Available online at www.sciencedirect.com

SCIENCE DIRECT

Int. J. Production Economics 92 (2004) 181–196

international journal of production economics

www.elsevier.com/locate/dsw

Reactive JIT ordering system for changes in the mean and variance of demand

Katsuhiko Takahashi*, Katsumi Morikawa, Nobuto Nakamura

Department of Artificial Complex Systems Engineering, Graduate School of Engineering, Hiroshima University, 4-1, Kagamiyama 1 chome, Higashi-Hiroshima 739-8527, Japan

Received 2 June 2001; accepted 15 October 2003

Abstract

This paper proposes a reactive just-in-time (JIT) ordering system for multi-stage production systems with unstable changes in demand, not only in the mean but also in the variance. For a JIT ordering system, that is, the Kanban system, a reactive controller of buffer size is proposed. In the proposed system, unstable changes in the mean and variance of demand can be detected by exponentially weighted moving average charts for mean and variance, and the buffer size at each inventory point is controlled as a reaction to the detected unstable change. The performance of the proposed system is analyzed by simulation experiments.

Keywords: Reactive system; JIT production; Variance change

1. Introduction

In order to realize agile control in just-in-time (JIT) environments, this paper proposes a reactive ordering system for multi-stage production systems with unstable changes in demand, not only in the mean but also in the variance.

JIT production is a concept for producing a required volume of a required item at a required point in time. Research on JIT production from various viewpoints is carried out all over the world, and, in recent years, the application of JIT concept to supply chain management is attempted (Zimmer, 2002; Kim and Ha, 2003). Some researchers, e.g. Price et al. (1994), provide reviews of the literature, and recently, Takahashi (2002) and Machuca (2002) edit special issues on JIT systems. In JIT production, the order release for each process is determined on the basis of the actual demand, that is, without demand forecasts. As JIT ordering systems, the Kanban system (Kimura and Terada, 1981) and the concurrent ordering system (Izumi and Takahashi, 1993; Takahashi et al., 1996a) have been proposed. In the Kanban system, not only demand arrival, but also parts supply and process availability are used for determining the order release. On the other hand, in the concurrent ordering system, only demand arrival is used. The relative performance of the systems has been analyzed and compared,

*Corresponding author. Tel.: +81-824-24-7705; fax: +81-824-22-7024.

E-mail address: takahasi@pel.sys.hiroshima-u.ac.jp (K. Takahashi).

0925-5273/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2003.10.014

and it has been shown that the differences do effect the performance (Takahashi et al., 1996b; Takahashi and Nakamura, 1998). In this paper, we focus on the Kanban system in JIT ordering systems.

Although they show different performance, all of the JIT ordering systems have suffered from similar influence of demand. They can avoid the influence of forecasting error because they have never utilized demand forecasts. However, the influence of changes in demand cannot be avoided even in the JIT ordering systems. In most of the previous literature on JIT ordering systems, stable changes in demand have been assumed, and the influence of unstable changes in demand has never been considered. Recently, product life cycles have become shorter and shorter due to the diversification of customer needs, and the duration of stationary demand has also shortened. Therefore, not only stable changes in demand, but also unstable changes, should be considered in designing an ordering system. Where stable changes mean stochastic changes that exclude changes of the structural parameters, such as mean or variance, unstable changes mean the changes of the structural parameters.

Taking this into account, Takahashi and Nakamura (1999) have proposed reactive JIT ordering systems for unstable changes in demand. In reactive JIT ordering systems, unstable changes in demand can be detected by exponentially weighted moving average (EWMA) charts, and the buffer size at each inventory point is adjusted according to the detected unstable changes. However, in the reactive JIT ordering systems, only unstable changes in the mean of demand were considered, and the constant and known variance of demand was assumed. It is a restrictive assumption, and unstable changes in demand, not only in the mean but also in the variance, should be considered as a more realistic situation. Suppose that not only the mean, but also the variance of demand fluctuates dynamically. Then, changes in the variance as well as changes in the mean must be detected, and detecting changes in the mean is affected by changes in the variance. Additionally, the response to the changes must be determined.

In this paper, a reactive JIT ordering system for multi-stage production systems with unstable changes in demand, not only in the mean but also in the variance, is proposed. For a JIT ordering system, that is, the Kanban system, a reactive controller of buffer size is proposed. In the proposed system, unstable changes in the mean and variance of demand can be detected by EWMA charts for mean and variance, and the buffer size at each inventory point is controlled in response to the detected unstable change. If unstable changes in demand, not only in the mean but also in the variance, can be detected, it becomes unnecessary to know the parameters of demand distribution, and the ability of the reactive JIT ordering system will be improved.

This paper is organized as follows: In the next section, the literature on the JIT ordering systems that have an ability to respond to changes is reviewed. Then, assumptions for the production system and the production ordering system are defined, and an order-release mechanism for the Kanban system is formulated. In order to obtain the fundamental information for developing a control rule, the performance of the JIT ordering system is analyzed under the stable-demand conditions by simulation experiments in Section 3. Based on the results, a reactive JIT ordering system is proposed in Section 4, and the performance of the proposed system is investigated in Section 5. Finally, the findings obtained in this paper are summarized in Section 6.

2. Literature review

In this section, the literature on the JIT ordering systems which have an ability to respond to changes is reviewed.

Monden (1981), in his article for introducing the Kanban system, showed various types of Kanbans. Among them, an emergency Kanban is introduced as a temporarily issued Kanban for defective work, extra insertions or for a spurt in demand. Also, he claimed that the Kanban system should be used only in adapting to small fluctuations in demand and that the Kanban system has no adaptability for sudden and large variations in demand. For the latter case, the number of Kanbans must be increased or reduced in response to the variations. Based on his claim, Rees et al. (1987) proposed an algorithm for adjusting the number of Kanbans in response to the variations in demand and lead time. The Kanban system which has the ability to adjust the number of Kanbans in response to demand data was called the flexible Kanban system by Gupta et al. (1999). For the flexible Kanban system, Gupta and Al-Turki (1997) proposed an algorithm for adjusting the number of Kanbans in response to the demand data. Moore and Gupta (1999) developed the model of stochastic coloured Petri net, analyzed the structural properties, and compared the performance with that of traditional Kanban system via simulation experiments. Also, Gupta and Al-Turki (1998a,b) proposed the flexible Kanban system for responding to uncertain interruptions caused by preventive maintenance (Gupta and Al-Turki, 1998a) or breakdown of material handling system (Gupta and Al-Turki, 1998b) in the successive stage. In the literature, information about changes in demand or lead time, maintenance and breakdown in material handling system is collected and utilized to adjust the number of Kanbans. In determining the number of Kanbans, various factors should be considered under the complicated situations. Wray et al. (1997) and Markham et al. (1998) proposed a system to determine the number of Kanbans at the next period on the basis of demand variability, processing time variability, and the number of Kanbans at the current period as well as the demand forecast. In proposing the system, they applied neural network (Wray et al., 1997) and rule induction (Markham et al., 1998). On the other hand, Tardif and Maaseidvaag (2001) proposed an adaptive Kanban system that monitors the actual inventory level and adjusts the number of Kanbans by releasing or capturing the extra Kanbans.

