An application of master schedule smoothing and planned lead time control

Graves, Stephen C.; Teo, Chee-Chong; Bhatnagar, Rohit

2012

Teo, C.-C., Bhatnagar, R., & Graves, S. C. (2012). An application of master schedule smoothing and planned lead time control. *Production and Operations Management*, 21(2), 211-223. doi:10.1111/j.1937-5956.2011.01263.x

<https://hdl.handle.net/10356/98150>

<https://doi.org/10.1111/j.1937-5956.2011.01263.x>

© 2012 John Wiley & Sons, Inc. All rights reserved. This paper was published in *Production and Operations Management* and is made available with permission of John Wiley & Sons, Inc.

Downloaded on 26 Apr 2023 21:18:28 SGT

An Application of Master Schedule Smoothing and Planned Lead-times Control

Chee-Chong Teo, Rohit Bhatnagar

Nanyang Technological University, Singapore.

Stephen C. Graves

Massachusetts Institute of Technology, USA.

Abstract

When faced with a surge in demand and an increase in demand variability, a producer of customized offshore oil-rigs had to rely heavily on subcontracting to handle production requirements that were in excess of its capacity. A team of university researchers together with the company's production personnel applied analytical methods to reduce the subcontracting cost. The team focused on the manufacture of steel panels, which are the main sub-assemblies for building the hull of an oil-rig. The team considered the two key tactical parameters for planning production in this system: the planning window of the master production schedule and the planned lead-time of each workstation. The improvement efforts include the creation of a master schedule for each product family of the steel panels, and the smoothing of the master schedule over its planning window. Under the constraint of a fixed internal delivery lead-time, the team determined both the optimal planning windows and the planned lead-times at each station. We report our experience in implementing the analytical model, the managerial insights gained and how the application benefits the oil-rig producer.

Keywords: production smoothing, master production schedule, planned lead-times, oil-rig building.

1. Introduction

We present a case study in which we improve production planning for a producer of customized offshore oil-rigs. At the time of the application, the company had experienced a sudden surge in demand due to the global increase in demand for energy. However, given the cyclical nature of the industry, the firm was hesitant to invest heavily in expanding its capacity because it could result in over-capacity during the trough of the cycle. The firm's strategy has been to use subcontracting to handle upswings in its production requirements, and to use production smoothing to minimize subcontracting costs. This project was part of an initiative to improve the company's production planning. In this paper, we report our experience in improving the production planning in the oil-rig building facility.

In this application, we consider the two key planning parameters for smoothing production in the oil-rig production environment, namely the planned lead-time (PLT) and the planning window. The PLT is a planning parameter that typically projects how much time each job spends at each production step, workstation or facility; a short PLT leads to a more variable workload at the workstation, while a long PLT results in more work-in-process (WIP) at the station. The planning window is the time difference between the delivery lead-time and the total planned lead-time for a job. A longer planning window allows more smoothing of the master production schedule, which results in a less variable work release. We apply our model in TBG (2009), which considers a multi-product, multi-station make-to-order (MTO) production system; the model determines the optimal PLTs and planning windows to minimize subcontracting cost.

This paper is organized as follows. In the next section, we describe the process flow and the problems in production planning. In section 3, we describe the initial study and the plans for improvement. In section 4, we present the optimization model, describe how we validate the

underlying model assumptions for the application and explain how we parameterize the model inputs. In section 5, we explain the recommendations and managerial insights based on the optimization solutions. Subsequently, we validate the model output in section 6. We conclude in section 7 with a discussion on how the application had influenced the company.

2. Process Flow and Planning Challenges

The company is a major producer of customized jack-up oil-rigs. Jack-up oil-rigs are offshore rigs that are mobile in water, and can anchor themselves by deploying jack-like legs. The fundamental structure of a jack-up oil-rig is the hull, which is the platform on which most facilities of the rig are built. A typical hull consists of steel blocks; each steel block is constructed by joining steel structures called panels. A block commonly consists of ten to more than sixty panels. The steel panel represents the most elementary structure in the hull construction. A panel is built by welding steel plates together and then fixing outfitting components to them to strengthen the structure.

This case study highlights the production planning issues for the panels. There are three sub-assembly families: *Big Panels*, *Small Panels* and *Outfits*. We define *Big Panels* as panels built by joining two or more steel plates, while the *Small Panels* are built using a single steel plate. The size of the panels varies widely with the weight of a panel ranging from 0.5 ton to 10 tons. The *Outfits* are the outfitting components welded onto the panels. Figure 1 shows the process flow map for the panel production.

Big Panels: Blast→NC Gas→Bridge Cut→Bevel→Tack→Join→Mark→Outfitting (Big Panels)

Small Panels: Blast→ NC Gas→ Bridge Cut→ Mark→ Outfitting (Small Panels)

Outfits: Blast→ NC Plasma or NC Gas→ Bridge Cut→Profile Shop→Outfits (both Big and Small Panels)

Figure 1: Process flow map for panel production

At the *Blast* station, raw plates are “blasted” by ball grids to remove surface impurities, followed by the application of a corrosion preventive coating. The typical raw steel plate is 1.5m wide, 8m long and has a thickness of $\frac{1}{4}$ to $\frac{3}{4}$ inch. The blasted plates are sent to the *NC Gas Cut* or *NC Plasma Cut* stations to be cut to the required dimensions. The *NC Gas Cut* station is capable of cutting both thin and thick plates, while the *NC Plasma Cut* station can only cut thin plates. The *Big Panels* require thick plates and are therefore sent to the *NC Gas Cut* station. The *Small Panels* require thin steel plates but are processed at the *NC Gas Cut* station to balance the workload between the two stations. For the same reason, although the *Outfits* also require thin plates, about 27% of the *Outfits* are cut in the *NC Gas Cut* station while the rest are processed at the *NC Plasma Cut* station. After cutting, the parts remain joined to the steel plate by connectors called “bridges”, and these are manually cut at the *Bridge Cut* station.

