



Calculate delivery dates with your workload

28 October 2026, 10:00

Pascal Pollet

Use current workload and output for reliable delivery commitments

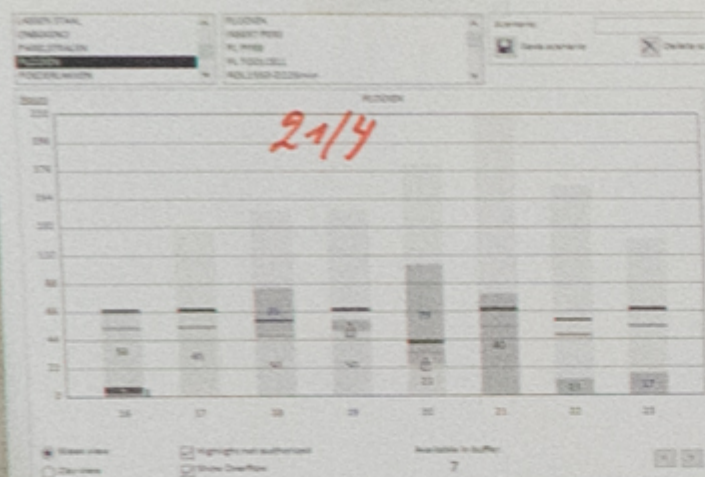
Historical lead times show what happened in the past. They say less about how busy your production is today. By linking the current workload to actual output, you can quickly set a reliable delivery date in simple production flows. This ensures that the timing you promise reflects the current production load more closely.

From fixed delivery time to current workload

In the [first article](#) in this series, we discussed three principles for better planning: make delivery reliability the starting point, do not treat rush orders as exceptions, and choose the simplest capacity model that works. In the [second article](#), we applied these principles to fixed delivery times. That method works well as long as demand and lead time are sufficiently stable.

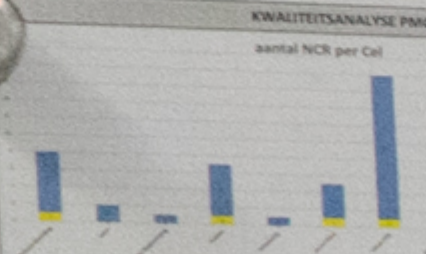
However, a problem arises when there are significant fluctuations in demand. During quiet periods, the fixed delivery time is unnecessarily long. During busy periods, the same delivery time becomes too optimistic. We therefore need a method that takes account of the current production situation.

EN OP VOORHAND
PLOOIEN
SPELLEN EN

PATTYN[®]

KWALITEITS CIJ

TER KENNIS GESTE
ALLE PMC MEDEWI

[illegible]

Look ahead, not back

Many companies set delivery dates based on historical lead times. In doing so, they assume that the future will unfold in roughly the same way as the past. But how much does the past tell us about an order that comes in today? The lead time for that order mainly depends on two factors:

- How much work needs to be processed before the new order
- How quickly the production system will complete that work

We call the first of these the current outstanding workload. This includes orders still waiting in an order pool, as well as physical WIP on the shop floor. Where that work is located is not important: all of it must be processed before the new order gets its turn.

You can therefore approximate the expected waiting time quite simply:

$$\text{Expected waiting time} = \text{outstanding workload} / \text{average output}$$

Suppose 1,200 units still need to be completed. The production system delivers an average of 50 units per day.

A new order that joins the back of the queue can be expected to wait 24 working days:

$$1,200 \text{ units} / 50 \text{ units per day} = 24 \text{ days}$$

You then add the actual processing time and a safety margin.

Principle 1: make delivery reliability the starting point

The ratio between workload and output gives an expected waiting time. It does not yet guarantee delivery within that time frame. After all, daily output is never completely constant. Breakdowns, material shortages, quality problems, absences, and variations in the product mix mean that the actual lead time is sometimes shorter and sometimes longer than forecast.

A safety margin is therefore needed:

$$\text{Promised delivery time} = \text{expected waiting time} + \text{processing time} + \text{safety margin}$$

Suppose the new order requires another day of actual processing after the expected waiting time. The expected lead time is then 25 working days. Historical data shows, for example, that 95 per cent of orders are ready no more than four days later than predicted. You can therefore promise a delivery time of 29 working days.

You do not simply add that safety margin for extra certainty. You determine it based on the desired delivery reliability. Historical deviations between predicted and actual lead times form the basis for this. This is an important difference compared with fixed delivery times. Historical data no longer wholly determines the promised delivery time: it simply helps determine the uncertainty around the current forecast. The current workload remains the starting point.

Use a realistic output

For the calculation, you need a realistic average output. For example, use the actual output over the past few weeks or months.

- **Do not use theoretical capacity:** a machine can theoretically produce 100 units per day, but that does not mean the entire production system delivers 100 units every day. Waiting times, breakdowns, material problems, and other constraints are reflected in actual output, but not in theoretical capacity.
- **Do not choose too short a measurement period:** an exceptionally good or bad week is rarely representative. A moving average over a longer period provides a more reliable picture.
- **Recalculate after structural changes:** are you adding a shift or a machine? Your representative output may also change.

