CrossMark

Modelling and analysis of the impact of correlated inter-event data on production control using Markovian arrival processes

Nima Manafzadeh Dizbin[1] · Barış Tan[2]

Abstract

Empirical studies show that the inter-event times of a production system are correlated. However, most of the analytical studies for the analysis and control of production systems ignore correlation. In this study, we show that real-time data collected from a manufacturing system can be used to build a Markovian arrival processes (MAP) model that captures correlation in inter-event times. The obtained MAP model can then be used to control production in an effective way. We first present a comprehensive review on MAP modeling and MAP fitting methods applicable to manufacturing systems. Then we present results on the effectiveness of these fitting methods and discuss how the collected inter-event data can be used to represent the flow dynamics of a production system accurately. In order to study the impact of capturing the flow dynamics accurately on the performance of a production control system, we analyze a manufacturing system that is controlled by using a base-stock policy. We study the impact of correlation in inter-event times on the optimal base-stock level of the system numerically by employing the structural properties of the MAP. We show that ignoring correlated arrival or service process can lead to overestimation of the optimal base-stock level for negatively correlated processes, and underestimation for the positively correlated processes. We conclude that MAPs can be used to develop data-driven models and control manufacturing systems more effectively by using shop-floor inter-event data.

✉ Barış Tan
btan@ku.edu.tr

Nima Manafzadeh Dizbin
ndizbin14@ku.edu.tr

[1] Graduate School of Business, Koç University, 34450 Rumeli Feneri Yolu, Istanbul, Turkey

[2] College of Administrative Sciences and Economics, College of Engineering, Koç University, 34450 Rumeli Feneri Yolu, Istanbul, Turkey

✷ Springer

1 Introduction

Technological advances allow manufacturers to access and collect data from the shop floor more easily and effectively. Extensive data collected from the shop floor can be analyzed to develop decision support systems for integrated product-production simulation, problem identification, and production control. As a result, data-driven modeling and control methods are now considered as enabling technologies to address technology challenges for implementing factory of the future (IEC 2015).

Empirical studies show that the inter-departure times of a production system demonstrate significant correlation (Schomig and Mittler 1995; Inman 1999). In addition, analysis of the output dynamics of production system with i.i.d. service time distribution demonstrates that the output process from a production system can be correlated (Hendricks and McClain 1993; Tan and Lagershausen 2017). Since the output from a production system is an input to another one, the inter-arrival times of work centers also exhibit autocorrelation. However, most of the studies in the literature make simplifying assumptions for the inter-event distributions. More specifically, they assume independence of the inter-event times in order to achieve analytical tractability. However, these simplifying assumptions may increase the risk of model misspecification that leads to errors in setting control parameters.

In this paper, we focus on the inter-event data such as the inter-arrival, inter-departure, and service completion times that can be collected from the shop floor. Our objective is to answer the following questions: *how can the collected inter-event data be used to represent the statistical properties of the flow dynamics of a production system accurately* and *what is the benefit of capturing flow dynamics accurately on the performance of a production control system?*

This paper addresses the first problem by presenting a modeling framework that faithfully captures the statistical properties of the shop-floor inter-event data. For example, Fig. 1 and Table 1 show the empirical inter-arrival time distribution, and the autocorrelation structure of the data collected from a specific equipment at the Robert Bosch Reutlingen semiconductor manufacturing plant and the fitted distribution, and the autocorrelations. As the figure shows, the inter-arrival times have significant autocorrelations. The modeling framework presented in this study that yields the fitted distribution and the autocorrelation structure shown in the figure allows developing an analytical model that captures these statistical properties for the production processes. These models can then be used to evaluate the performance by using analytical techniques and simulation and also to control the system.

We adopt Markovian arrival processeses (MAP) to model the inter-event times. MAPs are point processes that can approximate any inter-event process arbitrarily close enough (Asmussen and Koole 1993). There exist numerous MAP fitting methods in the literature of telecommunications systems that can be exploited to fit a MAP for the collected inter-event data. This approach yields accurate models that can generate flow dynamics that is statistically very close to the collected
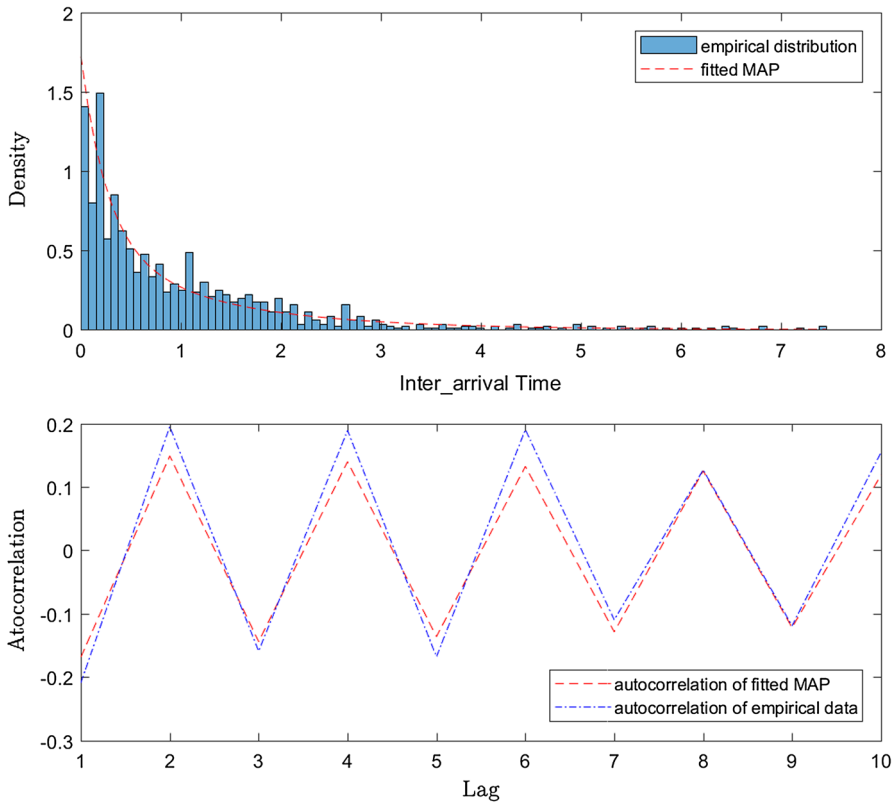
Fig. 1 The observed and fitted distribution and autocorrelations (ρ_k) of inter-arrival time of a specific equipment at the Robert Bosch Reutlingen semiconductor manufacturing plant

Table 1 Statistical descriptors of the inter-arrival time of an equipment and its fitted MAP

	Moments			Autocorrelation (lag)				
	1	2	3	1	2	3	4	5
Real	1.00	2.37	8.87	− 0.2088	0.1950	− 0.1593	0.1902	− 0.1690
MÂP (Horváth and Telek 2017)	1.00	2.65	11.78	− 0.1684	0.1488	− 0.1439	0.1398	− 0.1360

data. These models can be used in process visualization as building blocks for analytical and numerical simulation. Since the inter-event data can be collected for a single machine, for a group of machines, or for a production area, this approach can also be used to develop aggregate models for planning. The MAP representation of a production system can also be used to determine the parameters of the control policies for a given production system.

We first present a comprehensive review on MAP modeling and MAP fitting methods applicable to manufacturing systems and present results on the effectiveness of these fitting methods to provide an answer to the first research question on *how* the collected inter-event data can be used to represent the flow dynamics of a production system accurately.

In order to answer the second research question on the *impact* of capturing the flow dynamics accurately on the performance of a production control system, we then focus our analysis to the control of a manufacturing system that uses a single-level base-stock policy. This policy is fully specified with a single threshold, referred as the base-stock level where production stops when inventory hits the threshold. This policy (or its alternative representations) is commonly implemented in practice.

We utilize the MAP models to generate flow dynamics with the desired distribution and autocorrelation structure analytically. We investigate how representing the inter-event time with its first and second moments, with its distribution, and with the complete autocorrelation structure affects the determination of the base-stock level and the performance of the system that uses the base-stock level determined by these different representations.

We use the structural properties of MAPs to generate models that approximate the inter-event time process with i.i.d inter-event times that have exponential distribution which uses the first moment, acyclic phase-type (APH) distribution with the same first-two moment of the correlated process, and phase-type (PH) distribution of the correlated process. For each representation, we use the structural properties of MAPs to generate the processes and calculate the optimal base-stock level by using matrix analytic methods.

We demonstrate that capturing the inter-arrival time distribution more accurately while ignoring the autocorrelations does not improve the accuracy of the results. For example, modeling the correlated arrival process with a phase-type distribution does not necessarily result in a more accurate estimation of the optimal control parameter. In this case, modeling a positively correlated arrival with a coefficient of variation less than one, with an exponential distribution may result in better estimates of the base-stock level than using a phase-type distribution that matches the distribution of the observed inter-event times.