However, in the literature, unstable changes such as mean and/or variance changes, are not distinguished with stable changes such as random fluctuations, and the proposed systems in the previous literature are not only for responding to unstable changes. Under the conditions of demand with stable and unstable changes, the system may be sensitive to changes if the system responds not only to unstable changes but also to stable changes. Therefore, the systems may have an insufficient ability to deal with demand with stable and unstable changes. For the problem, Takahashi and Nakamura (1999, 2000a, b, 2002) proposed a reactive Just-in-Time ordering systems on the basis not only of the Kanban system but also of an alternative, the concurrent ordering system. In the reactive JIT ordering systems, by utilizing a process control chart, the EWMA chart, only unstable changes are detected from demand with stable and unstable changes, and the number of Kanbans and buffer size are adjusted in response to the detected unstable changes (Takahashi and Nakamura, 1999), or the JIT ordering system is switched from one system to the other (Takahashi and Nakamura, 2000b). Furthermore, the reactive JIT ordering systems were applied to a logistics system (Takahashi and Nakamura, 2000a), and a decentralized control for the reactive Kanban system was proposed (Takahashi and Nakamura, 2002).

The literature showed the proposed system is effective in detecting unstable changes and reacting to the detected changes, however, in the literature, only unstable changes in the mean of demand are analyzed while the variance is assumed to be constant. As unstable changes, changes not only in the mean but also in the variance can be pointed out. Wray et al. (1997) and Markham et al. (1998) considered demand variability in determining the number of Kanbans at the next period, however, they assumed two levels of the demand variability, high and low. The performance may be affected by how to evaluate demand variability, and it seems difficult to respond to a slight change in variance. Therefore, in this paper, a reactive JIT ordering system for multi-stage production systems with unstable changes in demand, not only in the mean but also in the variance, is proposed based on the system proposed by Takahashi and Nakamura (1999).

As the other literature on reactive JIT ordering systems, the works of Chang and Yih (1998) and Hopp and Roof (1998) can be pointed out. The Kanban system itself is an ordering system for make-to-stock environments, however, Chang and Yih (1994) proposed a Kanban system for job-shop systems, that is for make-to-order environments, and they called it generic Kanban system. For the generic Kanban system, Chang and Yih (1998) constructed a fuzzy rule base and proposed a system for dynamically controlling the number of Kanbans in response to changes not only in demand but also in waiting jobs, machine utilization, and processing time. Additionally, for the CONWIP system, an alternative to the Kanban system, Hopp and Roof (1998) proposed a system for monitoring the work-in-process (WIP) levels and controlling the throughput by adjusting the total WIP level.

3. Modeling JIT ordering systems

In this section, assumptions for the multi-stage production system considered are defined, and a mathematical model of the Kanban system is formulated as an ordering system for the multi-stage production system.

3.1. Assumptions

The production system considered in this paper is assumed as follows:

1. The production system produces a standard product that can be made to stock.
2. The product demand includes stable and unstable changes, and the ith inter-arrival time of demand (x_i) is distributed stochastically with unstable changes in the mean (μ_i) and variance (σ_i^2).
3. The product is produced through a serial production system with N stages. Each stage has a production process, called 1st, 2nd, …, or Nth stage accordingly, as the process proceeds.
4. The production time at each production stage is distributed stochastically.
5. The transportation process between the $(n-1)$th and the nth production stages is called the nth transportation stage.
6. Each production stage has two inventory points, one before and one after the production stage.
7. The buffer inventories, $S_B^{(n)}$ and $S_A^{(n)}$, are stocked at the inventory points before and after the nth production stage, respectively. The buffer size is not fixed but controlled dynamically in response to unstable changes in demand.
8. A backorder of product demand can be allowed.

As unstable changes in the variance of demand are considered in this paper, the demand with unstable changes, not only in the mean but also in the variance, is assumed as shown in the second assumption, which differs from that in Takahashi and Nakamura (1999).

3.2. The Kanban system

There are two kinds of Kanban systems: the well-known and the pure Kanban systems. In the Kanban system, an order is released on the basis of the following three kinds of information: a demand arrival from the succeeding stage, the parts supply from the preceding stage, and the process availability at the current stage. In the model of the Kanban system developed by most of the researchers, such as Mitra and Mitrani (1990, 1991) and Tayur (1993), the orders are released not on the basis of all the three kinds of information, but on the basis of only the first two kinds of information. On the other hand, in the model developed by Spearman (1992), the orders are released on the basis of all the three kinds of information. Takahashi and Nakamura (1998) called the former the well-known Kanban system and the latter the pure Kanban system. Then, for the two kinds of the Kanban system, Takahashi and Nakamura (1998) analyzed the effect of the difference and showed that the pure Kanban system is superior to the well-known Kanban system. Therefore, in this paper, the pure Kanban system is dealt with as the Kanban system.

Let $OP_i^{(n)}$ and $OT_i^{(n)}$ be the ith order release time for the nth production stage or the nth transportation stage, respectively. Then, they can be formulated as follows:

$$OP_i^{(n)} = \max\,\{OT_i^{(n+1)}, P_{i-S_A^{(n)}}^{(n)}, T_{i-1}^{(n+1)}\} \quad (n = 1, 2, \ldots, N-1), \tag{1}$$

$$OP_i^{(N)} = \max\{D_i, P^{(N)}_{i-S_A^{(N)}}\}, \tag{2}$$

$$OT_i^{(n)} = \max\{OP_i^{(n)}, T^{(n)}_{i-S_B^{(n)}}, P^{(n)}_{i-1}\} \quad (n = 1, 2, \ldots, N). \tag{3}$$

Here, the first, the second and the third terms in the maximum function in each equation correspond to the first, the second and the third conditions concerning order release described above, respectively. As shown in the assumptions, the buffer inventories $S_X^{(n)}$ ($X = B, A$) are stocked at the inventory points before and after the nth production stage, and are assumed to have been prepared and stocked from the beginning. Then, it can be assumed that $P^{(n)}_{i-S_A^{(n)}} = T^{(n)}_{i-S_B^{(n)}} = 0$ if $i \leqslant S_X^{(n)}$.

Also, in the above equations, D_i is the time when the ith demand for product arrives, $P_i^{(n)}$ is the completion time of the ith production at the nth production stage, and $T_i^{(n)}$ the completion time of the ith transportation at the nth transportation stage. Then, $P_i^{(n)}$ and $T_i^{(n)}$ are formulated as follows:

$$P_i^{(n)} = \max\{OP_i^{(n)}, T^{(n)}_{i-S_B^{(n)}}, P^{(n)}_{i-1}\} + p_i^{(n)} \quad (n = 1, 2, \ldots, N), \tag{4}$$

$$T_i^{(1)} = \max\{OT_i^{(1)}, T^{(1)}_{i-1}\} + t_i^{(1)}, \tag{5}$$

$$T_i^{(n)} = \max\{OT_i^{(n)}, P^{(n-1)}_{i-S_A^{(n-1)}}, T^{(n)}_{i-1}\} + t_i^{(n)} \quad (n = 2, 3, \ldots, N). \tag{6}$$

Here, $p_i^{(n)}$ is the ith production time at the nth production stage, and $t_i^{(n)}$ the ith transportation time at the nth transportation stage.

