

Performance measurement and lumped parameter modelling of single server flowlines subject to blocking : an effective process time approach

Citation for published version (APA):

Kock, A. A. A., Wullems, F. J. J., Etman, L. F. P., Nijse, F., Adan, I. J. B. F., & Rooda, J. E. (2006). *Performance measurement and lumped parameter modelling of single server flowlines subject to blocking : an effective process time approach*. (SE report; Vol. 2006-09). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2006

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Systems Engineering Group
Department of Mechanical Engineering
Eindhoven University of Technology
PO Box 513
5600 MB Eindhoven
The Netherlands
<http://se.wtb.tue.nl/>

SE Report: Nr. 2006-09

Performance measurement and
lumped parameter modelling of
single server flowlines subject to
blocking: an effective process time
approach

A.A.A. Kock, F.J.J. Wullems, L.F.P. Etman, I.J.B.F. Adan, F. Nijssse, J.E. Rooda

ISSN: 1872-1567

SE Report: Nr. 2006-09
Eindhoven, December 2006
SE Reports are available via <http://se.wtb.tue.nl/sereports>

Abstract

The present paper extends the so-called Effective Process Time (EPT) approach to single server flowlines with finite buffers and blocking. The power of the EPT approach is that it quantifies variability in workstation process times without the need to identify each of the contributing disturbances, and that it directly provides an algorithm for the actual computation of EPTs. It is shown that EPT realizations can be simply obtained from arrival and departure times of lots, by using sample path equations. The measured EPTs can be used for bottleneck analysis and for lumped parameter modeling. Simulation experiments show that for lumped parameter modeling of flowlines with finite buffers, in addition to the mean and variance, offset is also a relevant parameter of the process time distribution. A case from the automotive industry illustrates the approach.

1 Introduction

Single server workstations with finite buffer sizes in a tandem flowline are an important class of manufacturing systems. Examples of such flowlines are semi-synchronous lines and assembly lines, as e.g. encountered in the automotive industry.

The performance of a flowline is commonly expressed in terms of throughput and flow time. Both performance indicators are influenced by blocking. The finite capacity of the buffers in the single server flowlines considered in this paper introduces blocking in the line.

Blocking causes suspension of service to a lot (which implies loss of production capacity) since a finished lot cannot be sent on due to a saturated downstream buffer. Starvation refers to the situation where processing of the next lot is suspended due to an empty upstream buffer.

Variability in process times is the main reason that blocking and starvation occur. The variability of process times can be traced to several common sources. First of all, natural process times are variable due to differences in product types, machine states at product entry, operator behaviour, etcetera. Furthermore, disturbances such as setups, preventive maintenance, machine failures and absence of operators occur. These disturbances cause loss of production capacity effectively available at the workstation and increase the variability of process times, which in turn decreases the throughput. Subsequent workstations affect one another more prominently as the variability of process times increases. Variability of process times on workstation j can cause starvation on workstation $j + 1$. Furthermore, in a flowline with finite buffers, variability of process times of workstation j can cause blocking on workstation $j - 1$.

Obviously, for performance analysis of a finitely buffered flowline, an analysis tool that quantifies both the production losses and the level of variability of process times is required. A commonly applied performance analysis metric is the overall equipment efficiency, OEE. However, OEE can only be used for quantifying production losses. Therefore an alternative analysis tool will be used in this paper.

[5] introduced this alternative concept to account for irregularities in process times of workstations. The new concept, effective process time (EPT), is defined as the total time seen by a lot at a workstation from a logistical point of view. Here, total time indicates the total time that the lot has effectively consumed production capacity of the workstation. EPT is based on the notion that, from a logistical perspective, a workstation does not care whether production capacity is claimed since the server is processing the lot or whether production capacity is claimed by other influences. These other influences are included in the EPT of the workstation.

[5]'s notion of including processing disturbances in the effective process times is not new, see e.g. the work of [2, 4, 1]. The aforementioned authors all assume, or measure, distributions for the various processing disturbances and combine these into one single distribution. However, from industrial practice, it is often hard, if not impossible, to identify and quantify all individual disturbances. Starting from the concept of EPT, [7, 8] presented a method to translate lot arrivals and lot departures at an infinitely buffered workstation into an EPT distribution. In automated manufacturing environments, arrival and departure data is often available.

The obtained EPT distributions can be used for performance optimization. Based on the characteristic parameters of the EPT distributions, i.e. the mean effective process time t_e and the squared coefficient of variation c_e^2 , a bottleneck analysis can be performed, after which an approximating model can be used to predict the changes in system performance. [6] also use t_e and c_e^2 as workstation parameters in an open queueing network model for flowline optimisation. However, they compute t_e and c_e^2 from the theoretical process time values by assuming that outages are adequately represented by exponentially distributed failures and repairs. The EPT framework

presented in this paper follows the concept of [7, 8], which does not require the characterisation of the various contributing disturbances. This is a clear advantage of the EPT–approach since, as mentioned above, it is in practice often hard or impossible to quantify all individual sources of disturbances.

This paper aims to generalize the EPT–approach for application to single server flowlines subject to blocking. That is, the paper considers finite buffers rather than infinite ones. Workstations can no longer be analysed in isolation due to the dependencies introduced by blocking. Therefore, an EPT–algorithm for the blocking case is presented. Furthermore, the effect of the distribution shape on the accuracy of the EPT lumped parameter (ELP) model is investigated. Two theoretical examples and a case from automotive industry is used to illustrate the EPT–approach. Note that throughout the paper, mainly the effects of blocking are discussed since starvation also occurs in infinitely buffered workstations.