The MAP representation also allows us to investigate the effects of autocorrelation on the performance of a production system controlled with a base-stock policy. We demonstrate the effectiveness of employing MAP in modeling shop-floor inter-event data by evaluating its performance in estimating the optimal base-stock level. We simulate inter-arrival times from a given process. Then, we use the simplest MAP fitting method, which generates a MAP with the same first-two moments and first-lag autocorrelation of data, to fit a MAP into the simulated data. We adopt the fitted MAP to estimate the optimal base-stock level of the system. We compare its performance in fitting base-stock level with that of the original correlated process, and exact marginal distribution of the correlated process. Our analysis shows that using MAP fitting methods allows setting the base-stock level in a way that performs at least as good as the performance obtained by using the exact marginal distribution. Employing more sophisticated fitting methods can result in more accurate estimated optimal base-stock levels.

This study contributes to the literature in three areas. First, we present a comprehensive review of MAPs and the existing MAP fitting methods to build MAPs from inter-event shop-floor data for the manufacturing systems literature. Second, we present results that show the impact of correlation in inter-event times on the production control of a system for the first time in the literature. We show that ignoring autocorrelation of a correlated arrival process results in setting the base-stock level at a higher or lower level. Finally, we present the effectiveness of the existing MAP fitting methods as they are used to build MAP models from the inter-event time data and then used to determine the base-stock levels. As a summary, this study shows that MAPs can be used to develop data-driven models and control manufacturing systems more effectively by using shop-floor inter-event data.

The rest of this paper is organized as follows: in Sect. 2, we review the literature of presence of correlation, impact of correlation on performance evaluation and MAP fitting methods. In Sect. 3, we introduce the model we use for our evaluation and present the methodology that we incorporate in the numerical experiments. In Sect. 4, we evaluate the impact of correlation in inter-arrival times on the optimal base-stock level, and performance of the renewal estimation of a correlated process in setting the base-stock level. We evaluate the performance of modeling a correlated process by means of commonly adopted distributions ignoring autocorrelation in Sect. 5. In Sect. 6, we evaluate the performance of MAP fitting methods that are used to fit a MAP model to real, and simulated inter-event time data and then to set the base-stock level. We summarize our findings from these numerical experiments in Sect. 7. Finally, we conclude our study and state possible future research directions in Sect. 8.

2 Literature survey

We divide our discussion of how this work fits into the existing research literature into two areas. First area is related to papers that demonstrate the presence of autocorrelation in manufacturing systems, and evaluate the impact of autocorrelation on performance measures of the system. Second area is related to the existing MAP fitting methods.

2.1 Presence of autocorrelation in inter-event times of manufacturing systems

Schomig and Mittler (1995) analyze the cycle time of semiconductor manufacturing systems and show that the cycle time is highly correlated. Inman (1999) demonstrates the presence of autocorrelation in the output data of some manufacturing stations in automotive industry. Altiok and Melamed (2001) present data depicting the presence of significant empirical first lag autocorrelation in times between machine alignment (registration) failures, aviation equipment times to failure, and inter-arrival times of packaged food items. Tan and Lagershausen (2017) present evidence of correlated inter-departure times of cars leaving an assembly line of an automotive manufacturer. Hendricks and McClain (1993) demonstrate that

the output process of a production line with i.i.d. arrival and service processes can be correlated. Tan and Lagershausen (2017) propose an analytical method to calculate the inter-departure time correlation analytically and discuss how buffer capacity and process variability affect autocorrelation structure.

2.2 Impact of correlation on the performance of queuing and manufacturing systems

The impact of correlation on the performance of a system has been investigated by using simulation and by using analytical methods in the literature. Simulation studies introduce correlation to queues by methods like Transform Expand Sample and Minification (Livny et al. 1993), vector-auto-regressive-to-anything (Biller and Nelson 2003) and Markovian arrival processeses. On the other hand, analytical studies use Markov Renewal Processes, Markovian arrival processeses, and Supplementary Variables to take dependence into account (Distefano and Trivedi 2013).

2.2.1 Simulation studies

Livny et al. (1993) demonstrate that autocorrelation in service or inter-arrival times may severely affect the queue lengths and consequently waiting time distribution of a system. Takahashi and Nakamura (1998) analyze a production system controlled by Kanban cards with concurrent and sequential ordering processes. They demonstrate that correlated demand arrivals can have significant impact on the WIP inventories and on the expected waiting time of the products. Altiok and Melamed (2001) study the effects of correlated job arrivals in an M/M/1 queue, the effects of correlated production process in an M/G/1 queue, and in a pull-type production system. Resnick and Samorodnitsky (1997) study the impact of arrival dependence on the performance measures of G/M/1 queues. They demonstrate that ignoring long range dependence can significantly alter behavior of the system. Dahl and Willemain (2001) analyze the impact of long-memory arrivals on the performance measures of queuing systems, such as utilization and expected number of customers in the system. Resnick and Samorodnitsky (1997) implement a similar study, stating positive autocorrelation among inter-arrival times increases the waiting time and queue size of the process. Civelek et al. (2009) study the impact of dependence in a single server queue in different scenarios and conclude that positively correlated arrivals increase the expected waiting time of a customer in the system. Brickner et al. (2010) conduct a simulation study with MAP arrival, PH service with infinitely many servers and finite buffer capacities. They show that positively correlated arrivals spent more expected time in the system. Pereira et al. (2012) evaluate the impact of correlation on assembly processes, serial lines and disassembly processes. They demonstrate that considering autocorrelation in modeling the process enhances the accuracy of estimated performance measures of the real system.

2.2.2 Analytical approaches

Runnenburg (1961, 1962) study the effect of dependence on the expected waiting time of a system with integer Markov dependent arrival intervals and independent exponential service. Hadidi (1985) studies the impact of positive correlation on the waiting time distribution of a M/M/1 queue. He shows that by increasing the correlation, the pace of convergence of waiting time distribution to its liming value increases. Patuwo (1989) is an extensive numerical study of a correlated arrival process which is represented by a two-state Markov Renewal process. Patuwo et al. (1993) investigate the impact of correlation on mean queue length of queues with correlated Markov renewal arrival process, and exponential service time (MR/M/1). They show that mean queue length of a system with correlated arrival, independent of the arrival distribution, can be over 30 times greater than renewal processes' queue length. Szekli et al. (1994a, b) study the the impact of positively correlated arrival times on the queuing measures of a MR/GI/1 queues. They demonstrate that the higher correlation in arrival streams may result in more variability of the waiting times and higher mean queue lengths. Bauerle (1997) study the effect of the transition matrix of the environmental process on the waiting time of the nth customer and on the stationary waiting time. They generalize the results of Szekli et al. (1994b) and state that the more dependency in the arrival process, the larger is the stationary waiting time with respect to the increasing convex order. Xu (1999) studies the correlation among the jobs at separate facilities and evaluate the effect of correlation on a variety of system performance measures such as queue length. Adan and Kulkarni (2003) study a single server queue with MAP inter-arrival and generally distributed service times with a cross-correlation between arrival and service process. They evaluate the impact of autocorrelation and cross-correlation on the mean waiting time. Hwang and Sohraby (2003) consider a discrete time queuing system with two-state discrete time Markov modulated batch arrival with autoregressive input. They demonstrate that the mean queue length of the processes is quite different in correlated arrival processes.

Our study differs from the existing literature in the sense that we evaluate the impact of autocorrelation on the optimal control of a production system. In particular, we evaluate the impact of correlation in inter-arrival or service times in a make-to-stock system controlled by a base-stock policy. The influence of demand variability on the performance of a make-to-stock systems has already been investigated (Jemai and Karaesmen 2005). However, the impact of correlation in demand or service times on the optimal control of the system is not known.

2.3 MAP fitting methods

MAP estimation is a quite new research topic such that most of the available methods have been developed during the last decade. In this section, we review the major MAP fitting approaches. We exclude papers fitting special structures of MAP, such as MMPP or MAPs of order two. The interested reader can refer to Gerhardt and

Table 2 Classification of MAP fitting methods

Paper	Fitting approach	Input data	Output process	Algorithm specification
Horváth et al. (2005)	Moment based	Inter-event time	MAP	Autocorrelation fitting
Bause et al. (2010)	Moment based	Inter-event time	MAP	Joint moment fitting
Casale et al. (2010)	Moment based	Inter-event time	MAP	Kronecker product composition
Telek and Horváth (2007)	Moment based	Inter-event time	MAP	Minimal representation
Horváth (2013)	Moment based	Inter-event time	MAP	–
Buchholz (2003)	Maximum likelihood	Inter-event time	MAP	Uniformization technique
Breuer (2002)	Maximum likelihood	Inter-event time	Batch MAP	Uniformization technique
Klemm et al. (2002)	Maximum likelihood	Inter-event time	Batch MAP	–
Kriege and Buchholz (2014)	Maximum likelihood	Inter-event time	MAP	Aggregated data
Buchholz and Panchenko (2004)	Maximum likelihood	Inter-event time	MAP	Faster convergence
Okamura et al. (2013)	Maximum likelihood	Inter-event time	MAP	Deterministic annealing
Horváth and Okamura (2013)	Maximum likelihood	Inter-event time	Marked MAP	–
Okamura et al. (2009)	Maximum likelihood	Group Data	MAP	–
Breuer and Kume (2010)	Maximum likelihood	Group Data	MAP	–
Buchholz et al. (2010)	Moment based	Inter-event time	Marked MAP	–

Nelson (2009) and Buchholz et al. (2014) for a comprehensive review of those methods.