3.3. Performance measures

As performance measures, the mean waiting time of demand (*wt*) and the total of the mean WIP inventories (*twip*) are evaluated in this paper. The former measure, *wt*, is for evaluating the service to the customers. Customers in a JIT production system want their needs to be satisfied just in time, that is, without waiting, and this should be the goal of such a production system. Therefore, decreasing the *wt* means a better ability to serve the customers. On the other hand, the latter measure, *twip*, is for evaluating the efficiency of production system in improving the former measure. When we consider only improving customers' satisfaction, a naive answer is to possess sufficient WIP inventories. However, this is inefficient and impractical. Decreasing the former measure with less increase in the latter measure, or decreasing the latter measure while maintaining the former measure at less than the required level, is required. Therefore, in this paper, the two measures are dealt with, and the trade-off relationship between the two measures is investigated.

4. Analyzing the stable performances

In order to develop a control rule in response to unstable changes in demand, the performance of JIT ordering systems is investigated under various demand conditions in this section. Also, the effects of buffer size upon the performance measures are investigated.

4.1. Experimental conditions

The simulation experiments in this section are performed under the following conditions:

1. The inter-arrival time of demand: a normally distributed variable. The mean and variance are fixed at μ_x and σ_x^2, respectively. The means investigated in the experiments are $\mu_x = 0.90, 0.92, \ldots, 1.18, 1.20$. Also, the variances are $\sigma_x^2 = 0.05^2, 0.10^2, \ldots, 0.25^2, 0.30^2$.
2. The number of stages: $N = 5$.
3. The production time at each production stage: an independent and identical gamma-distributed variable with a mean of 0.8 and a variance of 0.8.
4. The transportation time at each transportation stage: $t_i^{(n)} = 0$.
5. The buffer size at each inventory point: identical. The sizes investigated in the experiments are $s = 1, 2, \ldots, 20$.
6. The simulation run-length: 100,000 time units, excluding the warm-up run of 5,000 time units.

7. The number of replications for each simulation run: 1.

In the first condition, the inter-arrival time of demand is assumed as a normally distributed variable, and the experiments are designed for investigating the influence of the parameters of the distribution, mean and variance. Also, the production time is assumed as a gamma-distributed variable, and the parameters are fixed in the third condition. The distribution of production time affects the performance of JIT ordering system, this influence is investigated by Takahashi and Nakamura (1998). According to their findings, the qualitative influences do not depend on the distribution, although the quantitative ones do. Therefore, in this paper, the distribution of production time is fixed in the experiments. In the fourth condition, the transportation time is assumed as negligible. In more realistic and general production systems, the transportation time would be positive, however, this paper concentrates on the production time and neglects the transportation time. This represents transportation in a factory or a production base that consists of the production processes located near to each other. In spite of this assumption, the buffer for the transportation stage is necessary for absorbing the fluctuations at the preceding and succeeding production stages.

4.2. Results and implications

In the results of the simulation experiments, Fig. 1 shows the influence of the mean inter-arrival time of demand upon the two performance measures at the variance of the inter-arrival time of demand $\sigma_x^2 = 0.05^2$. Also, Fig. 2 shows the influence of the variance of the inter-arrival time of demand upon the two performance measures at the mean inter-arrival time of demand $\mu_x = 1.0$. In the figures, the results at the same buffer size are connected by a solid line, and the influence of the mean and variance of demand at each buffer size is shown.

In Fig. 1, it can be seen that, as the mean inter-arrival time of demand decreases, *wt* at the same

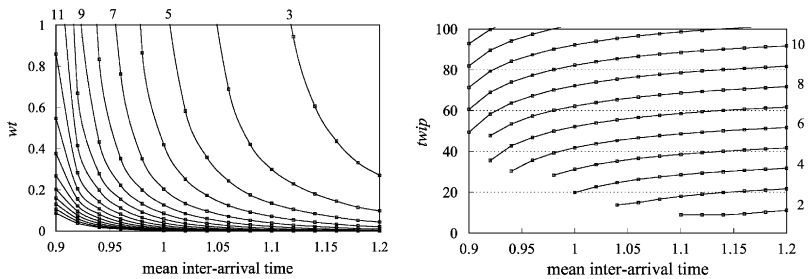

Fig. 1. The influence of the mean inter-arrival time of demand (μ_x) upon the two performance measures at each buffer size and the variance of the inter-arrival time of demand $\sigma_x^2 = 0.05^2$.

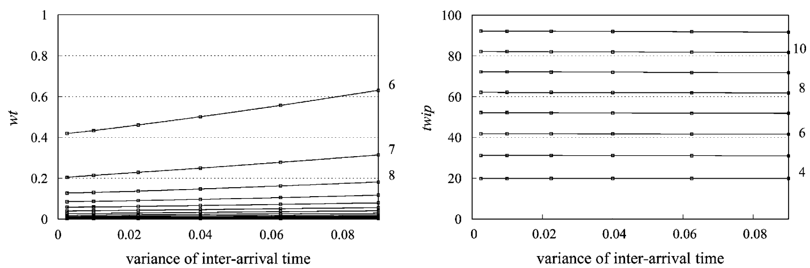

Fig. 2. The influence of the variance of the inter-arrival time of demand (σ_x^2) upon the two performance measures at each buffer size and the mean inter-arrival time of demand $\mu_x = 1.0$.

buffer size increases, and *twip* at the same buffer size decreases. Similarly, with Fig. 2, it can be seen that as the variance of the inter-arrival time of demand increases, *wt* at the same buffer size increases, whereas *twip* at the same buffer size decreases just a little. Therefore, if the mean or variance of the inter-arrival time of demand changes and the buffer size is fixed at a certain size, *wt* suffers from a significant influence. In order to avoid this influence, it is necessary to consider the adjustment of buffer size. With Figs. 1 and 2, it can be understood that, by increasing the buffer size, we can decrease *wt* whereas *twip* increases. Therefore, in order to minimize *twip* while maintaining *wt* at less than a certain level, the buffer size should be adjusted in response to the changed mean or variance of the inter-arrival time of demand. For example, in order to minimize *twip* while maintaining $wt \leqslant 0.6$ at $\sigma_x^2 = 0.05^2$, the buffer size should be altered to $12, 11, \ldots, 4$, and 3 in response to the mean inter-arrival time of demand of $[0.90, 0.91)$, $[0.91, 0.92), \ldots, [1.07, 1.15)$, and $[1.15, 1.20]$, respectively (see Fig. 1). Also, in order to minimize *twip* while maintaining $wt \leqslant 0.6$ at $\mu_x = 1.0$, the buffer size should be altered to 6 and 7 in response to the variance of the inter-arrival time of demand of $[0.025, 0.079]$, $(0.079, 0.090]$, respectively (see Fig. 2).

5. Proposing a reactive JIT ordering system

Based on the stable performance analyzed above, a reactive controller for detecting unstable changes and controlling the buffer size is proposed as a reactive JIT ordering system in this section. Also, in this section, the performance of the proposed system is estimated on the basis of the simulation results under the stable-demand conditions.