The paper is organized as follows. In Section 2, an outline of the EPT–approach is presented. Subsequently, computation of EPT–realisations for single server workstations with finite buffers is considered in Section 3. EPT–based lumped parameter modelling in the context of finitely buffered flowlines is discussed in Section 4. The concepts discussed throughout this paper are illustrated using the aforementioned examples and case in Sections 5 and 6. Finally, Section 7 concludes the paper.

2 A framework for implementing EPT

The EPT–approach, based on the concept of [8], consists of four stages, as visualized in Figure 1.

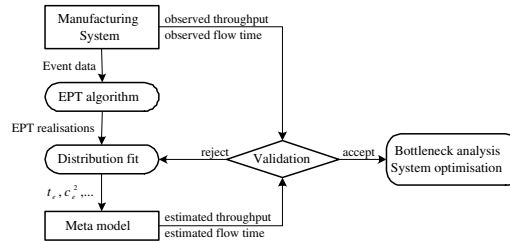


Figure 1: Schematic representation of the EPT–approach

First, EPT–realisations are obtained from the discrete manufacturing system. An EPT realisation is defined by [8] as: ‘the time a lot was in process plus the time a lot (not necessarily the same lot) could have been in process’. EPT–realisations can be computed from event data, such as arrivals and departures of lots on workstations. The EPT–realisations are computed by means of an EPT–algorithm. The EPT, or similar concepts (such as completion time) are used in sample path analyses of queueing systems. Sample path equations are typically used to determine lot departures from lot arrivals and the effective process time. The EPT–concept presented in this paper uses the sample path equations differently, that is, effective process times are determined from arrival and departure data. The sample path equations are thus a means to obtain EPT–realisations from an operating production system. The operation time as defined by [11] is very similar to EPT; however, [11] do not use it to quantify the level of variability.

Next, the EPT–realisations are fitted to distributions. Here, distributions are fitted based on relevant workstation properties, such as the mean EPT t_e and the squared coefficient of variation c_v^2 .

3 A framework for implementing EPT

Parameter t_e quantifies the mean effective capacity used for a lot by the workstation, c_e^2 quantifies the effective variability.

Subsequently, a so-called EPT lumped parameter (ELP) model can be built using the fitted distributions. This ELP model can be used for performance prediction and optimisation. The structure of the ELP model follows the original system in terms of the number of servers on each workstation, the buffer sizes of workstations, the flow of materials between workstations, etcetera. In this model, detailed modelling of shop-floor realities such as failures, repairs, setups, operators and lot sizes is avoided. The various sources of variability are aggregated into the EPT-distributions of the workstations. [8] used the term ‘meta model’ rather than ‘lumped parameter model’. However, the phrase ‘meta model’ may suggest that a simplified model is derived from another model. Since this is certainly not the case, the terminology ‘lumped parameter model’ is used in this paper. Here, the lumped parameters refer to the distribution parameters of the EPT-distributions.

Before the ELP model is accepted, it is validated by comparing the throughput and flow time as estimated by the model to those observed in the actual system, since one is interested in how well the lumped parameter model describes the behaviour of the actual system. If the estimated throughput and flow time are accurate enough, the ELP model and the EPT-distributions are accepted. If they are rejected, distribution fitting and model building are reconsidered. Possible changes include enhancing the level of detail of the model or using more parameters to fit more accurate distributions.

If the EPT-distributions and the ELP model are accepted, they can be used for performance analysis and optimisation. A bottleneck analysis can be carried out based on the distribution parameters t_e and c_e^2 of the various workstations. The effect of suggested improvements can be evaluated using the ELP model by accordingly adjusting the EPT distribution parameters in the model.

Implementation of the EPT-approach provides several significant advantages. First of all, many shop-floor realities are included in the EPT-distributions and thus do not have to be included explicitly in the ELP model. Now, an ELP model can be obtained that is accurate, yet simple when compared to the detailed models that are typically used. Second of all, since the processing disturbances are included in the EPT-distributions, directly obtained from industrial data, the EPT parameters t_e and c_e^2 readily give insight in the behaviour of the flowline, allowing for straightforward bottleneck analysis.

3 Measuring EPT

The EPT was introduced by [5] to be used in queueing models. Similar concepts, such as completion time, are used in sample path equations. In all references, the respective distributions are assumed to be known *a priori*, and then the sample path equations are used to derive properties concerning flow time, throughput, etcetera. None of the authors, however, specifies how these distributions should be estimated from industrial data.

[8] presented a method to compute EPT-realizations for infinitely buffered, isolated workstations from industrial data. Their method does not assume the effective process time distributions *a priori*, but, in a way similar to using a sample path equation, determines these distributions. For a single machine workstation, the sample path equation is:

$$\mathbf{EPT}_{i,j} = \mathbf{AD}_{i,j} - \max \{ \mathbf{AA}_{i,j}, \mathbf{AD}_{i-1,j} \}, \quad (1)$$

where $\mathbf{EPT}_{i,j}$ denotes the EPT realisation of lot i on workstation j , $\mathbf{AD}_{i,j}$ is the departure of lot i from workstation j and $\mathbf{AA}_{i,j}$ is the arrival of lot i on workstation j . From Equation (1), one sees

that an EPT realisation encompasses all time during which the server could have been processing the lot. For the events holds that $\mathbf{AA}_{i,j} \leq \mathbf{AD}_{i,j}$. In case of timeless transport, $\mathbf{AD}_{i,j-1} = \mathbf{AA}_{i,j}$.

