Hedenstierna, C.P.T., Disney, S.M. and Holmström, J., (2016), "Order book management in 3D printing service operations: A design science approach", 23rd Annual EurOMA Conference, Trondheim, Norway, 17th – 22nd June, 10 pages.

Order book management in 3D printing service operations: A design science approach

Carl Philip T. Hedenstierna (cpthed@gmail.com)
Department of Production and Quality Engineering
NTNU, Norwegian University of Science and Technology
Trondheim, Norway

Stephen M. Disney
Logistics System Dynamics Group, Cardiff Business School, Cardiff University
Cardiff, UK

Jan Holmström
Department of Industrial Engineering and Management
Aalto University
Helsinki, Finland

Abstract
3D printing services are appearing, allowing consumers or businesses to order custom parts for delivery within a quoted lead-time, regardless of design specification. The manufacturer's problem is to ensure that customers receive their goods on time, while capacity costs (including overtime costs) are kept low. We introduce a performance measure called the service rate, analogous to the fill rate of inventory control. Using design science, we identify and evaluate order book control policies to manage the service rate – capacity cost trade-off. Our results show that production smoothing is compatible with high service rates for customized products with short delivery times.

Keywords: Service levels, Make to order (MTO), Production smoothing

Introduction
Additive manufacturing or 3D printing is predicted to revolutionize supply chains, as a single manufacturer now can make customer-specific parts for the mass market. The variety of products is so large, that a service operation is necessary, producing to specific customer orders. This means that a 3D-printing operation cannot use inventory to decouple demand variability from production – any demand variability must be absorbed by capacity, or by the delivery time. Take Shapeways as an example; with physical production times of minutes or hours, they offer a manufacturing throughput time of six days (or two days with expediting) for white plastic products with no dimension exceeding 250mm (Shapeways, 2016). This leaves some slack time for decoupling demand variability from production.

For a delivery to occur on time, the promised delivery time must be longer than the total time spent in the order book (not yet released to production) plus the physical production time (Kingsman et al., 1989), as Figure 1 illustrates. Comparing the promised delivery time of 3D printing to other service operations, we find it relatively short, leaving

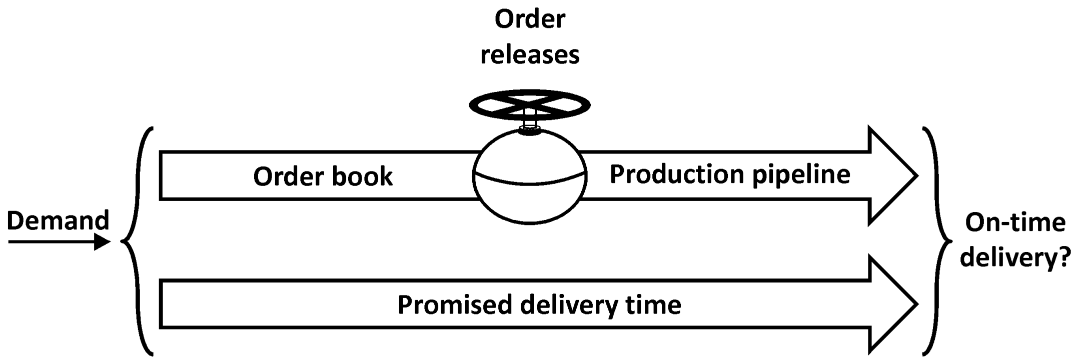

Figure 1 – The relation between lead time and delivery performance

only moderate slack time for buffering (Table 1).

Most of the make-to-order theory focuses on scheduling and workload control, often requiring coordination between multiple machines in job shops (Stevenson et al., 2005). Additive manufacturing differs, because a single machine can achieve complex part geometries, and because each customer may want a unique product, making the sequencing of products less of an issue than in traditional job-shop environments (Holmström et al., 2016). For additive manufacturing, production sequencing may thus be less critical than production smoothing. Nevertheless, the smoothing problem in service operations is both practically important and academically understudied.

The first reference about smoothing an order-based-operation may be Forrester (1961, p. 144), who implemented a proportional policy in a system dynamics model, where a fixed fraction of the order book was shipped in every period. The same policy was later used by Sterman (2000, pp. 723–725) in a similar context. An alternative proportional policy was suggested by Wikner et al. (2007), who also considered an agile policy for capacity adaption. In relation to this, Anderson et al. (2005) used control theory and a system dynamics model to show that bullwhip can occur in multi-stage service operations where the capacity is adapted over time. This suggests that the smoothing problem is as important in service operations as it is in inventory settings. Note that none of these studies investigated the resulting delivery performance.

