INTRODUCTION

This document addresses the question of how many spare parts (or any consumable resource) should be ordered to keep the maintenance line moving.

In short, the answer is that you order according to the ***average*** consumption, **BUT** you maintain enough "*Safety Stock*" to compensate for variability in usage.

Safety Stock is the number of excess serviceable parts (or consumables) kept on-hand to prevent "stock-outs": that is, needing one of a particular item but not having it available. Most of this document deals with calculating the correct amount of safety stock to maintain.

SAFETY STOCK

There are three primary reasons why month-to-month demand is uncertain, and thus why safety stock should be maintained. They are:

1) The inherent variability in demand. This variability results directly from unplanned or unscheduled "failures" of fielded items, or other unscheduled events. Since we do not know exactly when unscheduled events will occur, we cannot account for them with absolute certainty: the best we can do is to model them as a random variable. This in turn gives rise to variability in the parts requirements forecasted by any simulation run.
2) Incorrectly forecasted usage or maintenance practices. For example, customers tell us they will operate the equipment 100,000 hours next year, but instead, they operate it 130,000 hours. Another example would be a management-directed increase in the build-policy that was unanticipated by the forecasters.
3) Unpredictable issues, such as a cleaning process gone bad that damages a batch of parts; or an entire pallet load of parts getting dropped. Still another is an unexpected change in the percentage of parts being discarded for out-of-inspection-limits.

Another unpredictable issue that can diminish parts availability is a disruption in the upstream supply chain. Safety stock should (sometimes – depending on the criticality) be maintained to deal with this unknown as well.

In this document we address only the inherent variability in demand (the first reason given above) and refer to the number of parts that should be maintained to address this as "*Parts Cache*". The other reasons are maintenance or operations issues that cannot be predicted by the planner/forecaster. Nevertheless, planners can – and probably should maintain safety stock levels BEYOND Parts Cache to help avert stock-outs caused by those other reasons.

CALCULATING REQUIREMENTS

In the following discussion, we assume that a Monte Carlo simulation model – such as the model(s) available from Reliable Forecasts™, are being used to predict maintenance requirements and spare-parts requirements.

A Monte Carlo model runs the simulation multiple times, and gathers the results of the multiple runs into statistics. For example, a single run of the simulation model may indicate that 100 end-items will generate to the central repair facility for overhaul. Another run, with exactly the same input parameters, may indicate that 120 end-items will generate. The difference is because we cannot account for random (unscheduled) events deterministically – we can only account for them as random variables, leading to probabilistically different answers. That is, if we run the simulation several hundred times, instead of a single answer (like the 100, or 120 above) we will get an *AVERAGE* (or *mean*) answer, along with a measure of the spread of the answer – like the *Standard Deviation*. Again, for example, after running the simulation hundreds of times we may get that the AVERAGE number of end-items that generate for overhaul will be 105, with a standard-deviation of 10 (about plus-or-minus 15).

As noted in the introduction, we will order spare-parts and consumables according to the AVERAGE, but we must maintain enough safety stock so that if an unusually high number of end-items require maintenance, we will be able to keep up with demand. This might dictate that we ORDER MORE THAN WE THINK WE NEED for several months, and allow the excess to build-up in our internal supply system.

CALCULATION DETAILS

As noted earlier, we order spares according to the average demand, but we cache enough spares to accommodate fluctuations in demand. The question is: how much should we cache?

Suppose we run the Monte Carlo forecast 100 times. The highest demand run is the 99'th percentile run. But what does "highest" mean? One run may have the highest demand over the full simulation period, but another run may have the highest demand over a 12-month period. And which 12-month period are we talking about? Each 12-month period may have a different run that produces the highest demand for that period. And why did we choose 12-months? Shouldn't we be concerned about the highest demand for each month?

The answer is: Yes, we need to be concerned about the highest demand for each month. But merely summing the highest demand from month-to-month will produce an unrealistically high demand. In reality (as in the simulation runs as well), high-demand months tend to be followed by low-demand months. Just picking the highest demand run for each month is the wrong way to compute the required parts-cache.

Instead, the process is as follows: Beginning now (the date of the forecast run), we look for the run that gives the highest demand for the first month. Then we look for the run that gives the highest ***cumulative*** demand over the first two months, then for the highest cumulative demand over the first three months, etc. We do this for every number of months, always beginning with the date of the forecast run. This series of numbers tells us the maximum number of parts (resources) that are likely to be needed from now through that month. We base our purchasing decisions on this series of numbers.

Question: Where do we stop? 24-months? 36-months?

Next question: What do we do with these numbers? How do we use them to compute the required Parts Cache levels?

Final question: What should we do when the demand spikes (and the number of on-hand assets falls)?

To answer the first question, consider this: If you knew TODAY that your on-hand levels were too low, when would your first opportunity to correct those levels be? Answer: lead-time away, if you ordered today.