There are two types of MAP fitting methods in general: moment-based, and maximum-likelihood-based fitting methods. Table 2 gives a classification of the MAP fitting methods discussed in this section.

The moment-based approaches employ statistical properties of the trace in different stages. Two of the main methods of this approach estimate the associated PH distribution at first stage. Then, they introduce the empirical dependence into the PH distribution. Horváth et al. (2005) propose the first moment-based fitting method. They construct a nonlinear optimization problem that matches the theoretical autocorrelation of the MAP into the empirical autocorrelation structure of the observed data. This method gives an exact solution to the first-lag autocorrelation if the parameters are within an acceptable range. Bause et al. (2010) employ the joint moments of the empirical data to construct a nonlinear optimization problem with linear constraints. The objective function of this problem is minimizing the squared difference between the empirical and theoretical joint moments. Casale et al. (2010) determine the most important statistical descriptors by conducting a sensitivity analysis over MAP/M/1 queue. Then, they construct MAPs of order two with the same coefficient of variation and autocorrelations as empirical data. They employ Kronecker product to build MAPs with higher orders. Telek and Horváth (2007) fit a MAP with a specified number of parameters. They utilize the minimal representation of the MAP to determine the minimum number of parameters. Horváth (2013) suggest an algorithm that fits a MAP to the mean, coefficient of variation and the first lag autocorrelation of the data. Their method fits the third-moment and takes the autocorrelation decay parameter into account if the parameters are within an acceptable range.

The second group of the fitting methods are the maximum-likelihood-based approaches. They employ Expectation Maximization (EM) algorithms to increase the maximum likelihood of the fitted MAP at each iteration of the algorithm. EM algorithms were first applied to PH and Markov Modulated Poisson Processes. Buchholz (2003) extends application of the EM algorithm on PH distribution and hidden Markov models into MAP. He applies the uniformization technique into EM algorithm to reduce its time complexity. The method generates acceptable estimations for trace lengths of some thousands elements. Klemm et al. (2002) and Breuer (2002) provide similar studies that applies EM algorithm for parameter fitting of batch MAP (BMAP) process. Kriege and Buchholz (2014) improve the convergence speed of the EM algorithm by employing aggregated data. The main advantage of the method over former EM algorithms is truncating data and preserving the statistical properties at the same time. Buchholz and Panchenko (2004) improve the EM algorithm for MAP fitting by adopting a two step algorithm. The first step of the algorithm fits a PH distribution and the second step captures the empirical auto-covariance. Okamura and Dohi (2009) propose uniformization based EM algorithm to improve the time complexity of the existing algorithms. Okamura et al. (2013) provide a deterministic annealing PH and MAP fitting algorithm that deals with local maxima problem of the conventional algorithms. Horváth and Okamura (2013) extend the EM algorithm for fitting MAPs with different types of arrivals or MMAPs. Buchholz and Kriege (2017) exploit both EM and

moment fitting methods to fit processes with serial and cross correlation at the same time.

Okamura et al. (2009) extend the EM algorithm for fitting a MAP into a group data. The method is useful for situations where the collected data contains the number of arrivals in a given time interval instead of the inter-arrival times. Their method is computationally more expensive than others. However, it contains other EM approaches as its subclass. Breuer and Kume (2010) propose an EM algorithm that fits an empirical counting process, observed at discrete times, to a MAP. The input data for this algorithm are the numbers of observed events in disjoint time intervals as well.

In order to give an example to the performance of fitting methods, we fit MAPs by using different fitting algorithms to the real data of processing, cycle, and inter-arrival time collected from three different equipments at Robert Bosch Reutlingen plant. Figures 1, 7, 8 show the observed and a fitted distribution and autocorrelation for these processes.

3 Model and evaluation methodology

We now present the basic model and the evaluation methodology used to answer the main research questions: *how can the collected inter-event data be used to represent the flow dynamics of a production system accurately* and *what is the benefit of capturing flow dynamics accurately on the performance of a production control system?*

3.1 Markovian arrival processeses

We model the correlated demand arrival or service processes as MAPs (Neuts 1979). MAPs are generalization of PH distributions (Neuts 1975) that can capture correlated inter-event times. They contain most of the commonly used arrival processes such as Erlang processes, Coxian distributions and Markov-modulated Poisson processes (MMPP) as subclasses. A MAP consists of two different sub-processes each of which has a discrete state space called phases. One sub-process represents the dynamics of the phase process denoted by D_0, i.e transition between phases without an event, while the other corresponds to the occurrence of an event denoted by D_1. A MAP can be interpreted as a continuous-time Markov-chain with the generator matrix $D = D_0 + D_1$, and $|D|$ states. Let D be an irreducible generator matrix with initial probability vector β_0. The Markov chain starts at state i with probability $\beta_0(i)$, spends an exponential time with rate $\lambda_i = -D_0(i, i)$ there, and moves to state j with probability p_{ij} defined as:

$$p_{ij} = \begin{cases} \frac{D_1(i,i)}{\lambda_i}, & j = i \\ \frac{D_0(i,j) + D_1(i,j)}{\lambda_i}, & j \neq i \end{cases} \tag{1}$$

when the Markov chain experiences a state transition from state i to j, an arrival occurs with probability $\frac{D_1(i,j)}{\lambda_i}$. Let $P = -(D_0)^{-1} D_1$ and $\beta P = \beta$ and $\beta \mathbb{1} = 1$ where $\mathbb{1}$ is

vector of ones with appropriate size. The phase distribution at arrival instants form a discrete time Markov chain with transition probability matrix P. A MAP can be fully specified with D_0 and D_1 if we let $\beta_0 = \beta$. We denote a MAP with these sub-processes with $\text{MAP}(D_0, D_1)$. For further detailed information regarding, distribution, moments, autocorrelation structure, and other features of MAPs the reader is referred to He (2013), Buchholz et al. (2014) and Lakatos et al. (2013).

3.1.1 MAP representation of the output process of a production system

We will represent the output process of the production systems by differentiating between transitions that lead to departure of a product and the rest of transitions. By capturing the transitions that do not lead to departure in matrix D_0, and transitions that lead to departure in matrix D_1, we represent the output process as $\text{MAP}(D_0, D_1)$.

For instance, the MAP representation of a two-station production line with exponential service times with rates μ_1 and μ_2, respectively, and no inter-station buffer is given below. The state of this system consists of a tuple (s_1, s_2) where s_i demonstrates the state of machine i, $s_i \in \{0, 1\}$ where $s_i = 1$ represent that machine i is working, $s_1 = 0$ represents machine 1 is blocked and $s_2 = 0$ represents machine 2 is idle. The D_0, and D_1 matrices of this line can be written as:

$$D_0 = \begin{bmatrix} -\mu_1 & \mu_1 & 0 \\ 0 & -(\mu_1 + \mu_2) & \mu_1 \\ 0 & 0 & -\mu_2 \end{bmatrix} \qquad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ \mu_2 & 0 & 0 \\ 0 & \mu_2 & 0 \end{bmatrix}$$

where the states are ordered as (1, 0), (1, 1), (0, 1). The D_1 matrix captures transitions that are related to process completions on machine 2 that generates an output from the production line. Tan and Lagershausen (2017) use this approach to determine the autocorrelation structure of the output process from open and closed queuing networks subject to blocking. The output process of the two-station production line with no inter-station buffers with the $\text{MAP}(D_0, D_1)$ representation given above demonstrates a negative first-lag autocorrelation.

3.2 Base-stock model

We consider a production/inventory system with correlated demand inter-arrival and service times that are modeled as MAPs. We consider demand arrivals generated by the output of a different production system and therefore they can exhibit positive or negative autocorrelation. An arriving demand will be satisfied from the inventory according to the first-come-first-served (FIFO) rule if a product is available in the inventory. Otherwise, if the inventory is empty, it will be backlogged until it is satisfied. We will assume that the production is controlled by a base-stock policy. Under this policy, the producer produces when there is an outstanding production order i.e., inventory is under a given threshold. We employ this control policy since it (or its equivalent representations) is commonly implemented control policy in practice. However, it is

not necessarily the optimal control policy in this setting. Dizbin and Tan (2017) state that the optimal policy to control a production/inventory system with correlated arrival and service times is a state-dependent base-stock policy.

We assume that raw materials are supplied from an unlimited stock with zero lead-time and the system is continuously reviewed. The cost structure of the system consists of inventory holding and backlog costs. The inventory holding cost is h per unit per unit of time and the backlog cost is b per unit per unit of time. The objective of the problem is minimizing the long run average cost of the system.

3.2.1 Optimal base-stock level of a system with MAP arrival and service times

In this section, we demonstrate how to calculate the optimal base-stock level of a system with correlated arrival and service processes modeled as MAP. Correlated arrival processes can be an output process from an earlier stage of the production line. Correlated service can be the result of production time variations of the machines. Let S be the base-stock level of the system, $O(t)$ be the number of outstanding production orders at time t to reach base-stock level S, $X(t) = S - O(t)$ be net inventory level at time t, and $G_X(S)$ be the stationary distribution of X given base-stock level S.

It is known that the optimal base stock level that minimizes the expected cost under the single threshold policy is given by:

$$S^* = \arg\min_s \left\{ G_X(S) \geq \frac{b}{b+h} \right\}. \tag{2}$$

The $G_X(S)$ of a system with MAP arrival and service processes is equivalent to the queue length distribution of the MAP/MAP/1 queue given in Eq. 8 which is a subclass of Quasi-Birth–Death (QBD) processes.