Algorithmic extensions have been presented for workstations with multiple parallel servers [8] and with batching [?]. However, the algorithms are only applicable to workstations with an infinitely large buffer. This paper studies finite buffers, which gives rise to blocking. Due to blocking, $\mathbf{EPT}_{i,j}$ depends on events occurring on workstation $j - 1$, rendering the previous algorithms inapplicable.

Considering finitely buffered workstations, the sample path equation for the departure of lots is given by (see page 184 of [1] or [?]):

$$D_i^{(j)} = \max \left[\max \left\{ D_i^{(j-1)}, D_{i-1}^{(j)} \right\} + EPT_i^{(j)}, D_{i-b_{j+1}}^{(j+1)} \right] \quad (2)$$

Herein, $D_i^{(j)}$ is the i th departure from workstation j ; the term $\max \left\{ D_i^{(j-1)}, D_{i-1}^{(j)} \right\}$ represents the time at which processing of the lot can start; $EPT_i^{(j)}$ represents the completion time and $D_{i-b_{j+1}}^{(j+1)}$ represents the earliest time at which the receiving workstation has sufficient capacity available to receive the lot. Thus, $\max \left\{ D_i^{(j-1)}, D_{i-1}^{(j)} \right\} + EPT_i^{(j)}$ equals the time at which the lot can leave the workstation, provided that the receiving workstation has sufficient capacity available. In this paper, this moment is denoted as $\mathbf{PD}_{i,j}$. However, the lot only actually leaves, $\mathbf{AD}_{i,j}$, if the receiving workstation is available. Thus, the difference between a possible departure and an actual departure is caused by blocking. Hence,

$$\mathbf{PD}_{i,j} = \max \left\{ \mathbf{AA}_{i,j}, \mathbf{AD}_{i-1,j} \right\} + \mathbf{EPT}_{i,j}, \quad (3)$$

which can be rewritten to

$$\mathbf{EPT}_{i,j} = \mathbf{PD}_{i,j} - \max \left(\mathbf{AA}_{i,j}, \mathbf{AD}_{i-1,j} \right). \quad (4)$$

As can be seen, one should replace $\mathbf{AD}_{i,j}$ in Equation (1) by $\mathbf{PD}_{i,j}$. Possible occurrences of blocking should not be included in the EPT realisation. They are a physical part of the finitely buffered flowline and will also appear in the EPT based lumped parameter (ELP) model.

4 Lumped parameter modelling

Distribution fitting is the second phase of the EPT–approach. The relevant distribution parameters are estimated based on the measured EPT–realisations and appropriate distribution functions are proposed.

Process time distributions based on the first two moments of the distribution are often used in models of manufacturing systems consisting of workstations with infinitely large buffers. The two–moment fits are supported by queueing theory, see [1] and [3].

For workstations in a flowline with finite buffer sizes, distribution fitting could be more complicated. Due to blocking, workstations are expected to affect one another more prominently. Therefore, extra information may be needed. Regardless, in queueing theoretical approaches, two moment distribution fits are used for computational reasons. However, in case of simulation, the use of higher order information may be reconsidered. A typical example thereof is presented by [9]. They study constant natural process times with exponentially distributed times to failure and times to repair. In the EPT–approach, the sources of disturbances are lumped. In this respect, no assumptions regarding the distribution of the process times or disturbances are made. The necessity of additional distributional information in ELP models of finitely buffered flowlines will

be studied here.

Using simulation, the influence of the offset parameter is investigated. The offset parameter is chosen since, in practice, many operations require at least a minimum amount of time. The offset refers to the smallest possible value of a distribution. The simulation model is a flowline consisting of three unbuffered single server workstations in which lots do not overtake. The three workstations have process times distributed according to a shifted Gamma distribution. The distributional parameters are $t_e = 1.0$, $c_e^2 = 1.0$ and offset Δ_e . The offset parameter is varied from $\Delta_e = 0.0$ to $\Delta_e = 0.9$.

The corresponding simulation results are presented in Figure 2. The results show that for large offsets, significant differences in throughput (δ) and flow time (φ) are observed. Increasing Δ_e from 0.0 to 0.9 results in a throughput increase of 50% and a flow time decrease of 21% (see Figure 2).

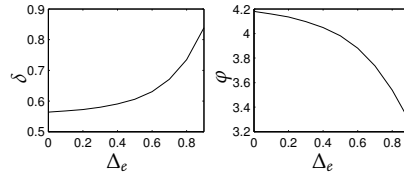


Figure 2: Influence of the offset parameter on δ and φ

The observed phenomenon can readily be explained by considering the nature of the offset. An offsetted distribution consists of a constant part, Δ_e , that is increased by a random variable with mean t_1 and squared coefficient of variation c_1^2 , where $t_1 = t_e - \Delta_e$. Since the variance of the process time distribution does not change, one knows that $t_e^2 c_e^2 = t_1^2 c_1^2$. Now, if $t_1 = 0.1t_e$ (i.e. $\Delta_e = 0.9$), $c_1^2 = 100c_e^2$. Due to the large c_1^2 , most process times will be small ($\gtrsim \Delta_e$), and sporadically a value greatly exceeding the average ($\gg t_e$) will occur. The sporadic large process time realisation therefore causes massive amounts of blocking on preceding and starvation on successive workstations. If $\Delta_e = 0.0$ however, all process times will be centred around t_e . Process times will thus often be larger than t_e , frequently causing some blocking and starvation on preceding or successive workstations.