5.1. Detecting unstable changes

The product demand data x_i includes not only stable changes but also unstable changes. Moreover, it includes not only changes in the mean but also those in the variance. In order to detect unstable changes with the original time series data for the product demand, the serial data are grouped into serial batches of size m, and the mean $\bar{x}_t$ and variance s_t^2 of the tth batch are utilized as new time series data.

At first, in order to detect unstable changes in the mean inter-arrival time of demand, the EWMA for mean, H_t, is calculated from the new time series data $\bar{x}_t$ as follows:

$$H_t = \alpha \bar{x}_t + (1 - \alpha) H_{t-1}, \tag{7}$$

where α is a smoothing constant. Suppose that x_t is an *i.i.d.* variable with mean μ_x and variance σ_x^2, and $\bar{x}_t$ becomes a variable with mean μ_x and variance σ_x^2/m. Then, H_t becomes a variable with mean μ_x and variance $(\alpha/(2-\alpha))(\sigma^2/m)$, and the EWMA chart (Lucas and Saccucci, 1990) is applied to H_t for detecting unstable changes in the mean as with Takahashi and Nakamura (1999). In the EWMA chart, H_t is compared with the upper and lower control limits $\mathrm{UCL}(\sigma_x^2, s)$ and $\mathrm{LCL}(\sigma_x^2, s)$, and the control limits can be calculated as follows:

$$\mathrm{UCL}(\sigma_x^2, s) = \bar{\mu}_x(\sigma_x^2, s) + \delta \sqrt{\frac{\alpha}{2-\alpha} \frac{\sigma_x^2}{m}}, \tag{8}$$

$$\mathrm{LCL}(\sigma_x^2, s) = \underline{\mu}_x(\sigma_x^2, s) - \delta \sqrt{\frac{\alpha}{2-\alpha} \frac{\sigma_x^2}{m}}, \tag{9}$$

where δ is a multiplier, and $[\underline{\mu}_x(\sigma_x^2, s), \bar{\mu}_x(\sigma_x^2, s))$ is the interval of the mean inter-arrival time of demand under which the buffer size s is appropriate. In contrast to Takahashi and Nakamura (1999), the variance σ_x^2 is unknown, and the estimate is obtained from the EWMA for variance, Z_t, described below.

On the other hand, in order to detect unstable changes in the variance of the inter-arrival time of demand, the new time series data s_t^2 are used for calculating the EWMA for variance, Z_t, as follows (see Gan, 1995):

$$Z_t = \lambda W_t + (1 - \lambda) Z_{t-1}, \tag{10}$$

where λ is a smoothing constant, and $W_t = \ln(s_t^2)$. Suppose that x_t is a *n.i.d.* variable (normal distribution is assumed in identical distributions) with mean μ_x and variance σ_x^2, and s_t^2 becomes a

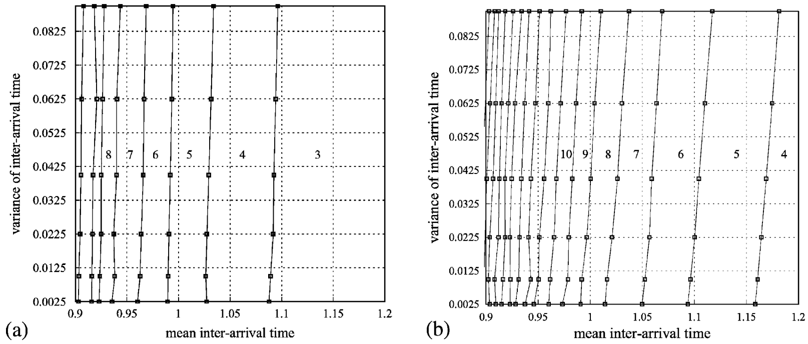

Fig. 3. The buffer size for minimizing the total of mean WIP inventories (*twip*) while satisfying the required level for the mean waiting time of demand (wt) $wt_0 = 2.5$ (a) and 0.15 (b).

gamma-distributed variable with shape parameter $(m-1)/2$ and scale parameter $2\sigma_x^2/(m-1)$. Then, W_t becomes a logarithmic gamma-distributed variable, and the mean and variance can be calculated as follows (for details, see Crowder and Hamilton, 1992):

$$E[W_t] = \ln\left(\frac{2\sigma_x^2}{m-1}\right) + \psi\left(\frac{m-1}{2}\right), \tag{11}$$

$$\mathrm{Var}\,[W_t] = \psi'\left(\frac{m-1}{2}\right), \tag{12}$$

where $\psi(y)$ and $\psi'(y)$ are the digamma and trigamma functions for y, respectively. Therefore, the upper and lower control limits for Z_t can be obtained as follows:

$$\mathrm{UCL} = E[W_t] + \delta\sqrt{\frac{\lambda}{2-\lambda}\mathrm{Var}[W_t]}, \tag{13}$$

$$\mathrm{LCL} = E[W_t] - \delta\sqrt{\frac{\lambda}{2-\lambda}\mathrm{Var}[W_t]}, \tag{14}$$

where δ is a multiplier. These control limits are designed for detecting unstable changes from $E[W_t]$. Based on these limits, the control limits for detecting the unstable changes for which the current buffer size s should be adjusted can be formulated as follows:

$$\mathrm{UCL}(\mu_x, s) = \overline{W}(\mu_x, s) + \delta\sqrt{\frac{\lambda}{2-\lambda}\mathrm{Var}[W_t]}, \tag{15}$$

$$\mathrm{LCL}(\mu_x, s) = \underline{W}(\mu_x, s) - \delta\sqrt{\frac{\lambda}{2-\lambda}\mathrm{Var}[W_t]}, \tag{16}$$

where $(\underline{W}(\mu_x, s), \overline{W}(\mu_x, s)]$ is the interval of W_t under which the buffer size s is appropriate. The interval is obtained from the interval of the variance of the inter-arrival time of demand under which the buffer size s is appropriate, $(\underline{\sigma}_x^2(\mu_x, s), \bar{\sigma}_x^2(\mu_x, s)]$. In obtaining the interval, the mean inter-arrival time μ_x is necessary, and the estimate is obtained from H_t.

For the EWMA charts, the intervals of the mean and variance of the inter-arrival time of demand for which the buffer size s is appropriate, $[\underline{\mu}_x(\sigma_x^2, s), \bar{\mu}_x(\sigma_x^2, s))$ and $(\underline{\sigma}_x^2(\mu_x, s), \bar{\sigma}_x^2(\mu_x, s)]$, can be obtained based on the simulation results under the stable demand conditions. Fig. 3 shows the buffer size for minimizing *twip* while satisfying $wt_0 = 2.5$ and 0.15, and the regions.

With Fig. 3, for example, $[\underline{\mu}_x(0.05, 6), \bar{\mu}_x(0.05, 6))$ and $(\underline{\sigma}_x^2(0.99, 6), \bar{\sigma}_x^2(0.99, 6)]$ can be obtained as shown in Fig. 4. Then, the interval of the variance is utilized for calculating the interval of Z_t, $(\underline{W}(\mu_x, s), \underline{W}(\mu_x, s)]$.

