ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.143

HVLV Manufacturing System Controlled with Heijunka-Kanban: A Simulation Study of Dead Time and Delay Constraints

Fatma Tangour [+] , Fidel Villalpando Rodriguez and Claude Martinez
IUT of Nantes University of Nantes and LS2N UMR CNRS 6004, Carquefou, 44475 France.

Abstract. In this work, we treat the Hight Variety Low Volume (HVLV) problematic using Heijunka lean manufacturing tool. In a manufacturing system, the principal objective is to satisfy the customer with quality and also by respecting the delivery delay. In this paper, we tackle the problem of sizing a Heijunka-Kanban production control system in order to cope with SLA delays. Unexpected delays that arise due to a variety of orders and dead time in the production loop are compensated with a suitable sizing of the Heijunka-Kanban control. Unfortunately, because of stock level constraints, unexpected delays may remain. Our work presents a typical case study of the automotive industry, a provider of rubber hoses.

Keywords: Heijunka, Kanban, Dead time, Delay, Simulation, HVLV

1. Introduction

In the work of I. Nasri et al [1], HVLV systems are characterized by a wide variety of products using shared machines, a weak and personalized demand, relatively long processing times and frequent change over and set-up times. In the work of S. Wilson [2], HVLV consept refers to the variability of the orders, they propose a classification based on a coefficient defined as the standard deviation of weekly demand divided by the average of the weekly demand. Productions systems that are controlled with tools of the Lean philosophy, as Heijunka-Kanban for instance, are sensitive to orders variability [3], thus researchers have provided some methods to model and take into account these variabilities. In his work [4], A. M. Deif presents a stochastic model to model uncertainties in lean production cells. Lean Manufacturing is a management philosophy focusing on reduction of the seven types of waste: over-production, waiting time, transportation, processing, inventory, motion and scrap in manufacturing or any type of business [5], [6]. To improve transparency and to reduce system variability, Lean manufacturing tools introduce and steadily improve flow production [7]. Indeed, Production Levelling (Heijunka) improves operational efficiency by means of flexibility, cost and service level [8], [9].

The assumption underlying Heijunka is that the producer has a choice concerning the amount of variability in the job arrival sequence to be admissible or not [10]. According to Matzka's article [11] and Korytkowski's [12], the Heijunka approach focuses on the queues of different process based on a statistical calculation. The orders are inserted in a Heijunka table. The rows represent product types and the columns represent specific production times. This model uses a notion called EPEI (Each Part Each Interval). Nevertheless, the adaptive lot sizing approach presented in [12] results in a fluctuating time period that is designed to compensate orders variation, on the other hand-side, the classical Heijunka approach results in a fixed period for each product type. This method defines the time interval of the production time for each product to be processed. We applied this method to determinate the initial values of parameters in the optimization tool of FlexSim in our case-study.

Several studies have been carried out where the determination of the number of Kanban cards was based on the phenomenon of queues. Many articles from the literature evoke three methods to size the number of Kanbans: simulation study, programming of the mathematical model and the Markov chain model. Davis and Stubitz, [13], developed a simulation to assess the performance of a Kanban system. They found and optimized certain parameters, tested different combinations in order to optimize queues of orders.

[+] Corresponding author.
E-mail address: fatma.tangour-toumi@univ-nantes.fr.

Another approach is to combine simulation and evolutionary principles in order to determine the optimal number of the Kanban cards. For example, the work of Shahabudeen et al. [14] with a multi-product application. Another approach concerns the mathematical simulation models of Kimura and Terada, [15]. They developed the first mathematical model for a Kanban system by formulating basic equilibrium equations for a mutli-production. Wang and Wang, [16], set up a continuous time Markov model. This model contains several items, several process steps, a double Kanban card with a simple withdrawal for each step of the process. Matzka's article [11] describes a dual Kanban card system. The concept of multi-product and large volume falls within this framework. This analysis shows us how to analyse a levelling box as well as Kanban control where customer orders are filtered and limited thanks to the Kanban loop.

In the remaining part of this article, we first examine the impact of an additional delay modifying a single synchronisation mechanism made of two Kanban loops studied in [3], [11]. In section III, we present a case study of the automotive industry. We conclude this article with a brief discussion of the straightforward application of general methods as Heijunka on HVLV constrained production systems as the presented case study.

2. The effect of decision delay in the Kanban loop

In the automotive industry, the use of a specific agreement that regulates the relations between provider and consumer is common. In such a Service Level Agreement (SLA), the delivery delays, quality compliance, service level rate are specified. We study in this article some effects of delays in the layout sizing of an automotive components' provider. We first define the Service Level Customer (SLC). For a given set U of orders u_i, each order is given with a delivery delay τi, we define a subset IU of orders delivered on time. The SLC is the ratio between the number of orders delivered on time and the number of orders in *U*:

$$\mathrm{SLC} = \frac{|{}^{I}\mathcal{U}|}{|\mathcal{U}|},$$

$|\cdot|$ denotes the cardinality of a set. We denote by $y_i(t)$ the instant of delivering the order u_i at the output buffer of the workshop, $u_i(t)$ is the instant of arriving for order u_i. The order u_i is delivered on time, say:

$$u_i \in {}^{I}\mathrm{U} \text{ if } y_i(t) \leq u_i(t + \tau_i).$$

Beside the durations of processing time and transportation times in a factory, some additional delays can appear in the decision process as for instance the time necessary to decide which fabrication order will be launched next. We denote η_i the delay associated to order u_i. In order to design the control system of a workshop that runs a Heijunka-Kanban scheme, one has to select the correct number of Kanbans at each stage of the process. Section I, gives a insight of methods to solve this kind of problems in the literature. Out of these methods, we have studied the approach presented in [7], [11] that is based on a Markov chain analysis of the synchronisation phenomena that occurs with Kanban controlled workshop.

