



Promising reliable delivery dates with Monte Carlo simulation

28 October 2026, 11:00

Pascal Pollet

Use Monte Carlo to factor uncertainty into your delivery commitment

A calculation based on average output takes no account of daily fluctuations in your production. With a Monte Carlo simulation, you make that uncertainty visible. In this way you can use historical output to calculate a delivery date that corresponds to the desired delivery reliability.

Want to start using Monte Carlo yourself?

Would you like to test the method yourself? You can request a simple Monte Carlo template in Excel from our expert Pascal Pollet.

[Contact our expert](#)

From average output to reliable delivery dates

In the [previous article](#), we calculated the expected waiting time based on two pieces of data: current workload and average output.

Expected waiting time = outstanding workload / average output

To arrive at a reliable delivery commitment, we added a safety margin to this. That is a deliberately simple approach. A planning rule must be easy to understand, use, and maintain; a more complex calculation only makes sense when it leads to better decisions.

However, a fixed safety margin remains an approximation. An order that waits five days is exposed to disruptions for a shorter period than an order that waits fifty days. Think of breakdowns, material problems or fluctuations in daily output. This means that the required safety margin is often related to the expected waiting time.

Even a fixed safety factor only partly solves this. Suppose you always multiply the waiting time by 1.2. You are then assuming that uncertainty grows in proportion to the waiting time. That too is an approximation.

We are therefore looking for a method in which the required safety margin automatically moves in line with the current situation. This is where Monte Carlo simulation comes in. The underlying idea is surprisingly simple: the method makes the uncertainty around the predicted lead time visible and backs up the required safety margin with historical data.

From an average to thousands of possible futures

An average conceals the daily fluctuations in your production. Suppose a production system processes an average of 50 units per day.

The actual daily output might look like this:

Day	Actual output
1	43 units
2	51 units

Day	Actual output
3	47 units
4	56 units
5	49 units
...	...

With a workload of 1,200 units, the average gives the following waiting time: **1,200 / 50 = 24 days**. But those 24 days are only an average expectation. If you happen to get lots of production days with low output, it will take longer. With a favourable sequence of production days, it will go faster. A Monte Carlo simulation mimics these different possible outcomes.

How does a Monte Carlo simulation work?

We start from the current workload and the historical daily output. For each simulated production day, we randomly select one output value from the historical data. We subtract that quantity from the remaining workload. We repeat this until the entire workload has been processed.

For example, a simulation run might look like this:

Simulated day	Randomly selected output	Remaining workload
Start	—	1,200
1	48	1,152
2	53	1,099
3	44	1,055
4	57	998
...	...	...

The number of days required is the predicted waiting time for that simulation run. A single simulation run tells you little. We therefore perform the calculation 10,000 times, for example. Each run uses a different random sequence of historical daily outputs. The result is not one predicted waiting time, but a distribution of possible waiting times.

Principle 1: make delivery reliability the starting point

Delivery reliability must not be an accidental result of the chosen delivery date. First determine what reliability you want to achieve, and then derive the corresponding delivery lead time from that. A Monte Carlo simulation makes that choice concrete. For example, the results might look as follows:

Desired delivery reliability	Simulated waiting time
50%	24 days
85%	27 days
95%	29 days
99%	32 days

The median here is 24 working days. In half of the simulations, the workload was processed within that period, but in the other half it took longer. Do you want 95 per cent delivery reliability? Then, in this example, you will use 29 working days in your calculation. The implicit safety margin is therefore five working days.

The difference from the previous method is important. We don't set the safety margin in advance as a fixed number of days. It follows from the simulated spread of waiting times. As a result, it moves in line with your current workload and historical variation.

The simulation described above calculates the waiting time until it is the new order's turn. To determine the full delivery lead time, the work for the new order must also be included. You can add the processing time afterwards. You can also add the size of the new order to the simulated workload straight away.

Principle 2: don't treat rush orders as exceptions

Do rush orders occur regularly? If so, your delivery commitments must take them properly into account. That also goes for Monte Carlo simulations. Future rush orders are not yet part of your workload today, but new rush orders will come in during the waiting time and take priority.

The simplest solution for normal orders is not to use the total daily output, but the historically achieved output that has actually been available for normal work. The effect of rush orders will then already be incorporated into the data. The daily distribution might, for example, consist of values such as **32, 41, 38, 47 and 39 normal units**.

The Monte Carlo simulation draws from these values each time and therefore takes account not only of the average share of urgent work, but also of its daily fluctuations. You can run a separate

simulation for rush orders, starting from the rush order queue and the available total output.

Principle 3: choose the simplest capacity model that works

Monte Carlo sounds advanced, but it doesn't necessarily require a complex production model. You don't need to build a separate probability distribution for each machine. Three basic data items may be enough:

1. The current relevant workload
2. The historical daily output
3. The desired delivery reliability

Your historical output thus forms a set of realistic production days. If you repeatedly choose from that data, you don't even need to assume a theoretical probability distribution. Only add more detail when it is truly necessary.

Do product families have systematically different production flows and outputs? If so, you can run a separate simulation for each family. Are those differences slight? If so, the main effect of such a breakdown will be to make the model harder to maintain.

Monte Carlo can also be run in a spreadsheet

You don't necessarily need specialised planning software for this. Even a simple spreadsheet can run thousands of simulation rounds. The 95th percentile of those results will then represent the waiting time that can be used with 95 per cent reliability. If you also include the new order in the simulation, you will immediately get the associated delivery lead time.

Use representative historical data

The simulation assumes that historical output is a good approximation of possible future output. Therefore, don't use data from a period that differs fundamentally from the current situation. Have you added an extra shift in the meantime? Has a substantial bottleneck been eliminated? Or has your product mix changed significantly? If so, check what historical data is still representative.

Choose a long enough period too. A few weeks may happen to be exceptionally good or bad; a longer period contains more normal variation. But don't go too far back either. Data that is too old may describe a production system that no longer exists today.

When does Monte Carlo work well?

The Monte Carlo simulation builds directly on the workload-based method from the [previous article](#). It is therefore suitable for the same production flows, in which most orders follow roughly the same production route, and the amount of outstanding work therefore largely determines the lead time.

If orders take very different routes, or each places a load on different shared work centres, this method becomes less suitable. In such a complex job shop, the total workload says too little about the waiting time for a specific order.

In summary

A Monte Carlo simulation translates historical fluctuations in your production into a reliable delivery lead time. You first choose the desired delivery reliability. The simulation then determines the corresponding waiting time. Your safety margin will then automatically move in line with the current workload and production variation.

Conclusion: make uncertainty visible

An average gives a single expected waiting time, but production never follows that average exactly. Monte Carlo makes that uncertainty visible and links it to your desired delivery reliability. This gives you a dynamic delivery commitment without the need to build a complex capacity model.

From total workload to future load

A Monte Carlo simulation refines the workload-based method by making uncertainty explicitly visible. The required safety margin is no longer set as a fixed value, but derived directly from the simulated probability distribution. However, this method too assumes that the total workload is a good measure of the work that must be carried out before a new order.

In the next article, we look at more complex production environments in which orders follow different routes and place loads on different work centres.

Recommended reading

Would you like to explore probabilistic forecasting in more depth? Daniel S. Vacanti explains the approach in detail in *When Will It Be Done?*. The book shows how historical flow data helps predict lead times with explicit reliability.

Authors



Pascal Pollet