5.2. Controlling the buffer size

If an unstable change in the mean or variance of the inter-arrival time of demand is detected, the buffer size should be adjusted in response to the change. We propose the following control rule for

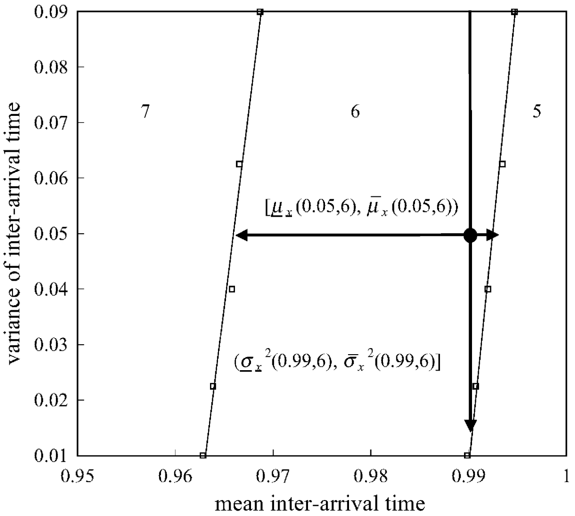

Fig. 4. An example of the intervals of the mean and variance of the inter-arrival time of demand for which the buffer size s is appropriate, $[\underline{\mu}_x(\sigma_x^2, s), \bar{\mu}_x(\sigma_x^2, s))$ and $(\underline{\sigma}_x^2(\mu_x, s), \bar{\sigma}_x^2(\mu_x, s)]$.

the buffer size as similar to the rule in Takahashi and Nakamura (1999).

$$S_A^{(n)} = S_B^{(n)} = \begin{cases} s + \theta^+ & (H_t \leqslant \mathrm{LCL}(\sigma_x^2, s) \text{ or } \\ & \mathrm{UCL}\ (\mu_x, s) \leqslant Z_t), \\ s & (\mathrm{LCL}(\sigma_x^2, s) < H_t \\ & < \mathrm{UCL}(\sigma_x^2, s) \text{ and } \\ & \mathrm{LCL}(\mu_x, s) < Z_t \\ & < \mathrm{UCL}(\mu_x, s)), \\ s - \theta^- & (\mathrm{UCL}(\sigma_x^2, s) \leqslant H_t \text{ or } \\ & Z_t \leqslant \mathrm{LCL}(\mu_x, s)). \end{cases} \tag{17}$$

Here, θ^+ or θ^- is the increase or decrease in the buffer size. If an unstable change in the mean inter-arrival time is detected with H_t, the increase or decrease is obtained by the interval of the mean inter-arrival time, $[\underline{\mu}_x(\sigma_x^2, s), \bar{\mu}_x(\sigma_x^2, s))$, in which H_t is included. Also, if an unstable change in the variance of the inter-arrival time is detected with Z_t, it is obtained by the interval of the variance of the inter-arrival time, $[\underline{\sigma}_x^2(\mu_x, s), \bar{\sigma}_x^2(\mu_x, s))$, in which e^{Z_t} is included.

In the control rule, the buffer sizes at all the inventory points are assumed to be equal to each other, and they are controlled simultaneously. Of course, the buffer size at each inventory point can be controlled individually, and this may lead to more sophisticated control. For individual control, however, allocating the buffer to each inventory point and constructing the control rules for the buffer size at each inventory point would be problematic.

5.3. Estimating the performance

Before analyzing the performance under the unstable-demand conditions, here we tried to estimate them based on the simulation results under the stable-demand conditions shown in Section 3.

Suppose that the buffer size is controlled as shown in Fig. 3 and there is no delay in detecting unstable changes and controlling the buffer size for the changes. Then, it can be estimated that *wt* will be maintained at less than the required level. At the same time, *twip* is estimated as shown in Fig. 5 at the required level for *wt*, $wt_0 = 2.5$ (Fig. 5(a)) and 0.15 (Fig. 5(b)). In the figure, the estimate at the two levels of the variance of the inter-arrival time of demand, 0.0025 (thin lines) and 0.09 (thick lines) are shown together.

In Fig. 5, it can be seen that *twip* can be reduced while satisfying the required level for *wt* by adjusting the buffer size in response to unstable changes in the mean and variance of the inter-arrival time of demand. By calculating the estimate for the mean inter-arrival time of demand (0.90, 1.20) and the variance (0.05^2, 0.30^2), the estimate of *twip* is obtained as follows: At $wt_0 = 2.5$, the estimate with assuming the minimum variance of the inter-arrival time of demand (thin line in Fig. 5(a)) is 30.56, and it is about 5% less than that assuming the maximum variance (thick line in Fig. 5(a)) of 32.56. These results show the maximum effects of controlling the buffer size in response to unstable changes in the variance of the inter-arrival time. Then, the estimate of *twip* for the interval of the mean inter-arrival time (0.90,1.20) and that of the variance ($0.05^2, 0.30^2$) can be calculated as 31.71 at $wt_0 = 2.5$, and as 64.60 at $wt_0 = 0.15$.

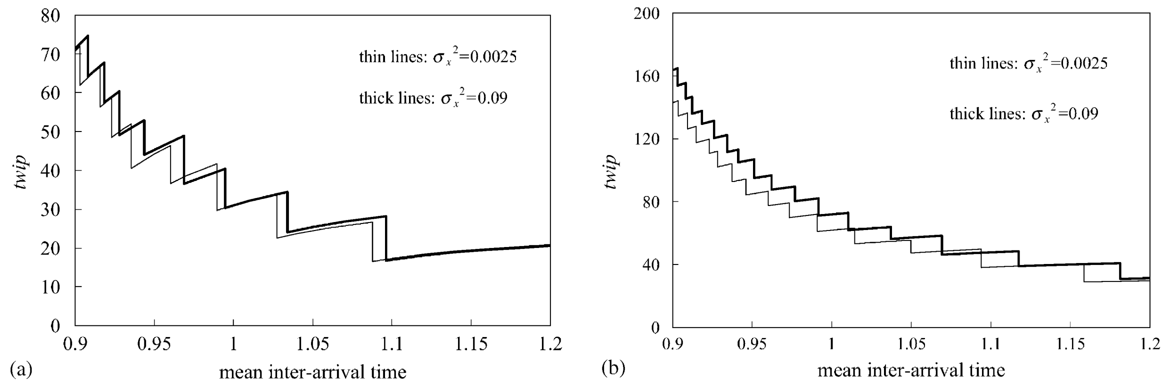

Fig. 5. The estimate of the total of mean WIP inventories (*twip*) at the two required levels for the mean waiting time of demand (wt_0), 2.5 (a) and 0.15 (b).

6. Analyzing the performance of the proposed system

The performance of the proposed system is analyzed by the simulation experiments under unstable-demand conditions in this section.

6.1. Analysis of variance

First, the analysis of variance (ANOVA) is performed for the parameters of the proposed system, and whether the influence is significant or not is analyzed. In the ANOVA, the influence of the following three kinds of parameters of the proposed system are analyzed, and the levels for each parameter are specified as follows:

1. The batch size of the EWMA charts: $m = 3$, 4, 5.
2. The number of the smoothing constants of the EWMA charts: 3 ($\alpha = \lambda = 0.1$, 0.3, 0.5), 4 (0.1, 0.233, 0.367, 0.5), 5 (0.1, 0.2, 0.3, 0.4, 0.5).
3. The multiplier of the EWMA charts: $\delta = 1.0$, 2.0, 3.0.