Principle 2: do not treat rush orders as exceptions

Do rush orders occur regularly? If so, your delivery commitments must take them properly into account. You can then process urgent work without constantly disrupting other delivery commitments. Future rush orders are not yet in the current workload, but they will come in, jump the queue, and take up capacity. This means that the full system output is not available for normal orders. You should therefore systematically factor the average share of urgent work into your calculation.



Calculating the delivery date for normal orders

Suppose the system delivers an average of 50 units per day and urgent work represents an average of 20 per cent of the total workload. For normal orders, this leaves, on average:

$$50 \times 80\% = 40 \text{ units of output per day}$$

This reduction to 40 units per day takes into account the new urgent work that will come in later on. Existing rush orders must still be counted separately. Suppose there are 1,000 normal units and 200 urgent units ahead of a new normal order.

The expected waiting time becomes:

$$(1,000 + 200) / 40 = 30 \text{ working days}$$

Why not 24 working days? A calculation of **$(1,000 + 200) / 50$** ignores future rush orders. They will continue to come in during the waiting time and be given priority.

Why not 25 working days? A calculation of **$1,000 / 40$** overlooks the 200 existing urgent units. They have to be processed first as well. The expected waiting time is therefore 30 working days.

Calculating the delivery date for rush orders

Rush orders are treated as a separate priority category with its own delivery commitment. Suppose there are currently 200 urgent units in the queue ahead of the new rush order. This category is given priority, so the existing rush order queue can be processed using the full system output.

The expected waiting time is then:

$$200 / 50 = 4 \text{ working days}$$

We assume here that new urgent orders join the back of the rush order queue. You then add the processing time and an appropriate safety margin again. For rush orders too, you should periodically monitor actual performance in order to keep the safety margin in line with your desired delivery reliability.

Prevent priority inflation

The proportion of rush work must remain sufficiently stable. If more and more orders are treated as urgent, the result is priority inflation. Ultimately, almost every order is urgent and the priority rule loses its effectiveness.

Principle 3: keep your capacity model simple

The strength of this method lies in its simplicity: you do not need to know exactly where each order is, and you do not need to model every future machine operation. You need to know two things: how much relevant work is ahead of the new order and how quickly your production flow will process that work.

Workload and output must use the same unit. If orders are roughly the same size, the number of orders may be an adequate unit. Where there are larger differences, the number of units is often more suitable. In other environments, kilograms, square metres, number of boards, or installation days may be more appropriate.

The most detailed unit is not necessarily the best. Hours seem accurate, but require reliable processing times. You also need to decide exactly which hours you will count. If there are different work centres and routings, such a calculation quickly becomes complex and error-prone, so start with the simplest unit that represents your amount of work well enough.

When are product families useful?

A division into product families will sometimes improve the forecast. Suppose the workload consists of 900 windows and 300 doors. Production processes an average of 30 windows and 20 doors per day. You then get:

- Windows: $900 / 30 = 30$ working days
- Doors: $300 / 20 = 15$ working days

In this way, you take account of consistent differences between product flows. You do not need to model every individual work centre to do this. Only divide into product families when each family genuinely has a different routing or output. This simple method will also become hard to maintain if there are too many categories.

When does this method work well?

This method works particularly well when orders follow roughly the same production route, with the lead time mainly being determined by the amount of work still to be processed. A second condition is that the work that is counted really does need to be carried out before the new order.

Orders that have already been entered but do not need to be delivered for several months do not automatically belong in this queue. Otherwise, you will overestimate the workload and promise unnecessarily long delivery times. This approach is therefore particularly suitable for production environments in which customers want delivery as quickly as possible. Do customers order well in advance? Or do they reserve capacity for specific future periods? Then you need a different planning method.

Materials and technical information must also be available on time. If a critical component only arrives after the calculated production start, you will need to check material availability separately. The delivery date will then be determined by the constraint that is resolved last: production capacity or material availability.

Finally, your average output must remain representative. If capacity undergoes a permanent change, you must also recalculate the expected output.

From average output to probability distribution

This method uses an average output, which makes the calculation simple and transparent, but hides the daily fluctuations in production. In the next article, we therefore look at how a Monte Carlo simulation can take account of the historical variation in daily output. Instead of a predicted lead time, we then directly obtain the delivery time that corresponds to, for example, 85, 95, or 99 per cent delivery reliability.

In summary

A reliable delivery date does not necessarily require a complex planning model. Combine the current workload with the actual average output and add a substantiated safety margin. Your delivery commitment will then automatically respond to busy and quiet periods and take account of the current state of your production.

Conclusion: ensure your delivery dates reflect the changing situation

Historical lead times remain useful, but do not tell you how busy your production is today. By combining the current workload and actual output, you get a simple and dynamic forecast. You can then promise delivery times that are better aligned with your capacity and desired delivery

reliability.

Do you have questions about production planning in variable environments?

Contact Sirris expert Pascal Pollet for more information.

[Get in touch](#)

Authors



Pascal Pollet