Subsequently, the plates for each sub-assembly family follow different process routes. The thick plates for the *Big Panels* are beveled at the *Bevel* station, and then go through a series of welding processes at the *Tack* and *Join* stations to join the plates to form the basic panel

structure. The panels are then paint-marked at the *Mark* station to indicate the positions of the *Outfits* to be fitted. Eventually, the *Big Panels* are moved to the *Outfitting (Big Panels)* station to weld the *Outfits*. Since the *Small Panels* are built using a single plate, no welding is needed and they are moved directly to the *Mark* station and subsequently to the *Outfitting (Small Panels)* station. The *Outfits* are sent to the *Profile Shop* from the *Bridge Cut*, where the cut steel parts undergo some simple metal forming operations to produce the outfitting components. Finally, the *Outfits* are transferred to either the *Outfitting (Big Panels)* or the *Outfitting (Small Panels)* to be fitted onto their corresponding panels. The completed panels are then ready for quality checks before moving downstream to construct the steel blocks.

The panel production shown in Figure 1 occurs in three separate shops, namely the *Blasting Shop*, *NC Shop* and *Panel Shop*, which are physically located next to each other. The *Blasting Shop* consists of one *Blast* station, the *NC Shop* has the two NC stations and the *Panel Shop* includes the rest of the processing stations. Production control for the panel production is based on the PLTs at the shop level. The PLTs for the *Blasting Shop*, *NC Shop* and *Panel Shop* were 5, 8 and 13 days respectively. In times of high loading, jobs are subcontracted to vendors to keep jobs from falling behind their PLTs. Most of these vendors are located close by or within the panel production facility itself. Production planning relied on a single MPS (master production schedule) for all sub-assembly families. In addition, there was no smoothing of the MPS; that is, jobs were released immediately upon the receipt of orders from downstream internal customers.

With the surge in demand, most stations in the production shop experienced a heavier and more variable workload, which resulted in rising subcontracting costs. The company sought to increase its utilization of in-house capacity and to reduce subcontracting costs.

The management identified that one cause of the variable workload was the highly fluctuating MPS. The MPS was highly variable for three reasons. First, the production schedules of panels for the different oil-rig projects were not well coordinated. As a result, the total internal demand for panels was quite variable and resulted in variable work release. Second, raw steel plates of required thickness and grades were frequently unavailable and this delayed the release of some panels. Third, the raw steel plates were stacked to conserve space and therefore, finding the required plate often took substantial time; this added considerable variability to the picking time of the raw plates. In addition to the varying MPS, another cause for the large workload variability was the diverse processing requirements at the workstations, which are due to the different sizes of plates as well as the different number of plates and outfitting components required for each panel. The challenges in the panel production are illustrated in Figure 2.

Figure 2. Challenges in panel production

The management recognized that addressing these causes would lead to a reduction in production variability. However, they believed that this would require a long-term effort and would depend on factors external to the company. Coordination of the project schedules would

involve coordination between the company's various departments as well as with its customers on delivery schedules. Improving availability of raw steel plates would need better coordination with suppliers and more accurate forecast of global steel supply. Shorter picking times for steel plates would require considerable capital investment and time to devise new ways or equipment for material handling. Reducing the diverse processing requirements would need standardization of panel types, which would involve setting restrictions on oil-rig customization that might be detrimental to customer satisfaction.

The management decided that in order to reduce subcontracting cost within the short term, they should focus on tactical improvements, namely improving the way they did production planning so as to increase the utilization of available production resources.

3. Initial Study and Recommendations

The company learned about our model in TBG (2009) from a seminar and they found it relevant to their problem, since the model also considers a multi-station, multi-product system (the different sub-assemblies in this context) with the aim to minimize expediting cost.

TBG (2009) addresses the production planning challenge faced by MTO manufacturers with a fixed delivery lead-time in response to demand uncertainty. Setting a fixed delivery lead-time is common in markets where customers do not tolerate variable delivery lead-times, e.g. in markets for industrial products, where unpredictable lead-times from suppliers would complicate the customers' own production planning. However, to meet the delivery lead-time, a capacity-constrained MTO manufacturer that faces a time-varying demand must expedite production in periods of high demand, e.g., by overtime production or subcontracting. In particular, the model prescribes the PLT at each production station and the planning window of the MPS of each

product, with the objective to minimize production expediting costs and WIP inventory holding cost. The model determines the variance of production requirements (which is a measure of production smoothness) as well as the expected WIP level at each workstation; these are embedded in an optimization program to determine the optimal PLTs and planning windows.

A key aspect of the model is that it operates on two time scales. It models in discrete time the production planning process that is typically defined in time buckets, usually on the order of a day or a week. In addition, the intra-period workflow is modeled at a finer time scale, thus permitting multiple job movements between the workstations within each time period. The model incorporates the translation of a stochastic demand process (defined in planning time buckets) into job release, where the smoothness of the release is controlled by the length of the planning window. The jobs in the multi-station system are allowed to make multiple moves between the stations throughout each time bucket, with the PLTs controlling the production at each station. Thus, the model establishes the link between discrete time planning and intra-period workflow, and as such, the model is able to depict how the discrete-time planning process affects the workflow and production of individual workstations.

