Practical Boundary Case Approach for Kanban Calculation on the Shop Floor Subject to Variation

Christoph Roser, Daniel Nold

Practical Boundary Case Approach for Kanban Calculation on the Shop Floor Subject to Variation

Christoph Roser[1], and Daniel Nold[2]

[1] Karlsruhe University of Applied Sciences, Moltkestrasse 30, Karlsruhe, Germany, christoph.roser@hochschule-karlsruhe.de
[2] Dr. Ing. h.c. F. Porsche AG, Porscheplatz 1, 70435 Stuttgart, Germany

Abstract: Determining the number of kanban in a pull production is essential for a good system performance. While this is simple in a static case, it becomes much more challenging if the system is subject to variations, as all practical systems are. The paper presents a novel boundary case approach to determining the number of kanban subject to a desired delivery performance. The approach is compared with the well-known Toyota formula. An extensive simulation-based approach verification for different coefficients of variations for the replenishment time and the customer takt was used to verify this boundary case approach also in comparison with the Toyota formula. The presented approach handles variation much better and determines a much more suitable number of kanban. This approach can also be adapted to practical everyday use in industry with little additional effort and with only a limited understanding of the underlying randomness of the key variables.

Keywords: kanban formula, pull production, random variation, replenishment time, lean manufacturing

1   Introduction

Lean production is one of the best approaches for promoting business excellence through continuous improvement [1]. One of the key concepts is pull production. The idea of pull production is based on limiting the maximum number of products or jobs processed in the system at the same time. The most common approach to implement pull production is through kanban, especially for high-volume low-mix products. In setting up a kanban system, the number of kanban cards has to be determined. There exist a variety of different kanban formulas. All of them are based on the relation between the replenishment time and the customer takt as shown in (1), where RT is the replenishment time, TT is the customer takt, NPK are the number of parts per kanban, and K is the number of kanban.

$$K = RoundUp \left[\frac{RT}{TT \cdot NPK} \right] \qquad (1)$$

As we will see below, this equation works well for a static system, but presents a challenge for dynamic systems that include random variations. Unfortunately, all real-world systems are dynamic, and these random variations cannot be ignored. This paper will

look at the relation between replenishment time, customer takt, and the number of kanban using both theoretical and experimental approaches.

2   Theoretical Approach

2.1   Static Situation

In the first part we assume a static situation with no random fluctuations. This of course is a widely optimistic and unrealistic assumption for practical use. However, it is useful to understand the theoretical foundations. A key variable here is the customer takt. It represents the average demand of the customer during a time period. To calculate it, you divide the available working time by the customer demand during that time, as shown in equation (2) [2], where TT is the customer takt, AT is the available work time, and CD is the customer demand during that time:

$$TT = \frac{AT}{CD} \tag{2}$$

Customer Takt slower than or equal to replenishment time

If the customer takt is larger than the replenishment time, one Kanban card will suffice for production. If the customer wants a product, he receives the single piece in stock, and then we reproduce it to increase stock again to one piece. Hence, the behavior of the inventory over time may look like Figure 1. As the RT is faster than the TT, the product is always completed before the customer requests the next one. This concurs with equation (1) above.
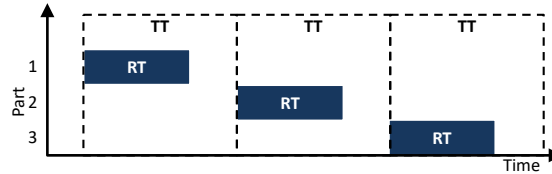
Figure 1: Theoretical behavior if customer takt is slower than the replenishment time

Customer Takt faster than replenishment time

If the customer takt is larger than the replenishment time, we would need multiple Kanban cards as calculated in equation (1). An example is shown in Figure 2, where the replenishment time is almost three times as long as the customer takt. Hence, we would need three kanban to have one part ready before the customer demands it.
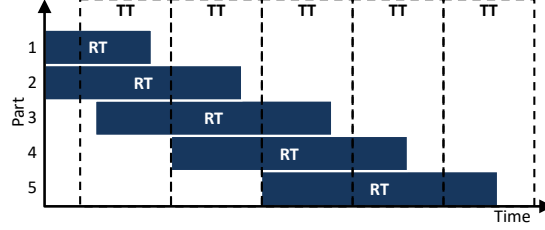
Figure 2: Theoretical behavior if customer takt is faster than the replenishment time