Our contribution is the development of the *service rate*, a metric comparable to the fill rate popular in inventory theory, which measures customer satisfaction against promised delivery times. In addition, we investigate how production-smoothing policies influence service levels in the additive manufacturing context. We show that high service levels are compatible with production smoothing, and present strategies for its implementation.

While our motivation for this paper came from the 3D printing problem highlighted above, our modelling methodology and model solution could be applied to other service operations settings, see Table 1 for examples.

Methodology

Operations management often finds it useful to adopt a critical realist line of thinking, as it assumes an objective world to be understood or improved (Mingers, 2015). To achieve this improvement, we select appropriate models or frameworks: Here, we use the CIMO framework, explained in Denyer et al. (2008) as well as in Pawson and Tilley (1997), to guide the development of an analytical model. The main steps are illustrated in Figure 2.

First, every situation or problem that we seek to improve is embedded in a *context*, or a set of circumstances. Although the overall context is simply service operations, a more

Table 1 – Examples of service operations with promised delivery times.

Company	Product or service	Promised delivery time
British Gas	Heating system repair	Same-day
Moonpig	Greeting cards	24h
Shapeways	3D Printing	2–16d
Amazon	Supersaver delivery	5d
Dell	Customized computers	7d
Lenovo	Customized computers	14d
Nationwide	Loan application processing	14d
Anonymous	Industrial equipment	1–6w
Anonymous	Material handling equipment	Several weeks

precise description includes customer requirements, the cost structure of the operation, and the characteristics of the product and the process. Depending on the context, we may entertain a set of feasible *interventions*, which are modifications expected to improve the situation or to rectify the problem. Based on the context, there is a systems *mechanism* by which the intervention produces *outcomes*. The mechanism may loosely be regarded as the physics of the situation, while the outcomes are the results of interest. These can be divided further into *expected* outcomes, which is a change in the parameters, and *unexpected* outcomes, reflecting those outcomes that cannot be anticipated prior to implementation and testing. As this is a theoretical piece, we can only investigate expected outcomes.

Context – Order book management in the 3D printing industry

The introduction highlighted the context of our study. We do not repeat it here for brevity. Similarly, the intervention is production smoothing via the order book, which we have specified now, and will revisit after defining the mechanism and the expected outcomes.

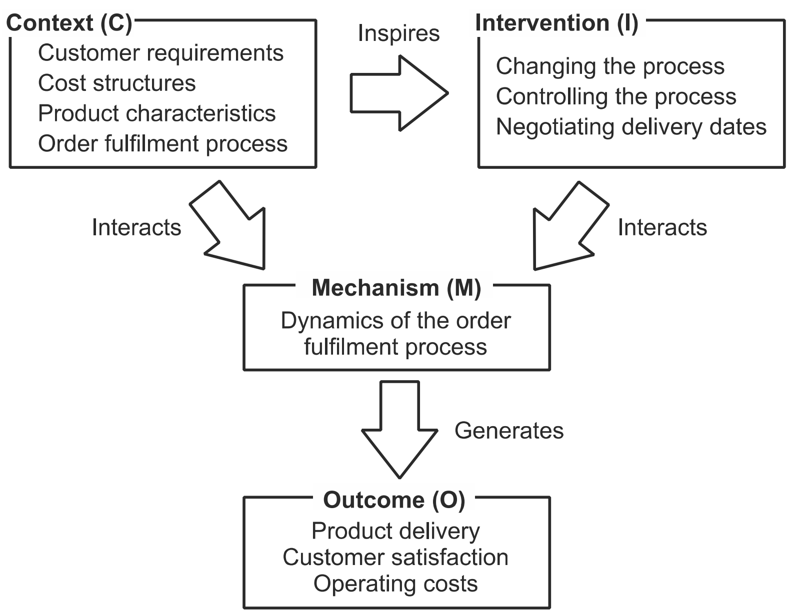

Figure 2 – Embedding the order book management problem in the CIMO framework

Mechanism – Dynamics of the order book

Consider a production system where temporally independent periodic demand, d_t, is drawn from a normal distribution, $d_t \in N(\mu_d, \sigma_d)$. The variable d_t represents the total work content (in hours) ordered by customers in period *t*. The demand must be released as production orders within *Q* periods to be delivered on time (immediate releases are necessary when $Q = 0$). Let o_t denote the production orders released at time *t*. The order book b_t contains all received customer orders that have not yet been released to production. It has the difference equation