To further explore how the model is able to help the company, a team was formed, consisting of the authors and personnel from the production department. After a careful study of the model, the production personnel were particularly interested in two potential improvement areas, namely the smoothing of the MPS and the production smoothing at the stations.

Smoothing of MPS

The production personnel became aware that a smoother MPS results in a less variable release and eventually a smoother production at the workstations. They recognized that smoothing of the MPS is possible only if the delivery lead-time (*DLT*) is longer than the *product planned lead-*

time (*PPLT*). We define the *PPLT* as the sum of the planned lead-times of all stations in the route of the sub-assembly family; thus the *PPLT* is the planned duration for a job to move through the shop. The planning window is the difference between the *DLT* and the *PPLT*. The MPS can be smoothed over this planning window, which spans from period $t + PPLT$ to period $t + DLT$. Thus the planning window has a length of $DLT - PPLT + 1$ and the extent of smoothing depends on its length, as a longer planning window permits a smoother spread of workload.

One seemingly obvious solution is to increase the *DLT* to allow a longer planning window and planned lead-times at the stations. However, the management did not wish to change the current *DLT* for panel production of 26 days, as they did not want to affect the production schedule for the downstream stages. This implies that the panel facility would commit to deliver each order exactly 26 days after the receipt of an internal order from downstream facilities. Therefore the team had to find other ways to smooth production given the fixed *DLT*.

Initial recommendation 1. Upon studying the process flow, the team identified that the *Small Panels* and *Outfits* require fewer processing steps than the *Big Panels*, and thus, there was an opportunity to smooth the release for both the sub-assembly families if each family had its own MPS.

Smoothing at Stations

The team recognized that the production requirements at each station fluctuate due to the dissimilar processing requirements of the jobs as well as the varying job arrivals. The variability of the production requirements at a station depends on the length of its PLT. We refer hereafter to the PLT of each individual station as the *station planned lead-time (SPLT)*. A longer *SPLT* permits a relatively smoother production as one would have greater flexibility to level out the

production requirements over the *SPLT* by early production of some orders. However, with a longer *SPLT*, jobs stay longer at the station, which leads to a higher WIP inventory.

Initial recommendation 2. The original values of the PLTs for the three shops were set based upon the experience of the production planners. The team identified that there was an opportunity to achieve a smoother workflow by determining the values of PLTs using analytical methods. In addition, production control in the *Panel Shop* was based on the PLT of the entire shop, rather than for individual stations. As a result, subcontracting decisions were “guesswork” of predicting if a job could be completed in time to meet its due-date. With a *SPLT* for each station in the *Panel Shop*, the company would be able to track the progress of jobs more precisely. The initial recommendations are illustrated in Figure 3.

Figure 3. Initial recommendations

Planning Decisions

Given the two initial recommendations, the team identified the key planning decisions, which are summarized as follows:

- Given the fixed *DLT* of 26 days and the constraint $PPLT \leq DLT$, how does one set the *SPLT* of each station to minimize total subcontracting cost?

- How does one divide the *DLT* between the planning window and that of the *PPLT*?
Increasing the planning window allows more smoothing of the MPS, which results in a less variable job release to the shops. However, a longer planning window results in a shorter *PPLT*, which allows for less smoothing at the stations.
- In this multi-station, multi-product setting, the team must take into account the interdependence of workflow among the stations as well as among the sub-assembly families.

The team recognized that the key features of the production shop could be captured by our model in TBG (2009), namely production control using planned lead-times, influence of discrete time MPS smoothing on production workflow, variability at the workstations as well as subcontracting production to meet the capacity shortfall. However, before the team decided to apply the model, the team also considered other simpler alternatives. For example, the team considered setting the *SPLTs* of *all stations* to be proportional to each station's utilization rate while satisfying the *DLT* constraints. The rationale behind this method was that the more heavily utilized stations would need longer *SPLTs*. However, this method would not be able to account for many important aspects of the scenario, e.g., the variability of demand and processing requirements, the difference in subcontracting cost between stations and the interdependencies of workflow between stations. The team also considered the alternative of using a discrete-event simulation. However, they found that simulation was not suitable for the extensive "what-if" analysis they would like to perform as it would be slow to make the numerous simulation runs, especially the runs requiring optimization. Furthermore, the management preferred a method that did not require an extensive learning process for its personnel. As our model was formulated and solved in MATLAB, the management thought that the planners and engineers could readily learn

how to use it. In addition, the model parameters, e.g., unit subcontracting cost, workflow and capacity levels, could be easily altered according to changes in the actual situation. In addition, MATLAB provided an optimization toolbox that could be used to find the optimal planning windows and *SPLTs*.

We note that the characteristics of the panel production encompass various operational aspects that can be found in many production systems: a highly variable MPS, multiple product families and process routes, dissimilar processing requirements and expediting actions taken to meet delivery schedules. Hence, this case study provides an assessment for the applicability of the model in TBG (2009) in capturing real-life scenarios.