2.2 Dynamic Situation

In reality both the replenishment time and the customer takt have random variations. The distribution of customer takts can often be approximated using exponential or log-normal distributions. For replenishment times we also often find log-normal distributions, Weibull, Gamma, or Pearson type 5 distributions. Overall, the equation (1) would now consist not of discrete numbers but random variables. Hence, equation (1) would change to equation (3).

$$K = \frac{RT}{TT \cdot NPK} \qquad (3)$$

There are methods available to calculate the division of probability functions. The general ratio distribution is shown in equation (4), where $p_K(k)$ is the probability density function of the number of kanban needed [3].

$$p_K(k) = \int_0^\infty |TT| \, p_{RT,TT}(k \cdot tt, tt) dtt \quad (4)$$

However, this requires independence of the variables, which is not given in our case. The replenishment time depends on the customer takt, as more orders increase the replenishment time. Furthermore, the replenishment time also depends on itself, as, for example, a part with a long replenishment time is usually followed by another part with a long replenishment time. Finally, the option of buffering parts in a supermarket is not included in this simplified model. Hence, the theoretical approach has unrealistic assumptions and cannot be used. Therefore the problem in determining the number of kanbans is how to deal with these uncertainties.

3 Simulation Approach

We have also analyzed manufacturing systems using a simulation-based approach. In this simulation, a single process is controlled using a kanban system in order to supply parts to a customer. The customer demand is modeled using a lognormal distribution. The processing time is modeled using a Pearson Type V distribution. Due to the large number of possible variables, an exhaustive search is prohibitive. Nevertheless, we simulated a wide range of different systems, varying three primary variables:

4

- The relation of the customer takt to the process cycle time, which effectively changes the utilization of the process
- The standard deviation of the customer takt
- The standard deviation of the process cycle time.

Additionally, the number of kanban cards was varied exhaustively between 1 and 50, with additional simulations having up to 1000 kanban cards to also test for systems with extremely large number of kanban cards. Each simulation was repeated 30 times for statistical validity and to allow the calculation of confidence intervals. Two primary results were measured:

- Work in process inventory: how many parts were in the system
- Delivery performance: what percentage of the customer request were fulfilled immediately. Not immediately fulfilled request were fulfilled when the next part became available.

The full details of the simulation experiment can be found in [4]. Figure 3 shows the location of the kanban cards in the system for different number of kanban cards for both the customer takt and the cycle time having a coefficient of variation of 1, with a mean customer takt of 12 and a mean cycle time of 10. Showing all results from [4] would exceed the scope and space available within this paper.
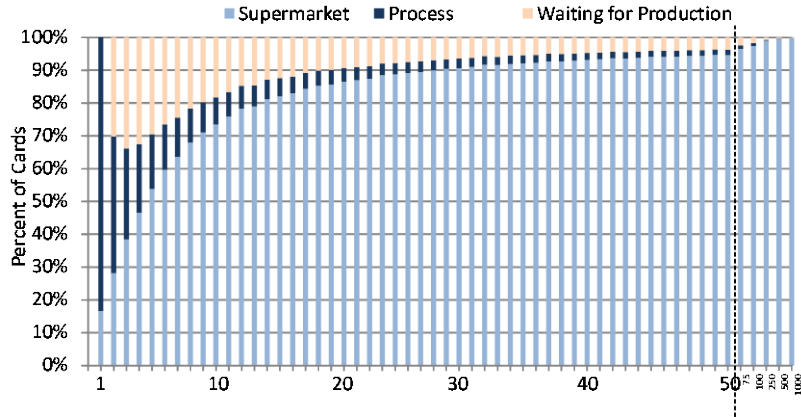
Figure 3:Location of the kanban cards for a system with the customer takt and the cycle time having a coefficient of variation of 1, a mean customer takt of 12 and a mean cycle time of 10. The system was simulated for all kanbans up to 50, and then for selected kanbans up to 1000.