$$b_t = b_{t-1} + d_t - o_t. \tag{1}$$

Since we cannot release orders that we have not yet received, the order book will never turn negative. This means that $\sum_{t=1}^{T} o_t \leq b_0 + \sum_{t=1}^{T} d_t$ must hold. We assume that orders are released according to a First-In-First-Out (FIFO) policy implying all tardy orders must be processed before any orders that are not yet late. The *schedule adherence* a_t in a given period describes the difference between the cumulative actual deliveries and required deliveries according to

$$a_t = a_{t-1} - d_{t-Q} + o_t. \tag{2}$$

The schedule adherence behaves much like the order book, with the special property that it is positive when all deliveries are on time, and negative when there are tardy deliveries. Thus, a_t can be understood as the amount of orders released ahead of their due date and is of most interest when negative as it then quantifies tardiness.

Outcome – Service delivery performance

Let *availability,* S_1, denote the fraction of periods in which all expiring demand has been satisfied on time. This is analogous to inventory systems where S_1 measures the probability of not experiencing a stockout in an order cycle (Axsäter, 2006, p. 94). According to this definition, in a given period *t*, $S_1(t) = 1$ if $a_t \geq 0$, otherwise $S_1(t) = 0$. The expectation of $S_1(t)$ is simply,

$$S_1 = \mathrm{P}(a_t \geq 0) = \Phi(\mu_a / \sigma_a), \tag{3}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution and μ_a and σ_a are the mean and standard deviation of *a*, respectively.

In inventory control, the fill rate, S_2, denotes the long-run fraction of demand that can be filled immediately from stock (Sobel, 2004). The equivalent for service operations, which we term the *service rate*, measures the fraction of demand satisfied on or before the promised delivery date. Therefore, we define the service rate as $S_2 = E[\mathrm{DOT}] / E[d_{t-Q}]$, where DOT is the demand *Delivered On Time* in period *t*. DOT depends on both the expiring demand d_{t-Q} and the schedule adherence a_t. Let us introduce the variable *modified schedule adherence* $w_t = a_t + d_{t-Q}$, which describes DOT when $0 \le w_t \le d_{t-Q}$. Should $w_t < 0$, none of the demand that expires in *t* is satisfied, and if $a_t > 0$ all demand that expires in period *t* is satisfied. The on-time deliveries can then be expressed as,

$$\text{DOT} = \begin{cases} 0 & w_t < 0 \\ w_t & 0 \le w_t \le d_{t-Q} \\ w_t - a_t & 0 < a_t \end{cases}. \tag{4}$$

The expectation of DOT is provided by Hedenstierna (2016, pp. 120–122) as $E[\text{DOT}] = E\left[(w)^+\right] - E\left[(a)^+\right]$ for arbitrary nonnegative demand distributions. This can be used as an approximation when demand is normally distributed with a low probability of negative demand; then it takes the form (Disney et al., 2015):

$$S_2 = \frac{E[\text{DOT}]}{E\left[d_{t-Q}\right]} = \frac{\sigma_w\left(-\mu_w/\sigma_w\right) - \sigma_a G\left(-\mu_a/\sigma_a\right)}{\mu_d}, \tag{5}$$

where $G(x) = \int_x^\infty (v-x)\varphi(v)\,dv = \varphi(x) - x\left[1-\Phi(x)\right]$ is the unit normal loss function (Axsäter, 2006). A useful property of the service rate is that when expressed as a function of *Q* it reflects the cumulative density function of the time spent in the order book. Before identifying expressions for the variables required in (5), we shall now introduce a capacity cost model.