4. Model Application

Optimization model

We present the optimization model based on TBG (2009) with the objective to minimize the expected total subcontracting cost. The decision variables are the planning windows of sub-assembly family k and *SPLTs* of station i , denoted by W_k and n_i respectively. We define the discrete time period t of the model to be one day, which is the time bucket used in the existing planning system. The random variable for the production requirement at station i in day t is denoted by P_{it} . We let c_i be the expediting cost per workhour for station i and M_i as the nominal capacity for station i (in workhours per day). We denote the exogenously determined *DLT* of each sub-assembly family by DLT_k . The notation ω_{ki} represents the number of times a job from sub-assembly family k visits station i .

$$\begin{aligned} \text{Min} \quad & \sum_i c_i E[P_{it} - M_i]^+ \\ \text{s.t.} \quad & \sum_i \omega_{ki} n_i + W_k - 1 = DLT_k, \quad \forall k \end{aligned} \tag{1}$$

$$W_k \geq a_k, \quad \forall k \tag{2}$$

$$n_i \geq b_i, \quad \forall i \tag{3}$$

where $x^+ = \max(x, 0)$. The objective function is the total expected subcontracting cost per day across all stations. In order to characterize the production random variables, we assume that P_{it} is normally distributed. We constructed normal probability plots for the daily production output from a two-month data which validated this assumption. In the appendix we gave the characterization of the first two moments of P_{it} as a function of the decision variables. With the first two moments of P_{it} , the objective function can be computed using the normal linear loss integral.

Constraint (1) defines the relationship between the planning windows, *SPLTs* and *DLTs*, and the first term represents the *PPLT* of sub-assembly family k . Constraints (2) and (3) assure a lower bound of at least a_k and b_i on the planning windows and the *SPLT* of each station i , respectively. Lower bounds of *SPLTs* and planning windows are needed to avoid excessively frequent monitoring to track the job progress and the MPS. We set both a_k and b_i to be one day, as the planning time bucket of one day was the minimum duration that the management perceived to be suitable for planning within this highly dynamic system. The decision variables were not restricted to be integers, as non-integer values were acceptable for production control.

We have not been able to determine whether the objective function is convex. Therefore, we cannot assert that the solutions are global optimal. In this exercise, we solved each test problem using many different starting points. We attained the same solution for each test problem, which increased our confidence that we had obtained a global optimum. We solved the nonlinear program by the optimization toolbox in MATLAB.

We note a few characteristics of the problem. First, the team assumed that the *SPLT* of the *Profile Shop* which produces the *Outfits* is fixed at 2 days, as the station's processing time is relatively stable and reasonably independent of the workload. Second, the team did not consider

the holding cost of the WIP inventories in our analysis. To explain the reasoning for this, we need to consider the two main components of inventory holding cost, i.e., the material cost and the value added to the jobs through processing. Since the *DLT* is fixed at 26 days, the raw plates stay in the production system for approximately the same duration regardless of how the plates are scheduled. In addition, the management perceived that the value added to the jobs through processing (and hence the incremental holding cost) is significantly lower than the subcontracting cost.

Validation of model assumptions

The team studied the key model assumptions to determine if they were valid for the application:

Capacity assumptions

The model assumes that when the production requirement surpasses the nominal capacity of a station, the station is always able to complete the production requirement within the *SPLT* but it incurs an additional expediting cost for all production beyond the nominal capacity. In the context of panel production, subcontracting is routinely used to expedite work to meet the *SPLT* requirement. Since the subcontracting is performed in nearby shops and by on-site contract workers, relatively little time is wasted in transporting the jobs. Furthermore, from the management's experience, there are very few occurrences where the subcontractors failed to produce the outsourced demand within the *SPLT*. Thus, the team was satisfied that the assumption that the outsourced output always occurs within the *SPLT* is valid.

The *Panel Shop* operates with skilled workers, some of which are cross-trained to work at more than one station. Thus, there is some flexibility in allocating workers to the heavily loaded stations. Nevertheless, the team found the assumption of fixed capacity to be reasonable as such re-deployment of workers was infrequent after the company experienced the demand surge; the

stations with cross-trained workers were usually heavily loaded at the same time, this preventing labor re-allocation.

Workflow assumptions

We assume, in the model, that work arrives at each station uniformly within each time bucket. However, in the actual panel production, the jobs move to the downstream station upon completion, causing the arrival at the downstream station to be non-uniform. TBG (2009) validate this assumption via a simulation model wherein jobs move to the next station immediately after completion. The study shows that, despite relaxing the assumption, the simulation results are close to the model output.

The model has two production functions, the continuous time function and the sub-interval function (equations A2 and A3 respectively in appendix). The sub-interval function is used for stations with low arrival rates, since the continuous time assumption required for the development of the continuous time function would be less reasonable. The *Join*, *Mark* and both the *Outfitting* stations assemble or process panels rather than plates and *Outfits*, and therefore have lower arrival rates; we employ the sub-interval function for these stations.

Production assumption

By the linear production function (A1 of appendix), the model assumes that the station does not produce up to the nominal capacity level even when there is available work-in-queue. As such, jobs are not dispatched downstream earlier than scheduled and hence, avoid unnecessary

inventory holding cost. However, as stated earlier, the incremental WIP inventory holding cost in the panel production is relatively insignificant compared with the subcontracting cost. In spite of this, the team found that the assumption is still justified. Due to the large physical size of the jobs and the space constraints in the facilities, the stations usually produce just to meet the *SPLT* and it only produce jobs earlier for smoothing production, so as to avoid taking up the downstream shop space unnecessarily.

Demand assumption

One assumption of the model is that the demands are stationary. In contrast, the observed demand for each sub-assembly family is generally non-stationary over time. However, the team observed that the demand is stationary if the time horizon is considered as successive time segments, with each time segment representing a constant number of oil-rigs in production in their respective stable project phase (each time segment typically ranges from about 15 to 45 days). That is, the demand is rather stationary within each time segment. In addition, the demand data showed a fairly constant coefficient of variation (standard deviation/mean) of demand within each segment. We show the demand data for the *Big Panels* over a three-month period in Figure 4.