4   The Toyota Formula

The formula used by Toyota to calculate the number of kanban is called the Toyota Formula, and is widely used in literature, e.g. [5] [6] [7]. This approach calculates the number of kanban based on the average daily demand *DD*, the replenishment time (called lead time, but includes the times for collecting kanban), the container capacity (i.e., the number of parts per kanban *NPK*), and a safety factor α as shown in equation (5).

$$K = \frac{DD \cdot RT}{NPK} \cdot (1 + \alpha) \tag{5}$$

As the daily demand DD is merely the inverse of the customer takt TT, equation (5) translates into equation (3) with an additional safety margin. The safety factor α is to account for uncertainties, and is generally recommended to be no more than 10%. We verified this Toyota formula using the simulation and the resulting data from [4]. For full details on the simulation data please refer to [4]. The average for the simulated systems was a delivery performance of 71.3%, which is still far from acceptable for a functioning production system.

5 A practical Boundary Case Kanban Approach

5.1 Description of the Boundary Case Kanban Approach

The boundary case approach for calculating the number of kanban is based on equation (1). However, instead of using the mean value for the replenishment time and the customer takt, we now use boundary cases. This is not the extreme case, as customer takts and process times are time-based and hence have boundaries of [0,∞], which would give us an infinite number of Kanban in the worst case. More practically, we will be using the values for selected percentiles of the distribution. For the replenishment time we want to cover all but the slowest replenishment times. Similar is true for the customer takt, but now we want to cover all but the fastest customer takt. The selected percentile depends on the desired reliability and delivery performance of the kanban system.

5.2 Simulation Verification of the Boundary Case Kanban Approach

Using the same simulation data as from the verification of the Toyota formula, we now verified the boundary case approach. We selected to verify this approach for 80% percentiles, 90% percentiles, and 95% percentiles on both all but the slowest replenishment time and all but the fastest customer takts. Table 1, Table 2, and Table 3 show the resulting number of kanban for a coverage of 80%, 90%, and 95% respectively for the customer takt and the replenishment time for different coefficients of variation.

Table 1: Number of kanban for different coefficient of variations, covering 80% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	1	2	3	7	14
	25%	2	2	4	8	16
	50%	3	4	6	11	20
	75%	5	6	9	15	23
	100%	6	8	13	19	31

Table 2: Number of kanban for different coefficient of variations, covering 90% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	1	2	4	9	19
	25%	2	3	6	11	23
	50%	3	5	9	16	31
	75%	5	8	12	23	38
	100%	8	11	19	30	51

Table 3: Number of kanban for different coefficient of variations, covering 95% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	1	2	5	12	26
	25%	2	3	7	15	33
	50%	4	6	11	22	44
	75%	6	9	17	33	56
	100%	9	14	25	43	78

The resulting delivery performance is shown in Table 4, Table 5, and Table 6. Overall, the performance is much more satisfactory. If both the customer takt and the replenishment time is set to cover 80% of the slowest and fastest times respectively as shown in Table 4 the average delivery performance across all tested situations is 87,8%. Similarly, for a coverage of 90% as shown in Table 5, the average delivery performance is 93.3%. Finally, for a coverage of 95% as shown in Table 6, the average delivery performance is 95.4%.

Table 4: Delivery performance for different coefficient of variations using the boundary case approach covering 80% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	100,0%	96,9%	82,4%	89,4%	92,9%
	25%	96,2%	84,1%	86,3%	90,2%	94,1%
	50%	82,2%	85,8%	86,1%	90,2%	95,3%
	75%	78,3%	81,4%	85,8%	89,0%	92,0%
	100%	71,6%	80,0%	85,2%	87,9%	92,4%

Table 5: Delivery performance for different coefficient of variations using the boundary case approach covering 90% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	100,0%	96,9%	92,2%	93,7%	96,9%
	25%	96,2%	96,0%	95,9%	95,8%	98,2%
	50%	82,2%	91,2%	95,4%	96,8%	99,1%
	75%	78,3%	89,1%	91,3%	96,5%	98,6%
	100%	78,2%	87,0%	93,1%	95,4%	97,6%

Table 6: Delivery performance for different coefficient of variations using the boundary case approach covering 95% of the customer takt and replenishment time