Imagine this scenario: Today you have 8 (extra) on-hand assets. You consume an average of 5/month, and you have 5/month on-order scheduled to be delivered. You run the forecast and it agrees that for the next 24 months the average consumption will be 5/month. Then you look for the highest cumulative demand run for the next 24 months. It says that the maximum anticipated consumption over the next 24 months will be 132, which is 12 over the expected value of 24*5 = 120, and 12 over what you expect to receive since you only have 5/month on order. Since you only have 8 on-hand right now, there is a chance that you'll be negative 4 by month number 24, unless you can get more assets before then. If the asset's lead time is 14 months, you still have 10 more months to plus-up your order by 4 assets. Since you probably don't want to suddenly order 5+4=9 assets for one month, you probably want to plus-up by one asset each month until you have 12 (extra) assets on-hand.

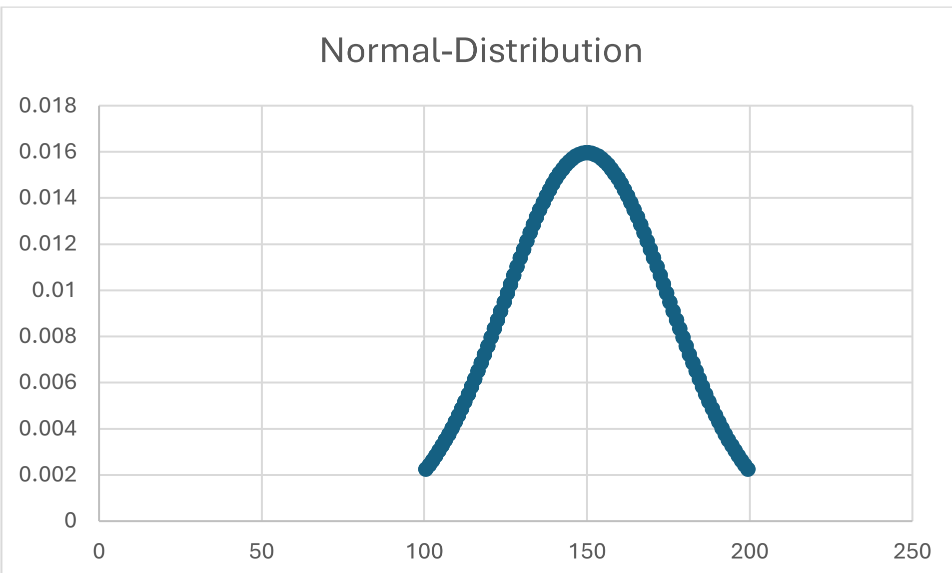
That scenario kind-of answered both question:

- How far out should we compute max-cumulative demand? Answer: At least lead-time out.
- How do we translate max-cumulative demand into required Parts Cache levels? The Parts Cache level should be equal to the MAX-CUMULATIVE CONSUMPTION minus the AVERAGE-CUMULATIVE CONSUMPTION.

We said "kind-of" answered the questions because there is still a critical part missing – the *CONFIDENCE* of our answer. How confident are we that this "MAX minus AVERAGE" is the correct answer?

The "MAX-CUMULATIVE minus AVERAGE-CUMULATIVE" formula is correct for the Parts Cache, but how confident are we that the MAX-CUMULATIVE value is correct? **IF** we have setup the model correctly, and **IF** some outside effect doesn't throw a wrench into the works (like the operators not using the equipment the way they told us they would), then the simulation model should actually be simulating reality, including the random aspects of reality. If we have captured the randomness of reality accurately, then each run of the simulation model represents one *ACTUALLY POSSIBLE OUTCOME*. Conceptually, if we were to run the model billions of times, eventually one simulation run would exactly match what will actually happen in the future. Of course this is unrealistic, and we wouldn't know which one of the billions of runs is the correct one anyway. Instead, the results should form something like a *Normal-Distribution*, and we anticipate that the *MOST LIKELY* outcome of reality will be something near the mean of the simulation run results.

But the MOST LIKELY outcome is what we use for our AVERAGE-CUMULATIVE value in the above formula, not the MAX-CUMULATIVE. The max-cumulative value comes from the right-tail of the results distribution curve.



Normal Distribution with Mean = 150 and Std-Dev = 25

The above graph is from 100 data points. In this case, the "99th percentile" point is the last value (highest value) at approximately 199. The 95th percentile would be the 95th point, at about 195.

This is how we simulate the *Confidence Level* of our Max-Cumulative. If we want a 99% confidence answer, we run the model 100 times and take the highest answer. A better approach would be that we run the model 1000 times, and take the 990th highest result. Similarly, if we wanted a 95% confidence answer, we would run the model 100 times (or 1000 times) and take the 95th highest (or the 950th highest) run results. Since the model is simulating actual reality, only 5% (one in twenty) of all possible real outcomes should be above the 95% confidence answer. A section below addresses what confidence level YOU should use.

WHEN TO PANIC

Now to answer the "Final question" above: what to do when on-hand levels of safety stock begin falling. Well, you could *immediately* panic and order more parts. But to CORRECTLY answer the question, keep this in mind: if your forecast includes good "as of today" fleet-data, and you are running a correct forecast methodology, then your uncertainty will be low in the early forecasted months, and will grow logarithmically the further out in time you go. Use this fact to come to the following conclusion: if your on-hand levels begin to worry you, just re-run the forecast with today's data. This will tell you with high certainty whether you are in trouble within the next few months, or whether you have just undergone a period of high demand that should be settling down soon, and your cache will soon begin returning to acceptable levels.

