



**HAL**  
open science

## Practical Estimation of the Impact of a Reduction of the Number of Kanban Cards on the Delivery Performance

Christoph Roser, Yannik Regending, Bernd Langer, Claas-Christian Wuttke

### ► To cite this version:

Christoph Roser, Yannik Regending, Bernd Langer, Claas-Christian Wuttke. Practical Estimation of the Impact of a Reduction of the Number of Kanban Cards on the Delivery Performance. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2021, Nantes, France. pp.690-696, 10.1007/978-3-030-85874-2\_75 . hal-04030419

**HAL Id: hal-04030419**

**<https://inria.hal.science/hal-04030419>**

Submitted on 16 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Practical Estimation of the Impact of a Reduction of the Number of Kanban Cards on the Delivery Performance

Christoph Roser<sup>1</sup>, Yannik Regending<sup>1</sup>, Bernd Langer<sup>1</sup>, Claas-Christian Wuttke<sup>1</sup>

<sup>1</sup> Karlsruhe University of Applied Sciences, Moltkestrasse 30, Karlsruhe, Germany,  
Corresponding author: christoph.rosen@hochschule-karlsruhe.de

**Abstract:** Kanban systems are the best-known variant of pull systems and a cornerstone of lean manufacturing. Kanban systems help to maintain a good material availability in relation to the inventory. One of the main adjustments is the number of kanban cards in a kanban system. The number of kanban is a tradeoff between the cost of inventory, and the cost of a stock-out. Since kanban are used for make-to-stock production, a high availability of completed products is desired, often measured as the delivery performance (on time in full). At the same time excessive inventory is to be avoided to reduce inventory cost and to improve throughput. This paper describes a practical approach to estimate the impact of a reduction of kanban on the delivery performance, based on an analysis of the supermarket inventory.

**Keywords:** Kanban, Supermarket, Delivery Performance

## 1 Introduction

Lean manufacturing is one of the significant frameworks for manufacturing optimization. It includes a multitude of different methods and philosophies, from eliminating waste, fluctuations, and overburden to improving quality. Lean was also either the basis or a significant inspiration for other process performance related frameworks like six sigma, Agile, and others, and is often related to topics like Industry 4.0 [1, 2].

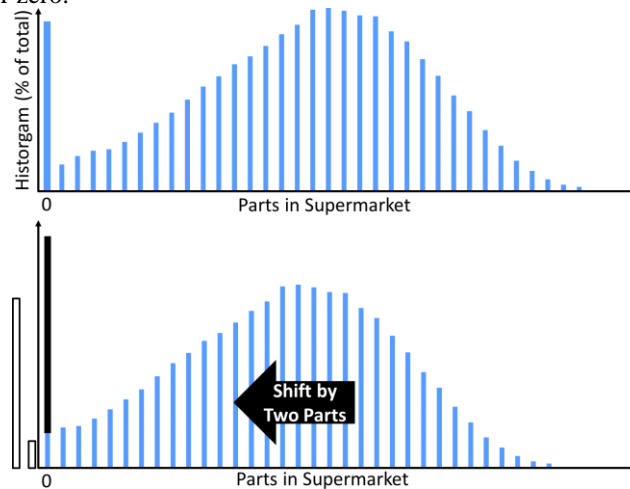
One of these significant philosophies is the idea of flow, with material moving from process to process rather than idling in inventories, improving the lead time. As per Little's law [3], the lead time is directly related to the inventory. Subsequently, inventory reduction is a key aspect of lean manufacturing [4]. Pull production is one of the significant tools to manage an inventory. While there are different definitions of pull systems, it is a limit on inventory coupled with a signal to refill this inventory. Kanban is the best known version of pull, and an excellent tool for a tradeoff between inventory cost and stock outs in make-to-stock production [5]. These stock outs are often measured as the delivery performance, i.e. the percentage of deliveries that were both on time and delivered in full, sometimes abbreviated as OTIF (On Time In Full). A kanban system aims to replenish the target inventory, often known as a supermarket. Setting up a kanban system for a good tradeoff requires among other things to determine the number of kanban for each product type in the pull system. Different versions of

kanban formulas [6, 7] help with determining the number of kanban, but all of these are merely an estimation. Depending on the assumptions, the number of kanban may be less than ideal for the system. It may be much easier to adjust the number of kanban based on the actual performance of the system. While this can be done using trial and error exploration of the system, this is impractical for real world systems and usually limited to simulations. Different authors propose a dynamic adjustment [8, 9], although it may be difficult to adjust kanban dynamically.