		Customer Takt				
	CV	0%	25%	50%	75%	100%
Production Process	0%	100,0%	96,9%	96,2%	97,9%	99,0%
	25%	96,2%	96,0%	97,5%	97,9%	99,8%
	50%	90,2%	94,4%	97,4%	98,9%	99,8%
	75%	83,0%	90,5%	96,5%	98,4%	99,5%
	100%	80,5%	87,2%	96,6%	95,9%	98,4%

Overall, the calculated number of kanban is much more suitable for the needs of the industry. Delivery performances of 87%, 93%, or 95% for the different percentile boundaries are much more acceptable than an average delivery performance of 71% using the Toyota formula. The level of coverage can also be influenced by the percentile of the boundary case. Covering 95% of the fastest and slowest cases for the replenishment time and customer takt respectively will give a higher number of kanban and hence an overall better delivery performance than a coverage of only 90%, which in turn will again be better than a coverage of 80%.

5.3 Practical Application of the Boundary Case Kanban Approach

The boundary case approach to determine the number of kanban works very well as verified using extensive simulation data. For a practical application, however, these 80%, 90%, or 95% percentiles are difficult to determine. While the mean customer takt or mean cycle time are usually well known on the shop floor, the shape of the distribution of this random variable is usually unknown. Hence, there is usually insufficient data to determine an 80%, 90%, or 95% percentile.

However, there is usually either historic data or – more commonly – personal experience of "bad days," where either the customer demand or the replenishment time was exceptionally difficult, meaning very fast or very slow respectively. Therefore, while sacrificing some accuracy, we propose for practical purposes to select the worst case that has to be covered either from data or from the experience of the personnel.

In personal experience, shop floor supervisors seek safety in numbers and may want to cover worse situations but change their opinion once they see the resulting number of kanban and hence potential inventory. Therefore, we also propose that this selection of the boundary cases to be covered is not done in independence, but should involve a back and forth of selecting boundary cases and determining the number of kanban until a suitable compromise of boundary case coverage and number of kanban is achieved. Thus we achieve a suitable tradeoff between a high delivery performance (lots of kanban) and low inventory (fewer kanban).

6 Summary

The presented boundary case approach uses the well-known relation of the replenishment time to the customer takt as in equation (1) to determine the number of kanban.

However, rather than mean values, this approach uses boundary cases (i.e., the largest replenishment time and the smallest customer takt that the system has to cover). This can be either expressed as a percentile of the distribution function if the function is known, or more practically based on the boundary cases known from experience or historic data that the system has to cover. The approach is intuitive, and requires little knowledge of statistics, hence making it suitable for everyday shop-floor use. Nevertheless, the approach has been extensively verified based on simulations, and returns much better delivery performances with much less outliers than the conventional Toyota formula. This information can be used to aid with the implementation of the Kanban system, regardless if it is a paper Kanban, a digital Kanban (where the time for the information flow is usually significantly less), or any other form of Kanban.

7   References:

[1]   J. Kurilova-Palisaitiene and E. Sundin, "Challenges and Opportunities of Lean Remanufacturing," *International Journal of Automation Technology*, vol. 8, no. 5, pp. 644–652, Sep. 2014.

[2]   A. B. Badiru and O. A. Omitaomu, *Handbook of Industrial Engineering Equations, Formulas, and Calculations*, New. edition. Boca Raton, FL: Industrial Innovation, 2010.

[3]   J. H. Curtiss, "On the Distribution of the Quotient of Two Chance Variables," *The Annals of Mathematical Statistics*, vol. 12, no. 4, pp. 409–421.

[4]   D. Nold, "Simulationsbasierte Analyse der Liefererfüllung und des Umlaufbestands verbrauchsgesteuerter Produktionssysteme in Abhängigkeit von der Anzahl verwendeter Kanban-Karten," Master Thesis, Karlsruhe University of Applied Sciences, Karlsruhe, Germany, 2017.

[5]   B. J. Berkley, "A review of the kanban production control research literature," *Production and Operations Management*, vol. 1, no. 4, pp. 393–411, 1992.

[6]   Y. Monden, *Toyota Production System - An Integrated Approach to Just-In-Time*. Engineering & Management Pre, 1998.

[7]   F. T. S. Chan, "Effect of kanban size on just-in-time manufacturing systems," *Journal of Materials Processing Technology*, vol. 116, no. 2–3, pp. 146–160, Oct. 2001.