In their work [11], authors have presented a detailed method to calculate the optimal number of Kanbans, i.e. production Kanbans, denoted K_p and withdraw Kanbans (K_w), in a workshop controlled with an Heijunka-levelled production system. The model they used do not include specific decision delay in the synchronization Kanban loop. In the present section, we present some experiments that conclude on the importance to take the delays into account in the sizing of a Kanban controlled workshop.

2.1. Problem statement

We analyse and discuss here the problem of a workshop that produces parts and deliver them to a specific customer in a given SLA framework. Specifically, we focus on the number of Kanbans that a single synchronization loop workshop should have in order to solve optimally the SLA compliance problem.

The objectives are:

- minimise the number of lost orders;
- minimise the cost of delayed orders (those that are not delivered on time as specified in the SLA);
- minimise the decision delay in the workshop.

The constraints are:

- respect the SLC specified in the SLA;
- take into account the stocks capacities of the workshop, i. e. $K_p \in [K_{pmin}, K_{pmax}]$;
- take into account the decision delay $\eta_i \in [\eta_{min}, \eta_{max}]$ in the workshop;
- take into account the number of orders $\omega \in [\Omega_{min}, \Omega_{max}]$ in the decision process.

2.2. Experiment

In this section, the workshop produces parts of a single type. The SLA specifies an objective SLC = 95%.

The simulation experiments have been done with the FlexSim simulation tool, a previous work with a similar aim has been presented in [17]. The Kanban loop that was studied is similar to the one presented in [11], it is composed of a single machine, a buffer of finished parts at the supplier place and a buffer at the consumer place. There is a milk-run between the supplier and consumer. The whole model is controlled with a pair of Kanban loops that are synchronized. At the supplier's place, the loop contains Kp Kanbans, and the milk-run loop contains Kw Kanbans. The processing time on the unique machine is constant and equal to one second. The average duration of a milk-run is one second. The orders arrive at the supplier and the time between arrival of two orders is modelled as a random variable with an exponential distribution, exp(λ) with λ = 1s. The main difference of the Kanban loop studied here and the one presented in [11] is that we have included a delay that models a decision step at the workshop. Indeed, orders that arrive at a workshop are received and prepared by a team leader before they are launched or picked in a Kanban buffer. In the present study, a maximum number of orders, say **Ω**max can be received and treated at the same time. The decision step delay is constant and equal to η in that section. Different values of η and Ω_{max} are tested in the simulation experiments and the resulting optimal (minimal) values of K_p, K_w that lead to a minimal SLC ≥ 95% are given as results for each configuration. Our simulation model is given in figure 1.

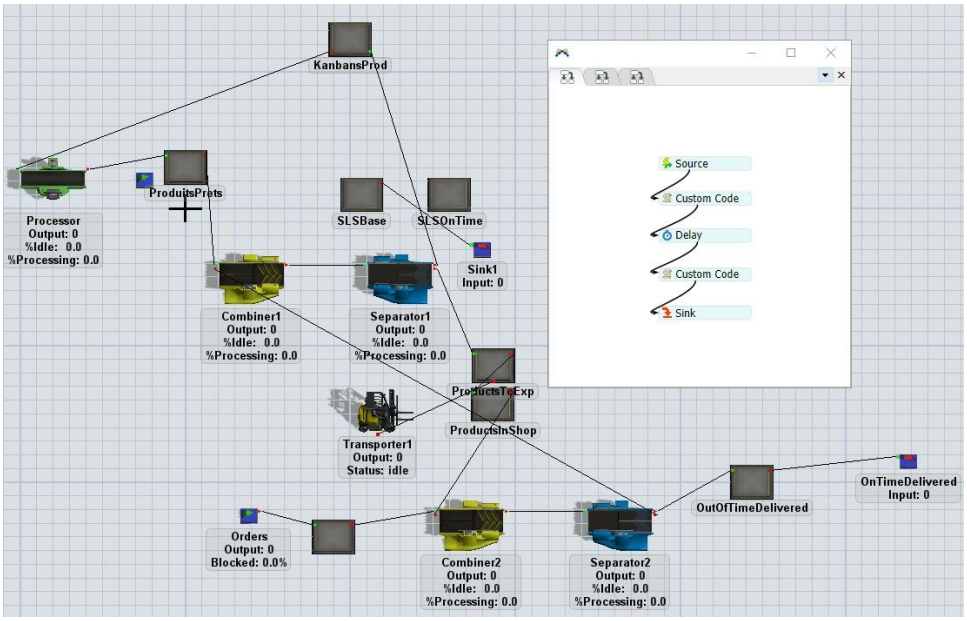

Fig. 1: The Kanban controlled workshop simulation model.

2.3. Experimental results

Each configuration has been tested as follows: ten simulation replications with random seeds, a simulation replication refers to 16 hours in the workshop. The first experiment concerns the original configuration similar to the one studied in [11]. The results are given in figure 2 and 3. The value of $K_w = 5$ as recommended in [11] and the values of K_p are taken from {72, 73, 74, 75, 76} around the optimal 73 with respect to the criteria of [11].