A new set of simulations is used to test how relevant the usage of Δ_e is. As stated above, one can expect the shape of the distribution to have more influence if the amount of blocking and starvation increases. This expectation is investigated using simulation. For a flowline consisting of three finitely buffered workstations with a single server, the number of bufferspaces between WS_0 and WS_1 and between WS_1 and WS_2 will be changed. In addition, the level of variability is changed. Process times on the workstations will have identical t_e . However, c_e^2 is chosen at 1.0 at the first workstation, but is varied from 0.5 to 2.0 at the other two workstations. The throughput and flow time will be evaluated at offset levels of $\Delta_e = 0.0$ and $\Delta_e = 0.9$. The corresponding simulation results are depicted in Figure 3.

Several observations can be made from Fig. 3. The first observation is that the mean throughput for $\Delta_e = 0.9$ approaches the throughput for $\Delta_e = 0.0$ as the amount of buffer spaces increases. A second important observation, from comparing Fig. 3(a) to Fig. 3(c), is that the difference in mean throughput between $\Delta_e = 0.9$ and $\Delta_e = 0.0$ becomes larger as the squared coefficients of variation are increased. A change in the squared coefficient of variation has more effect on performance for $\Delta_e = 0.0$ than for $\Delta_e = 0.9$. This observation can be explained by the fact that a flowline with $\Delta_e = 0.0$ is more likely to be blocked than a flowline with $\Delta_e = 0.9$. Since an increase in variability implies an increase in the amount of blocking, the flowline with $\Delta_e = 0.0$ is more heavily affected.

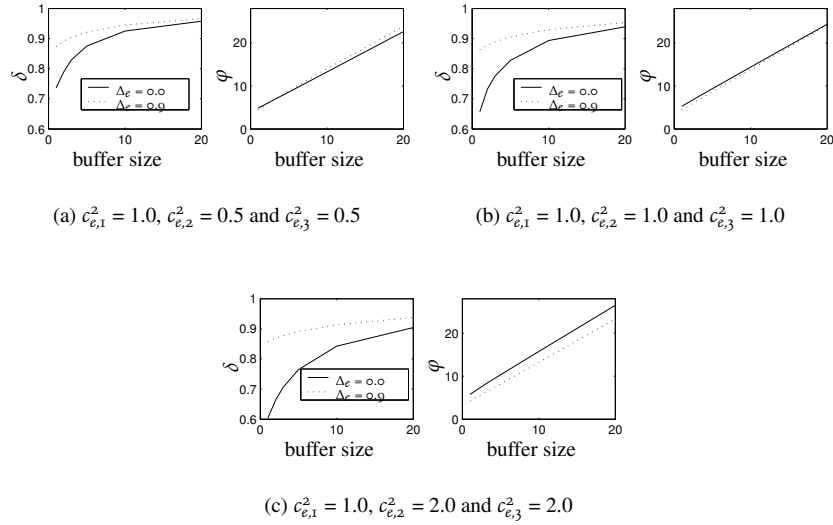


Figure 3: Influence of buffer spaces on throughput and flow time

These results, and additional simulation results presented in [10], imply that, as the amount of blocking and starvation in the flow line increases (by reducing the number of bufferspaces or by increasing the level of variability), the influence of higher order information of the distribution shape increases.

5 Example

Two examples are presented to validate the computation of EPT-realizations and to illustrate the EPT-approach.

Example I

Consider a flow line consisting of five workstations labelled WS_i for $i = 0, \dots, 4$. Each workstation has a buffer of size one and one server. The first workstation is never starved whereas the final workstation is never blocked. All workstations have exponentially distributed natural process times with mean $t_{o,i} = 1.0$ for all i . The servers are subject to operation dependent failures, with busy time between failures exponentially distributed with mean $t_{f,i} = 15$ for all i . Once a failure has occurred, the server is repaired. Repair times are exponentially distributed with mean $t_{r,i} = 2$ for all i . After the repair is finished, processing of the lot is continued for a period of time equal to the remaining process time. The flowline is represented using a detailed discrete event simulation model, explicitly modelling the failure and repair behavior. This model will be referred to as the ‘original’ model.

The first stage of the EPT approach is carried out by applying Equation (4) to the arrival and departure events generated by the original model. This leads to a large set of EPT realizations for each of the workstations. During the second stage of the approach, the realizations are translated into shifted Gamma distributions with mean $t_{e,i}$, squared coefficient of variation $c_{e,i}^2$ and offset $\Delta_{e,i}$ as presented in Table 1. The t_e and c_e^2 values of the table are verified using Equation (5) as presented by [5]. Herein, t_o is the mean natural process time, c_o^2 is the corresponding squared

coefficient of variation, c_r^2 is the squared coefficient of variation of the times to repair, and A is the availability. The equation gives values $t_{e,i} = 1.13$ and $c_{e,i}^2 = 1.42$ for all i , which corresponds to the measured equivalents in Table 1.

$$\begin{aligned} t_{e,i} &= \frac{t_{o,i}}{A_i}, \\ A_i &= \frac{t_{f,i}}{t_{f,i} + t_{r,i}}, \\ c_{e,i}^2 &= c_{o,i}^2 + (1 + c_{r,i}^2) A_i (1 - A_i) \frac{t_{r,i}}{t_{o,i}}. \end{aligned} \quad (5)$$

Since the natural process times are exponentially distributed, as are the failures and repairs, the effective process time distributions of the workstations do not have an offset, i.e. $\Delta_e = 0.0$. As can be seen in Table 1, the estimated value of Δ_e is indeed 0.0.

