

**How & Why**  
**Demand Driven Supply Chains**  
**Transform**  
**Operations Performance**

There is a major opportunity for the vast majority of companies to make significant improvements in their Supply Chain & Operations performance across

- Cost - reductions of c20%
- Inventory – reductions of between 30% and 50%
- Service – achieve planned levels

The opportunity can be taken by adopting new, but well established, replenishment processes in a robust and sustainable manner - and it has nothing to do with improving forecast accuracy!

Most Supply Chains have long been driven using forecasts for both Planning and Replenishment. While the former is obvious, use of the forecast to directly drive Replenishment (often called ‘forecast push’) is actually plain wrong and results in serious under-performance across the key metrics of cost and inventory – and often service too.

This paper explains why ‘forecast push’ is so ineffective and both why and how the use of demand driven replenishment can transform ‘end to end’ supply chain performance.

**Queuing Theory**

The relevant underpinning of Supply Chain Management and Operations is Queuing Theory which recognizes that any system involving at least one conversion stage, be it a supermarket checkout counter, a factory’s production line, a distribution channel or supplier, has finite throughput capacity (ie. is a constraint) and always suffers from queues. If demand exceeds capacity the queue will grow indefinitely. Even if demand is below 100% capacity, however, there is still a queue. This is because variability in the rate of arrivals at a work station and / or its processing time will cause both lost capacities and demand spikes and the development of a queue with finite average length (NB. if there is plenty of spare capacity, however, no queue will actually be observable for most of the time).

The average waiting time in one of these queues, for a given level of capacity utilization, is directly proportional to the flow variability, not the processing time. On the other hand, the relationship between the queue length and the level of capacity utilization is exponential, and very noticeably so when variability or utilization is high (see Fig. 1).

The relationship between queues, variability and capacity utilization can be modeled using the VUT equation

$$\text{Average wait in queue} = \text{V} \text{ariability} \times \text{U} \text{tilization} \times \text{T} \text{ime processing} \quad (1)$$

We know that queue time is a major component of lead time and through Little’s Law we also know that lead time is directly related to inventory level

$$\text{System Lead Time} = \text{System Inventory} / \text{System Throughput}$$

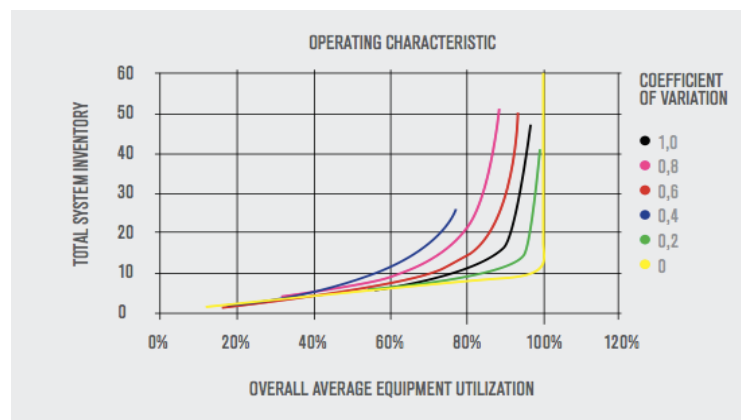


Fig. 1 Shows how queue / lead time/ WIP increases with variability and exponentially at high levels of capacity utilization

Little’s Law and the VUT equation are at the heart of Hopp & Spearman’s ‘Factory Physics’ (2) and are used to explain the underlying rationale for why the tools of Lean are so effective. Conceptually we can understand that if supply was unconstrained and totally flexible we could meet any demand (ie. volume and mix) and achieve perfect continuous flow without holding any static stock, either in process or as finished goods. In the real world, any inability by the supply conversion process to respond instantaneously to demand changes implies the existence of a constraint (eg. a processing machine) that inevitably has flow variability which causes the creation of cost generating buffer. The default buffer response to variability is an immediate increase in lead time due to queuing, ameliorated by use of spare capacity (if any); management may respond with an increase in capacity

(eg. overtime) and finished goods inventory up-sizing often follows to prevent service issues. Lean activities such as TPM, TQM, SMED, 5S, Standard Work, DFM etc eradicate not only wasted effort and cost, they also reduce flow variability, including that caused by excessively big batches (thereby supporting flexibility), which means less cost generating buffer is created or needed. And this is particularly valuable when working at high levels of capacity utilization (Fig. 1). In consequence, the Lean supply chain uses less unplanned capacity and can operate at higher levels of capacity utilization (ie. lower costs), with shorter queue / lead times (ie. greater responsiveness) and lower levels of stock. Reducing flow variability is actually synonymous with minimizing cost and inventory while increasing responsive flexibility, and is why Hopp & Spearman describe Lean as

“fundamentally about minimizing the cost of buffering variability” (3)

Lean’s potential has rarely been fully exploited because most companies, through their replenishment execution processes, are still unwittingly introducing huge amounts of variability into their supply chains and operations. The source of that variability is the use of forecasts to drive replenishment execution using DRP/MRP.