This paper proposes a practical tool for the estimation of a delivery performance for a reduction in the number of kanban cards based on the historical supermarket inventory data for a given system.

## 2 Prediction Model

The model to predict the reduction in delivery performance for a reduction in the number of kanban uses the supermarket histogram. **Fig. 1** shows the model using a generic histogram of a supermarket. We predict that a reduction in the number of kanban shifts the histogram a corresponding number of parts to the left (for our model one kanban represents one part, but this may be different for other systems). When shifting the histogram to the left, there can never be less than zero parts in the system. Hence, any value in the histogram that would go below zero is added to the value in the histogram for zero.



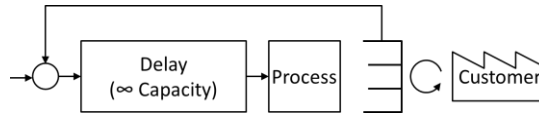
**Fig. 1.** Kanban Reduction Prediction Hypothesis

This updated histogram value for zero is the predicted likelihood of the supermarket being empty, and can be used to predict the delivery performance. This approach is mathematically easy enough to be feasible for most manufacturing companies. In mathematical terms, the new delivery performance  $D_{New}$  is approximately the remainder to 1 of the sum of the individual histogram values  $h_n$  from zero to the number of parts  $m$  that are represented by the removed kanban as shown in equation [1].

$$D_{New} \approx 1 - \sum_{n=0}^m h_n \quad [1]$$

### 3 The Simulated System

The behavior of the inventory in a supermarket depends on many factors. Relevant for us are the number of kanban, the replenishment time, and the magnitude of the fluctuations. We use simulations to model a simple system producing a single part as shown in **Fig. 2**.



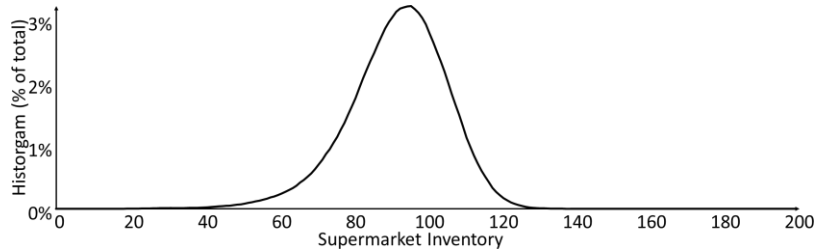
**Fig. 2.** The Simulated System

A part can only enter the system if a kanban is available, where one kanban represents one part. To model larger systems every part in the system has a fixed delay, where multiple parts can be delayed simultaneously. A separate process has an erlang distributed processing time. The erlang distribution has a finite lower bound of zero but an infinite upper tail, making it well suited to model processes. This process provides the fluctuations for the supplying system that fills up the supermarket.

A customer with an exponentially distributed interarrival time picks up one part from the supermarket. If there is no part available, the customer waits. We measure the percentage of customers that do not have to wait for a part, which is the delivery performance. A part departing the system also gives a signal (the kanban) to release the next part in the system from an infinite supply of parts. Overall, this system is a simplified version of complex systems, reducing the number of variables while still behaving comparable to larger systems.

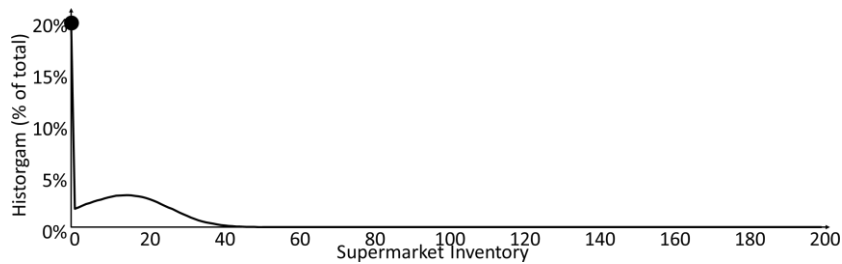
The exponentially distributed customer interarrival time was set to 10 time units. The exponentially distributed process was given a mean of 9 time units. This gives the process a utilization of 90%. The delay was set to 991 time units. Together with the mean process time this gives a mean replenishment time of 1000 time units. Dividing the average replenishment time by the customer takt of 10 time units gives us *in average* 100 kanban. However, to account for fluctuations this would be insufficient, and 100 kanban will result in a delivery performance of zero if customers wait until they are served. In any case, the number of kanban is the variable we will test.