The original simulation model has $\delta = 0.495 \pm 0.01\%$ and $\varphi = 14.15 \pm 0.01\%$. This implies that, with a probability of 95%, the range $(0.49495, 0.49505)$ contains the true value of δ and the range $(14.1486, 14.1514)$ contains the true value of φ .

WS_i	$\Delta_{e,i}$	$t_{e,i}$	$c_{e,i}^2$
WS_0	0.00	1.13	1.41
WS_1	0.00	1.13	1.42
WS_2	0.00	1.13	1.42
WS_3	0.00	1.13	1.42
WS_4	0.00	1.13	1.42

Table 1: Measured EPT parameters for example I

During the third stage of the approach, the approximated distributions are used as input for a discrete event EPT-based lumped parameter (ELP) model. The structure of the ELP model follows the structure of the original system, i.e. five workstations consisting of one buffer space and one server. Servers have process times distributed according to the shifted Gamma distribution, with parameters according to Table 1. The ELP model finds approximates $\delta = 0.491$ and $\varphi = 14.26$, which means that the difference between the EPT approximation and the original situation is 0.81% in throughput and 0.77% in flow time. The error in the approximation is computed by:

$$\left| \frac{\delta - \hat{\delta}}{\delta} \right| \cdot 100\% \quad \text{and} \quad \left| \frac{\varphi - \hat{\varphi}}{\varphi} \right| \cdot 100\% \quad (6)$$

Note that (6) is used in the remainder of this paper to compute the error in approximations.

If both the original system and the ELP model do not contain buffer spaces, the original model gives performance measures $\delta = 0.399$ and $\varphi = 9.23$, whereas the approximation is $\hat{\delta} = 0.393$ and $\hat{\varphi} = 9.34$, giving an error of 1.5% for throughput and 1.2% for flow time. Increasing the number of buffer spaces on all workstations to 5 leads to $\delta = 0.656$ and $\varphi = 29.72$ for the original model compared to $\hat{\delta} = 0.657$ and $\hat{\varphi} = 29.76$ for the EPT-based meta model. This is an error of 0.2% in throughput and 0.1% in flow time. Obviously, the error decreases as the number of buffer spaces in the line increases, which corresponds with the observations of section 4.

Example II

Consider a flowline consisting of five workstations WS_i for $i = 0, \dots, 4$. Workstation WS_i has b_i buffer spaces and one single server, where $[b_0, b_1, b_2, b_3, b_4] = [0, 2, 1, 2, 1]$. The flowline produces two product types, pt_0 and pt_1 in the deterministic sequence $[pt_0, pt_1, pt_0, pt_1, \dots]$. The first workstation is never starved whereas the final workstation is never blocked. At WS_0 , all

products are processed with exponentially distributed natural process times with mean 1. At WS_1 and WS_3 , natural process times for products of type pt_o are distributed according to a shifted Gamma distribution with $\Delta_{o,o} = 0.6$, $t_{o,o} = 1.5$ and $c_{o,o}^2 = 0.75$, whereas $\Delta_{o,i} = 0.2$, $t_{o,i} = 0.5$ and $c_{o,i}^2 = 0.75$ on these stations for products of type pt_i . On workstations WS_2 and WS_4 , products of type pt_o are processed with natural process times according to a triangular distribution with $\Delta_{o,o} = 0.4$, $t_{o,o} = 0.5$ and maximum 0.6 and thus, $c_{o,o}^2 = 6.67 \cdot 10^{-3}$; for pt_i however $\Delta_{o,i} = 1.2$, $t_{o,i} = 1.5$ and maximum 1.8 giving $c_{o,i}^2 = c_{o,o}^2$. On WS_i for $i = 1, 2, 3, 4$, a constant setuptime of 0.1 time units is required if the product type is changed. The servers are subject to operation dependent failures, with busy time between failures exponentially distributed with mean $t_{f,i} = 15$ for all i . Once a failure has occurred, the server is repaired. Repair times are exponentially distributed with mean $t_{r,i} = 2$ for all i . After the repair is finished, processing of the lot is resumed at the point where it was interrupted. Simulation results for the example have been obtained for 95% confidence levels with a relative width of 1% or less of the corresponding mean.

First, EPT realisations are computed for each of the workstations by applying Equation (4) to the arrival events (AA) and departure events (PD, AD) obtained from the simulation model. Next, the realisations are translated into shifted Gamma distributions with mean $t_{e,i}$, squared coefficient of variation $c_{e,i}^2$ and offset $\Delta_{e,i}$ as presented in Table 2. The t_e and c_e^2 values of the Table are verified using Equation (5). To properly apply these equations, the two natural process time distributions of a workstation are first translated into a general natural process time distribution. Let X denote the overall natural process time and X_o and X_i reflect the type specific natural process times. Then:

$$t_{o,i} = E[X_i], \quad \text{for } i = \{o, i\} \quad (7)$$

$$c_{o,i}^2 = \frac{E[X_i^2]}{(E[X_i])^2} - 1 \quad (8)$$

$$E[X^2] = 0.01 + 0.1(t_{o,o} + t_{o,i}) + 0.5(t_{o,o}^2(c_{o,o}^2 + 1) + t_{o,i}^2(c_{o,i}^2 + 1)), \quad (9)$$

