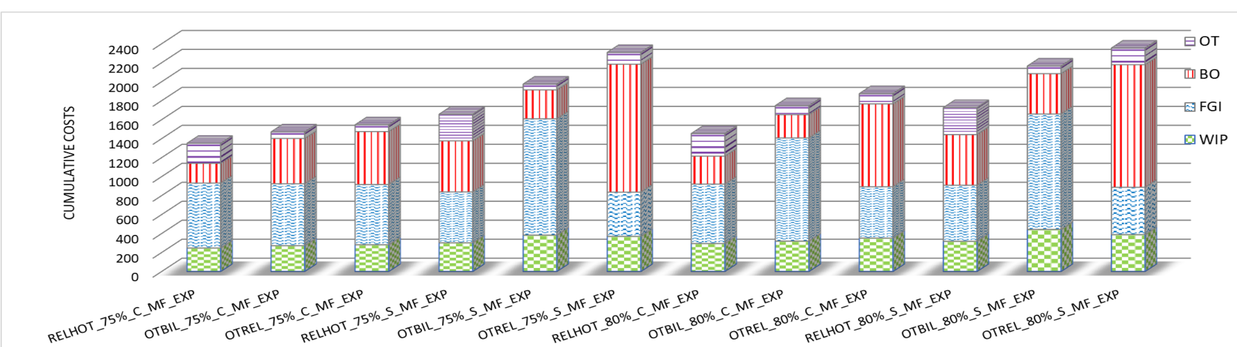

Figure 7: Results for both demand patterns, 75% and 80% machine utilization with high process uncertainty.

Table 3: All averaged results for 75% and 80% utilization with a planning horizon of 13 periods. Only significantly different scenarios are shown.

