QRM It's About TIME

The Competitive Advantage of
Quick Response Manufacturing

Appendix D

Simple System Dynamics

Calculations for Quick Insights

APPENDIX D

SIMPLE SYSTEM DYNAMICS CALCULATIONS FOR QUICK INSIGHTS

In this Appendix, I show you how to perform some calculations that will give you insights into the main drivers of Flow Time in your organization. This will include insights into capacity planning, batch sizing, and other decisions related to operating your QRM Cells.

First, I show you some simpler calculations that you can perform easily to get initial insights without needing to obtain a lot of detailed data. Next, if you are willing to engage in more data gathering and analysis, I show you how to get more accurate estimates of the impact of various decisions. After this I also show you a formula that can work in certain situations to help you evaluate your batch sizes. Finally, for a more complete analysis of system dynamics in your factory, you will need to use some type of dynamic modeling software and I give you a few suggestions to explore further. This Appendix continues in more depth the explanations I gave you in Chapter 3. I will assume you have reviewed that chapter before you continue with the material here. The basis for all the discussions in this Appendix will be the formula in Chapter 3 for the average Flow Time of a job visiting a resource:

$$\textit{Flow Time} = (AV \times M \times TJ) + TJ$$

As a reminder, in this formula *AV* is the average variability, which includes variability in both the arrivals and the job times, and I go into more detail on this in the "How to Calculate Variability Values" section. *M* is the magnifying effect of utilization and equals $u : (1 − u)$ where u is the utilization of the resource. Finally, *TJ* is the average time taken by the resource to work on a job. All these quantities were defined and explained in Chapter 3. In the rest of this Appendix I show you how to calculate these quantities for your situation and the types of insights you can get by using these formulas. Before proceeding with the calculations, you need to make an important decision.

FIRST DETERMINE WHAT TIME PERIOD YOU WILL USE

Since all the formulas here are based on statistical data from your operation, you need to be sure to use data for a time period that is reasonably representative of the operation you wish to analyze. On the one hand, the time period should not be too short, or your data might not be representative. On the other hand, if the time period is too long, then you might average out peak-demand months with low-demand months and you don't want to do that either because then you lose insights into how to operate in these two different situations. Common sense should help you pick a time period for your particular situation, but in general I have found that anything less

than a week is too short, while a month or a quarter are reasonable periods. If you have high- and low-demand periods, you should perform separate analyses for each one – for example, you could do an analysis for a month with peak demand, a second one for a month with normal demand, and a third analysis for a month with low demand. You can use historical data if you feel it is representative of expected operations in the near future, or you can use planned (estimated future) data if you feel that operations will be significantly different in the near future.

Finally, note that the purpose of this Appendix is to give you insights into the main causes of delays and the key decisions that will help you reduce them. The approach here is not aimed at accurate day-to-day running of your factory, but rather, at giving you the major direction you need to take. For this purpose, rough data is good enough, so don't spend too much time in data gathering. Get some initial numbers and get going on the analyses here! In most cases I have seen that there are some significant insights obtained right away. I will start with the calculation of utilization (u) for a resource as that is always the first step.

REMEMBER, IN QRM "UTILIZATION" INCLUDES ALL THE TIME THE RESOURCE IS OCCUPIED FOR ANY TASK

Chapter 3 included a detailed discussion of this issue. To calculate the value of *u* for a given resource, select the period you will analyze. For this period, add up all the times that this resource was occupied for all tasks plus any times it was not available for other reasons such as unexpected downtimes due to machine failures or absenteeism. Let's call this the Total Utilized Time. If you are doing this for a past period and have reasonable historical data in your MRP system, you can add up the times spent by jobs at this resource. If you feel that the actual data is not accurately recorded, you can also work with the standards in the MRP system if you feel they are reasonably representative. In either case, be sure that the data includes setup times or else you will need to add in estimated setup times. Finally, be sure to add in any downtimes and other times as just mentioned. Next, add up all the times that the resource was scheduled to work; call this Total Scheduled Time. This can usually be done just by checking the calendar and knowing how many shifts the resource operates. However, be sure to check past records and add in all instances of overtime that was used beyond the scheduled times that you included. With these calculations in place, the utilization of this resource is simply:

$$u = \frac{Total\ Utilized\ Time}{Total\ Scheduled\ Time}$$

As a check on whether you have included all the potential instances in the Total Utilized Time, I'd like to remind you of the alternative approach in Chapter 3, which helps to "back into" the value of utilization. If you are not sure whether a given category of time should be included in Total Utilized Time, then ask this question: Was this resource actually available to start an unexpected job if it had arrived at the resource during this

time? If the answer is no, then this time should be included in Total Utilized Time. I'll give you two examples:

- A machine breaks down during in the middle of a shift and is down for the rest of the shift (4 hours) and still down for the first 2 hours of the next shift. Then you definitely need to add 6 hours into the Total Utilized Time value.
- A machine is available but the operator unexpectedly has to go home and is gone for 5 hours during the shift and no one else is available to run the machine. Again, you need to add these 5 hours to the Total Utilized Time.

If you are concerned that your data isn't good enough to record such events, don't worry, just talk to managers or supervisors and get rough numbers for such instances. Now I'll give a different example. A machine needs preventive maintenance on its bearings and this is scheduled to be done on Thursday and Friday. This is known ahead of time, and the Planning Department does not include these 2 days when it is planning capacity and scheduling the machine. In this case, the 2 days need to be taken out of the Total Scheduled Time (or not included in the total in the first place).

In general, the rule is simple: If the time was scheduled and planned for, and then it becomes unavailable, it needs to go into the Total Utilized Time category. If the time was not scheduled to begin with (in the planning system), then just be sure not to include it in the Total Scheduled Time. You might think, if a machine fails and is unavailable for 10 hours, why not just take these 10 hours out of the Total Scheduled Time? Why are they included in the Total Utilized Time? The answer has to do with the system dynamics effect of planned versus unplanned capacity utilization, as this example shows.

Suppose the Total Scheduled Time for a machine is 40 hours a week, but during one week the machine unexpectedly breaks down and is not available for 10 hours of the working time. Also, say that 20 hours of work are scheduled on this machine for that week. If you take the breakdown time out of the scheduled time, then you get a utilization of $u = 20/30$ or 67% for that week. But if you add the breakdown time to the Total Utilized Time and leave the Total Scheduled Time at 40 hours, then you get $u = 30 : 40$ or 75% for that week! This difference is even more pronounced when you calculate its impact on Queue Time through the value of M as explained in Chapter 3. In the first calculation you get $M = 2$, while in the second case you get $M = 3$, a 50% increase in Queue Time! So this calculation procedure is important. But which one is right? The answer is the second one (with the higher values of u and M). The reason is the system dynamics effects occur based on what you plan, and then how the real world evolves. If you originally expected the time to be available and planned for that, and now the time is not available, this creates dysfunctional effects that are estimated by the system dynamics formula.

GET INITIAL INSIGHTS WITH A QUICK FIRST ANALYSIS

Now that you understand how to calculate utilization, get rough estimates of its value

for key resources. You can do this for most of your resources if your data gathering can be automated (e.g., using your MRP system), or else pick some resources that you know are bottlenecks and causing long MCT. For example, there may be some cells that still have large MCT values and you can start your analysis with the resources in these cells – or if you suspect that there are a few specific resources within the cells that are problematic, you can restrict your first analysis to those resources and then expand to other resources later. Note that "resource" can be a machine or a person such as an assembly operator. It is important to try this approach on a few resources first as you will also discover how to gather and use your data and whether it is good enough or if more effort is needed in getting accurate data.

I'll use an example to show you how to conduct this first analysis. Let's say you suspect that in a particular QRM Cell, a large milling machine is a bottleneck and has a long Flow Time, thus contributing to the large MCT value for the cell. You need to calculate only two data values for this machine (I'll call it the Large Mill) to get initial insights. First, calculate the utilization (u) as previously described. Next, for the time period you are analyzing, also calculate TJ, the average time that it takes for a job on this machine (this includes setup time plus the run time for all the pieces in the job).