$$t_o = E[X] = 0.1 + \frac{t_{o,o} + t_{o,i}}{2}, \quad (10)$$

$$c_o^2 = \frac{E[X^2]}{(E[X])^2} - 1. \quad (11)$$

Equations (5), (10) and (11) yield $t_{e,o} = 1.27$, $c_{e,o}^2 = 1.66$, $t_{e,i} = t_{e,3} = 1.39$, $c_{e,i}^2 = c_{e,3}^2 = 1.59$ and $t_{e,2} = t_{e,4} = 1.39$, $c_{e,2}^2 = c_{e,4}^2 = 0.82$. As can be seen in Table 2, the estimated EPT parameters are correct. When considering the input distributions, one knows that $\Delta_{e,o} = 0.0$, $\Delta_{e,i} = \Delta_{e,3} = 0.3$ and $\Delta_{e,2} = \Delta_{e,4} = 0.50$, which also corresponds to the values presented in the Table.

WS_i	$\Delta_{e,i}$	$t_{e,i}$	$c_{e,i}^2$
WS_o	0.00	1.27	1.66
WS_1	0.30	1.39	1.59
WS_2	0.50	1.39	0.82
WS_3	0.30	1.39	1.59
WS_4	0.50	1.39	0.82

Table 2: Measured EPT parameters for example II with a single EPT distribution

The observed flowline performance is $\delta = 0.462 \pm 0.01\%$ and $\varphi = 15.70 \pm 0.01\%$. This implies that, with a probability of 95%, the range (0.46195, 0.46246) contains the true value of δ and the range (15.69843, 15.70157) contains the true value of φ .

Next, shifted Gamma distributions with parameters as presented in Table 2 are used as input for an ELP model. The ELP model approximates $\delta = 0.444$ and $\varphi = 16.74$, which means that the

difference between the EPT approximation and the original situation is 4.0% for throughput δ and 6.6% for flow time φ .

Part of these errors can be explained as follows. Firstly, the ELP model assumes identically and independently distributed (iid) process times on all workstations. In the case considered here, each lot is of a different type than the preceding one. Since $t_{i,0}$ differs from $t_{i,1}$ $i = 1, 2, 3, 4$, a correlation is expected for successive process times on a workstation. Due to the assumption of iid process times in the ELP model, these correlations between successive process times on a workstation are neglected. Secondly, in the ELP model, the process times of one lot on the successive workstations are assumed to be independent. In the original model however, process times for one lot on successive workstations are correlated due to the type-specific natural process times. The lumped parameter model again does not incorporate this correlation.

The error in the approximation can be reduced by fitting EPT-distributions for each product type per workstation. The new distributional properties are presented in Table 3. Comparing these values with Equations (10) through (11) again shows that the estimated values are correct. Inserting the distribution properties of Table 3 into the lumped parameter model yields $\delta = 0.460$ and $\varphi = 15.78$, which is an error of 0.4% for throughput and 0.5% for flow time.

WS_i	pt_0			pt_1		
	$\Delta_{e,i}$	$t_{e,i}$	$c_{e,i}^2$	$\Delta_{e,i}$	$t_{e,i}$	$c_{e,i}^2$
WS_0	0.00	1.27	1.66	0.00	1.27	1.66
WS_1	0.70	2.03	1.07	0.30	0.76	1.63
WS_2	0.50	0.76	1.11	1.30	2.03	0.42
WS_3	0.70	2.03	1.07	0.30	0.76	1.63
WS_4	0.50	0.76	1.11	1.30	2.03	0.42

Table 3: Measured EPT parameters for example II with deterministic lot type sequence and product type specific EPT-distributions

The latter procedure is repeated for different levels of buffering. If both the original system and the lumped parameter model contain no buffer spaces, the original model gives performance measures $\delta = 0.364$ and $\varphi = 10.16$, whereas the approximation finds $\delta = 0.358$ and $\varphi = 10.29$, giving an error of 1.7% for throughput and 0.3% for flow time. Increasing the number of buffer spaces on all workstations to 5 leads to $\delta = 0.565$ and $\varphi = 25.06$ for the original model compared to $\delta = 0.565$ and $\varphi = 25.05$ for the approximation. This is an error of less than 0.1% for both throughput and flow time. These results correspond to the observations of section 4.

Implications

Two main observations can be derived from the examples presented here. First, the measured EPT parameters comply with the analytically calculated parameters. Secondly, adding detail to the ELP model, by using product type specific EPT-distributions, results in more accurate approximations.

6 Industrial case

A case from an automotive manufacturing plant will be used to illustrate the practical applicability of the EPT-approach.

System description

Experimental data has been obtained from one of the clients of Steelweld B.V. This particular client produces two types of cars, called pt_o and pt_i in the remainder of this section. Focus is on a small semi-synchronous flowline within the manufacturing plant. On this flowline, referred to as FL in the remainder of this section, lots are produced according to a constant product mix, i.e. $pt_o/(pt_o + pt_i) = 0.57$. The actual sequence of lots is determined by an overhead scheduler. Since the scheduler is not considered in this case, the stream of lots entering the system will have a random lot type sequence.