The simulation had a warm up period of 100 000 time units and a run time of 5 000 000 time units, producing around 500 000 parts per simulation. We measure the delivery performance and the histogram of the supermarket inventory. An example supermarket histogram for an excessive number of 200 kanban is shown in **Fig. 3**. The mean is around 100 as expected from the division of the mean replenishment time by the mean customer takt. The delivery performance is 100%, as the system never ran out of inventory.



**Fig. 3.** Supermarket Histogram for 200 Kanban

If we reduce the number of kanban, this histogram will shift to the left. **Fig. 4** shows the histogram for 120 kanban. The delivery performance is now only 80.61%. The supermarket is empty for 19.79% of the time, marked with a black dot in **Fig. 4** for better visibility. The rest of the histogram is almost identical with **Fig. 3** but shifted 70 parts to the left (note the different percentage scale).

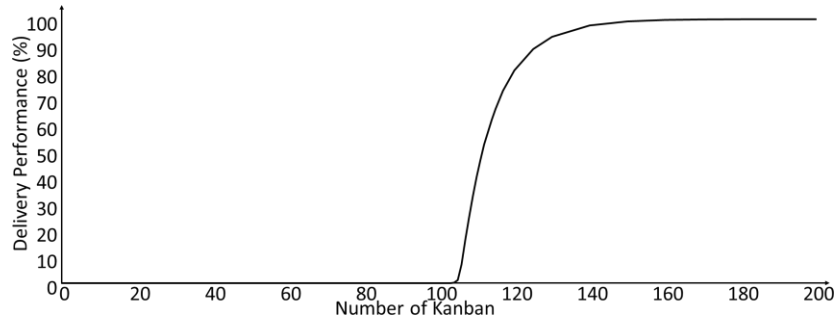


**Fig. 4.** Supermarket Histogram for 130 Kanban

The percentage of the time the supermarket was not empty corresponds almost perfectly to the delivery performance. The supermarket was empty 19.79% of the time, and 19.39% of the customers had to wait for a part. The difference of 0.4% represents the few cases where the supermarket was empty, but filled up again before a customer arrived. For all practical purposes, the percentage the of supermarket being empty is a good estimator of the delivery performance.

## 4 Model Verification

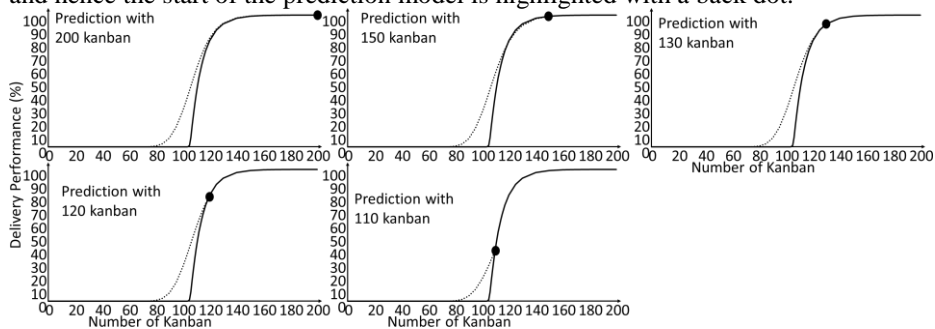
We used this simulation model to establish the delivery performance for different number of kanbans. We measured the delivery performance after a warm up time of 100 000 time units for an additional simulation duration of 1000 000 time units. Each simulation was repeated 30 times to determine the mean and the 95% confidence interval. **Fig. 5** shows the behavior of the delivery performance. The confidence intervals are not shown, as they would graphically overlap almost completely with the mean delivery performance.



**Fig. 5.** Delivery Performance of Simulation System for different Number of Kanban

The delivery performance was consistently 100% for over 190 kanban. This rapidly dropped to zero for 100 kanban. 100 kanban would be the average replenishment time of 1000 time units divided by the customer takt of 10 time units. Due to the customers waiting, the system is unable to recover with 100 kanban or less, and the subsequent delivery performance is zero.

We predicted the delivery performance based on selected supermarket histograms at 200, 150, 130, 120, and 110 kanban. The result is shown in **Fig. 6**, where the solid line is the true delivery performance and the dotted lines are the predictions. To make the graph easier to read, the number of kanban used to generate the supermarket histogram and hence the start of the prediction model is highlighted with a back dot.