If it turns out that you ARE in trouble: that is, your current cache levels and the most recent forecast indicate that you WILL run out of parts, then something has gone wrong. There is really nothing you can do at this point except to pay for expedited parts delivery or take the production hit. However, this *is* a good time to analyze what went wrong, and to try to put corrective actions into place so that it doesn't happen again. Here is a list of good places to look for the trouble area:

- Did the operators or maintainers operate or maintain the parts differently than they said they would?
- Did the inherent reliability of the assemblies or systems drop unexpectedly? This can happen if a new failure mode is exposed due to operating beyond previously experienced ages, or within new environments.
- Did YOU fail to properly maintain the simulation model or to run it often enough that you could have seen the trouble coming and taken preemptive action while there was still time. It requires effort and diligence to stay on top of failure trends and to adjust levels within the model to accurately reflect reality.

DETERMINING CONFIDENCE LEVELS

Now to address the question of what confidence levels you should be using to compute your required Parts Cache. This is a difficult question, and the answer depends on a LOT of things. Ultimately, any answer that can be given will have some subjectivity embedded in it. Understanding that, however, the key to correctly answering the question is to have the *right people* asking and answering the *right questions*. Here is a list of some of the ***right questions***:

- How critical is it that you do not run out of parts?
- - How much does it cost to idle, or to slow the production line?
- - How many spare assets does your customer have: will slowing your production kill your customer?
- - Will failure to meet production RUIN your reputation, possibly leading to loss of future contracts?

If you are finding that a moderately high confidence is leading to excessively high levels of Parts Cache, you might want to accept a greater risk of a stock-out, and use a lower confidence level. Before you do, however, ask these questions:

- Can you "borrow" parts from elsewhere, perhaps another facility, or even a competitor? If you have a known buffer, you can probably use a lower-confidence Parts Cache.

- How quickly can parts vendors respond to increased demand. Is it possible for them to greatly accelerate deliveries? And if so, at what cost? If they promise accelerated deliveries, can you trust them?

- Can you slow inductions, possibly by having your customers temporarily extend field limits. For example, instead of removing assets from service at strict vibration limits or oil leak limits, perhaps they could relax the limits just a little for a month or two; just until you could get by a possible bump-in-the-road?

- Can you relax your "build-policy" a little to not discard so many of the parts that are in short supply? This will imply that your product will be lower in value since it will need to return to maintenance sooner than it normally would. You may need to reduce your sales price to compensate for the decreased intrinsic value of your product.

- Can you take preemptive mitigation actions? For example, rather than waiting until a critical month is upon you, maybe you can force-generate some (old) assets early, during slow times. You may find that doing this causes the Parts Cache levels in the critical month to be lower than if no action is taken. Thus, you can still use a high confidence level, but not have to cache so many parts.

- How much manpower do you maintain. A team of people *might* be able to come up with a workaround to an improbable parts problem. But that's the trade-off: do you want to spend money on excessive manpower, or excessive inventory? It might be cheaper to cache more parts than more people.

- What percentage of your total Safety Stock is Parts Cache? In other words, if the real drivers for variability will be reasons 2 and 3 identified at the beginning of this paper, does it even make sense to be worrying about confidence levels within the forecast? Perhaps it makes more sense to just pick a very high confidence level (say 99.7%), and then to focus on the other issues.

With these questions answered, the next step is to perform a "sensitivity analysis". If you suspect (intuitively, after reflecting on the questions above) that the right answer is 98% confidence, then you should compute the Parts Cache at 97.5% confidence, and at 98.5% confidence. This may indicate that with small increases in cache levels you can significantly improve confidence. Or it might indicate the converse: that small changes in confidence lead to large changes in cache and the associated costs of maintaining that cache.

Then, after reflecting on the questions above and the results of the sensitivity analysis, if you still need a better answer, you will be forced to attempt to generate a numeric answer. The only way to get a numeric answer is to run an actual *optimization*. This involves making accurate estimates regarding "costs". Cost numbers must be assigned to every variable, including:

- cost to idle the production line
- cost to cache each additional on-hand part
- cost of not delivering to your customer on-time
- cost of building finished products with lower life-expectancy
- cost to force generate (pull in repairable items ahead of schedule)
- cost of extra manpower
- etc.

This paper does not cover the subject of optimization. Many books have been written on the subject of “Linear Programming” as well as non-linear optimization techniques. We would refer the curious reader to those.

FINAL NOTES REGARDING RUNNING FORECASTS

To compute Parts Cache to any particular confidence level, you only MUST know the requirements lead-time away. But it would be helpful to have some advance notice. Plus, you may not be running the forecast every month; so, you should run the forecast for lead-time, plus a couple months, plus however long between forecast runs. Then, you should use that data and plan to look out beyond the lead-time a few months each month to make sure you're not going to get into trouble, and to give some time for your supplier(s) to ramp-up production.