In the second condition, the number of the smoothing constants of the EWMA charts is specified together with the levels. The levels vary between 0.1 and 0.5, and they are determined by the preliminary analysis (see Takahashi and Nakamura, 1999).

Also, in the ANOVA, the conditions of simulation are specified as follows:

1. The inter-arrival time of demand: a normally distributed variable. The mean and variance of the ith inter-arrival time is assumed to change uniformly with $(0.90, 1.20)$ and $(0.05^2, 0.30^2)$, respectively, at every τ arrivals, where $\tau = 20$, 40, 60, 80, and 100.
2. The required level for wt: $wt_0 = 2.50$ (loose) and 0.15 (severe).
3. The simulation run-length: 1,000,000 time units, excluding the warm-up run of 5,000 time units.
4. The number of replications for each simulation run: 1.

In the first condition, uniformly distributed unstable changes in the mean and variance of the inter-arrival time of demand are assumed. Also, in this condition, unstable changes at a constant interval are assumed for investigating the influence of the interval. As a more realistic situation, it can be pointed out that the interval of unstable changes is not constant but uncertain, however, the assumption can be considered as more severe than the assumption of a sufficiently long and constant interval, and not so severe as the assumption of the minimum interval, that is, $\tau = 1$. The simulation run-length, in the fourth condition, is ten times as long as that in the simulation experiments in Section 3. The other conditions

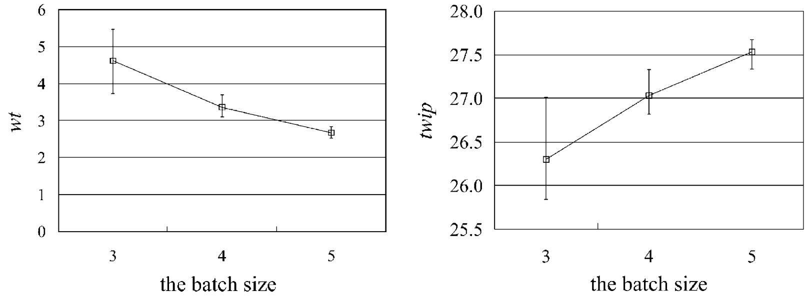

Fig. 6. The influence of the batch size of the EWMA charts upon the two performance measures at $\tau = 100$ and $wt_0 = 2.5$.

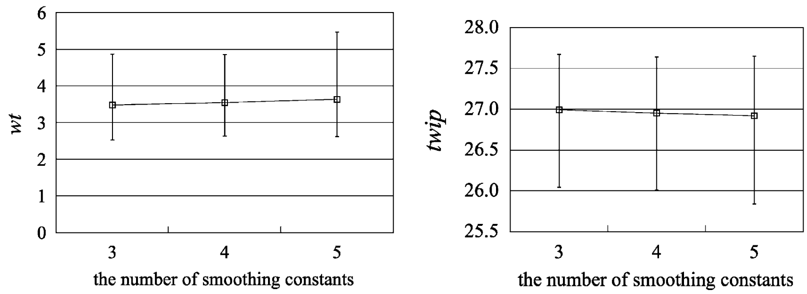

Fig. 7. The influence of the number of smoothing constants of the EWMA charts upon the two performance measures at $\tau = 100$ and $wt_0 = 2.5$.

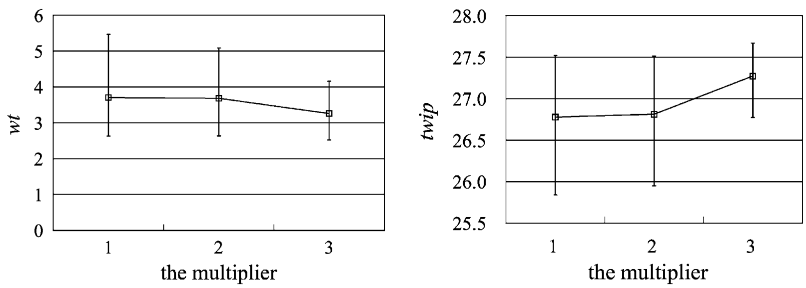

Fig. 8. The influence of the multiplier of the EWMA charts upon the two performance measures with $\tau = 100$ and $wt_0 = 2.5$.

that are not mentioned here are specified as similar to the conditions in Section 3.

As each of the three parameters has three levels, $3^3 = 27$ experiments are performed for the simulation experiment at each condition. Then, as five levels of τ and two levels of wt_0 are considered, $27 \times 5 \times 2 = 270$ experiments are performed totally in this simulation.

At each of the conditions, the simulation result for each of two performance measures is analyzed by ANOVA, and it can be clarified that the influence of m and δ is highly significant (with a significance level $p < 0.01$) and that of the number of the smoothing constants is not significant at all of the conditions. Figs. 6–8 show the influence of the parameters upon the two performance

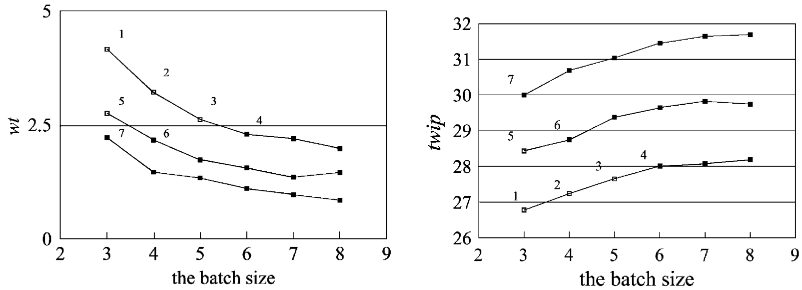

Fig. 9. An example of setting the safety stock and the batch size of the EWMA charts for minimizing *twip* while satisfying $wt_0 = 2.5$.

measures at $\tau = 100$ and $wt_0 = 2.5$. In the figures, the average of each performance measure at each condition is shown with the maximum and minimum of the measure at the condition.

With Fig. 6, it can be seen that, as m increases, *wt* decreases, *twip* increases, and the difference between the maximum and minimum of *wt* or *twip* decreases. Also, with Fig. 8, it can be seen that δ has a similar, but not as great as m in Fig. 6, influence on *wt* and *twip*. On the other hand, with Fig. 7, it can be seen that the number of smoothing constants has little influence on *wt* and *twip*, and *wt* and *twip* hardly change even if it increases. Therefore, based on this analysis, it can be claimed that m has the most significant influence on *wt* and *twip*, and the influence of the other parameters decreases as m increases.

6.2. Influence of input parameters

Next, the influence of the interval τ of unstable changes in demand and the required level wt_0 for the mean waiting time of demand is analyzed by simulation experiments. In the experiments, the number of smoothing constants is fixed at 5 and the multiplier of EWMA charts is fixed at 3, because they do not have so significant influence as the batch size. On the other hand, the batch size as well as the safety stock at the final inventory point is determined as minimizing *twip* while satisfying the required level for *wt*. The safety stock at the final inventory point is for absorbing the delay in detecting unstable changes in demand and controlling the buffer size, and it has an obviously significant influence on *wt* and *twip*. As either of the batch size or the safety stock increases, *wt* decreases and *twip* increases. Therefore, the batch size and the safety stock are initially set at 3 and 0, respectively, and the batch size is increased one by one until *wt* satisfies the required level. The procedure is repeated, increasing the safety stock one by one until *wt* satisfies the required level at the minimum batch size $m = 3$. Then, in the investigated settings, the setting with the minimum *twip* while satisfying the required level for *wt* is selected, and the batch size and the safety stock are determined. Fig. 9 shows an example to illustrate the procedure.