**Fig. 6.** Prediction quality for different Number of Kanban

The prediction model is spot on for delivery performances over 70% with an average error of less than one percentage point, and still very good for delivery performances above 50% with an average error of less than 10 percentage points. For delivery performance below 50% however, the model no longer takes the cumulative effect into account of more and more customers waiting for their products.

For practical purposes, however, this is a very good prediction, as most companies strive to achieve higher delivery performances above 50%. Few companies would reduce the number of kanban to reduce inventory if the delivery performance is already below 50%.

Hence, this method is suitable for use in real world manufacturing systems. Two issues, however, can reduce the accuracy. First, an accurate supermarket histogram is needed,

which may not always be available. Secondly, the system should not change over the duration for which the histogram was taken. However, especially if there is a risk of a stock out, extraordinary measures are often taken and firefighting tries to prevent a stock out. Therefore, in reality the frequency of stock-outs may be less than in this theoretical model as emergency measures are taken to prevent such a stock out, whereas this theoretical model lets the system simply run its course.

## 5 Discussion and Summary

Overall, a supermarket histogram is an accurate tool to predict the delivery performance for a reduction of kanban for delivery performances in the range between 50 and 100%. The prediction accuracy worsens significantly for delivery performances below 50%, although a reduction in kanban is rarely desired for such inadequate delivery performances.

While kanban in general and pull in particular are not the only methods to manage make-to-stock production, it is usually the preferred approach in industry due to its reliability and ease of use. Other methods like push control requires much more management attention, which is often difficult due to the large number of different products produced in parallel in many factories. Hence, push often has an inferior performance to pull.

Overall, such a prediction can significantly help practitioners to fine-tune their kanban systems. A kanban system aims to have a trade-off between the availability of finished goods and the inventory needed to achieve this availability. Determining a suitable number of kanban is difficult. It is usually recommended to start with a sufficient or slightly excessive number of kanban, and reduce over time while observing the running system. This paper presents a tool to estimate the impact of the reduction of the number of kanban on the material availability, reducing the uncertainty in the management decision of this reduction.

Further research will look into increases in the number of kanban to improve the delivery performance as well as analyze systems where the customers are not waiting until they are served but will leave the system.

## 6 References:

1. Buer S-V, Semini M, Strandhagen JO, Sgarbossa F (2020) The complementary effect of lean manufacturing and digitalisation on operational performance. *International Journal of Production Research* 0:1–17. <https://doi.org/10.1080/00207543.2020.1790684>
2. Lai NYG, Wong KH, Halim D, et al (2019) Industry 4.0 Enhanced Lean Manufacturing. In: 2019 8th International Conference on Industrial Technology and Management (ICITM). pp 206–211
3. Little JDC (1961) A Proof for the Queuing Formula:  $L = \lambda W$ . *Operations Research* 9:383–387. <https://doi.org/10.1287/opre.9.3.383>



4. Marodin GA, Tortorella GL, Frank AG, Godinho Filho M (2017) The moderating effect of Lean supply chain management on the impact of Lean shop floor practices on quality and inventory. *Supply Chain Management: An International Journal* 22:473–485. <https://doi.org/10.1108/SCM-10-2016-0350>
5. Roser C (2021) All About Pull Production: Designing, Implementing, and Maintaining Kanban, CONWIP, and other Pull Systems in Lean Production. AllAboutLean.com, Offenbach, Germany
6. Co HC, Sharafali M (1997) Overplanning factor in Toyota's formula for computing the number of kanban. *IIE Transactions* 29:409–415. <https://doi.org/10.1080/07408179708966346>
7. Roser C, Nold D (2019) Practical Boundary Case Approach for Kanban Calculation on the Shop Floor Subject to Variation. In: *Advances in Production Management Systems Production Management for the Factory of the Future*. Springer, Austin, Texas, USA, pp 12–20
8. Gupta SM, Al-Turki YAY (1997) An algorithm to dynamically adjust the number of Kanbans in stochastic processing times and variable demand environment. *Production Planning & Control* 8:133–141. <https://doi.org/10.1080/095372897235398>
9. Tardif V, Maaseidvaag L (2001) An adaptive approach to controlling kanban systems. *European Journal of Operational Research* 132:411–424. [https://doi.org/10.1016/S0377-2217\(00\)00119-3](https://doi.org/10.1016/S0377-2217(00)00119-3)