Outcome – Capacity costs

We assume that we pay for a fixed amount of regular capacity from labour, even if the actual production requirements are sometimes less than this. Overtime charges are incurred if order releases exceed the normal capacity in any single period. The capacity cost C_t in a period can be expressed as $C_t = c_1 \cdot z + c_2\left(o_t - z\right)^+$, where c_2 is the labour cost per hour under overtime hours, c_1 is the hourly labour cost during normal hours (with $c_1 < c_2$), and *z* is the nominal (guaranteed) capacity per period, also expressed in hours, μ_o is the average production rate, and σ_o^2 the corresponding variance. The cost function we use follows Hosoda and Disney (2012), and its expectation is

$$\text{E}(C) = c_2\sigma_o\varphi\left[\Phi^{-1}\left(\tfrac{c_2-c_1}{c_2}\right)\right] + c_1\mu_o, \tag{6}$$

when using the optimal capacity is $z^* = \mu_o + \sigma_o\Phi^{-1}\left(\frac{c_2-c_1}{c_2}\right)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative density function (Hosoda and Disney, 2012). Note that the capacity costs are a linear function of both the mean μ_o and the standard deviation σ_o. When $\sigma_o \to 0$, $z^* \to \mu_o$, furthermore, z^* is an increasing function of σ_o when $c_2 > 2c_1$ and a decreasing function of σ_o when $c_1 < c_2 < 2c_1$.

Intervention – Order book control

As we have a certain time to fulfil demand, we will entertain a policy that pools demand in the order book, to enable a steady release rate. We choose the general linear policy of Balakrishnan et al. (2004):

$$o_t = \sum_{n=0}^{\infty} \theta_n d_{t-n}, \tag{7}$$

where θ_n is simply the covariance function between orders and demand, and $\sum_{n=0}^{\infty} \theta_n = 1$. As d_t is a sequence of independent random variables, the expectation and variance can be taken directly, providing (8) and (9) in Table 2. For the schedule adherence, we may use (2) to express a_t as a weighted sum of past demands $(\ldots, d_{t-1}, d_t)$ and then take the expectation and the variance, providing (10) and (11). Here $H(\cdot)$ is the Heaviside step function. The modified schedule adherence w_t is directly obtainable from a_t, leading to (12) and (13). These results are general, but not concrete. We shall now consider two pragmatic policies that are special cases of (7): The moving-average policy and the proportional policy.

Table 2 – Properties of the general linear policy

	General linear policy	
μ_o	μ_d	(8)
$\frac{\sigma_o^2}{\sigma_d^2}$	$\sum_{n=0}^{\infty} \theta_n^2$	(9)
μ_a	$\mu_d \sum_{n=0}^{\infty} \left[\left(\sum_{j=0}^{n} \theta_j \right) - H(n-Q) \right]$	(10)
$\frac{\sigma_a^2}{\sigma_d^2}$	$\sum_{n=0}^{\infty} \left[H(n-Q) - \sum_{j=0}^{n} \theta_j \right]^2$	(11)
μ_w	$\mu_d \sum_{n=0}^{\infty} \left[\left(\sum_{j=0}^{n} \theta_j \right) - H(n-Q-1) \right]$	(12)
$\frac{\sigma_w^2}{\sigma_d^2}$	$\sum_{n=0}^{\infty} \left[H(n-Q-1) - \sum_{j=0}^{n} \theta_j \right]^2$	(13)

Under the *moving-average (MA) policy*, orders are simply a moving average of the last $\beta \in \mathbb{N}^+$ periods demand (including the current). One benefit of this policy is that *all* demand will be satisfied if $\beta \le Q+1$, but if β is greater there is a risk for tardiness. The *proportional policy (ES, for exponential smoothing)* releases a fixed fraction of the order book each period, i.e. $o_t = \alpha(b_{t-1} + d_t)$, where $\{\alpha \in \mathbb{R},\ 0 < \alpha \le 1\}$. Table 3 shows the properties of these policies, following directly from the general linear policy in Table 2.

For the moving-average policy with perfect service, $\beta = Q+1$, the optimal capacity level is increasing in *Q*, and tends to μ_d as $Q \to \infty$. For both policies (MA and ES), assuming that only α or β are altered, S_1 and S_2 are decreasing functions of σ_o.