Math Model	UTIL	Demand Pattern	Machine Failures	Processing Time Dis.	BIL	WIP	FGI	BO	OT	SUM	Service Level	% to Best
OTBIL	75%	CON	NOMF	Uniform	BIL2	93.87	187.73	92.80	25.47	399.87	0.96	100.00%
RELHOT	75%	CON	NOMF	Uniform		97.83	1228.67	0.00	3.33	1329.83	1.00	332.57%
OTREL	75%	CON	NOMF	Uniform		98.50	1226.00	0.00	21.07	1345.57	1.00	336.50%
OTBIL	75%	SINE	NOMF	Uniform	BIL4	233.63	691.33	71.47	18.27	1014.70	0.97	100.00%
OTREL	75%	SINE	NOMF	Uniform		213.93	756.67	17.60	63.60	1051.80	0.99	103.66%
RELHOT	75%	SINE	NOMF	Uniform		167.80	936.80	0.00	138.67	1243.27	1.00	122.53%
OTBIL	75%	CON	NOMF	Exp	BIL3	195.00	403.20	412.80	18.13	1029.13	0.85	100.00%
OTREL	75%	CON	NOMF	Exp		198.23	848.13	84.27	29.60	1160.23	0.97	112.74%
RELHOT	75%	CON	NOMF	Exp		178.90	914.67	41.07	128.40	1263.03	0.98	122.73%
RELHOT	75%	SINE	NOMF	Exp		242.77	682.80	183.47	223.87	1332.90	0.93	100.00%
OTBIL	75%	SINE	NOMF	Exp	BIL5	287.83	1034.67	168.00	17.33	1507.83	0.94	113.12%
OTREL	75%	SINE	NOMF	Exp		291.17	599.73	625.60	59.33	1575.83	0.80	118.23%
OTBIL	75%	CON	ALLMF	Uniform	BIL3	169.23	435.47	129.60	68.67	802.97	0.94	100.00%
RELHOT	75%	CON	ALLMF	Uniform		168.23	947.60	2.13	21.07	1139.03	1.00	141.85%
OTRE	75%	CON	ALLMF	Uniform		164.67	961.87	2.13	69.60	1198.27	1.00	149.23%
RELHOT	75%	SINE	ALLMF	Uniform		244.87	657.47	115.73	189.33	1207.40	0.95	100.00%
OTREL	75%	SINE	ALLMF	Uniform		288.83	550.00	389.33	122.40	1350.57	0.83	111.86%
OTBIL	75%	SINE	ALLMF	Uniform	BIL5	312.40	922.13	110.93	50.67	1396.13	0.95	115.63%
RELHOT	75%	CON	ALLMF	Exp		247.07	683.73	208.00	210.67	1349.47	0.92	100.00%
OTBIL	75%	CON	ALLMF	Exp	BIL4	271.97	651.47	477.33	68.80	1469.57	0.86	108.90%
OTREL	75%	CON	ALLMF	Exp		281.07	635.07	557.33	71.07	1544.53	0.84	114.46%
RELHOT	75%	SINE	ALLMF	Exp		301.80	535.33	538.13	280.53	1655.80	0.81	100.00%
OTBIL	75%	SINE	ALLMF	Exp	BIL6	382.63	1225.60	304.53	60.80	1973.57	0.90	119.19%
OTREL	75%	SINE	ALLMF	Exp		369.90	465.47	1348.27	123.20	2306.83	0.66	139.32%
OTBIL	80%	CON	NOMF	Uniform	BIL2	126.03	149.33	309.87	41.07	626.30	0.87	100.00%
RELHOT	80%	CON	NOMF	Uniform		129.10	1211.60	0.00	10.67	1351.37	1.00	215.77%
OTREL	80%	CON	NOMF	Uniform		133.27	1194.93	0.00	40.80	1369.00	1.00	218.59%
OTREL	80%	SINE	NOMF	Uniform		260.77	712.00	108.27	94.27	1175.30	0.95	100.00%
OTBIL	80%	SINE	NOMF	Uniform	BIL4	281.73	666.27	260.80	34.53	1243.33	0.89	105.79%
RELHOT	80%	SINE	NOMF	Uniform		208.13	896.40	3.73	161.07	1269.33	1.00	108.00%
RELHOT	80%	SINE	NOMF	Exp		268.70	711.87	234.67	242.13	1457.37	0.91	100.00%
OTREL	80%	SINE	NOMF	Exp		328.93	596.67	737.60	88.93	1752.13	0.79	120.23%
OTBIL	80%	SINE	NOMF	Exp	BIL5	360.97	972.40	449.07	31.47	1813.90	0.87	124.46%
OTBIL	80%	CON	ALLMF	Uniform	BIL3	207.60	385.20	254.40	102.53	949.73	0.90	100.00%
RELHOT	80%	CON	ALLMF	Uniform		218.13	859.60	16.53	47.07	1141.33	0.99	120.17%
OTREL	80%	CON	ALLMF	Uniform		209.90	891.60	12.80	104.27	1218.57	0.99	128.31%
RELHOT	80%	SINE	ALLMF	Uniform		296.53	605.33	253.87	217.47	1373.20	0.90	100.00%
OTBIL	80%	SINE	ALLMF	Uniform	BIL5	374.53	880.13	297.07	98.13	1649.87	0.88	120.15%
OTREL	80%	SINE	ALLMF	Uniform		335.47	541.87	622.93	170.53	1670.80	0.77	121.67%
RELHOT	80%	CON	ALLMF	Exp		294.03	625.60	294.93	237.60	1452.17	0.89	100.00%
OTBIL	80%	CON	ALLMF	Exp	BIL5	319.60	1086.93	245.33	94.27	1746.13	0.93	120.24%
OTREL	80%	CON	ALLMF	Exp		351.30	541.07	873.07	107.33	1872.77	0.75	128.96%
RELHOT	80%	SINE	ALLMF	Exp		317.93	589.73	533.87	284.13	1725.67	0.82	100.00%
OTBIL	80%	SINE	ALLMF	Exp	BIL6	442.43	1216.53	425.07	80.00	2164.03	0.88	125.40%
OTREL	80%	SINE	ALLMF	Exp		388.13	498.40	1291.73	174.93	2353.20	0.68	136.36%

4.3 Centralized Release Planning and Flexible Overtime Planning: RELHOT

Finally, the model that plans order release quantities and flexible schedules overtime on the shop floor (RELHOT) should perform best in cases of high process uncertainty independent of demand variability. Figure 7 shows that indeed RELHOT yields the lowest total costs for all scenarios with high process uncertainty. This algorithm benefits from the central release plan by yielding the lowest WIP levels and yielding the best timing performance by balancing FGI and BO costs inspite of the highest overtime costs.