Now to make this first analysis simple and to get quick insights, you will use the value 1.0 for both the arrival variability and the job variability. I will explain both of these later in this Appendix, but here is a short explanation for now. For arrival variability, the value of 1.0 is commonly used in system dynamics models to represent random arrivals such as customer orders when there isn't better data available, and is accepted as a good starting point for analysis. For job variability, the value of 1.0 also is commonly used to denote a reasonable amount of variability (e.g., some very short jobs, some long ones) – it represents a statistical compromise between very high variability and very little variability. Next, as explained in Chapter 3, the average variability is obtained by squaring these two variability values, adding the squared values and dividing by two. Since both these values are 1.0, when you do this calculation the average variability AV just comes out at 1.0 also. When you put these values into the formula for Flow Time it simplifies to

$$Flow\ Time = \frac{TJ}{1 - u}$$

Now let's see what this formula tells us. Suppose for the Large Mill you find that TJ = 6 hours and u = 0.88. Then the formula predicts a Flow Time of 50 hours. So you see that while on average there are only 6 hours of work to be done on a job, it takes 50 hours to get through the Large Mill. You can do similar analyses for other resources that appear to be problematic. The results of these analyses will enable you to target a few resources that need improvement. You can start working on two aspects of these resources right away – namely, reducing the values of TJ and u by applying some of the improvement ideas discussed in Chapter 3. For the Large Mill as an example, suppose you think you can reduce the average job time by 5% to 5.7 hours and this reduction, along with improved maintenance and uptime, would bring the utilization down to 81%,

what would that do to the Flow Time? The formula predicts a new Flow Time of 30 hours – a 40% reduction – which may help you to get the investments needed to make these improvements. After you have targeted a few such opportunities for improvement, you can then proceed to conduct more detailed analyses as explained next, in order to pinpoint additional opportunities for improvement.

How to Calculate Variability Values
In order to get to the next level of accuracy with your system dynamics analysis, you need to estimate the arrival time variability and the job time variability for the resource being analyzed. I'll start with arrival time variability. To calculate this you need data on the times between job arrivals. If a job arrives at the resource at 9:00 a.m. and then the next job arrives at 2:00 p.m. then the time between the first and second arrival is 5 hours. It is important that this data is calculated by using only working hours. Let's say the resource works one shift, from 8:00 a.m. to 4:00 p.m. Continuing with the same example where the last job arrived at 2:00 p.m., let's say the next job arrives at 9:00 a.m. the next day. Then the time between this new arrival and the previous one should be recorded as only 3 hours (2 hours of work time on the previous day and 1 hour of work time on this day). If your MRP system has such data on job arrival times, you can automate this calculation, as long as you can remove nonwork times including nights and weekends. It's possible that you don't have this detailed data for the resource in question, or it is not accurate enough, or it is hard to extract the values I described. In that case you have a few alternative options:

• You can look at data for initial job releases in the MRP system, that is, the times when jobs are released to the shop floor. Be sure to look only at jobs that are destined for this resource, and then use the times between these job releases. Essentially, you are assuming that as various jobs work their way through the shop floor, the arrivals to this resource have a similar pattern. If a bunch of jobs destined for this resource are all released at the same time, you give the first job in this bunch the interarrival time value of the actual time from the previous job release, and all the other jobs get an interarrival time of zero.

• If jobs on the shop floor are closely related to customer orders, then you can look at arrival patterns for customer orders for the relevant jobs and use those in the same way as previously described.
• Use manual observations for a period of time, such as a week. Ask a cell team member or a planner or scheduler for that area to record data.
• As a last resort, if you are really having difficulty getting good enough data, you can again resort to using the value 1.0 for the arrival variability. As I explained earlier, this is a commonly accepted starting point for system dynamics analysis. This will allow you to get going on some of the other calculations in the following paragraphs. However, this is a stopgap measure so that you can do some initial number crunching. Do not accept this as an end solution. Start the process of gathering data so you can get a more accurate number in the near future.

Once you have all the data you need, list these values of interarrival times in a spreadsheet. (In the following section, I suggest additional calculations that need a spreadsheet, so you can keep in mind that you could integrate those into this spreadsheet.) Use the spreadsheet functions to calculate the average of all the interarrival times:; say this value is TA (for average Time between Arrivals). Next also use the spreadsheet functions to calculate the standard deviation of all these values; say this is SA (for Standard deviation of time between Arrivals). Then the final value that you need, namely the variability of arrivals (I'll call it VA) for the resource is given simply by

$$VA = \frac{SA}{TA}$$

Next I'll show you how to calculate the variability in job times. Here you need data on how long it takes the resource to work on each job. Remember that this must include the setup time plus the time to complete all the pieces in the job. In order to do some of the "what-ifs" that I will discuss in the following section, it would be best if you would record separately, for each job, the setup time, the batch size used, and the run time. The run time can be either the total for the whole batch, or the average time per piece – either is fine because if you know the batch size you can calculate either value from the other. For example, if you use actual data, you might have only the total run time recorded. On the other hand, if you use standards from the MRP system, you will probably have the run time per piece. As in the previous case, you should organize all these values in a spreadsheet as shown in Table D.1.