Table 3 – Properties of two popular policies from the literature

	Moving average (MA)	Proportional (ES)
θ_n	$1-H(\beta-n)$	$\alpha(1-\alpha)^n$
μ_o	μ_d	μ_d
$\frac{\sigma_o^2}{\sigma_d^2}$	β^{-1}	$\alpha(1-\alpha)^{-1}$
μ_a	$\mu_d\left[Q-(\beta-1)/2\right]$	$\mu_d\left[Q-(1-\alpha)/\alpha\right]$
$\frac{\sigma_a^2}{\sigma_d^2}$	$\frac{1+2\beta^2}{6\beta}+Q\left(\frac{1+Q}{\beta}-1\right)-\frac{1}{2}$	$1+Q+\frac{3-2\alpha}{\alpha(\alpha-2)}+\frac{2(1-\alpha)^{1+Q}}{\alpha}$
μ_w	$\mu_d\left[Q+1-(\beta-1)/2\right]$	$\mu_d\left[Q+1-(1-\alpha)/\alpha\right]$
$\frac{\sigma_w^2}{\sigma_d^2}$	$\frac{13+2\beta^2}{6\beta}+Q\left(\frac{3+Q}{\beta}-1\right)-\frac{3}{2}$	$2+Q+\frac{3-2\alpha}{\alpha(\alpha-2)}+\frac{2(1-\alpha)^{2+Q}}{\alpha}$

Numerical analysis

Consider normally distributed demand with $\mu_d = 10$ and $\sigma_d = 1$. We shall study the effect of *Q*, so the physical lead-time is of no import, as is the case with the cost factors c_1 and c_2, because the capacity cost is proportional to the standard deviation of orders.

The first trade-off to consider is that between availability and capacity costs, illustrated in Figure 3. A prominent feature is the miniscule difference in cost between some near-zero and near-perfect service configurations. Similar results hold for the service rate, as Figure 4 illustrates. Both of these figures suggest that for a fixed service level, the variable *Q* offers diminishing returns in terms of capacity cost reduction. We can show this analytically for the moving average policy with perfect service, as the cost reduction achieved by incrementing *Q* is $\sqrt{(Q+1)/(Q+2)}$, which tends to zero as *Q* increases. This is made clear in Figure 5, which illustrates the capacity cost required to maintain very high service levels for both policies. In this setting, the proportional smoothing policy offers a significant cost advantage over the moving-average policy.

Managerial implications

The evidence so far suggests three possible configurations for the 3D-printing operation: plan for perfect service with the moving-average policy, at a considerable cost; plan for near-perfect service with the proportional policy, at significantly lower cost than perfect service; or set a completely level schedule with α close to zero, accepting poor service levels. The proportional alternative seems to be the most realistic, as it ensures high service with significant production smoothing. If, as in Shapeways case, there is an option for expediting, separate order book and control policies (with different α-values) must be maintained for each delivery mode. As the same capacity can be used for different promised lead times, a capacity pooling effect can be exploited, see Hedenstierna and Disney (2012).