The generator matrix of a QBD process consists of three types of matrices. Forward matrices (F) capture the transitions with new arrival. Local matrices (L) capture the transition between phases of the process without any arrivals or departure from the system. Backward matrices (B) capture transitions leading to a service completion. These matrices for a system with MAP (D_0, D_1) arrival, and MAP (A_0, A_1) service processes are

$$\begin{aligned} F &= D_1 \otimes I_A \\ B &= I_D \otimes A_1 \\ L &= D_0 \oplus A_0 \\ L_0 &= D_0 \oplus I_A \end{aligned} \tag{3}$$

where $\otimes$, $\oplus$, I_D, and I_A are Kronecker product, Kronecker sum, and square identity matrices with size $|D_0|$ and $|A_0|$, respectively. The generator matrix of a process with MAP arrival and service can be written as follows:

$$Q = \begin{pmatrix} L_0 & F & & \\ B & L & F & \\ & B & L & F \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4}$$

The limiting probability distribution of QBD processes with block matrices F, L, B and level zero local matrix L_0 can be calculated as

$$\pi_n = \pi_0 R^n \tag{5}$$

where R is the solution of quadratic matrix equation

$$F + RL + R^2 B = 0 \tag{6}$$

and π_0 is the solution of the following set of equations:

$$\begin{aligned} \pi_0(L_0 + RB) &= 0, \\ \pi_0(I - R)^{-1} \mathbb{1} &= 1. \end{aligned} \tag{7}$$

There exist a number of computational algorithms for computing the geometric matrix R. We refer the reader to Bini et al. (2006) for a review of the methods and their algorithmic implementations. The stationary distribution of the X as a function of π_0 and R can be written as:

$$G_X(S) = \sum_{i=0}^{S} \pi_0 R^i \mathbb{1}. \tag{8}$$

In our study, we first generate the matrices D_0 and D_1 for the given arrival process and the matrices A_0, A_1 for the given service process with their autocorrelation structures and inter-event time distributions. Then, we generate the block matrices F, L, B, L_0 from D_0, D_1, A_0, A_1 by using Eq. (3). We determine the steady-state distribution for a given base-stock level S by using Eqs. (5), (6), (7), and (8). Finally, we determine the optimal base-stock level by using Eq. (2). Once the optimal base-stock level, S^* is determined, the performance measures can be evaluated from the distribution of $G_X(S^*)$. The expected inventory level ($E[X^+]$), the expected backlog level ($E[X^-]$), and the probability of not having inventory in the system ($Pr[X < 0]$) can be written as:

$$Pr[X < 0] = \sum_{s=S^*+1}^{\infty} \pi_s \mathbb{1} = \pi_0 R^{S^*+1}(I - R)^{-1} \mathbb{1}, \tag{9}$$

$$E[X^-] = \sum_{s=S^*+1}^{\infty} \pi_s(s - S^*)\mathbb{1} = \pi_0 R^{S^*+1}(I - R)^{-2}\mathbb{1}, \tag{10}$$

$$E[X^+] = \sum_{s=0}^{S^*-1} (S^* - s)\pi_s\mathbb{1} = S^*\pi_0(I - R)^{-1}\mathbb{1} - \pi_0(I - R)^{-2}(I - R^{S^*})R\mathbb{1}. \tag{11}$$

3.3 Evaluation methodology

3.3.1 Impact of autocorrelation on the optimal base-stock level

In order to measure the impact of autocorrelation on the optimal base-stock level, we employ processes with the same marginal distribution and different magnitude of autocorrelation. In order to generate processes with the same marginal distribution and different magnitudes of autocorrelation, we use the fact that first-lag autocorrelation of a MAP is a linear function of the elements of D_1 matrix (Horváth et al. 2005). We use processes with the same D_0 (which is associated with the marginal distribution of the process) and different D_1 matrices. We generate D_1^{new} matrices in the following form :

$$D_1^{new} = \theta D_1 + (1 - \theta)D_1^{ren} \tag{12}$$

where $\theta \in [0, 1]$, and D_1^{ren} is the D_1 matrix of a MAP with the same distribution and zero autocorrelations (renewal MAP). The D_1^{ren} of the renewal MAP can be calculated as:

$$D_1^{ren} = D_1\mathbb{1}\beta \tag{13}$$

where β is the phase distribution immediately after an arrival of the MAP. We use MAP(D_0, D_1^{new}) as an arrival process to measure the effect of autocorrelation on the optimal control policy. The first-lag autocorrelation of the MAP(D_0, D_1^{new}) is

$$\rho_1^{new}(\theta) = \theta\rho_1 \tag{14}$$

where ρ_1 is the first-lag autocorrelation of the MAP(D_0, D_1).

As an example, consider the system with two machine and no buffer in between, presented in Sect. 3.1.1, with rates $\mu_1 = \mu_2 = 1.5$. The D_0, D_1, and D_1^{ren} matrices of this system can be written as follows

$$D_0 = \begin{bmatrix} -1.5 & 1.5 & 0 \\ 0 & -3 & 1.5 \\ 0 & 0 & -1.5 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & 1.5 & 0 \end{bmatrix} \quad D_1^{ren} = \begin{bmatrix} 0 & 0 & 0 \\ 0.75 & 0.75 & 0 \\ 0.75 & 0.75 & 0 \end{bmatrix}$$

Note that as we increase the first lag autocorrelation, the magnitude of the higher lags increases as well. In addition, MAPs have a decaying magnitude of autocorrelation

structure (either negative or positive) which is a plausible autocorrelation structure for processing and cycle times of the manufacturing systems setting.

3.3.2 Performance of renewal approximations in estimating correlated processes

In the numerical experiments, we evaluate the performance of commonly adopted distributions in the literature in modeling a correlated arrival process of a system controlled by the single threshold base-stock policy. In particular, we consider the performance of modeling the correlated arrival process with processes that capture the exact marginal distribution, the first-two moments, and the first lag autocorrelation of the correlated process. We evaluate the performance of the system controlled by the threshold calculated by using the approximated processes by comparing their performance measures to the values obtained by using the threshold set by using the original correlated process.The performance measures that we consider are the total cost, the expected inventory and backlog, and the probability of not having an inventory in the system. For brevity, we present an exponentially distributed service and correlated arrival processes. However, the results are similar for the correlated service processes (Dizbin 2016). We calculate the performance measures of a system with the correlated arrival and exponentially distributed service process, by employing the queue length distribution of a MAP/M/1 queue.

We employ PH distribution with the D_0 equal to that of the correlated process and the column vector β to model a correlated process by means of its marginal distribution. A PH distribution can be fully specified with matrix D_0, and entry probability vector β. We employ the queue length distribution of the PH/M/1 queue in order to calculate the optimal base-stock level. Then, we calculate the performance measures of the original system controlled by this base-stock level. We employ a PH distribution with the same first-two moments (PH_2) to model the arrival process by means of it's first-two moments. We employ the method presented in Horváth and Telek (2017) to generate a PH distribution with the first-two moments of the correlated arrival process. We employ the queue length distribution of the PH_2/M/1 to calculate the base-stock level of the system. Finally, we model the correlated arrival process by means of its first moment by using exponential distribution. We employ the queue-length distribution of the M/M/1 queue to calculate the optimal base-stock level.

4 Impact of correlation in inter-arrival times on the control of a production/inventory system

In this section, we analyze the impact of first-lag autocorrelation on optimal base-stock level. We employ two types of processes with positive and negative first-lag autocorrelations. We consider autocorrelation structures with positive and negative first-lag and decaying magnitude of higher lags. We analyze the impact of autocorrelation with different values of coefficient of variation of the arrival and service processes, and the traffic intensity of the system.

Table 3 The parameters used in the numerical experiments

cv_a	{0.2, 0.5, 0.8, 1, 1.3, 1.8}
cv_s	{0.2, 0.5, 0.8, 1, 1.3, 1.8}
v	{0.5, 0.65, 0.8}
ρ_1^+	{0, 0.02, 0.04, …, 0.68, 0.70}
ρ_1^-	{0, − 0.02, − 0.04, …., − 0.48, − 0.50}

4.1 Impact of positive correlation in inter-arrival times on the control of a production/inventory system

In this part, we consider a production system with positively correlated arrival and PH distributed service processes. In order to analyze the effect of autocorrelation on the optimal base-stock level, we design an experiment with different first-lag autocorrelations, coefficient of variation of arrival and service processes, and traffic intensities. Table 3 gives the range of the coefficient of variation of the arrival (cv_a) and service (cv_s) times, first-lag autocorrelation of negatively (ρ_1^-) and positively (ρ_1^+) correlated processes, and traffic intensity (v) used in the numerical experiments. This experiment setup includes $3 \times 6 \times 36 \times 6 = 3888$ different cases analyzed.