FL consists of a transport system and eleven workstations in tandem (i.e. sequential). The workstations are labelled WS_o to WS_{10} . Here, WS_1 and WS_2 are manual workstations, served by one operator. Workstations WS_7 and WS_8 are single buffer spaces. Workstation WS_{10} is used for (occasional) manual quality checks. All other workstations in the line are used for hotmelting.

First stage of the EPT-approach

The event data needed for the EPT analysis is obtained from the programmable logic controllers (PLCs) within FL . In their present configuration, only possible departures and actual arrivals can be measured using the PLCs; the actual departures thus would have to be reconstructed. However, since the workstations can contain at most one lot at a time, one knows that $AA_{i,j}$ will always exceed $AD_{i-1,j}$, hence $AD_{i-1,j}$ is not required for determining EPT realisations. However, $AD_{i,j}$ should be known on the last workstation so that flowtimes can be computed for validation.

The actual arrival occurs only after transport from the sending workstation to the receiving workstation has ended. Therefore, if the logged actual arrival and possible departure are used, transport is excluded from the EPT realisation. However, the workcycle of these unbuffered workstations always begins with transport. Therefore, the actual arrival should be adapted so that the EPT realisation will include transport. Transport takes a fixed, known amount of time Δ_{min} , the value of which will not be reported here for reasons of confidentiality. By decreasing $AA_{i,j}$ with Δ_{min} , transport is included in the EPT.

No data was available for WS_7 and WS_8 . Therefore, WS_5 is the last workstation on which actual departures can be computed. Hence, workstations WS_6 and above will not be studied in the case.

Since not all gathered events are useable, the data must be filtered. First of all, a number of the events result in EPT-realizations that are unrealistically low or even negative if either possible or actual arrivals are registered too late. Furthermore, since the machines are reliable, large EPT-realizations due to failures and repairs only occur sporadically. Since only a few of these realizations occur within the considered time period, no reliable statistics concerning these high realizations can be obtained. The EPT realisation for lot i on workstation j is thus only used during the analysis if it satisfies Equation (12), hence machine failures are excluded.

$$\Delta_{min^*j} \leq \text{EPT}_{ij} \leq \Delta_{max^*j} \quad (12)$$

Second stage of the EPT-approach

Distribution fitting, the second stage of the EPT-approach, is done by computing the values for Δ_e , t_e and c_e^2 per workstation from the obtained filtered EPT realizations, as presented in Table 4. The data in Table 4 have been slightly rescaled, in order to respect the confidentiality of the data. Based on this data, shifted Gamma distributions were fitted for all workstations.

WS_i	t_e	c_e^2	Δ_e	$\frac{t_e}{\Delta_e}$
WS_0	82.73	0.106	57.19	1.45
WS_1	76.78	1.259	27.61	2.78
WS_2	94.32	0.765	19.72	4.78
WS_3	116.61	0.149	90.71	1.29
WS_4	112.09	0.077	78.88	1.42
WS_5	130.82	0.021	114.37	1.14

Table 4: Fitted distributions for the industrial case

Third stage of the EPT–approach

In the third stage, the shifted Gamma distributions with parameters as presented in Table 4 are used as input for an ELP model, a discrete event simulation model in this case. The structure of the model is identical to the structure of FL , i.e., six unbuffered single server workstations in a flowline.

A distribution capturing the starvation observed on the first workstation has been obtained from the data to model the starvation of the first workstation in the flowline. In order to obtain this starvation distribution, a filter similar to Equation (12) has been applied. The starvation distribution has properties $t_s = 63.63$, $c_s^2 = 2.564$ and $\Delta_s = 29.58$. If it is starving, the first workstation requests a lot from the generator. The generator sends a lot on to the first workstation after an appropriate period of starvation. Similarly, for the final workstation in the flowline, a distribution capturing the observed blocking is obtained. The parameters of this blocking distribution are $t_{b,5} = 15.10$, $c_{b,5}^2 = 8.04$ and $\Delta_{b,5} = 1.97$.

The true mean flow time φ of FL is determined by computing the individual flow times from the obtained data and deleting the unrealistic flow times. Flow time realisations are thus again filtered using a filter similar to Equation (12). Due to filtering, some EPT–realisations are discarded during data analysis. Consequently, the mean throughput cannot be computed as the amount of bodies produced during the measured time period. Instead, mean throughput δ will be computed by determining the mean interdeparture time of bodies on workstation WS_0 .

The ELP model underestimates the throughput δ by less than 1.0%, whereas the flow time φ is overestimated by 3.7% (simulation results presented in this section have a confidence level of 99% and a relative width of less than 0.1% of the mean). As can be seen, only a small error remains in the approximation. This error can partially be explained using the inter- and intra-correlations of workstations, as was presented in Example II of Section 5.

To improve on this, type specific EPT–distributions can be fitted, as presented in Table 5. The new distributions are used in the ELP model. The model now overestimates both δ and φ by less than 1.0%. By adding more detail, the approximations have become more accurate.

WS_i	t_e	pt_0			$\frac{t_e}{\Delta_e}$	pt_1		
		c_e^2	Δ_e	c_e^2		Δ_e	$\frac{t_e}{\Delta_e}$	
WS_0	86.01	0.139	59.16	1.45	78.42	0.041	57.19	1.37
WS_1	40.46	1.400	27.61	1.47	127.52	0.362	67.05	1.90
WS_2	138.68	0.157	86.76	1.60	38.08	0.308	19.72	1.93
WS_3	112.59	0.243	90.71	1.24	121.87	0.036	92.68	1.31
WS_4	105.67	0.021	78.88	1.34	121.89	0.119	110.43	1.10
WS_5	134.26	0.016	120.29	1.12	126.30	0.023	114.37	1.10

Table 5: Type specific fitted distributions for the industrial case

Fourth stage of the EPT–approach

A bottleneck analysis is performed, after which the suggested improvements are simulated by accordingly changing the EPT–distributions. It is used to determine which workstations are the major restrictions on throughput and flow time. Workstations with high t_e or c_e^2 are potential bottlenecks since they may cause starvation or blocking.