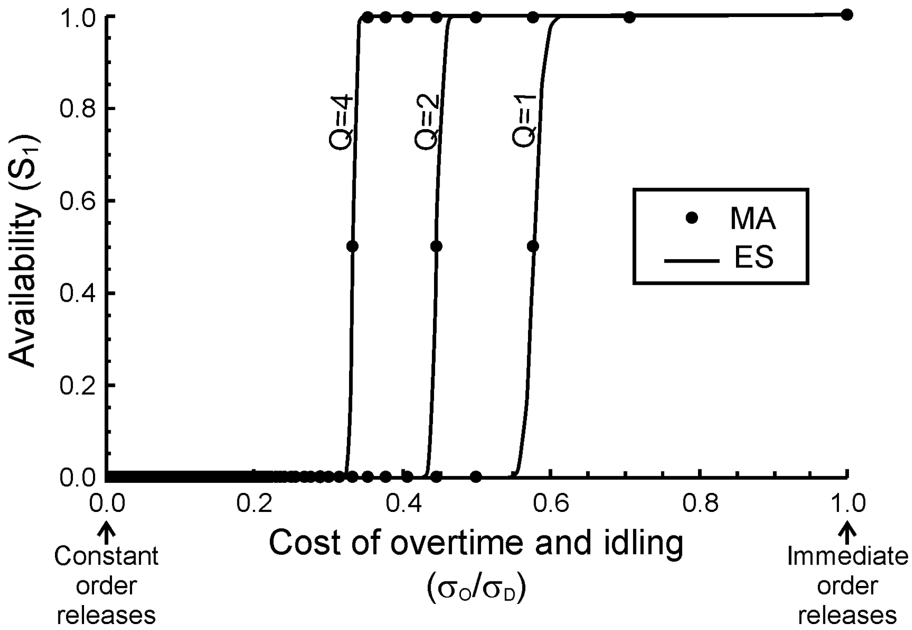

Figure 3 – Availability versus capacity cost

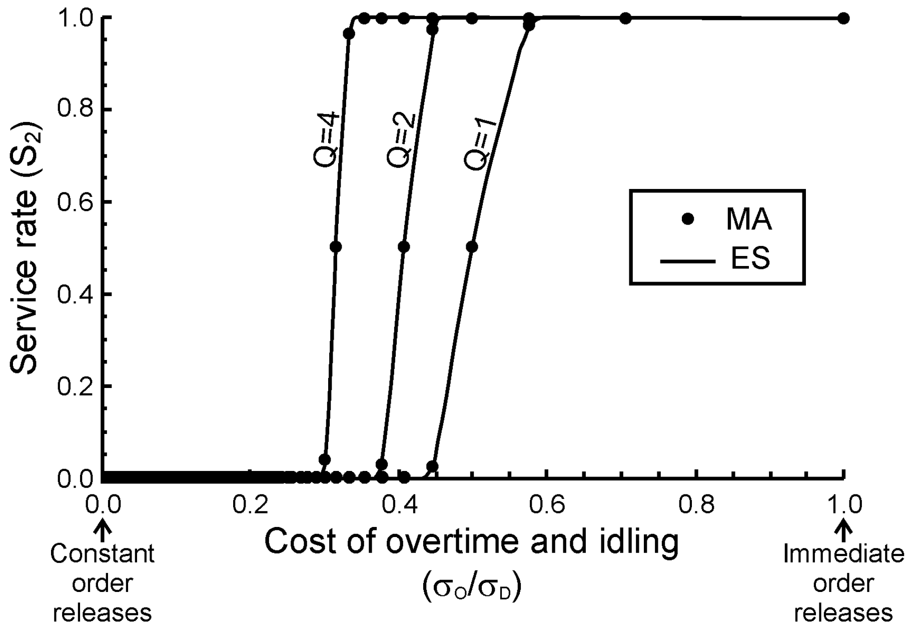

Figure 4 – Service rate versus capacity cost

Conclusion

At first sight, it would appear that service operations require an agile production system. However, an order book can release work orders to the production system smoothly, allowing one to make better use of capacity in a lean production mode. We have provided expressions for the first and second order moments of a general smoothing policy for order book management. We have also provided the moments for the moving average and proportional order book management policies.

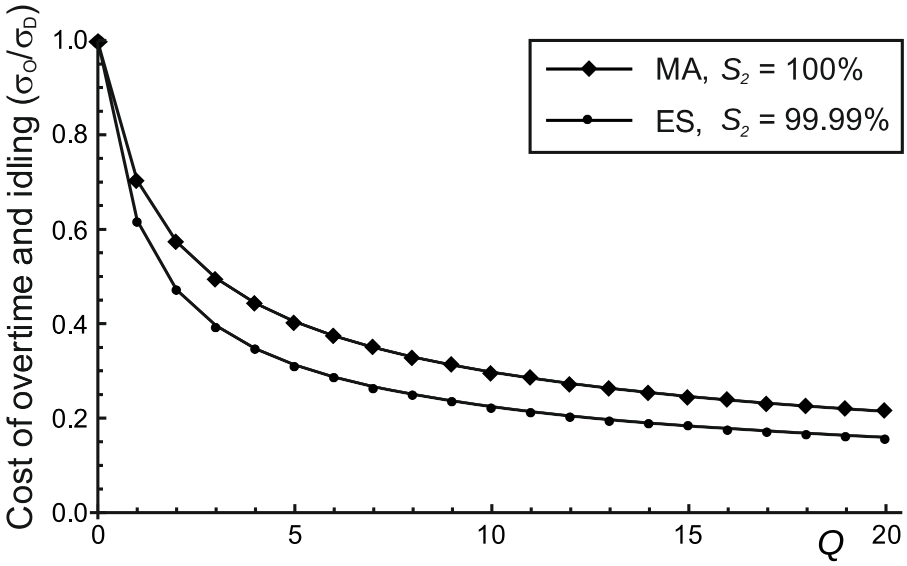

Figure 5 – The cost associated with high service levels for each of the policies

We have developed a unique metric for measuring on-time delivery called the service rate. Furthermore, we have provided a measure of capacity costs based on guaranteeing workers a nominal wage each week as well as the opportunity to gain overtime during peaks of high demand. Our analysis reveals that as Q decreases, the cost to maintain a given service rate increases. However, as the target service rate decreases, the capacity costs decrease. Interestingly, the optimal capacity level can be an increasing or decreasing function of σ_o depending on whether $c_2 > 2c_1$.