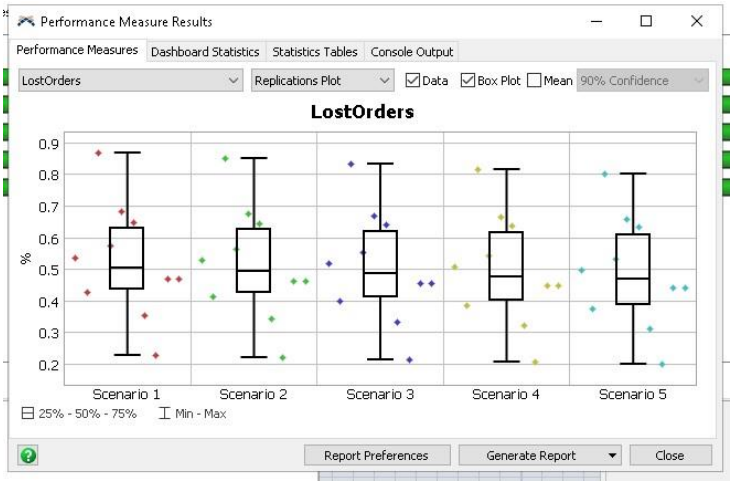

Fig. 2: Lost orders for $K_w = 5$, $K_p \in \{72,73,.74,75,76\}$, no delay in the decision process.

Notice that there is a slight difference between our results and those of [11], in their work, SLC ≥ 95% is achieved with $K_w = 5$, $K_p = 73$, in our study, with these parameters we obtain only 93.81% for that configuration. Recall that our results are given by a terminating simulation without warmup period and theirs result from an analytical computation. The second experiment concerns the inclusion of the decision step delay in a configuration similar to the one studied in [11]. The results are given in figure 4 and figure 5. The value of $K_w = 5$ as recommended in [11] and $K_p = 73$. The delay at the decision step is compensated with an extra delay τ at delivery as indicated at the top of the section. The simulations are run with a limit of five orders at the decision step.

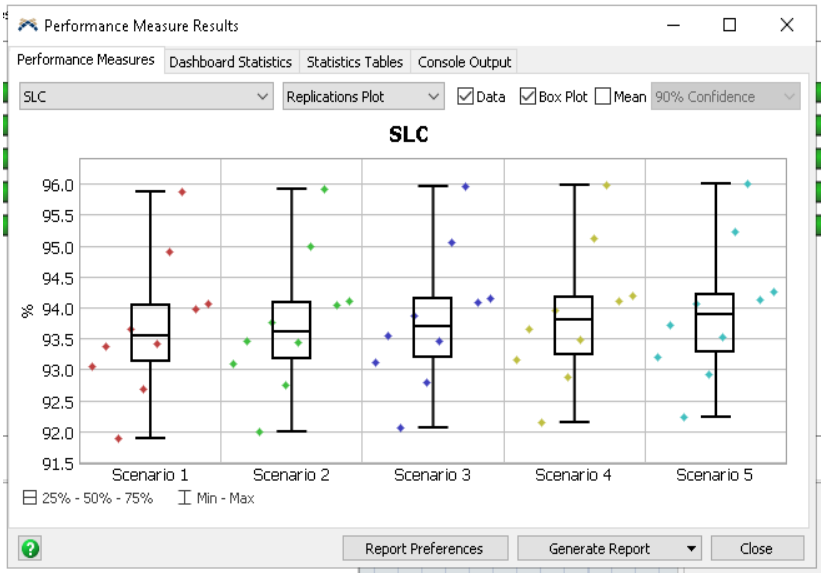

Fig. 3: SLC at Customer for $K_w = 5$, $K_p \in \{72,73,.74,75,76\}$, no delay in the decision process.

Notice that with these parameters 'values, one can achieve a SLC ≥ 95% yet with an extra delay $\tau = 2s$ but the counterpart here is that lost orders increase dramatically with greater τ. This limitation comes from the limit size of the decision step: in these simulations $\Omega_{max} = 5$. Finally, the third experiment concerns the same configuration as in the second experiment, i. e. with extra delay, decision step duration and decision step size limit, but with some different selected parameters: $K_w = 3$ $\eta = \tau = 5$, $\Omega_{max} = 5$ and $K_p \in \{20,30,40,50,60,70,80\}$.

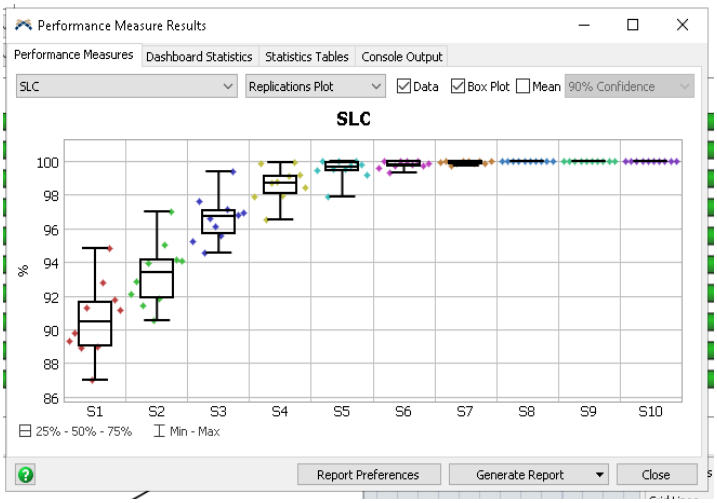

Fig. 4: SLC at Customer for $K_w = 5$, $K_p = 73$, delay $\eta \in \{72,73,.74,75,76\}$, five places in the decision step.