The last column in the table is the total time to complete each job. From this column you can use the spreadsheet functions to calculate the average time to work on a job (I've already called this *TJ*) and the standard deviation of the time to complete a job (I will call this *SJ*). Then the variability of job times (I'll call it *VJ*) for the resource is given simply by

$$VJ = \frac{SJ}{TJ}$$

Table D.1

Example of Spreadsheet Needed to Gather Job Data for System Dynamics Analysis
List of Jobs for NC Mill in Housings Cell during December

Job Number	Quantity	Setup Time (hrs)	Run Time per piece (hrs)	Total Job Time (hrs)
13011928	16	1.25	0.33	6.53
18121952	13	0.67	0.20	3.27
52133910	4	0.83	1.12	5.31
26522007	18	1.45	0.27	6.31

Finally, you now have the data to calculate the average variability (AV), which is needed in the Flow Time formula. This is given by

$$AV = (VA^2 + VJ^2) : 2$$

Note that if you set up a spreadsheet to calculate *TJ* and *VJ*, you can use the same data to get the utilization *u* as well. Just remember to add any downtimes or other times as I explained earlier.

HOW TO USE THE DETAILED FLOW TIME FORMULA TO GET INSIGHTS

Let's say you want insight into a QRM Cell that was put in with a target MCT of 3 days but currently has an MCT value of over 6 days. The cell team feels that a particular vertical turret lathe (VTL) in the cell is causing long Flow Times and needs your help in planning the load on this VTL. You've done some initial simplified analysis as shown previously but now you want to do some more detailed analysis to see what other improvements might be possible. Your analysis of the data for this VTL produces these values: u = 0.87, *TJ* = 4 hours, and *VJ* = 0.73; you also realize that you don't have an easy way to get interarrival time data, so as explained previously you assume *VA* = 1 for this analysis. Now you can calculate *AV* using the preceding formula and you get *AV* = (12 + 0.732)/2 = 0.766. You can also calculate *M*, which equals $u/(1 - u)$ = 6.69. For convenience, the Flow Time formula is:

$$Flow\ Time = (AV\ X\ M\ X\ TJ) + TJ$$

Putting all these values into the formula, you find that the average Flow Time through this VTL is 24.5 hours. If the cell works one shift a day (8 hours), this is a Flow Time of over 3 working days, so this one machine alone accounts for a Flow Time that exceeds the target for the whole cell. Note that the average job time on the VTL is only 4 hours in this example, so it is taking 3 days to do a 4-hour job – the cell team is right in its gut feeling that the main problem is the VTL. Even before I discuss what action you could take, note that just doing this calculation helps you rationally ascertain whether there is a problem at the resource rather than having to go on people's opinions.

Now that you know there is an issue with the VTL, you can look at potential solutions. In Chapter 3, I gave you a three-pronged strategy to reduce Flow Time along with several practical suggestions. Let's use a couple of these ideas here. You find that some parts are being made to stock in large quantities and then some of those pieces don't get sold – with your shorter lead times you need less safety margin and you can run roughly the same number of batches but with smaller quantities; there will be the same number of setups but less run time so this will reduce some load on the VTL. You also get management to agree to spend money on 4 hours of overtime per week for the cell (essentially a 10% increase over the 40-hour week). Finally, you find that the value of *VJ* is high because of occasional long breakdowns of the VTL. You discuss this with the maintenance crew and come up with a preventive maintenance strategy. You also get estimates for what the new downtime data might be in the future. You put this revised data into your spreadsheet and get a new value for *VJ*. With all these improvements, you estimate that the new values will be: u = 0.7, *TJ* = 3.9 and *VJ* = 0.5, and *VA* remains