Table 4 demonstrates the percentage of the optimal base-stock levels calculated for the correlated processes (S^{corr}) that are equal to the base stock level that is calculated by using the renewal approximation that ignores the autocorrelation ($S^{renewal}$) for different traffic intensities. When the utilization is 0.5, in 42% of the cases, the optimal base-stock levels calculated by using the autocorrelation structure and by using the renewal approximation are the same. However, as the traffic intensity of the system increases, the percentage of the base-stock levels of correlated processes equal to that of the renewal process decreases. When the utilization is 0.8, only in 13% of the cases, the optimal base-stock levels calculated by using the autocorrelation structure and by using the renewal approximation are the same.

Table 5 demonstrates the range of the optimal base-stock levels for correlated processes with their first-lag autocorrelation in the range [0, 0.7] for traffic intensity $v = 0.8$. The base-stock level of a system with $cv_a = cv_s = 0.2$ and zero first-lag autocorrelation is equal to 2. The optimal base-stock level for the same processes and first-lag autocorrelation 0.7 is equal to 10. The optimal base-stock level increases as a function of coefficient of variation of the arrival and service process. This result is in line with the findings of Jemai and Karaesmen (2005) who study the effect demand variability on performance measures of make-to-stock systems.

The range of the optimal base-stock level for processes with no autocorrelation is between 2 and 14. For the correlated processes, the range of the optimal base-stock level is from 2 to 80. This is an indication of significant impact of positive autocorrelation in inter-arrival times on the optimal base-stock levels.

Our numerical analysis shows that the impact of ignoring autocorrelation increases as a function of the first-lag autocorrelation. Figure 2 demonstrates the optimal base-stock level of the system as a function of the first-lag autocorrelation, for some of the cases. In all of the figures, the optimal base-stock level increases as a function of the first-lag autocorrelation. This behavior is due to

Table 4 Accuracy of the renewal estimation in estimating the optimal base-stock level of the correlated system

	% cases: $S^{corr} = S^{renewal}$
$v = 0.50$	41.98
$v = 0.65$	23.07
$v = 0.80$	12.96

the fact that positive autocorrelation increases the probability of having higher queue lengths in the output processes of the system since a short inter-arrival time is expected to be followed by a short inter-arrival time, and similarly a long inter-arrival time is expected to be followed by a long inter-arrival time in positively correlated processes. In other words, the process creates clusters of short and long inter-arrival times. This leads to an increase in the probability of higher queue lengths in comparison with independent inter-arrival times. Figure 3 demonstrates this effect for systems with exponential service time and correlated arrival process with first-lag autocorrelations of 0, 0.1, 0.2, 0.3, 0.5, coefficient of variation equal to 0.5, and traffic intensity 0.8. Note that these processes have the same distribution. The difference in the queue length distribution of these systems is due to the impact of positive autocorrelation.

4.2 Impact of negative correlation in inter-arrival times on the control of a production/inventory system

In this part, we analyze a production system with negatively correlated arrival and PH distributed service processes. Table 3 gives the values of the parameters used in the analysis. Figure 4 demonstrates the impact of negative autocorrelation on the optimal base-stock level for some of the cases. In all of the cases, the optimal base-stock level of the system decreases as a function of the first-lag autocorrelation. In other words, ignoring correlation results in overestimation of the optimal base-stock level of a system with negatively correlated arrival process. This is due to the fact that in negatively correlated processes a short inter-arrival time is expected to be followed by a long inter-arrival time. Therefore, the probability of having higher queue lengths in the system decreases, which results in having lower optimal base-stock level in comparison with the same process with lower magnitude of correlation. Figure 3 demonstrates the effect of negative autocorrelation on the queue length distribution of systems with exponential service time and correlated arrival process with the first-lag autocorrelations of 0, − 0.1, − 0.2, − 0.3, − 0.5, the coefficient of variation equal to 0.5, and the traffic intensity of 0.8. These processes have the same distribution. The difference in their queue length distribution is due to the impact of autocorrelation. We conjecture that queue length distribution of a process with negative autocorrelation, is stochastically dominated by queue length distribution of a processes with the same distribution and lesser magnitude of autocorrelation.

Table 5 Range of the base-stock levels of systems with different coefficient of variation of the arrival (cv_a) and service (cv_s) times ($v = 0.8$, $\rho_1^+ \in [0, 0.7]$)

	cv_s					
	0.2	0.5	0.8	1.0	1.3	1.8
cv_a						
0.2	2–10	3–11	4–12	5–12	6–13	7–15
0.5	4–23	4–24	5–25	6–25	7–26	8–28
0.8	5–35	6–36	7–37	7–38	8–39	10–40
1.0	6–43	6–44	7–45	8–46	9–47	10–48
1.3	7–55	8–56	9–57	9–58	10–59	12–60
1.8	9–75	10–76	11–77	11–78	12–79	14–80

We conclude that main drivers of the optimal base-stock level of a system with correlated arrival (or service) are the autocorrelation structure, coefficient of variation of arrival and service processes, and traffic intensity of the system. Our numerical analysis demonstrate that autocorrelation structure, significantly impacts the optimal base-stock level. The optimal base-stock level increase as the first-lag autocorrelation of the arrival (or service) process becomes more positive.

5 Effects of using different renewal processes to approximate correlated processes on the performance

In this section, we evaluate the performance of modeling a correlated process by using different distributions ignoring autocorrelation. As we saw earlier, ignoring autocorrelation in positively (negatively) correlated processes, underestimates (overestimates) the optimal base stock level of a production system controlled by basestock policy. In this section, we analyze the performance of modeling a correlated process by means of its marginal distribution (PH), first-two moments (PH_2), and first moment (M) in estimating the optimal base-stock level of the system. We consider processes with negative and positive first-lag autocorrelation, and coefficient of variation (cv_a) greater and less than one for the arrival process. For simplicity, we assume that service times are exponentially distributed and traffic intensity is $v = 0.8$. The performance measures that we consider are total cost of the system (TC), expected inventory ($E[X^+]$), expected backlog ($E[X^-]$), and probability of not having inventory in the system ($Pr[X < 0]$). We let the the inventory holding cost, and backlog cost to be $h = 1$, and $b = 5$, respectively.
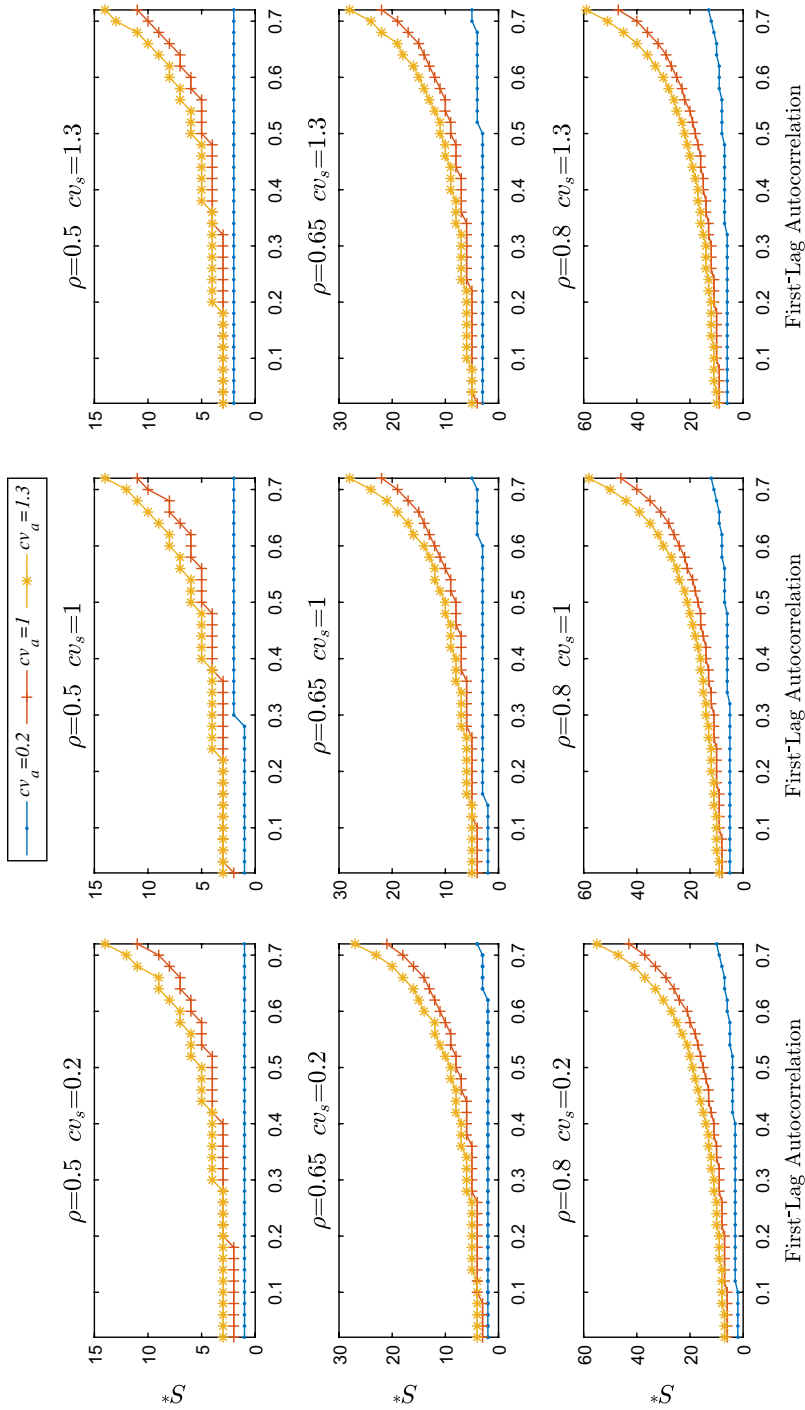
Fig. 2 Impact of the first-lag autocorrelation of a positively correlated arrival process on the optimal base stock levels of a manufacturing system
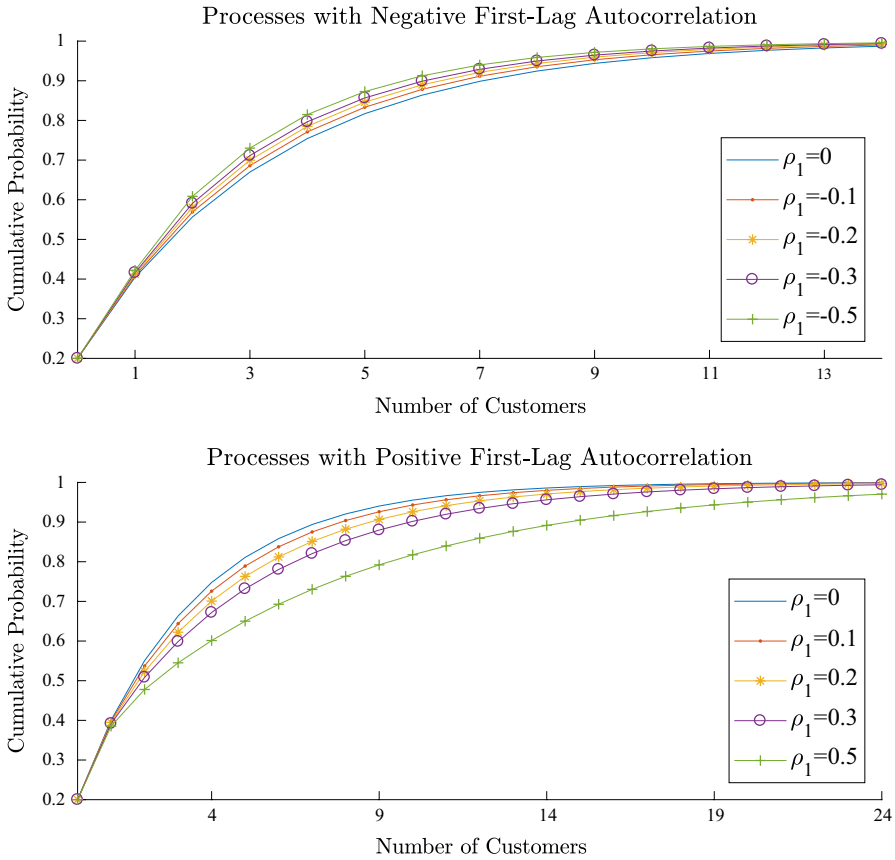
Fig. 3 Cumulative probability distribution of the queue length of MAP/M/1 queues with traffic intensity 0.8