To test the sensitivity of our solution to the non-stationary demand, the mean demand for *each* sub-assembly family was set to be either 100% or 125% of the original mean demand, while keeping the demand coefficient of variation constant. The analysis showed that for all test scenarios, the subcontracting cost resulting from the original demand's optimal *SPLTs* is no more than 4% higher than the test scenarios' minimum subcontracting cost.

Figure 4. Demand for *Big Panels* over three-month period

Data Collection

A large amount of data is required to parameterize the model inputs. We discuss the model inputs and how we obtained the relevant data.

- *Demand*: We obtained the mean and standard deviation of the internal demand for each sub-assembly family from an eight-month demand record. We found a low correlation coefficient of 0.13 between the demands of *Big Panels* and *Small Panels*. However, the correlations between *Big Panels* and *Outfits* as well as between *Small Panels* and *Outfits*, are high with correlation coefficients at more than 0.70.
- *Effective Processing Times*: We acquired the mean and variance of the effective processing times (in hours) at each station from data collected in a two-month period. We needed the mean processing times to define the workflow matrix in the model, which defines the amount of work that each unit of production at a station generated for each downstream station, on the average. We note that the stations from *Blast* to *Bevel* process steel plates while the rest of the stations process panels and outfits. Thus for the stations that process plates, we need to

convert the time taken per plate to per panel and per outfit. The variance of the processing time at each station is required to calculate the noise arrivals at each station; the variance of noise arrival in each day equals the variance of the workload for the mean number of jobs processed at the station per day (i.e., mean number of jobs processed per day multiply by the variance of processing time per job).

- Capacity: We observed the throughput rate, in periods of high demand, when most stations are operating at full capacity. We approximate the nominal capacity levels M_i by the mean throughput (in hours) per day observed in these periods.
- Subcontract Cost: The vendors quote the subcontracting costs in terms of cost per metric ton. Since workload in our model is measured in hours, we had to convert the subcontracting cost at each station into average subcontracting cost per workhour c_i . We approximate the average subcontracting cost based on the average weight of the jobs at the station (in metric tons) and the mean processing time of the jobs (in hours).

5. Results and Insights

We report the base-case solution, as well as the corresponding expected subcontracting work and costs, in Table 1. The results presented here have been disguised to protect the company's confidential data, but the insights drawn are identical to the conclusions based on the actual data.

With the separate MPS for each sub-assembly family, there is a greater flexibility to adjust the planning windows of each sub-assembly family for a smoother release. The solution suggests that there should be substantial smoothing for the release of *Small Panels* and *Outfits*, with

planning windows equal to 10.8 and 9.8 days respectively. The planning window for *Big Panel* is one day and thus, there is no smoothing of its MPS.

Table 1: Optimization results

Station	Original W_i or n_i	Optimal W_i or n_i	Expected	Expected
			Subcontracting Work (hours/day)	Subcontracting Cost (\$/day)
Dummy for Big Panels	1.0	1.0	-	-
Dummy for Small Panels	1.0	10.8	-	-
Dummy for Outfits	1.0	9.8	-	-
Blast	5.0	1.0	0.13	73
NC Gas Cut	8.0	4.4	0.37	155
NC Plasma Cut	8.0	4.4	0.12	60
Bridge		3.1	0.12	143
Bevel		1.0	<0.01	<1
Tack		1.0	<0.01	<1
Join		7.8	5.96	775
Mark		1.0	<0.01	<1
Outfitting (Big Panels)		6.6	5.83	874
Outfitting (Small Panels)		6.6	4.39	351

The solution suggests that the *SPLT* of *Blast* should be reduced from 5.0 days to 1.0 day, and *NC Gas Cut* and *NC Plasma Cut* from 8.0 days to 4.4 days. A proportion of the original *SPLT* at these stations acts as safety lead-time to buffer against the uncertainties of unavailable steel plates and the long picking times. For the *Small Panels* and *Outfits*, most of the excess days

from this reduction are reallocated to their planning windows. Here, the planning windows would not just smooth the MPS but also act as the safety lead-times for acquiring steel plates. For the *Big Panels*, the *SPLT* of the *Blast* station is mainly reallocated to the other workstations.

At the *Join* station, the optimal solution for the *SPLT* is the largest among the workstations at 7.8 days. The team observed that the utilization rate at the *Join* station is more than 90%; its processing requirements are also highly variable, with a daily coefficient of variation of 0.76. The optimal *SPLTs* of *Outfitting (Big Panels)* and *Outfitting (Small Panels)* are also high at 6.6 days, but still lower than *Join*, despite both stations having higher utilization rates and coefficient of variation for their processing requirements. The reason is that the larger *SPLT* at the *Join* station would smooth the production output, which in turn would lead to smoother arrival at the downstream *Outfitting (Big Panels)* and *Outfitting (Small Panels)*. The solution also suggests that *Bevel*, *Tack* and *Mark* stations have the shortest *SPLT* of 1.0 day due to their comparatively lower utilization; even with a *SPLT* of 1.0 day, the expected subcontracting costs are low at these stations.

By exercising the model for different what-if scenarios, the team developed the following insights for setting the planning windows and *SPLTs*:

- The team gained an understanding of how the discrete-time MPS planning influences the workflow of the production network. Most importantly, they gained insights on the interaction between the planning windows and the *SPLTs*; that is, the planning window smoothes the MPS (and thus the arrival at the stations) while the *SPLT* at a workstation smoothes the variability due to both job arrivals and the diverse processing requirements at the station. Therefore, a longer planning window is preferred if a sub-assembly family faces a highly variable demand and relatively lower workload variability at the stations; on

the other hand, if the variability of processing requirements at the stations is relatively larger, longer *SPLTs* are preferred to smooth production at the workstations.