In the figure, the settings of the safety stock and the batch size are investigated in order of the number in the figure, and the settings with the numbers 4, 6, and 7 satisfy $wt_0 = 2.5$. Then, in this example, the setting with the number 4 is selected as it minimizes *twip* while satisfying wt_0.

The other conditions are specified as the same as in the experiments in the preceding subsection. Also, in the experiments, the ideal system that assumes the known unstable changes in demand and controls the buffer size by the proposed control rule for the known unstable changes is considered. In the ideal system, there is no delay in detecting unstable changes in demand, however, there is a delay in controlling the buffer size as in the proposed system. In order to absorb the influence of the delay, the safety stock at the final inventory point is determined as in the proposed system.

Figs. 10 and 11 show the result of simulation experiments. In the figures, the plots with box and triangle show the proposed and the ideal systems,

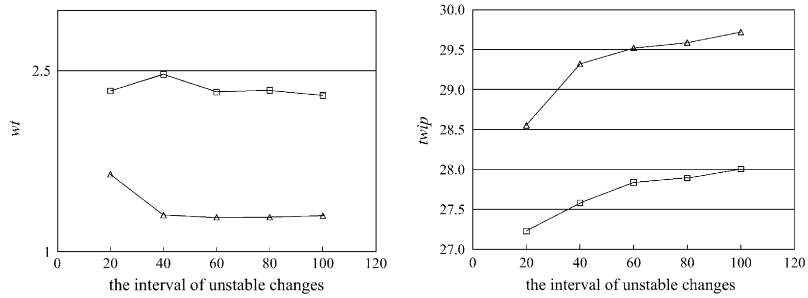

Fig. 10. The influence of the interval of unstable changes in demand upon the two performance measures at $wt_0 = 2.5$ without safety stock (□; the proposed system, △; the ideal system).

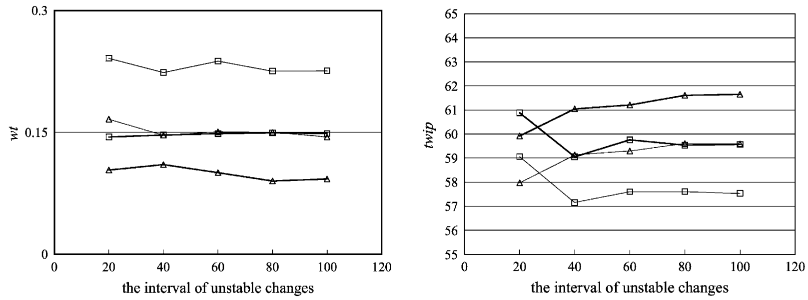

Fig. 11. The influence of the interval of unstable changes in demand upon the two performance measures at $wt_0 = 0.15$ (□; the proposed system, △; the ideal system, thin line; without safety stock, thick line; with safety stock).

respectively. Also, in the figure, thin lines show the performance without safety stock.

In Fig. 10, it can be seen that, without safety stock, *wt* satisfies the required level 2.5 regardless of τ, and *wt* of the proposed system is greater than that of the ideal system. On the other hand, *twip* of the proposed system is less than that of the ideal system. This is because of the delay in detecting unstable changes in demand. Because of the time lag, responses to unstable changes delay, especially that in responding to unstable increases in demand and increasing the buffer size is critical. As increasing the buffer size takes much more time than decreasing it, and the delay leads to an increase in *wt* and a decrease in *twip*. In the ideal system, there is no delay in detecting unstable changes, and *wt* becomes less than that of the proposed system and *twip* becomes greater. As τ decreases and unstable changes in demand occur frequently, the influence of unstable changes becomes more significant, and the influence of the delay in responding to them becomes more significant too. Then, as τ decreases, *wt* increases and *twip* decreases. However, as *wt* satisfies the required level 2.5, it can be claimed that both the proposed and the ideal systems are effective in responding to the unstable changes in demand. Furthermore, as *twip* of the proposed system is less than that of the ideal system regardless of τ, it can be claimed that the proposed system is superior to the ideal system at $wt_0 = 2.5$.

On the other hand, at $wt_0 = 0.15$, the required level for *wt* is more severe than $wt_0 = 2.5$, and the delay in detecting unstable changes and controlling the buffer size becomes important. Then, *wt* of the proposed system without safety stock does not satisfy the required level as shown in Fig. 11. As τ decreases to 20, *twip* of the proposed system increases, and it can be guessed that the proposed system cannot respond to the frequent unstable changes in demand under the severe requirement for *wt*. However, by holding a little safety stock (only one) at the final inventory point, *wt* can satisfy the required level, and it can be claimed that the proposed system is effective in responding to unstable changes even at the severe requirement for wt. Also, in the ideal system without safety stock, wt does not satisfy the required level depending on $\tau(= 20, 60, 80)$, and, by holding a little safety stock (only one) at the final inventory point, wt can satisfy the required level as in the proposed system. Then, at the condition of τ, *twip* of the ideal system becomes greater than that of the proposed system except for $\tau = 20$. At the other conditions of $\tau(= 40, 100)$, *twip* of the proposed system becomes greater than that of the ideal system.

Compared with the estimate of *twip* calculated at the preceding section (31.71 at $wt_0 = 2.5$ and 64.6 at $wt_0 = 0.15$), *twip* of the proposed system without safety stock is about 12–14% less than the estimate at $wt_0 = 2.5$, and about 9–12% at $wt_0 = 0.15$. The reduction in *twip* is caused by the delay in detecting unstable changes in demand and controlling the buffer size. On the other hand, *twip* of the ideal system without safety stock is about 6–10% less than the estimate at $wt_0 = 2.5$, and about 8–10% at $wt_0 = 0.15$. The reduction in *twip* in the ideal system is caused by the delay only in controlling the buffer size, and it can be understood that the reduction caused by the delay in controlling the buffer size (about 6–10% at $wt_0 = 2.5$, and about 8–10% at $wt_0 = 0.15$) is greater than the reduction caused by the delay in detecting unstable changes in demand (about 4% at $wt_0 = 2.5$, and about 1–2% at $wt_0 = 0.15$).

7. Conclusions

In order to realize agile control for multi-stage production systems, this paper proposed a reactive JIT ordering system based on the Kanban system. In the proposed system, unstable changes in demand, not only in the mean but also in the variance, can be detected by EWMA charts, and the appropriate buffer size for the detected unstable changes is calculated on the basis of the simulation results under the stable-demand conditions. Then, the buffer size is adjusted to the calculated size in response to the detected unstable changes. By simulation experiments, the performance of the proposed system was analyzed. The results showed that the batch size in grouping the time series data for detecting unstable changes in demand and the multiplier of EWMA charts, especially the former, have a significant influence on the performance. Also, the results showed the influence of the interval between unstable changes and the required level for the mean waiting time of demand. As a result, it was clarified that the required level for the mean waiting time of demand can be satisfied at the loose requirement, but it cannot be satisfied at the severe requirement because of the delay in detecting unstable changes and controlling the buffer size, especially the latter. However, by holding a little safety stock only at the final inventory point, the required level can be satisfied. Based on the results, it can be claimed that the proposed system is effective in reacting to unstable changes in demand. Also, it can be claimed that it is necessary to consider some mechanisms to counter the delay in detecting unstable changes and controlling the buffer size, especially under the severe requirement for the mean waiting time of demand. The results showed that the safety stock is valuable in countering the delay.