5.1 Impact of approximating positively correlated arrival process with different renewal processes on the performance

5.1.1 The low variability case: $cv_a < 1$

We consider an arrival process with $cv_a = 0.5$ and first-lag autocorrelation $\rho_1 = 0.3$. The performance measures of a system with such an arrival process are shown in Table 6. Estimating the optimal base-stock level using the marginal distribution and first-two moments of the process results in underestimating the optimal base-stock level. Controlling the system by a base-stock level calculated using these models results in a higher cost, and lower service level. Interestingly, estimating the correlated arrival process by means of exponential distribution results in better estimation of the optimal base-stock level. This is due to the fact that coefficient of variation of the exponential distribution is greater than coefficient of variation of the correlated process. Higher coefficient of variation increases the optimal base-stock level which
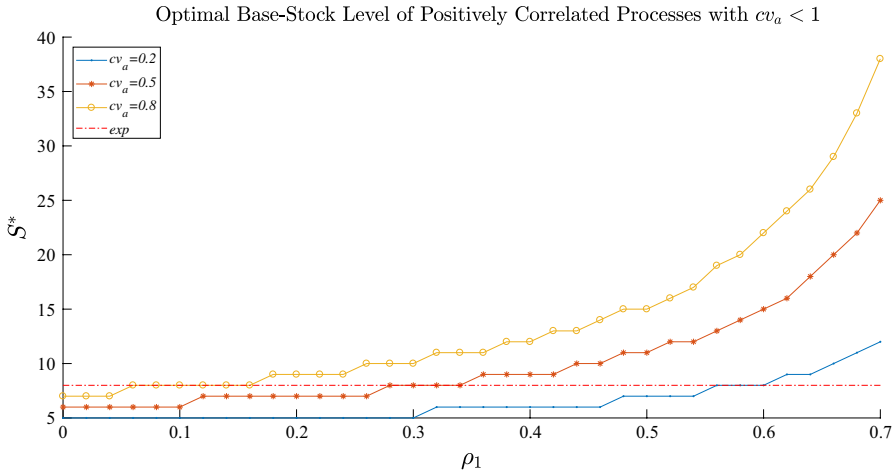
Optimal Base-Stock Level of Positively Correlated Processes with $cv_a < 1$

Fig. 4 Optimal base-stock level of positively correlated processes with $cv_a < 1$ and their estimation with the exponential model (*exp* exponential distribution)

results in a better estimation of base-stock level than the exact marginal distribution. Hence, in positively correlated arrival processes with $cv_a < 1$, employing the exact distribution of the process does not necessarily result in a better estimation of the optimal base-stock level than the exponential distribution. Figure 5 demonstrates the optimal base-stock level of positively correlated systems with $cv_a < 1$ and traffic intensity $v = 0.8$. The optimal base-stock level associated with modeling the correlated arrival with exact marginal distribution is equal to that of the a process with zero first-lag-autocorrelation. Modeling the arrival process with exponential distribution gives a better approximation of the optimal base-stock level as the first-lag autocorrelation increases.

5.1.2 The high variability case $cv_a > 1$

We consider an arrival process with $cv_a = 1.3$ and first-lag autocorrelation $\rho_1 = 0.3$. The performance measures of a system with such an arrival process is shown in Table 7. Modeling the arrival process by means of renewal processes result in underestimation of the optimal base-stock level of the system. In contrary to the case with $cv_a < 1$, the marginal distribution and first-two moments gives a better approximation of the correlated process than exponential distribution. This result holds for all of the cases with $cv_a > 1$ that we evaluated.
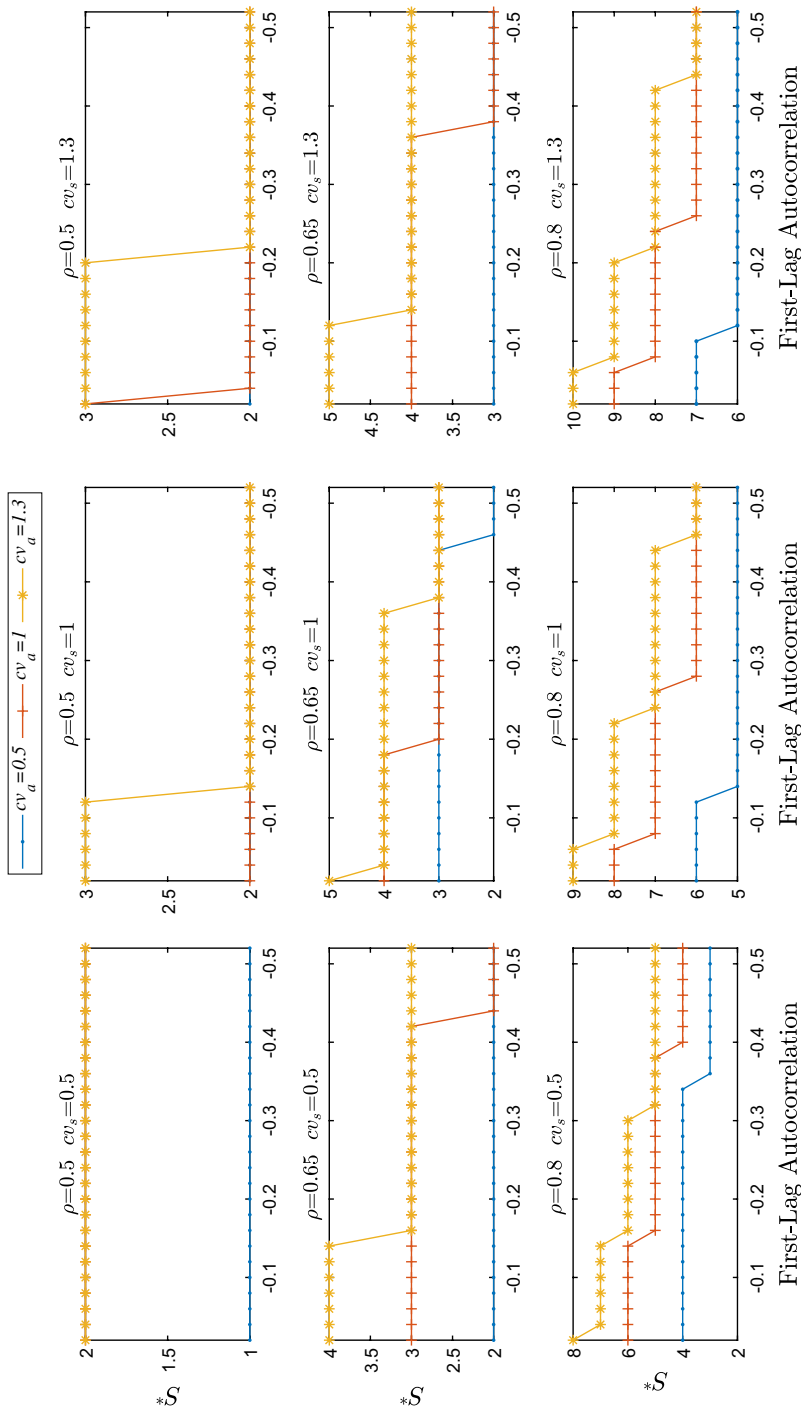
Fig. 5 Impact of the first-lag autocorrelation of negatively correlated arrival process on the optimal base stock levels of a manufacturing system