the same for now. Putting these estimates into the formula gives you the new Flow Time of 9.6 hours, or just over one shift. So the next insight you get from the formula is that these improvements can help you cut almost 2 days out of the Flow Time through this cell. There is one more important decision that you would normally explore and that is to do with the batch sizes being run on the VTL. Here I will show you one way of analyzing this issue, and then I will give you an alternative approach later. Let's say that in looking at the historical data that you used to get the TJ and VJ values (the data in your spreadsheet similar to that in Table D.1), you note that there are occasionally very large total job times; you find that these are usually associated with jobs being run in large batch sizes. So you can investigate the impact of cutting these down as the following examples show. If a part is being made to stock in batches of 20, you could consider running it in batches of 10. This would mean that during the time period you analyzed you would need to run two orders of 10. So in your spreadsheet you replace one order of 20 with two orders of 10. Remembering that each of these will take a setup time, you can't just cut the job times in half, so make sure to calculate the new total of setup and run times and enter the right total times for each of these two jobs. If cutting a batch size in half seems too drastic, you can be more conservative. Say that during the period you had two orders of the same part with 30 pieces in each order for a total of 60 pieces. You could consider cutting the batch size to 20 and going with three orders during the period. In this fashion, you can pick out some of the batches with the longest job times and revise the data for them. With modern spreadsheets you can automate this calculation to make your job easier. After you revise the data, three values will change: the utilization, u, will become larger due to more setups, TJ will become smaller as some of the job times will be much smaller than before, and SJ will become a bit smaller too since the very long jobs will be cut down so there will be less variability overall.

Let's say after you revise the VTL data with your "what-ifs" for the batch sizes, these new numbers are: $u = 0.72$, $TJ = 3.1$, and $VJ = 0.45$. Putting these estimates into the formula gives you the new Flow Time of 7.9 hours, so now the VTL Flow Time is reduced to under a shift. While this is not a big reduction from the previous value of 9.6 hours, it does set the stage for the next improvement. If long runs are an issue on this machine, clearly one item to explore at this point is whether setup times can also be reduced, enabling even further reduction in batch sizes. So you look at the impact of a 50% reduction in setup time for some of the parts with big batch sizes, and also reduce the batch sizes some more. You play around with some numbers and find batch sizes that will keep the utilization from going over its last value. With this new data in your spreadsheet, you get $u = 0.72$, $TJ = 2.7$, and $VJ = 0.4$, and the formula predicts a Flow Time of 6.7 hours. Again, this may seem like a small improvement over the previous value, but remember where you started – at 24.5 hours! The system dynamics analysis here has helped you come up with a series of improvements that will reduce the Flow Time through the VTL by 75%. If this data manipulation seems too time consuming, note that using spreadsheets and with the availability of query methods on modern ERP systems, one of your systems analysts could help you automate much of this "what-if" analysis. Alternatively, you could look into using one of the modeling packages mentioned at the end of this Appendix.

DETERMINING A GOOD BATCH SIZE

While the previous analysis will help you look at any resource, performing the "what-ifs" can be time consuming. If you feel that poor choice of batch sizes may be affecting the performance at a particular resource, there is a simpler rule that you can use for quick insights in certain cases. The resource must satisfy the condition that for all products going through this resource, the setup times are about the same across products, and similarly, the run times per piece are reasonably similar across products. In that case you can use the averages of these setup times and run times to come up with a good batch size. Specifically, the rule I give you will tell you the batch size that should result in the shortest Flow Time through that resource.

To apply the rule you will need to calculate some more quantities using the data you already have in the spreadsheet discussed previously. I will use a time unit of hours for illustration; you can pick any other time unit as well. The quantities you need to calculate are:

- S = Average setup time for a job, in hours
- H = Total Scheduled Time at the resource for the selected time period, in hours
- Q = Total number of pieces (of all products) made by the resource during the same time period
- R = Total of all run times at the resource during the period
- Z = Total of all "other times" at the resource during the period (see following explanation)

I have used the phrase "other times." These are the times that the resource is not in setup nor is it running a product, but it is still not available to run another job for some reason (such as when a machine is broken down). In Chapter 3 and at the beginning of this Appendix, I discuss the importance of including the total of such times in your utilization calculation.