- A workstation would require a longer *SPLT* if it faces greater variability in processing requirements, and has higher utilization rate and unit subcontracting cost. The management had previously thought that the *SPLT* should be based only on the utilization rate.
- Smoothing at an upstream station has the added advantage of smoothing the arrivals to downstream stations. The management learned that looking at each individual station in isolation is suboptimal.

We note that in situations where WIP holding cost is significant, one also has to consider that longer *SPLTs* lead to higher WIP inventory levels. The above insights are useful for understanding the tradeoffs among capacity, lead-time and production smoothness.

6. Validation

Before implementing the recommendations, the team attempted to quantify the potential cost savings that would result from these changes. The team compared the optimal total subcontracting costs with the actual cost and estimated that the recommendations would result in a 20% to 30% cost reduction. In another effort to estimate the potential cost savings, the team input the current *SPLTs* into the model and compared the resulting total subcontracting cost with the optimal cost. However, the individual stations in the *Panel Shop* had not been assigned *SPLTs*; to overcome this, the team estimated the current cost by setting the *SPLTs* of the stations in the *Panel Shop* proportional to the station's utilization rate while satisfying the *Panel Shop*'s planned lead-time of 13 days. The results from this comparison showed that the optimal solution would reduce the cost by about 21.8%. The team regarded this as a conservative estimate of the

actual savings because, in the actual setting, the production control of the *Panel Shop* was based on the aggregate planned lead-time of the shop, rather than on the *SPLTs* of the individual stations.

The team also compared the optimal cost with that of another alternative in which the *SPLTs* of *all* stations are set proportional to each station's utilization rate while satisfying the *DLT* constraints. The team input the *SPLTs* based on this alternative into the MATLAB program and found that the total expected subcontracting cost for this alternative is 14.4% higher than that of the optimal expected cost.

The team faced difficulties in validating the predictive capability of the model. The most useful validation would be to compare the actual amount of work subcontracted out with that predicted by the model. However, the subcontracting decisions in the current *Panel Shop* were based on the aggregate PLT of the shop; in our model, production control is set based on the *SPLTs* of individual stations. Thus, the team did not make such a comparison for the *Panel Shop* and unfortunately, this is where most of the workstations are located. For the *Blasting Shop* and *NC Shop*, the team found that the amount of subcontracted work predicted by the model is about 26% and 21% less than the actual data for the two stations. The team viewed this as a reasonable validation given the presence of high variability in the system. After the recommendation to assign the *SPLTs* to individual stations was adopted, new shop data became available and we re-validated the predictive capability of the model; we found a difference of about 9%, which is significantly lower than the partial validation using data only from *Blasting shop* and *NC Shop*.

7. Influences on Company

The company adopted the recommendations of creating a separate MPS for each sub-assembly family and individual planned lead-times for each station. However, before implementing the recommendations for the values of the planning windows and *SPLTs*, the company decided to add production capacity as it projected a growth in oil-rig demand, while maintaining the strategy of subcontracting to buffer against peak demands. Instead of implementing the precise recommendations for the values of planning windows and *SPLTs* as a static solution, the company utilized the model to support the monthly updating of the values in response to the continual changes in production capacity. Furthermore, since the model requires large amounts of data as inputs, the use of the model is helped by the company's new initiative to regularly update shop data.

The company also employs the model for what-if analysis to support planning; for instance, the model is used by the company to determine how capacity addition affects subcontracting cost. In addition, the model is also used as a guide for focusing improvement efforts. The model helped to identify steel availability and production scheduling as two potential areas in which subcontracting cost could be reduced; consequently, there were projects carried out to address these issues (see Huang 2006 and Tan 2006 for details of these projects). Another significant result is that the project has led to greater awareness of the importance of considering the interrelationship among lead-time, capacity and production smoothness. Furthermore, there is now a greater emphasis on reducing variability of production. However, we were not able to quantify the resulting savings in subcontracting cost because of the ongoing changes in the production shop and oil-rig demand, as well as fluctuations in global steel supply and subcontracting charges.

Appendix

We summarize the characterization of the first two moments of P_{it} required to compute the objective function in the optimization model in section 4. Refer to TBG (2009) for more details. The model is an extension of the Tactical Planning Model (TPM) from Graves (1986). The extension permits work to flow between workstations throughout each period, and not only at the start of the period as in Graves (1986). This enables the establishment of the link between discrete-time planning and intra-period job flows, where jobs are able to make multiple moves between the stations within a single planning time bucket.

Production at workstation

We first consider a workstation producing a single product family. We model the arrivals and production at the station in each sub-interval of time period t so as to permit arrivals and production throughout the period. We note that period t is equivalent to the planning time bucket. We sub-divide each time period t into p equal subintervals, and use index s , where $s = 1, 2, \dots, p$, to denote the sub-periods. We define the length of each sub-period $\Delta = 1/p$. We assume that workflow is continuous (e.g. in hours) and work arrives at the start of each sub-period. We assume that the production requirement in each sub-period is a linear function of the amount of WIP at the station, which is given by

$$Y_i(\Delta, s) = (\Delta / n_i) X_i(\Delta, s) \quad \text{for } s = 1, 2, \dots, p, \quad (\text{A1})$$

where $Y_i(\Delta, s)$ is the production requirement at station i in sub-period s of length Δ , $X_i(\Delta, s)$ is the queue length at start of sub-period s and n_i is the *SPLT* at station i . We assume n_i can assume any positive values. We express the production requirement as a function of Δ to allow arrivals to the station throughout the time period. We interpret (A1) as follows: to realize the *SPLT*, the workstation must process close to (Δ / n_i) of the work-in-queue in each sub-period s . By Little's

Law, the workstation must process an average of (Δ/n_i) of the work-in-queue in each sub-period to achieve an average lead-time of n_i . We approximate this by requiring the station to produce exactly (Δ/n_i) of the queue. We note that by (A1) the station levels its work-in-queue evenly over the *SPLT* and thus the longer the *SPLT*, the smoother the production requirement at the station.