Table 6 Impact of positively correlated arrival process with $cv_a < 1$ on the performance measures of a production system controlled by base stock policy ($v = 0.8$)

	S^*	TC	Error (%)	$E[X^-]$	Error (%)	$E[X^+]$	Error (%)	$Pr[X < 0]$	Error (%)
MAP/M/1	8	8.7344		0.8062		4.7035		0.1468	
PH/M/1	6	9.1283	5	1.2052	49	3.1024	− 34	0.2195	49
PH$_2$/M/1	6	9.1283	5	1.2052	49	3.1024	− 34	0.2195	49
M/M/1	8	8.7344	0	0.8062	0	4.7035	0	0.1468	0

Table 7 Impact of positively correlated arrival process with $cv_a > 1$ on performance measures of a production system controlled by base stock policy ($v = 0.8$)

	S^*	TC	Error (%)	$E[X^-]$	Error (%)	$E[X^+]$	Error (%)	$Pr[X < 0]$	Error (%)
MAP/M/1	14	15.8010		1.5021		8.2907		0.1586	
PH/M/1	9	17.5352	11	2.6244	75	4.4130	− 47	0.2771	75
PH$_2$/M/1	9	17.5352	11	2.6244	75	4.4130	− 47	0.2771	75
M/M/1	8	18.3945	16	2.9343	95	3.7229	− 55	0.3099	95

5.2 Impact of approximating negatively correlated arrival process with different renewal processes on the performance

5.2.1 The low variability case $cv_a < 1$

We generate negatively correlated MAP from the output process of a production line with three stations and zero buffers. We let the distribution of each machine to be acyclic PH distribution. The statistical descriptors of this output process is shown in the first row of Table 15. Its coefficient of variation and first-lag auto-correlation are $cv_a = 0.29$, $\rho_1 = -0.30$, respectively. We adopt this process as the arrival process. The performance measures of a system with such an arrival process is given in Table 8. Modeling the arrival process by means of renewal processes results in overestimation of the optimal base-stock level of the system. Estimating the arrival process by means of its marginal distribution and first-two moments results in better estimation of the base-stock level than the exponential distribution. This result holds for all of the cases with $cv_a < 1$ that we evaluated.

Table 8 Impact of negatively correlated arrival process with $cv_a < 1$ on performance measures of a production system controlled by base stock policy ($v = 0.8$)

	S^*	TC	Error (%)	$E[X^-]$	Error (%)	$E[X^+]$	Error (%)	$Pr[X < 0]$	Error (%)
MAP/M/1	4	4.2646	0	0.4417	0	2.0559	0	0.1535	0
PH/M/1	5	4.3439	2	0.2883	− 35	2.9024	41	0.1001	− 35
PH$_2$/M/1	5	4.3439	2	0.2883	− 35	2.9024	41	0.1001	− 35
M/M/1	8	6.0949	43	0.0801	− 82	5.6942	177	0.0278	− 82

Table 9 Impact of negatively correlated service process with $cv_a < 1$ on performance measures of a production system controlled by base stock policy ($v = 0.8$)

	S^*	TC	Error (%)	$E[X^-]$	Error (%)	$E[X^+]$	Error (%)	$Pr[X < 0]$	Error (%)
MAP/M/1	7	7.2548	0	0.6719	0	3.8956	0	0.1475	0
PH/M/1	9	7.6795	6	0.4093	− 39	5.6330	45	0.0898	− 39
PH$_2$/M/1	9	7.6795	6	0.4093	− 39	5.6330	45	0.0898	− 39
M/M/1	8	7.3701	2	0.5244	− 22	4.7481	22	0.1151	− 22

5.2.2 The high variability case $cv_a > 1$

We consider an arrival process with $cv_a = 1.3$ and first-lag autocorrelation $\rho_1 = -0.30$. Table 9 shows the performance measures of a system with such an arrival process and its renewal approximations. The results demonstrate that modeling negatively correlated arrival process by means of its first-tow moments, and marginal distribution does not necessarily result in a better estimation of the optimal base-stock level than exponential distribution.

Estimating the arrival process by an exponential distribution gives a better approximation than using the first-two moments or marginal distribution of the process. This result is similar to the result of estimating the optimal base-stock level of a positively correlated arrival process with $cv_a < 1$ by its mean. Figure 6 demonstrates the optimal base-stock level of positively correlated systems with $cv_a > 1$ and traffic intensity $v = 0.8$. The optimal base-stock level associated with modeling the correlated arrival with exact marginal distribution is equal to that of the a process with zero first-lag-autocorrelation. Modeling the arrival process
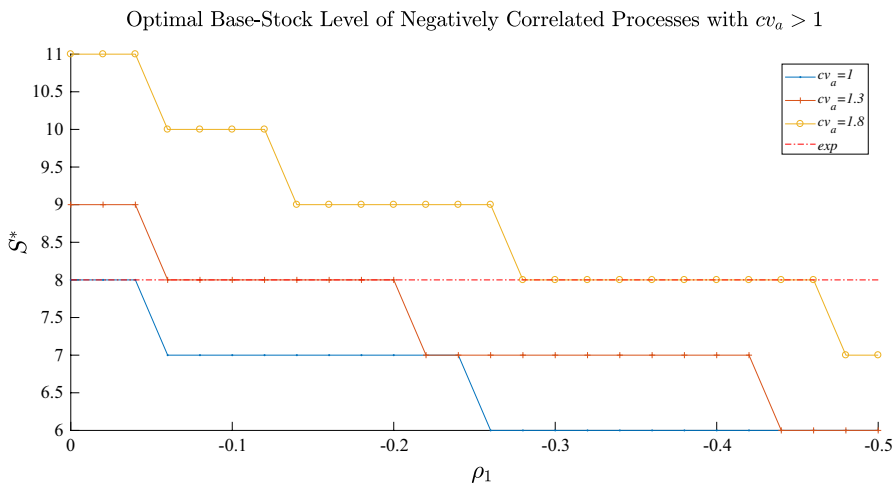

Optimal Base-Stock Level of Negatively Correlated Processes with $cv_a > 1$

Fig. 6 Optimal base-stock level of negatively correlated processes with $cv_a \geq 1$ and their estimation with exponential model

with exponential distribution gives a better approximation of the optimal base-stock level as the first-lag autocorrelation decreases.

Finally, it worth mentioning that, estimating the arrival process by means of its first-two moments and marginal distribution results in the same optimal base-stock levels in our numerical experiments.

6 Fitting MAP to data from shop-floor

6.1 Performance of MAP-fitting methods

In this section we evaluate the performance of MAP-fitting methods that fit a MAP into data collected from shop-floor. We use processing, cycle, and inter-arrival time data of different equipments collected from shop-floor. We use the methods implemented in Bause et al. (2010), and Horváth and Telek (2017) in fitting MAP. The goodness-of-fit of these methods have been investigated thoroughly in the telecommunications literature (Kriege and Buchholz 2010) for processes with positive autocorrelations. Their performance is shown to be satisfactory for data traces observed in telecommunications literature. However, their goodness-of-fit has not been investigated for negatively correlated processes since such autocorrelation structures are not observed in the telecommunications literature. Our numerical analysis demonstrates that fitting inter-event data with monotone increasing negative autocorrelations may result in less accurate fitted MAPs (Dizbin 2016).

Figure 7 demonstrates the distribution and autocorrelation structure of the processing time of 1756 lots processed in a specific equipment. Some of the statistical descriptors of the data and fitted MAP is given in Table 10. The fitted MAP captures the distribution shape and autocorrelation structure with an acceptable accuracy.

Figure 8 demonstrates the distribution and autocorrelation structure of the cycle time of 3908 lots in a specific equipment. Some of the statistical descriptors of the data and fitted MAP are given in Table 11. The fitted MAP captures the distribution shape and first-five lag autocorrelation with an acceptable accuracy.

Figure 1 demonstrates the distribution and autocorrelation structure of the inter-arrival time of 1068 lots to a specific equipment. The statistical descriptors of the data and fitted MAP is given in Table 1. The arrival process demonstrates a zigzag autocorrelation structure. The fitted MAP captures the distribution shape and autocorrelations structure with an acceptable accuracy. We note that we were not able to fit a MAP that captures the autocorrelation structure of processes with monotone increasing negative autocorrelations with acceptable accuracy.