5 DISCUSSION AND CONCLUSION

In this paper we compare three different designs of production planning architecture for overtime and order release planning. We analyze how different optimization models representing a central planning level and a corresponding lower level (a rule or a heurisitc) can cope with different degrees of demand and process uncertainties. The aim of this article is to determine, under different demand variability and process uncertainties, to what extent overtime and order release decisions should be integrated at a central planning entity. This means that, different to earlier literature we do not only compare monolithic and hierarchical models (Vogel et al. 2017; Ghadimi et al. 2022), but we empirically compare different designs of planning architectures that either explicitly plan overtime and/or order release quantities over a certain planning horizon or use simplifying assumptions for planning. Therefore, we execute optimal decisions within a simulation model of a multi-stage, multi-product stylized flow shop. Our results show that (i) our fully centralized planning model is not even a viable alternative for low process uncertainty levels, (ii) that centrally planned overtime paired with a backward infinite loading (BIL) release rule performs best for constant demand and low processing time uncertainty (i.e, uniformly distributed) and (iii) in cases of machine failures and exponentially distributed processing times the combination of centrally planning releases with a flexible overtime planning heuristic yields the lowest costs. This shows that, in our simulated case, a hierarchical planning architectures is superior to an integrated one and that it is not straightforward to decide which decision should be taken centrally or decentrally. Overall, our findings are in line with theory, with the only exception that our central monolithic model is always outperformed by other, more flexible, architectures which highlights the importance of flexibility in production planning to react to the uncertain manufacturing environment (Haeussler et al. 2019; Schneckenreither et al. 2022). Our results are of course only limited to the simulated case and the validity of the results for job shops must be assessed in future studies. Furthermore, it would be interesting to test other planning architectures including more sophisticated optimization models (Hackman and Leachman 1989; Kacar and Uzsoy 2010; Haeussler et al. 2020) or more advanced order release or overtime approaches (Thuerer et al. 2012; Kacar and Uzsoy 2010; Haeussler et al. 2019; Schneckenreither et al. 2021).

REFERENCES

Abdel-Aal, M. A. 2019. "A Robust Capacitated Lot Sizing Problem with Setup Times and Overtime Decisions with Backordering Allowed under Demand Uncertainty". *IFAC-PapersOnLine* 52(13):589–594.

Adam, N., and J. Surkis. 1977. "Note—A Comparison of Capacity Planning Techniques in a Job Shop Control System". *Management Science* 23(9):1011–1015.

Aurich, P., A. Nahhas, T. Reggelin, and J. Tolujew. 2016. "Simulation-based Optimization for Solving a Hybrid Flow Shop Scheduling Problem". In *Proceedings of the 2016 Winter Simulation Conference*, edited by Theresa M.K. Roeder, Peter I. Frazier, Robert Szechtman, Enlu Zhou, Todd Huschka, and Stephen E. Chick, 2809–2819. Piscataway, New Jersey.: Institute of Electrical and Electronics Engineers, Inc.

Bergamaschi, D., R. Cigolini, M. Perona, and A. Portioli. 1997. "Order Review and Release Strategies in a Job Shop Environment: A Review and a Classification". *International Journal of Production Research* 35(2):399–420.

Bertrand, J. W. M., and J. C. Wortmann. 1981. *Production Control and Information Systems for Component Manufacturing Shops*, Volume 1. New York: Elsevier Science Inc.

Chen, C.-S., S. Mestry, P. Damodaran, and C. Wang. 2009. "The Capacity Planning Problem in Make-to-Order Enterprises". *Mathematical and computer modelling* 50(9-10):1461–1473.

Cigolini, R., M. Perona, and A. Portioli. 1998. "Comparison of Order Review and Release Techniques in a Dynamic and Uncertain Job Shop Environment". *International Journal of Production Research* 36(11):2931–2951.

Costa, F., A. Portioli-Staudacher, D. Nisi, and M. Rossini. 2019. "Integration of Order Review and Release and Output Control with Worker's Allocation in a Pure Flow Shop". *IFAC PapersOnLine* 52(13):2632–2637.

de Kok, T. G., and J. C. Fransoo. 2003. *Planning Supply Chain Operations: Definition and Comparison of Planning Concepts*, Volume 11, Book section 12, 597–675. Amsterdam, The Netherlands: North-Holland Publishing Company.

Enns, S. T., and P. Suwanruji. 2004. "Work Load Responsive Adjustment of Planned Lead Times". *Journal of Manufacturing Technology Management* 15(1):90–100.

Gelders, L. F., and L. N. Van Wassenhove. 1982. "Hierarchical Integration in Production Planning: Theory and Practice". *Journal of Operations Management* 3(1):27–35.

Ghadimi, F., T. Aouam, S. Haeussler, and R. Uzsoy. 2022. "Integrated and Hierarchical Systems for Coordinating Order Acceptance and Release Planning". *European Journal of Operational Research* 67(5):1277–1289.

Hackman, S. T., and R. C. Leachman. 1989. "A General Framework for Modeling Production". *Management Science* 35(4):478–495.

Haeussler, S., P. Neuner, and M. Thürer. 2022. “Balancing Earliness and Tardiness within Workload Control Order Release: An Assessment by Simulation”. *Flexible Services and Manufacturing Journal* 35(2):1–22.

Haeussler, S., M. Schneckenreither, and C. Gerhold. 2019. “Adaptive Order Release Planning with Dynamic Lead Times”. *IFAC-PapersOnLine* 52(13):1890–1895.

Haeussler, S., C. Stampfer, and H. Missbauer. 2020. “Comparison of Two Optimization Based Order Release Models with Fixed and Variable Lead Times”. *International Journal of Production Economics* 227:107682.