**‘Forecast Induced’ Variability**

It is well known that all forecasts are wrong and 80% portfolio accuracy (ie. 20% wrong) is considered ‘world class’. Due to the 20:80 rule, such performance means that most medium and low volume sku’s (usually the majority) actually achieve accuracies that are somewhat worse, not least because at lower volume levels, variability tends to be higher and so too is forecast inaccuracy. Manufacturing or purchasing schedules based upon inaccurate sku forecasts lead to the production of unbalanced stocks with potential service issues, and expediting inevitably follows as Planners respond to exception messages. Changing schedules at short notice is often perceived as a sign of ‘agility’. Unfortunately they cost money and every factory schedule that is interrupted by an expedited ‘hot list’ item suffers from ‘forecast induced’ variability.

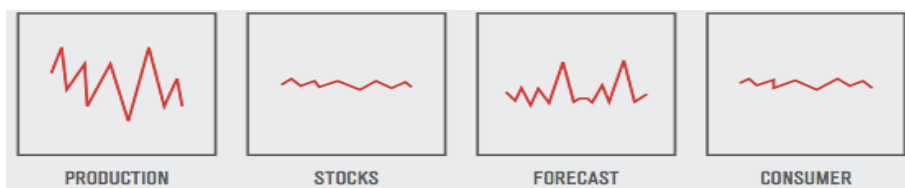


Fig.2 Production schedules suffer ‘Planner induced variation’ when driven by forecasts

Production schedule changes cause costly unplanned machine change overs, interrupt flow, cause congestion and increase lead times with knock on effects upon other schedules up and down the factory routings. In consequence, average lead times increase and become volatile (contrary to the DRP/MRP assumption of fixed lead times, therefore causing a service risk) and stock becomes both excessive and unbalanced with service issues often continuing to occur. The further up the supply chain, and away from end customer demand one goes, these problems amplify due to latency (ie. bullwhip caused by batching, response delays and misinterpreted demand signals).

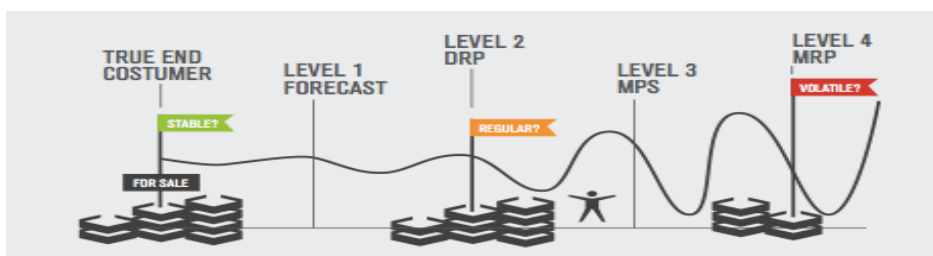


Fig. 3 Variability is amplified by the forecast as it is cascaded up the supply chain

A typical reaction to this set of problems is for management to increase time buffers with ERP system parameters such as lead time and time fences. Unfortunately these just make matters worse by leading to more work being released to the factory floor; this reduces capacity buffer which significantly increases both lead time and stock while reducing responsiveness. Another common activity is for companies to spend even greater sums upon their forecasting technology: the most recent incarnation being that of ‘demand sensing’ and its associated technique of ‘demand shaping’, the latter of which is implicit recognition that there is a major limit on what the

former can actually achieve! And, finally, some companies use ‘optimisation’ technology to manage their schedules not realising, of course, that they’re actually optimising the wrong replenishment model and significantly under performing.

**Transforming Supply Chain & Operations Performance with ‘Demand Driven’ Replenishment**

In addition to the ‘shop floor’ Lean that minimises factory supply variability, Operations and Supply Chain leaders can significantly reduce the performance destroying impact of the variability that they generate themselves through their use of inaccurate forecasts to drive replenishment execution.

True ‘agility’ is designed into a supply chain and is its ability to be genuinely flexible and autonomously responsive to real demand, and its variations, with minimal cost generating buffers that are of the right size and in the form that best serves both the company and its customers.