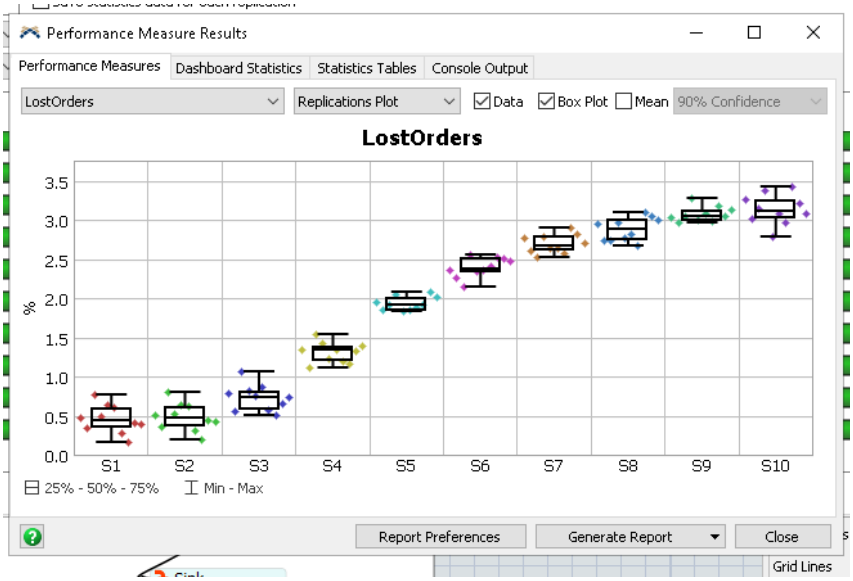

Fig. 5: Lost orders for $K_w = 5$, $K_p = 73$, delay $\eta \in \{72,73,.74,75,76\}$, five places in the decision step.

Notice that with these parameters' values, one can achieve a SLC $\geq 95\%$ with an extra delay $\tau = 5s$ with a reduced quantity of lost orders. This section does not provide any optimal value for the parameters K_p, K_w, τ, η and Ω_{max}. Its aim is to illustrate that the design problem of a Kanban controlled industrial plant is not straightforward when considering additional delays in the loop. We remark that the delay η and the size limit of the decision step Ω_{max} act together as a feedforward controller, see [18], [19] for details.

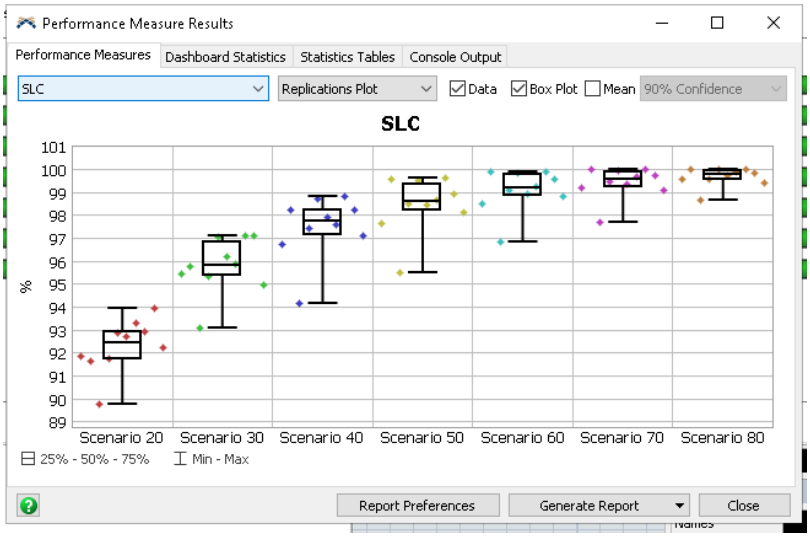

Fig. 6: SLC at Customer for $K_w = 5$, $K_p \in \{20, 30, 40, 50, 60, 70, 80\}$, $\eta = \tau = 5s$, five places at the decision step.

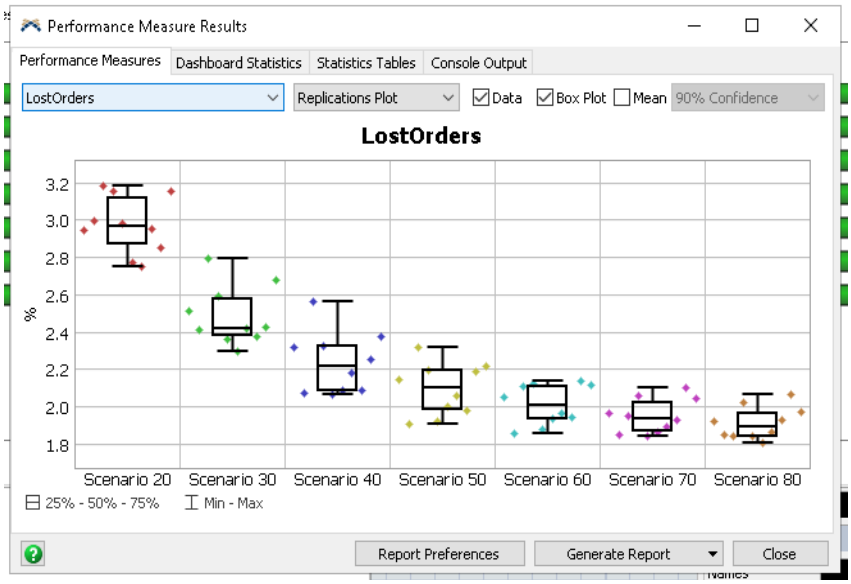

Fig. 7: Lost orders for $K_w = 5$, $K_p \in \{20, 30, 40, 50, 60, 70, 80\}$, $\eta = \tau = 5$s, five places at the decision step.

3. Case study: an industrial plant controlled with Heijunka-Kanban loop

We focus our analysis on a simplified model of a workshop dedicated to the production of automotive rubber hoses, as in [20]. The number of products family from the original workshop has been reduced to three types of products, the structure of the layout has been simplified also. We study the feasibility of a reduction of the number of Kanbans and its impact to the Service Level Customer (SLC).