While the CIMO framework guided this research, we had to revisit the intervention stage after having defined the mechanism and the output in unambiguous terms. In this way, the mechanism and the intervention gave rise to an integrated system for providing the intended outcomes. To identify unintended outcomes, we would have to test the policy on actual order book data from a service operation. This might reveal new insights, not predicted by the model, that could aid in the development of a refined intervention.

Acknowledgements
Carl Philip Hedenstierna gratefully acknowledges the support of the Norwegian Research Council and the BIA program for financial support of the RSC2020 project, which enabled this study.

References

Anderson, E.G., Morrice, D.J. and Lundeen, G. (2005), “The ‘physics’ of capacity and backlog management in service and custom manufacturing supply chains”, *System Dynamics Review*, Vol. 21 No. 3, pp. 217–247.

Axsäter, S. (2006), *Inventory Control*, Springer, New York.

Balakrishnan, A., Geunes, J. and Pangburn, M.S. (2004), “Coordinating Supply Chains by Controlling Upstream Variability Propagation”, *Manufacturing & Service Operations Management*, Vol. 6 No. 2, pp. 163–183.

Denyer, D., Tranfield, D. and Van Aken, J.E. (2008), “Developing design propositions through research synthesis”, *Organization Studies*, Vol. 29 No. 3, pp. 393–413.

Disney, S.M., Gaalman, G.J.C., Hedenstierna, C.P.T. and Hosoda, T. (2015), “Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand", *International Journal of Production Economics*, Vol. 170, Part B, pp. 501–512.

Forrester, J.W. (1961), *Industrial Dynamics*, MIT press, Cambridge, MA.

Hedenstierna, C.P.T. (2016), *Staggered Deliveries in Production and Inventory Control*, PhD thesis, Cardiff University, available at: http://orca.cf.ac.uk/87502/ (accessed 21 March 2016).

Hedenstierna, C.P.T. and Disney, S.M. (2012), "Impact of scheduling frequency and shared capacity on production and inventory costs.", *Pre-Prints of the 18th International Working Seminar on Production Economics*, Vol. 2, Innsbruck, Austria, pp. 277–288.

Holmström, J., Holweg, M., Khajavi, S.H. and Partanen, J. (2016), "The direct digital manufacturing (r)evolution: definition of a research agenda", *Operations Management Research*, Vol. 9 No. 1–2, pp. 1–10.

Hosoda, T. and Disney, S.M. (2012), "On the replenishment policy when the market demand information is lagged", *International Journal of Production Economics*, Vol. 135 No. 1, pp. 458–467.

Kingsman, B.G., Tatsiopoulos, I.P. and Hendry, L.C. (1989), "A structural methodology for managing manufacturing lead times in make-to-order companies", *European Journal of Operational Research*, Vol. 40 No. 2, pp. 196–209.

Mingers, J. (2015), "Helping business schools engage with real problems: The contribution of critical realism and systems thinking", *European Journal of Operational Research*, Vol. 242 No. 1, pp. 316–331.

Pawson, R. and Tilley, N. (1997), *Realistic Evaluation*, Sage, Thousand Oaks, CA.

Shapeways. (2016), "Materials Shipping Status Page - Shapeways", *Shapeways.com*, available at: http://www.shapeways.com/materials/material-status (accessed 28 May 2016).

Sobel, M.J. (2004), "Fill rates of single-stage and multistage supply systems", *Manufacturing & Service Operations Management*, Vol. 6 No. 1, pp. 41–52.

Sterman, J.D. (2000), *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill Education, New York.

Stevenson, M., Hendry, L.C. and Kingsman, B.G. (2005), "A review of production planning and control: the applicability of key concepts to the make-to-order industry", *International Journal of Production Research*, Vol. 43 No. 5, pp. 869–898.

Wikner, J., Naim, M.M. and Rudberg, M. (2007), "Exploiting the order book for mass customized manufacturing control systems with capacity limitations", *Engineering Management, IEEE Transactions on*, Vol. 54 No. 1, pp. 145–155.