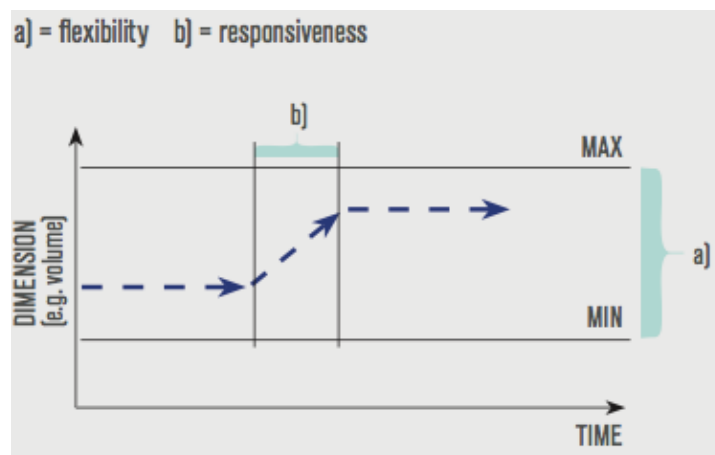


Fig. 4. True agility is the combination of “designed in” flexibility and responsiveness

The demand driven approach uses replenishment processes that are consistent with this description and companies that adopt these techniques sustainably achieve their desired service levels with inventory reductions of between 30% to 50% and cost savings of c20%.

The demand driven process uses a segmented approach to replenishment technique selection based upon the item’s demand profile in terms of volume and variability (see Fig 6). The key options are, broadly, the following:

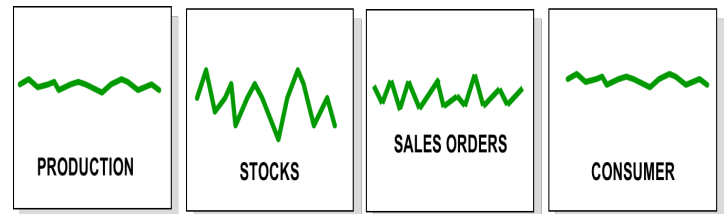


Fig 5. Demand driven processes support stability with high capacity utilization

- High Volume / Low Variability – where demand is high and relatively stable, as it often is for mature products and for upstream items before sku specific customisation takes place, supply can be leveled at a suitable fixed rate, subject to periodic review or using min-max logic. As with all ‘make to stock’ replenishment techniques, this rate-based or level schedule technique requires some inventory buffer.
- Medium Volume / Medium Variability – these, usually the majority of items in a portfolio, use consumption based pull whereby inventory is positioned in the supply chain to decouple processes and minimize lead times. Supply activities at each work station up and down the supply chain are scheduled according to an efficient sequence / cycle, and stock targets (buffer plus average demand over lead time) calculated so that the quantities triggered, rounded as necessary, replace what has been taken from the location immediately down stream without stock outs occurring. The consequence is that both the timing and quantity of replenishment responds autonomously in line with demand, requiring minimal average inventory, while capacity utilisation is kept high and level loaded. The sequence predictability, virtually guaranteed availability of components and lack of schedule interruptions (see Fig. 5) mean that this technique brings significant cost saving stability to Operations. The technique can be used even if demand has trend or is seasonal so long as the replenishment parameters reflect future demand patterns appropriately.

- High Volume / High Variability - when demand is high and genuinely volatile so that it is uneconomic to provide a stock buffered service (eg. response to tenders, significant promotions and other events), time buffered responses such as ‘make to order’ or ‘assemble to order’ are options for event management. Both these ‘postponement’ techniques, if supported by appropriate ‘design for manufacture’ and asset configuration, can deliver very cost effective and quick response. In these situations, use of an event forecast to drive advance stock build is also an option, but this is very different from traditional ‘forecast push’ MRP.
- Low Volume / High Variability – low volume items with or without high variability can usually be serviced ex stock from high demand coverage batch volumes (eg. 2 bin systems), especially if the item is of low value. Otherwise, and depending on the demand pattern, MTO, ATO or Poisson based buffer techniques can be used.

These techniques are frequently found in combination at different levels of the supply chain. Level schedule is often used upstream where materials are common across many SKU’s which themselves are replenished using consumption based pull. Similarly, in a contract manufacturing business, material supply may be managed using consumption based pull to support a fast ATO response to customer orders. Such companies might also add value by providing their customers with a VMI service using collaborative demand driven techniques. This also has the benefit of allowing them to respond to relatively stable end user demand, instead of lumpy customer orders, which enables them to more efficiently level load their entire supply chain using level schedule or consumption based pull.

**The New Planning**

The implementation of demand driven techniques has a significant impact upon the Planner’s role. Instead of constantly expediting and re-cutting inaccurate forecast driven replenishment schedules, the Planner can concentrate properly upon value add activities such as Supply Chain Conditioning.

‘Conditioning’ enables the supply chain to autonomously respond to demand, it is undertaken at each stock level echelon through demand profile analysis and replenishment technique selection (see Fig.6), and calculation of rate and stock targets. These activities are performed regularly (eg. technique selection annually and parameters monthly) but changes are very much by exception – only around 5% of targets may need changing at any one time.