3.1. Problem statement

The problem under study is similar to the problem discussed in section II. The objectives are the same as in section II for each type of products, say A, B and C. The Service Level Agreement (SLA) that links the workshop's factory and its customer 'Y' is composed of several rules. Among these SLA rules, this section is concerned with five of these important rules R_1, R_2, R_3, R_4 and R_5. R_1: Orders arrive at the workshop at any time from the opening and two hours before the closing of the workshop. R_2: Orders are due at least two hours after they have arrived and acknowledged at the workshop. R_3: The Service Level for acknowledged orders is above 95%. R_4: Unacknowledged orders are to be minimized and a penalty is applied. R_5: Acknowledged orders that are terminated exceeding due time are sent to the customer before closing time of the workshop, the delivery cost is charged to the workshop's factory.

3.2. Heijunka

Heijunka, also referred to as production smoothing or levelling the production schedule, has played an integral role in just-in-time and lean production since its inception, [6], [21], [22], [23], [24]. Heijunka is generally used in combination with other key Lean principles to stabilize value flow and this is a core concept that helps bring stability to a manufacturing process [25]. The objective of Heijunka is to avoid peaks and valleys in the production schedule [10]. In 1962, Taiichi Ohno, Japanese engineer at TOYOTA, developed a company management system dedicated to achieve goals as follows: Reduce waste; Maintain optimal product quality throughout; Avoid oversupply; Take into account the field experience; Enter into a continuous improvement process. It is a part of lean methodology of process improvement that helps organizations match unpredictable customer demand patterns and eliminate manufacturing waste by levelling the type and quantity of production output over a fixed period of time. Heijunka converts uneven Customer Pull into even and predictable manufacturing process. To reduce inventory, the factory must be able to produce each reference with the shortest possible recurrence. So, the more the recurrences are short, the smaller the lots, the less stocks and work-in-progress will have to be kept for cover the resulting queues.

minimized and a penalty is applied. R_5: Acknowledged orders that are terminated exceeding due time are sent to the customer before closing time of the workshop, the delivery cost is charged to the workshop's factory.

3.3. Experiments

The simulation experiments have been done with the FlexSim simulation tool. The workshop model is composed of three basic Kanban loops as presented in section II, a conveyer has been added before the single machine. Also, similarly to the model in section II there are a buffer of finished parts for each product type at the output of the machine place and a buffer at the expedition place in the workshop where the client picks its finished orders. The model does not include the milk-run between the workshop and the client's factory. The whole model is controlled with three pairs of Kanban loops. The production loops contain each K_{pA}, K_{pB}, K_{pC} Kanbans, and the withdraw loops contain K_{wA}, K_{wB}, K_{wC} Kanbans. The processing time on the single machine is either 33s; 26s; 26s, depending on the product type. There exists a setup time in order to prepare production for a subsequent different type. These setup times are respectively 120s; 240s; 180s for types A, B and C. The initial values of the Kanbans are K_{pA} = 800, K_{pB} = 800 and K_{pC} = 800, K_{wA} = 500, K_{wB} = 500, K_{wC} = 500 Kanbans. The objective is to reduce the number of Kanbans with an acceptable degradation of SLC.

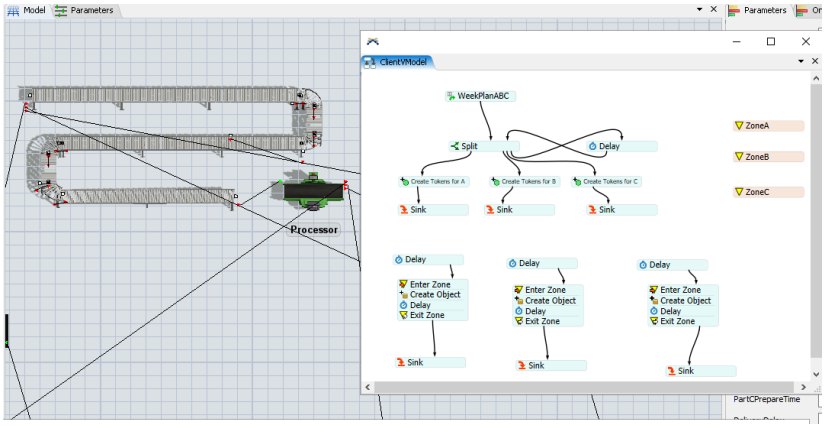

Fig. 8: The workshop model and the client process flow.

There exists a client model that emits the orders as follows: a set of prevision is randomly generated each week, say the length of a simulation run. Then a definitive set of orders that depends on the prevision set is again randomly generated, see figure 8. In this study, the workshop model receives only the definitive orders at the decision step for each product type.

3.4. Experimental results

Each configuration has been tested as follows: hundred simulation replications with random seeds, a simulation replication refers to 65 hours in the workshop, equivalent to five workdays with two turns of 6.5 effective hours. We first conserve all parameters as in the initial configuration except the number of production Kanbans K_{pA}, K_{pB}, K_{pC}, the delays at the decision step and the dedicated production time for each part type. We use the FlexSim optimizer tool to select suitable configurations, say K_{pA}, K_{pB}, K_{pC}, delays η_A, η_B, η_C and dedicated production time, in order to achieve an SLC $\geqslant$ Z%. The results are given in figures 9, 10, 11.