In this paper, the time series data on demand are grouped into batches and the batch mean and variance are utilized for detecting unstable changes in the mean and variance of demand. The delay in detecting unstable changes is also caused by the grouping. For solving the problem, an enhanced method to detect unstable changes would be an interesting problem for research. Also, in the proposed systems, not only detecting unstable changes but also controlling the buffer size is delayed, and the problem becomes significant under the short interval of unstable changes and the tight requirement for the mean waiting time of demand. It is expected that mechanisms for reducing or eliminating the delay would be effective in improving the performance of reactive JIT ordering systems. This suggests itself as a useful future area of research. Also, while the stochastic production time at each stage was assumed, unstable changes in the production time were not considered in this paper. Considering unstable changes in production time can be suggested as another future area of research.

Acknowledgements

The work in this paper was partially supported by the grant-in-aid for Scientific Research of the Japanese Ministry of Education, Science, Sports and Culture in 1998-2001.

References

Chang, T.-M., Yih, Y., 1994. Generic Kanban systems for dynamic environments. International Journal of Production Research 32 (4), 889–902.

Chang, T.-M., Yih, Y., 1998. A fuzzy rule-based approach for dynamic control of Kanbans in a generic Kanban system. International Journal of Production Research 36 (8), 2247–2257.

Crowder, S.V., Hamilton, M.D., 1992. An EWMA for monitoring a process standard deviation. Journal of Quality Technology 24 (1), 12–21.

Gan, F.F., 1995. Joint monitoring of process mean and variance using exponentially weighted moving average control charts. Technometrics 37 (4), 446–453.

Gupta, S.M., Al-Turki, Y.A.Y., 1997. An algorithm to dynamically adjust the number of Kanbans in stochastic processing times and variable demand environment. Production Planning and Control 8 (2), 133–141.

Gupta, S.M., Al-Turki, Y.A.Y., 1998a. Adapting just-in-time manufacturing systems to preventive maintenance interruptions. Production Planning and Control 9 (4), 349–359.

Gupta, S.M., Al-Turki, Y.A.Y., 1998b. The effect of sudden material handling system breakdown on the performance of a JIT system. International Journal of Production Research 36 (7), 1935–1960.

Gupta, S.M., Al-Turki, Y.A.Y., Perry, R.F., 1999. Flexible Kanban system. International Journal of Operations and Production Management 19 (10), 1065–1093.

Hopp, W.J., Roof, M.L., 1998. Setting WIP levels with statistical throughput control (STC) in CONWIP production lines. International Journal of Production Research 36 (4), 867–882.

Izumi, M., Takahashi, K., 1993. Concurrent ordering in JIT production system, in: Proceedings of 2nd China–Japan Symposium on Industrial Management, (J. Gu and H. Osaki (Eds.)), pp. 51–56. International Academic Publishers.

Kim, S.-L., Ha, D., 2003. A JIT lot-splitting model for supply chain management: Enhancing buyer–supplier linkage. International Journal of Production Economics 86 (1), 1–10.

Kimura, O., Terada, H., 1981. Design and analysis of pull system: A method of multi-stage production control. International Journal of Production Research 19 (3), 241–253.

Lucas, J.M., Saccucci, M.S., 1990. Exponentially weighted moving average control schemes: Properties and enhancements. Technometrics 32 (1), 1–12.

Machuca, J.A.D., 2002. JIT Facing the new millennium. International Journal of Production Economics 80 (2), 131–134.

Markham, I.S., Mathieu, R.G., Wray, B.A., 1998. A rule induction approach for determining the number of Kanbans in a just-in-time production system. Computers and Industrial Engineering 34 (4), 717–727.

Mitra, D., Mitrani, I., 1990. Analysis of a Kanban discipline for cell coordination in production line, I. Management Science 36 (12), 1548–1566.

Mitra, D., Mitrani, I., 1991. Analysis of a Kanban discipline for cell coordination in production line, II: Stochastic demands. Operations Research 39 (5), 807–823.

Monden, Y., 1981. Adaptable Kanban system helps Toyota maintain just-in-time production. Industrial Engineering 13 (5), 29–46.

Moore, K., Gupta, S.M., 1999. Stochastic coloured Petri net (SCPN) models of traditional and flexible Kanban systems. International Journal of Production Research 37 (9), 2135–2158.

Price, W., Gravel, M., Nsakanda, A.L., 1994. A review of optimization models of Kanban-based production systems. European Journal of Operational Research 75 (1), 1–12.

Rees, L.P., Philipoom, P.R., Taylor III, B.W., Huang, P.Y., 1987. Dynamically adjusting the number of Kanbans in a just-in-time production system using estimated values of leadtime. IIE Transactions 19 (2), 199–207.

Spearman, M.L., 1992. Customer service in pull production systems. Operations Research 40 (5), 948–958.

Takahashi, K., 2002. Guest editorial: Just-in-time: New applications and research directions. Production Planning and Control 13 (8), 679–680.

Takahashi, K., Nakamura, N., 1998. Ordering alternatives in JIT production systems. Production Planning and Control 9 (8), 784–794.

Takahashi, K., Nakamura, N., 1999. Reacting JIT ordering systems to the unstable changes in demand. International Journal of Production Research 37 (10), 2293–2313.

Takahashi, K., Nakamura, N., 2000a. Reactive logistics in a JIT environment. Production Planning and Control 11 (1), 20–31.

Takahashi, K., Nakamura, N., 2000b. Agile control in JIT ordering systems. International Journal of Agile Management Systems 2 (3), 242–252.

Takahashi, K., Nakamura, N., 2002. Decentralized reactive Kanban system. European Journal of Operational Research 139 (3), 262–276.

Takahashi, K., Nakamura, N., Izumi, M., 1996a. Concurrent ordering in JIT production systems. International Journal of Operations and Production Management 17 (3), 267–290.

Takahashi, K., Nakamura, N., Ohashi, K., 1996b. Order release in JIT production systems: A simulation study. Simulation 66 (2), 75–87.

Tardif, V., Maaseidvaag, L., 2001. An adaptive approach to controlling Kanban systems. European Journal of Operational Research 132 (2), 411–424.

Tayur, S.R., 1993. Structural properties and a heuristic for Kanban-controlled serial lines. Management Science 39 (11), 1347–1368.

Wray, B.A., Rakes, T.R., Rees, L.P., 1997. Neural network identification of critical factors in a dynamic just-in-time Kanban environment. Journal of Intelligent manufacturing 8 (2), 83–96.

Zimmer, K., 2002. Supply chain coordination with uncertain just-in-time delivery. International Journal of Production Economics 77 (1), 1–94.