From these values, you calculate u_R, which is the "run utilization" of the resource (proportion of time it was actually working on a product), and u_Z, which is its "other utilization" (proportion of time it spent in these "other times"):

$$u_R = \frac{R}{H}$$

$$u_Z = \frac{Z}{H}$$

With all these quantities in hand, the recommended batch size for this resource can be calculated by the formula

$$B^* = \frac{S \times Q \times (U_R + \sqrt{U_R \times (1 - U_Z)})}{H \times U_R \times (1 - U_Z - U_R)}$$

In case you are interested in the background to this formula, it is based on system dynamics analysis of batch sizing. As I explain in Chapter 3 (and illustrate there in Figure 3.6), there is a trade-off between the large batch sizes that cause long job times and the small batch sizes that create a lot of setups. Looking at Figure 3.6, B* is the batch size for which you get the shortest Flow Time through the resource. For more details on the formula derivation, see "For Further Reading" at the end of this Appendix. I'll walk you through an example to show you the application of this formula. Let's say that in a QRM Cell a particular CNC lathe runs jobs with typically 50 pieces at a time and you suspect that this batch size is too large and is causing long Flow Time. An initial look at the data for the lathe shows you that most jobs have similar run times for each piece, and discussions with the cell team lead you to believe that setup times are about the same across jobs. So you feel you can apply this formula to get some insight. For this lathe, you get the following data for a onemonth period.

- S = Average setup time = 1 hour
- H = Total Scheduled Time = 200 hours
- Q = Total number of pieces made = 520
- R = Total of all run times = 130 hours
- Z = Total of all "other times" = 10 hours

These values result in u_R = 130/200 = 0.65, u_Z = 10/200 = 0.05. Putting all these values into the formula (you can use a spreadsheet or a calculator for this) you get B* = 19.1. Even after rounding off to the next whole number, the recommended batch size is around 20 pieces, which is less than half the current batch size. Although this is an approximate analysis based on rough data, this is a significant enough difference from your current operating policy that it should convince you to start running smaller batches on the lathe. Even if you don't want to do anything as drastic as cutting all the batches in half, you could perhaps reduce them by 25%, observe the effects building up over the next month, and then reduce the batches some more after the Flow Time goes down as predicted. As a final general pointer on batch sizing, remember that even a few big batches (big in the sense that their total setup plus run time is long relative to other products) can have a disproportionately large effect on overall Flow Times, as I showed with the numerical example in Chapter 3. Hence always keep your eye out for such situations and find ways to justify reducing these big batches using various tools discussed in this book.

Consider Using Dynamic Modeling Software
The calculations here will enable you to get started and use system dynamics to get some initial insights. However, the analysis gets complicated once you try to include more details of your factory such as cross-trained people running multiple machines, products flowing through multistep routings, different products with different setup and run times on various machines, scrap and rework, and so on. If, after trying out some of these formulas, you see the potential of system dynamics and want to work

with a more realistic model of your factory, you will need to use some type of dynamic modeling software (see "For Further Reading").

FOR FURTHER READING

If you would like more insight into the system dynamics formulas and why utilization and variability impact Flow Time, but don't want too technical an explanation, an easy-to-read discussion along with step-by-step examples can be found in Chapter 7 of *Quick Response Manufacturing: A Companywide Approach to Reducing Lead Times, by R. Suri* (Productivity Press, 1998).

If you have a more theoretical background, the batch size formula is based on an analysis in "Lot Sizes, Lead Times and In-process Inventories," by U. S. Karmarkar, *Management Science*, March 1987, pages 409–418. However, note that I have modified the analysis to include the "other" utilization. I did this to provide more accurate predictions in the presence of significant downtimes, which often occur in real-world situations and need to be taken into consideration.

An easy-to-use software package for system dynamics analysis of factories is the MPX Rapid Modeling System from Network Dynamics, Inc. (www.networkdyn.com). It can model all the practical aspects of factories as mentioned in the text and at the same time takes just seconds to calculate the output values, even for large systems

There are more detailed modeling packages available for system dynamics analysis, and they are known under the general category of "discrete-event simulation" software. Several professional societies publish annual surveys of these packages and also hold regular tutorials and conferences on them. A few suggestions include the Institute of Industrial Engineers (www.iienet.org), INFORMS (www.informs.org), the Society for Computer Simulation (www.scs.org), and the Winter Simulation Conference (www.wintersim.org).