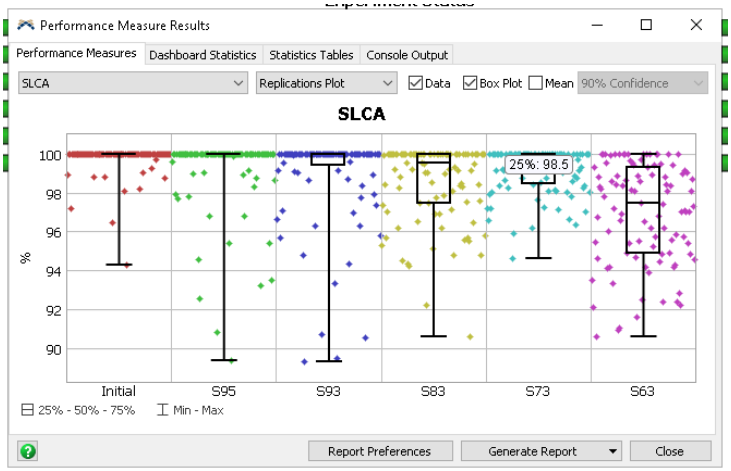

Fig. 9: SLC for product type A depending on the production Kanban numbers.

We can observe that the reduction of number of Kanbans and the selection of suitable dedicated production time for each type can lead to an SLC with respect to the SLA with the client. In figure 12, 13 and 14, we can observe the decreasing performance in terms of lost orders for product A type while the number of Kanbans decreases.

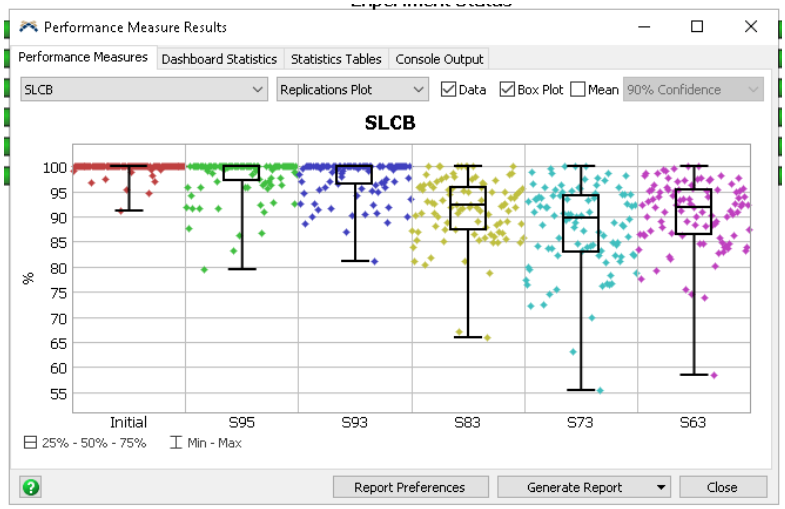

Fig. 10: SLC for product type B depending on the production Kanban numbers.

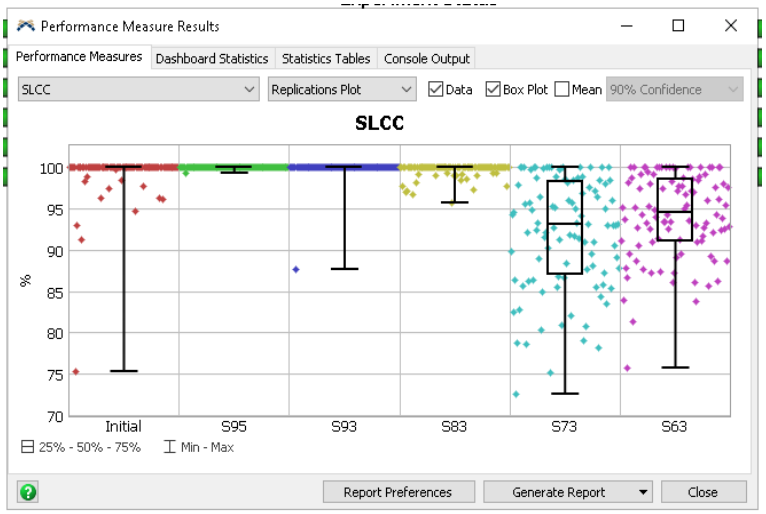

Fig. 11: SLC for product type C depending on the production Kanban numbers.

The second step of the simulation study is, conserving the production Kanbans K_{pA}, K_{pB}, K_{pC} selected in the previous simulations, we use again the FlexSim optimizer tool in order to determine suitable withdraw Kanbans K_{wA}, K_{wB}, K_{wC}, the delays at the decision step and the dedicated production time for each part type in order to achieve an SLC $\geq$ Z%. The results are given in figures 15, 16 and 17. We can observe that one can maintain a high level of service on the orders that have been acknowledged, even for configuration that include low number of Kanbans. For instance, the solution named 'S73' is obtained with the following set of parameters: K_{pA} = 268, K_{pB} = 238, K_{pC} = 230, K_{wA} = 123, K_{wB} = 15, K_{wC} = 44. The counterpart is that the number of lost orders increases for lower numbers of Kanbans. In the case under study, rule R4 of the SLA is designed to regulate this aspect. The last step of the simulation study is, conserving the previous selected parameters, we use again the FlexSim optimizer tool in order to determine suitable limitation size max, and the delays at the decision step and the dedicated production time for each part type in order to achieve an SLC $\geqslant$ Z%. The results again show that one can achieve a very high SLC for each part type and may have also a high dispersion on the lost orders. See for instance figure 18.

Fig. 12: Lost orders for product type A depending on the production Kanban numbers.