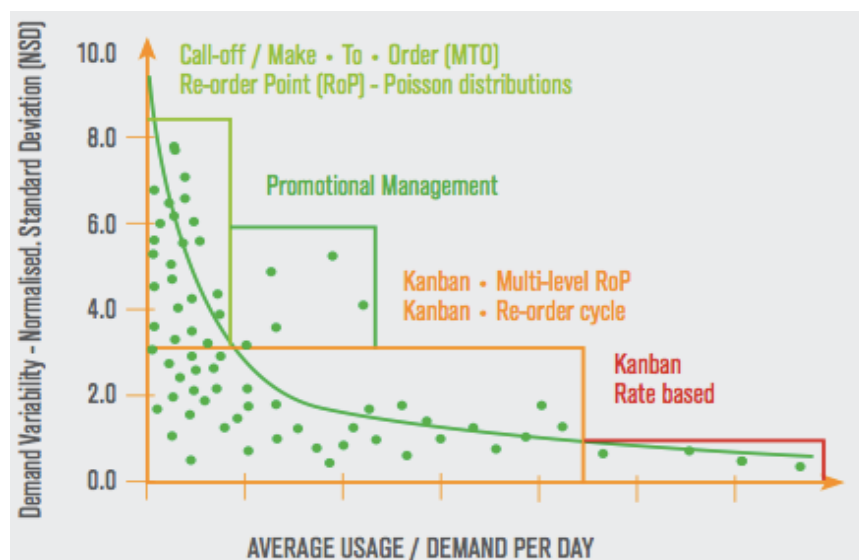


Fig 6. The criteria for echelon level replenishment technique selection is the items demand volume & variability profile

Planners will continue to be involved with NPL, Phase Out and Event Management, albeit the latter less frequently, and as the demand driven principles can be applied to manage replenishment across company boundaries, Planners will also have time to work a lot more collaboratively with key suppliers and customers.

In addition, of course, forecast based S&OP continues to be an essential supply chain support process. As it is forward looking it becomes part of the rate and stock target setting process as well as aligning material and capacity availability with the demand plan. In fact use of the demand driven approach improves the value of S&OP by eliminating 'fire fighting' and use of unplanned capacity thereby allowing the process to focus more accurately upon the future and to be less short term and review orientated in nature.

### **The Future for the Demand Driven 'Agile Supply Chain'**

The concepts behind demand driven replenishment are not new, they have been around since 1961 (and before at Toyota) when Kingman formulated the VUT equation, a proof of Little's Law was first published and Forrester and Burbidge initially wrote about 'bullwhip'. In recent years Cranfield University's Martin Christopher has also written extensively about them with Dennis Towill from the Cardiff Lean Enterprise Research Centre.

The reasons for the current relative scarcity of the demand driven approach are that it is counter intuitive and difficult to implement successfully across modern enterprises without appropriate and robust software support.

Its outlook, however, is very positive as its rationale in Queuing Theory and Factory Physics becomes more widely known. And, with the emergence of functionality rich, scalable and globally tested 'Software as a Service' systems, especially designed to support demand driven replenishment as well as forecast based planning (using simple FTP data exchange with ERP systems), it is now possible to quickly and inexpensively simulate, pilot and implement these techniques both within and across company boundaries. Some very major companies in the CPG and Life Science sectors are proving extremely successful pioneers in the adoption of the 'demand driven' approach.

### **References**

1. The full Kingman or VUT formula, valid for a single server queue, is

$$\text{Average waiting time in queue} = \left( \frac{\rho}{1 - \rho} \right) \left( \frac{c_a^2 + c_s^2}{2} \right) \tau$$

where T is the mean processing time, p is the utilization (ie. mean processing rate/mean arrival rate), Ca is coefficient of variation for arrivals and Cs is coefficient of variation for service/process times

2. Hopp & Spearman. 'Factory Physics'. McGraw Hill Intl. 1995
3. Hopp & Spearman. 'To pull or not to pull: what is the question' *M&SOM* 6.2. Spring 2004 p133-148

**Simon Eagle**

[simon@smartchainllp.com](mailto:simon@smartchainllp.com)

**John Earley**

[john@smartchainllp.com](mailto:john@smartchainllp.com)

[www.smartchainllp.com](http://www.smartchainllp.com)

**SmartChain International**, together with our software partners **Orchestr8**, have deep experience and the capability to help our clients make 'Demand Driven Agility' a sustainable reality delivering tangible benefits to their businesses and customers. We provide a complete suite of services covering: process design, change management, organisational structure and role design and software configuration specifically tailored for each client. If you are interested in learning how 'Demand Driven Agility' can transform your organisation's Operations and Supply Chain performance, please contact us using the details above.