We assume that each station can always output the production requirement. Each station has a nominal capacity limit and when the production requirement exceeds this nominal capacity, the station accomplishes the production requirement within the *SPLT* by subcontracting (albeit at an additional cost) all production in excess of the nominal capacity. We further assume that the arrival of work to the workstation in period t , which we denote as A_{it} , arrives uniformly over period t . That is, we assume that the arrival amount at the start of each sub-period is equal to A_t / p . We note that arrivals can vary *between* periods.

Based on (A1) and using the standard inventory balance equation for the sub-interval arrivals, we can then derive the production function, which is given by

$$P_{it} = \beta_i(\Delta)Q_{it} + \gamma_i(\Delta)A_{it}, \quad (\text{A2})$$

where P_{it} is the production requirement at station i in period t , Q_{it} is the queue length at the start of period t and A_{it} is the total work arrival in period t . The coefficients $\beta(\Delta)$ and $\gamma(\Delta)$ are given by

$$\beta_i(\Delta) = 1 - [1 - (\Delta/n_i)]^p$$

and

$$\gamma_i(\Delta) = 1 - \beta_i(\Delta) \left(\frac{1 - (\Delta/n_i)}{(1/n_i)} \right).$$

Equation (A2) expresses the production in each time period as functions of both the work queue at the start of the period and the arrivals during the period. The coefficients for the linear function depend on the size of the subintervals.

We consider the continuous time limits for $\beta(\Delta)$ and $\gamma(\Delta)$ as Δ goes to zero; this represents a fluid-like work arrivals and in effect, we assume that (A1) holds at every instant in time. We obtain the continuous time limit of $\beta(\Delta)$ and $\gamma(\Delta)$:

$$\beta_i = 1 - e^{-1/n_i}$$

$$\gamma_i = 1 - n_i \beta$$

We restate (A2) for the continuous-time function as:

$$P_{it} = \beta_i Q_{it} + \gamma_i A_{it} \tag{A3}$$

where β_i and γ_i are given above. The validity of the continuous time function depends on the frequency of job movements in each time bucket since a higher frequency would suggest a more fluid workflow. TBG (2009) shows that the continuous time function is a suitable approximation for systems with a frequency of workflow that can be found in most real-life production systems. However, in situations where job flow rates are low, we expect the sub-period coefficients $\beta_i(\Delta)$ and $\gamma_i(\Delta)$ to be more suitable, where we set Δ on the order of the average inter-arrival time. TBG (2009) test the suitability of the coefficients for different job flow rates using a simulation model and show that accurate model output can be achieved if the appropriate coefficients are employed.

Smoothing of MPS

We assume that the demand process of each product family is independent and identically distributed. We denote the demand in period t by D_t with mean $E[D_t]$. We regard these unreleased orders as a demand queue, waiting to be released into the shop.

We introduce a dummy workstation into the network of production stations to model the MPS smoothing and work release. Thus the queue at the dummy station is the order backlog and a job first “visits” the dummy station before it is released to the production system. The job

arrival to the dummy station in period t is the demand D_t and the production output of the dummy station corresponds to the job release. We model the MPS smoothing by

$$P_{0t} = \frac{Q_{0t}}{W}, \quad (\text{A4})$$

where station 0 is the dummy station; P_{0t} is the production output in period t , namely the job release into the shop; Q_{0t} is the queue of work at the start of period t and W is the planning window ($W \geq 1$). Equation (A4) models the smoothing of the release by spreading the on-hand orders evenly over the planning window. In each period, by Little's Law, the shop must release an *average* of $1/W$ of the order backlog. We approximate this MPS smoothing by assuming that the shop releases *exactly* $1/W$ of the on-hand orders in each period.

Workflow model

We model the work arrivals to station i by

$$A_{it} = \sum_j \phi_{ij} P_{jt}, \quad (\text{A5})$$

where ϕ_{ij} is a positive scalar. We assume that each time unit of production at station j generates ϕ_{ij} time units of input to station i on average. We note that $\phi_{0j} = 0$ for all j and that ϕ_{j0} is the average amount of work that starts at workstation j for each new job.