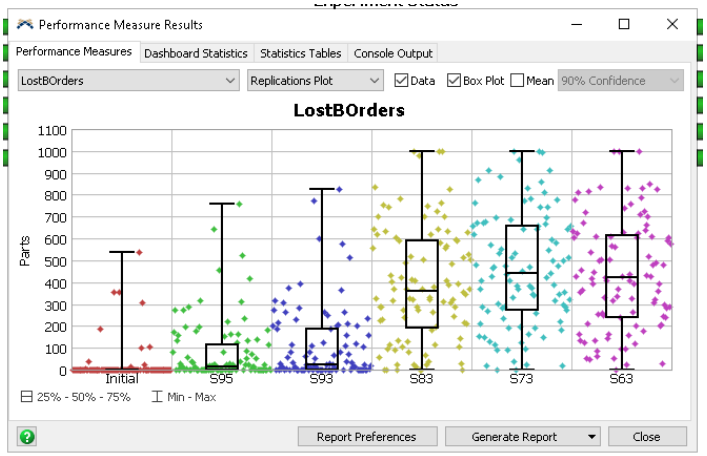

Fig. 13: Lost orders for product type B depending on the production Kanban numbers.

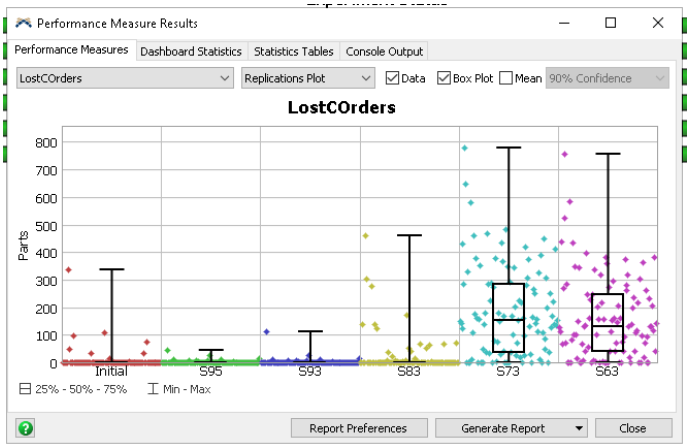

Fig. 14: Lost orders for product type C depending on the production Kanban numbers.

Fig. 15: SLC for product type A depending on the production Kanban numbers.

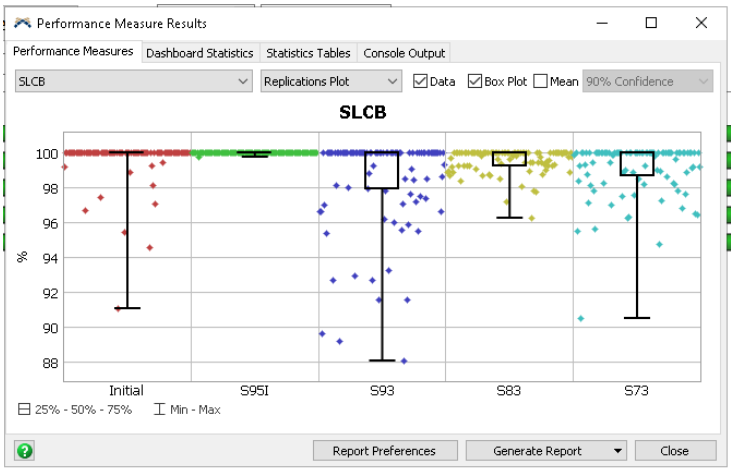

Fig. 16: SLC for product type B depending on the production Kanban numbers.

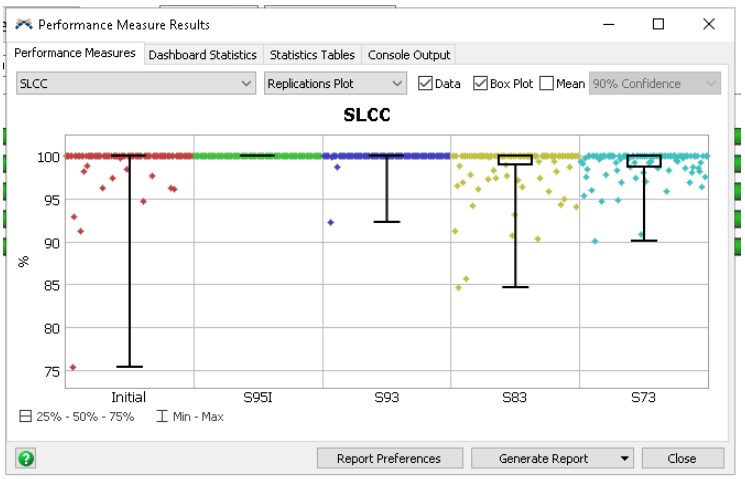

Fig. 17: SLC for product type C depending on the production Kanban numbers.

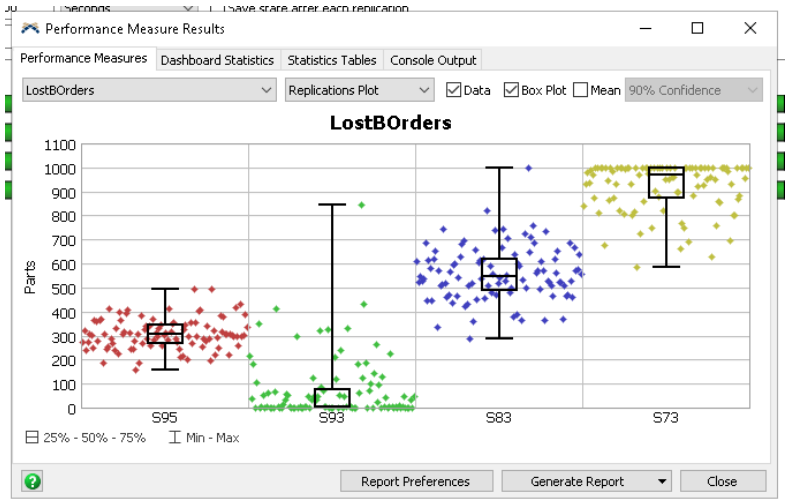

Fig. 18: Very high SLC for each part type versus dispersion on the lost orders (here type B).