Jaramillo, F., and M. Erkoc. 2017. “Minimizing Total Weighted Tardiness and Overtime Costs for Single Machine Preemptive Scheduling”. *Computers & Industrial Engineering* 107:109–119.

Kacar, N. B., and R. Uzsoy. 2010. “Estimating Clearing Functions from Simulation Data”. In *Proceedings of the 2010 Winter Simulation Conference*, edited by Björn Johansson, Sanjay Jain, Jairo Montoya–Torres and Enver Yücesan, 1699–1710. Piscataway, New Jersey.: Institute of Electrical and Electronics Engineers, Inc.

Kasper, T. A., M. J. Land, and R. H. Teunter. 2023. “Non-Hierarchical Work-in-Progress Control in Manufacturing”. *International Journal of Production Economics* 257:108768.

Lin, J. T., and C.-M. Chen. 2015. “Simulation Optimization Approach for Hybrid Flow Shop Scheduling Problem in Semiconductor Back-end Manufacturing”. *Simulation Modelling Practice and Theory* 51:100–114.

Puergstaller, P., and H. Missbauer. 2012. “Rule-based vs. Optimisation-based Order Release in Workload Control: A Simulation Study of a MTO Manufacturer”. *International Journal of Production Economics* 140(2):670–680.

Qi, X. 2011. “Outsourcing and Production Scheduling for a Two-Stage Flow Shop”. *International Journal of Production Economics* 129(1):43–50.

Schneckenreither, M., S. Haeussler, and J. Peiró. 2022. “Average Reward Adjusted Deep Reinforcement Learning for Order Release Planning in Manufacturing”. *Knowledge-Based Systems* 247:108765.

Schneckenreither, M., S. Windmueller, and S. Haeussler. 2021. “Smart Short Term Capacity Planning: A Reinforcement Learning Approach”. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems: IFIP International Conference, APMS 2021, Nantes, France, Proceedings, Part I*, 258–266. Springer.

Semini, M., H. Fauske, and J. O. Strandhagen. 2006. “Applications of Discrete-Event Simulation to Support Manufacturing Logistics Decision-making: A Survey”. In *Proceedings of the 2006 Winter Simulation Conference*, edited by Björn Johansson, Sanjay Jain, Jairo Montoya–Torres, and Enver Yücesan, 1946–1953. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Shi, S., H. Xiong, and G. Li. 2023. “A No-Tardiness Job Shop Scheduling Problem with Overtime Consideration and the Solution Approaches”. *Computers and Industrial Engineering* 178:109115.

Thuerer, M., M. Stevenson, C. Silva, M. J. Land, and L. D. Fredendall. 2012. “Workload Control and Order Release: A Lean Solution for Make-to-Order Companies”. *Production and Operations Management* 21(5):939–953.

Thürer, M., N. O. Fernandes, S. Haeussler, and M. Stevenson. 2023. “Dynamic Planned Lead Times in Production Planning and Control Systems: Does The Lead Time Syndrome Matter?”. *International Journal of Production Research* 61(4):1268–1282.

Vogel, T., B. Almada-Lobo, and C. Almeder. 2017. “Integrated Versus Hierarchical Approach to Aggregate Production Planning and Master Production Scheduling”. *OR Spectrum* 39:193–229.

Wang, K., S. Chen, Z. Jiang, W. Zhou, and N. Geng. 2021. “Capacity Allocation of an Integrated Production and Service System”. *Production and Operations Management* 30(8):2765–2781.

Welch, P. 1967. “The Use of Fast Fourier Transform for The Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms”. *IEEE Transactions on Audio and Electroacoustics* 15(2):70–73.

Wilcoxon, F. 1945. “Individual Comparisons by Ranking Methods”. *Biometris Bulletin* 1(6):80–83.

Wisner, J. D. 1995. “A Review of the Order Release Policy Research”. *International Journal of Operations & Production Management* 15(6):25–40.

Özdamar, L., and T. Yazgaç. 1997. “Capacity Driven Due Date Settings in Make-to-order Production Systems”. *International Journal of Production Economics* 49(1):29–44.

AUTHOR BIOGRAPHIES

JULIAN FODOR is currently research assistant and PhD candidate at the Department of Information Systems, Production and Logistics Management at the University of Innsbruck. He received his M.Sc. degree in Information Systems in 2022 and methodologically, he focuses on discrete event simulation and optimization. His email adress is julian.fodor@uibk.ac.at.

STEFAN HAEUSSLER is currently Associate Professor at the Department of Information Systems, Production and Logistics Management at the University of Innsbruck. His main research focus is on order release, lead time management, dispatching and their practical application. Methodologically, he focuses on discrete event simulation, optimization, economic experiments and machine learning methods. His email adress is stefan.haeussler@uibk.ac.at.