Using the information of Table 5, one can see that the values of t_e range from 38.08 to 138.68. Out of this range, acceptable values of t_e seem to lie between 100 and 125 seconds (although lower values are obviously desirable). Therefore, parameters $t_{e_{1,1}}$, $t_{e_{2,0}}$, $t_{e_{5,0}}$ and $t_{e_{5,1}}$ are reduced to 125.00 seconds

Furthermore, Table 5 illustrates that for most situations, $c_e^2 < 0.25$. Reduction of $c_{e_{1,1}}^2$ and $c_{e_{2,1}}^2$ to 0.25 is assumed to be feasible, whereas it is assumed that $c_{e_{1,0}}^2$ can be reduced to 0.75.

The suggested changes have been implemented in the ELP model. Implementation of these changes would, according to the ELP model result in an increase of 3.5% in δ and a decrease of 4.0% in φ . The simulation study with the unscaled data predicted improvements of the same order of magnitude; which was further confirmed (for the throughput) during implementation on the factory floor; the flow time was not studied during implementation.

7 Conclusions

A new method for performance analysis and lumped parameter modelling of single server flowlines subject to blocking has been proposed. The method is based on the effective process time (EPT). In previous work, EPT has only been considered for infinitely buffered, isolated workstations. Here, a calculation method for EPT-realizations for single server flowlines subject to blocking has been presented and validated. The method translates event data (actual and possible arrivals and departures of lots) into EPT–realisations using sample-path like equations.

The EPT of a lot is the time experienced by the lot on a workstation from a logistical perspective. It is implemented by means of an approach consisting of four stages, the so–called EPT–approach. In the first stage, EPT–realisations are gathered from industrial data. Next, the realisations are translated into distributions. Typically, distributions are fitted using the first two moments (t_e , c_e^2). Simulation results however show that for flowlines subject to blocking the offset Δ_e should be used as an additional distribution parameter. In the third stage, an ELP model can be built and validated. Finally, in the fourth stage, the flowline can be optimized.

The EPT–approach has been applied to a case study taken from automotive industry. The ELP model accurately estimated both throughput and flow time. Adding more detail to the ELP model (i.e., including product type specific shifted Gamma distributions) further reduced errors to less than 1.0%. Based on the EPT–approach, changes in t_e and c_e^2 were proposed to increase throughput and to decrease flow time.

Acknowledgment

The authors would like to thank Onno Boxma, Johan Jacobs, Ton de Kok, Erjen Lefeber, Marco Vijfvinkel, Marcel van Vuuren, Sven Weber, and Steelweld B.V. for their assistance.

Bibliography

- [1] J.A. Buzacott and J.G. Shanthikumar. *Stochastic models of manufacturing systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1 edition, 1993.
- [2] L. Chen and C-L. Chen. A fast simulation approach for tandem queueing series. In O. Balci, R.P. Sadowski, and R.E. Nance, editors, *Proceedings of the 1990 Winter Simulation Conference*, pages 539–546, 1990.
- [3] G.L. Curry, B.A. Peters, and M. Lee. Queueing network model for a class of material-handling systems. *International Journal of Production Research*, 41:3901–3920, 2003.
- [4] Y. Dallery and S.B. Gershwin. Manufacturing flow line systems: a review of models and analytical results. *Queueing Systems: Theory and Applications*, 12:3–94, 1992.
- [5] W.J. Hopp and M.L. Spearman. *Factory physics: foundations of manufacturing management*. London: Irwin McGraw-Hill, 2nd edition, 2001.
- [6] W.J. Hopp, M.L. Spearman, S. Chayet, K. Donohue, and E.S. Gel. Using an optimized queueing network model to support wafer fab design. *IIE Transactions*, 34:119–130, 2002.
- [7] J.H. Jacobs, L.F.P. Etman, E.J.J. van Campen, and J.E. Rooda. Quantifying operational time variability: the missing parameter for cycle time reduction. In *2001 IEEE/SEMI Advanced semiconductor manufacturing conference*, pages 1–10, 2001.
- [8] J.H. Jacobs, L.F.P. Etman, E.J.J. van Campen, and J.E. Rooda. Characterization of operational time variability using effective process time. *IEEE Transactions on Semiconductor Manufacturing*, 16:511–520, 2003.
- [9] D.S. Kim and J.M. Alden. Estimating the distribution and variance of time to produce a fixed lot size given deterministic processing times and random downtimes. *International Journal of Production Research*, 35:3405–3414, 1997.
- [10] A.A.A. Kock. Performance evaluation and simulation meta modelling of single server flow lines subject to blocking: an effective process time approach. Master’s thesis, Eindhoven University of Technology, Department of Mechanical Engineering, Systems Engineering Group, December 2003. SE 420367-A.
- [11] M.D. Rossetti and G.M. Clark. Estimating operation times from machine center arrival and departure events. *Computers and Industrial Engineering*, 33:493–514, 2003.