To analyze the workflow model, we combine (A3) and (A4), and express it in matrix notation:

$$\mathbf{P}_t = \mathbf{F}\mathbf{Q}_t + \mathbf{G}\mathbf{A}_t \quad (\text{A6})$$

where $\mathbf{P}_t = \{P_{0t}, P_{1t}, \dots, P_{mt}\}'$, $\mathbf{Q}_t = \{Q_{0t}, Q_{1t}, \dots, Q_{mt}\}'$, $\mathbf{A}_t = \{A_{0t}, A_{1t}, \dots, A_{mt}\}'$ are column vectors and m is the number of workstations. $\mathbf{F}$ is a diagonal matrix with its first diagonal element as $1/W$ and the remaining diagonal elements as β_i ($i = 1, 2, \dots, m$). $\mathbf{G}$ is a diagonal

matrix with zero as the top element and γ_i ($i = 1, 2, \dots, m$) as the next m diagonal elements. We express (A5) in matrix form:

$$\mathbf{A}_t = \mathbf{\Phi} \mathbf{P}_t. \quad (\text{A7})$$

$\mathbf{\Phi}$ is a square matrix with elements ϕ_{ij} . We express the inventory balance equation in matrix form:

$$\mathbf{Q}_t = \mathbf{Q}_{t-1} - \mathbf{P}_{t-1} + \mathbf{A}_{t-1} + \boldsymbol{\zeta}_t. \quad (\text{A8})$$

The term $\boldsymbol{\zeta}_t$ is a column random vector that combines both demand arrivals to the dummy station as well as the noise arrivals to the workstations; it has D_{t-1} in the first row and $\{\xi_{1t}, \dots, \xi_{mt}\}$ in the next m rows. The term ξ_{it} is a zero-mean noise term that represents any workload variability in the arrival stream to station i . We assume that the amount of noise is independent of the work arrivals. For the panel production, we use ξ_{it} to model variability due to the dissimilarity in processing requirements.

Using (A6) to (A8), we obtain $\mathbf{P}_t$ which we express as:

$$\mathbf{P}_t = \mathbf{B} \mathbf{P}_{t-1} + \mathbf{H} \boldsymbol{\zeta}_t, \quad (\text{A9})$$

where $\mathbf{H} = (\mathbf{I} - \mathbf{G} \mathbf{\Phi})^{-1} \mathbf{F}$, $\mathbf{B} = \mathbf{I} - \mathbf{H}(\mathbf{I} - \mathbf{\Phi})$ and $\mathbf{I}$ is the identity matrix. It can be shown that $(\mathbf{I} - \mathbf{G} \mathbf{\Phi})$ is invertible as long as the spectral radius of the workflow matrix $\mathbf{\Phi}$ is less than one.

Assuming an infinite history for the system, we get by repeated iteration:

$$\mathbf{P}_t = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{H} \boldsymbol{\zeta}_{t-s}. \quad (\text{A10})$$

We denote the expectation of vector $\boldsymbol{\zeta}_t$ by the vector $\boldsymbol{\mu}$, which is a column vector with $E[D_t]$ in the first row as the only non-zero element, since the noise terms $\{\xi_{it}\}$ have zero mean by definition. We show in Teo (2006) that the series in (A10) converges if and only if the spectral

radius of Φ is less than one. From (A10), we obtain the expectation and variance of the production requirements vector:

$$E[P] = (I - \Phi)^{-1} \mu. \quad (A11)$$

$$\text{Var}(P) = \sum_{s=0}^{\infty} B^s H \text{Var}(\zeta_t) H' B'^s, \quad (A12)$$

where $\text{Var}(\zeta_t)$ is a covariance matrix that consists of the variance of D_t and the covariances of the noise terms ξ_{it} . We note that (A12) provides the production variances for each station as well as the covariance for each pair of workstations. We can compute the power series in (A12) using the methods in Graves (1986). In addition to characterizing P_t , it is also useful to determine the distribution of queue levels to characterize the waiting time at each station; see Teo et al. (2007) for the derivation.

Extension to Multi-Family Model

In the panel production, each workstation has a single *SPLT*, which applies to all jobs that visit the workstation. However, the product families (i.e., sub-assembly families in the context of the panel production) are allowed to have different planning windows.

We want to determine the total production requirements from all the product families at each station. To extend the single-family linear-system model to accommodate the multiple product families, we model each product family the same way as in the single-family model, i.e., each product family k has a dummy station and has column vector ζ_{kt} and matrix Φ_k to model each family's demand, noise levels and routing. The single-family model characterizes the production requirements for product family k , i.e., $E[P_k]$ and $\text{Var}(P_k)$.

Assuming that demand is independent between the product families, we obtain the mean and variance of the total production requirements and the total queue lengths by aggregating across all product families:

$$E[\mathbf{P}_{total}] = \sum_k E[\mathbf{P}_k], \quad (\text{A13})$$

$$\text{Var}(\mathbf{P}_{total}) = \sum_k \text{Var}(\mathbf{P}_k), \quad (\text{A14})$$

We can incorporate correlated demands between product families for (A14) by incorporating the demand correlations into the covariance matrix for the noise vector ζ_{kt} . See Teo (2006) for details.

References

- Graves, S. C. 1986. A Tactical Planning Model for a Job Shop. *Oper. Res.*, **34**, 4, pp 522-533.
- Huang, F. 2006. Production Scheduling in a Make-To-Order Job Shop under Job Arrival Uncertainty. *M. Eng. Dissertation*. Department of Mechanical Engineering, MIT, Cambridge, MA.
- Tan, C. Y. 2006. Inventory Management of Steel Plates at an Oil Rig Construction Company. *M. Eng. Dissertation*. Department of Mechanical Engineering, MIT, Cambridge, MA.
- Teo, C. C. 2006. A Tactical Planning Model for Make-To-Order Environment under Demand Uncertainty. Ph.D. thesis, Singapore-MIT Alliance, Nanyang Technological University, Singapore.
- Teo, C. C., R. Bhatnagar, and S.C. Graves. 2007. A Tactical Planning Model for a Production Network with Continuous-Time Control. *Working Paper*, 22 pp.
- TBG. 2009. Setting Planned Lead-times for a Make-To-Order Production System with Master Schedule Smoothing. *Working Paper*, 34 pp.