4. Conclusion

We have presented a simulation study on the application of an Heijunka-Kanban control scheme of an HVLV production system. This study highlights the fact that a straightforward application of the Heijunka key principles on a constrained HVLV production system have to be further examined in order to achieve a suitable sizing of the number of Kanbans. This work also illustrates the limits of tight constraints that regulate the agreements (SLA) between provider and consumer in the automotive industry. At the same time, we claim that SLA can be refined on the basis of simulation case studies of the same kind of the one presented here in order to improve the agreement on feasible constraints and therefore that would improve the whole performance of the tandem provider consumer. A concurrent approach would be the design of an analytical model of the tandem provider consumer in the design of feasible constraints of the SLA. This could be developed following the Markov approach. This aspect needs further research.

5. References

[1] I. Nasri, G. Habchi, R. Boukezzoula, Use of (max, +) algebra for scheduling and optimization of HVLV systems subject to preventive maintenance, Simulation Modelling Practice and Theory, Vol.46, 149- 163, 2014.

[2] S. Wilson, Mix flexibility optimisation in hybrid make-to-stock / make-to-order environments in process industries, Cogent Engineering, 2018.

[3] M. Di Mascolo, Analysis of a synchronization station for the performance evaluation of a kanban system with a general arrival process of demands, European Journal of Operational Research, Vol. 89, 147-163, 1996.

[4] A. M. Deif, Dynamic analysis of a lean cell under uncertainty, International Journal of Production Research, 2012.

[5] T. Ohno, The Toyota Production System: beyond large scale production, Productivity Press, Portland, OR, Cambridge, Mass, 1988.

[6] Y. Monden, Toyota Production System Industrial Engineering and Management Press, Institute of Industrial Engineers, Norcross, GA., 1983.

[7] C. R. Lippolt and K. Furmans, in IFIP International Federation for Information Processing, processing times and consumption, which must be compensated by inventory, Vol. 257, Lean Business Systems and Beyond, Tomasz Koch, ed.; Boston: Springer, 11-19, 2008.

[8] L. Fonseca de Araujo and A. Alves de Queiroz, Production Leveling (Heijunka) Implementation in a Batch Production System: A Case Study, IFIP AICT 338, IFIP International Federation for Information Processing, 105-112, 2010.

[9] A. Smalley, Creating Level Pull, Lean Enterprise Institute, São Paulo, 2004.

[10] A. Hüttmeir, S. Treville, A. Ackere, L. Monnier, J. Prenninger, Trading off between heijunka and just-in-sequence, Int. J. Production Economics, Vol. 118, 501-507, 2009.

[11] J. Matzka, M. Di Mascolo, K. Furman, Buffer sizing of a Heijunka Kanban system, Intell Manuf 23: 49-60 DOI 10.1007, 2012.

[12] P. Korytkowski, F. Grimaud, A. Dolgui, Exponential Smoothing for Multi-Product Lot-Sizing with Heijunka and Varying Demand, Management and Production Engineering Review, Vol.5, 20-26, 2014.

[13] W. J. Davis, and S. J. Stubitz, Configuring a kanban system using a discrete optimization of multiple stochastic responses, International Journal of Production Research, 25/5, 721-740, 1987.

[14] P. Shahabudeen, R. Gopinath, K. Krishnaiah, Design of bi-criteria Kanban system using simulated annealing technique, Computers & Industrial Engineering, Vol. 41, Issue 4, 355-370, 2002.

[15] O. Kimura, H. Terada, Design and analysis of pull system, a method of multi stage production control International Journal of Production Research, Vol. 19, 241-253, 1981.

[16] H. Wang and H. Wang, Optimum number of kanbans between two adjacent workstations in a JIT system International Journal of Production Economics, Vol. 22, 179-188, 1991.

[17] F. Villalpando, C. Martinez, A study of delays in Heijunka Kanban system for perishable goods production, International Symposium on Precision Engineering and Sustainable Manufacturing PRESM, 2021.

[18] C. Cárdenas, J. Loiseau, and C. Martinez, Controlled invariance and dynamic feedback for systems over semirings, in Proceedings of the Conference on Control and its Applications, 1–8, 2015.

[19] C. Cárdenas, J. Cardillo, J. Loiseau, and C. Martinez, Control problem in max-plus linear model with temporal constraints, Revista Iberoamericana de Automatica e Informatica Industrial, Vol. 13, 438-449, 2016.

[20] C. Martinez, P. Castagna, Sizing of an industrial Plant using tight time constraints using complementary approaches: (max, +) theory and computer simulation, Simulation Practice and Theory, Elsevier, Vol. 11, 75-88, 2003.

[21] R. J. Schonberger, Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity. Free Press, New York, 1982.

[22] R.W Hall, Zero Inventories. Irwin, Homewood, IL, 1983.

[23] F.A. Abdulmalek, J. Rajgopal, Analyzing the benefits of lean manufacturing and value stream mapping via simulation: a process sector case study, International Journal of Production Economics Vol.107, 223-236, 2007.

[24] J.P. Womack, D.T. Jones, D. Roos, The Machine that Changed the World, Harper Collins Publishers, New York, 1990.

[25] K. S. Raguram and G Jayanthi, Implementation of heijunka for improving performance indicators: a process sector case study, Gorteria Journal Vol. 34, issue 7, 20-28, 2021