6.2 Setting the base-stock level by using shop-floor data

In this section, we evaluate the performance of setting the base-stock level by fitting a MAP to simulated inter-arrival data. In this evaluation, we first fit a MAP into inter-event times. Then, we use the fitted MAP, denoted by $\hat{MAP}$, to calculate the
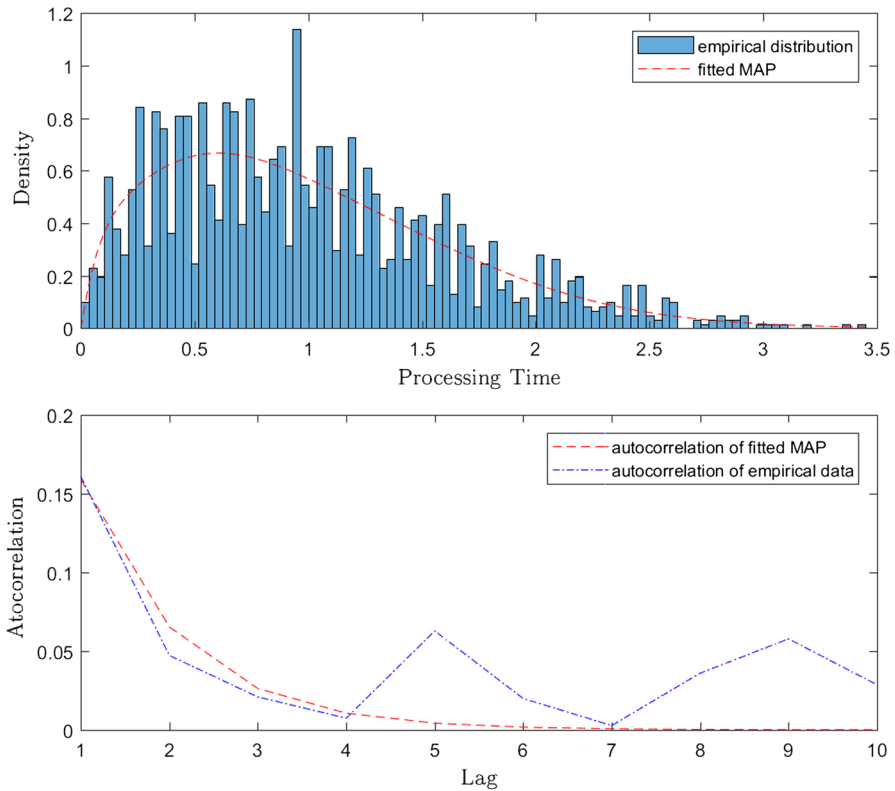
Fig. 7 The observed and fitted distribution and autocorrelations (ρ_k) of processing time of a specific equipment at the Robert Bosch Reutlingen semiconductor manufacturing plant

Table 10 Statistical descriptors of a processing time of an equipment and its fitted MAP

	Moments			Autocorrelation (lag)				
	1	2	3	1	2	3	4	5
Real	1.00	1.40	2.39	0.1606	0.0472	0.0209	0.0074	0.0628
MÂP (Bause et al. 2010)	1.00	1.40	2.41	0.1588	0.0653	0.0262	0.0106	0.0043

optimal base-stock level by employing Eq. (2). For simplicity, we assume that processing time is exponentially distributed with rate 1.25.

We first simulate 10,000 inter-arrival times for a given correlated process. Any of the MAP fitting methods reviewed in Sect. 2.3 can be used to fit a MAP into inter-event data.
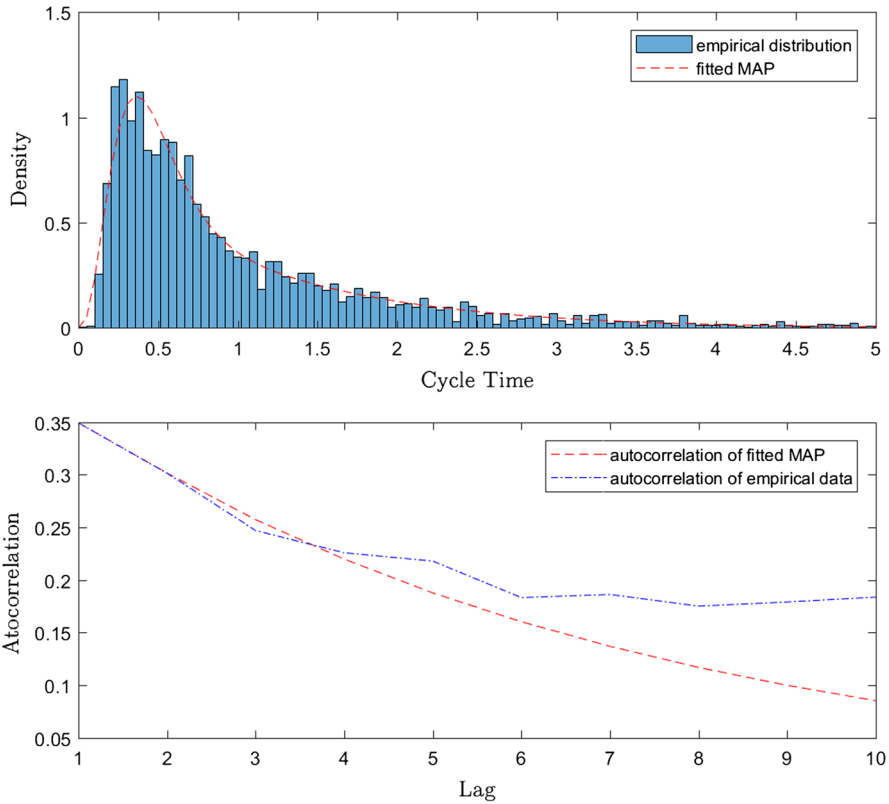
Fig. 8 The observed and fitted distribution and autocorrelations (ρ_k) of cycle time of a specific equipment at the Robert Bosch Reutlingen semiconductor manufacturing plant

Table 11 Statistical descriptors of the cycle time of an equipment and its fitted MAP

	Moments			Autocorrelation (lag)				
	1	2	3	1	2	3	4	5
Real	1.00	1.80	4.77	0.3498	0.3020	0.2474	0.2263	0.2184
MÂP (Bause et al. 2010)	1.00	1.80	4.72	0.3501	0.3025	0.2532	0.2225	0.2012

In this study, we use the simplest fitting method among the available methods, which fits a MAP with the same first-two moments and first-lag autocorrelation to inter-event data (Horváth 2013). We adopt such a MÂP to demonstrate the potential of employing MAPs in estimating the optimal base-stock level by using shop-floor inter-event data. Employing more sophisticated fitting methods can result in more accurate estimated optimal base-stock level. We compare the performance of MÂP in estimating the optimal base-stock level, by comparing it to the optimal base-stock

level of the original MAP, and exact marginal distribution of the process. Dizbin (2016) compares the performance of using different fitting methods.

6.2.1 Performance of fitted MAP in setting the base-stock level of a positively correlated inter-arrival times

In this part, we evaluate performance of the fitted MAP in estimating the optimal base-stock level of a system with positively correlated arrival process. The D_0 and D_1 matrices of the arrival MAP is given below. The statistical properties of the real MAP, simulated inter-arrival data, and $\hat{\text{MAP}}$ are given in Table 12. The simulated data with 10,000 arrivals captures the statistical properties of the real MAP closely. $\hat{\text{MAP}}$ captures the first two moments and first-lag autocorrelation of the simulated data accurately. However, the higher lags are different from simulated data and original MAP. Notice that the higher lags of the fitted MAP decay more rapidly than the real and simulated MAP. One can employ methods such as Horváth et al. (2005), and Bause et al. (2010), or EM-based algorithms which employ more information from data in building MAP to build more accurate $\hat{\text{MAP}}$.

$$D_0 = \begin{bmatrix} -1.2609 & 0 & 0.03589 \\ 0.0028 & -1.2921 & 1.1972 \\ 0 & 0 & -1.2922 \end{bmatrix} \qquad D_1 = \begin{bmatrix} 1.1973 & 0.0277 & 0 \\ 0 & 0 & 0.0921 \\ 0.1544 & 1.1378 & 0 \end{bmatrix}$$

The performance measures of systems with original MAP, $\hat{\text{MAP}}$ and exact marginal distribution (PH) are given in Table 13. The optimal base-stock level of a system with MAP arrival and exponential distribution is 10. $\hat{\text{MAP}}$ estimates the optimal-base-stock level to be 9. Controlling the system with base-stock level of 9 results in almost the same total cost and less service level for the system. Estimating the optimal base-stock level by means of its exact marginal distribution, results in base-stock level equal to 7. Controlling the system with base-stock level of 7 results in 6% higher cost for the system.

Table 12 Statistical descriptors of real, simulated, and fitted MAP of a positively correlated process with coefficient of variation less than one

	Moments			Autocorrelation (lag)				
	1	2	3	1	2	3	4	5
Real	1.00	1.90	5.23	0.0948	0.0695	0.0591	0.0496	0.0416
Simulation	1.00	1.93	5.38	0.1095	0.0632	0.0629	0.0684	0.0390
Horváth (2013)	1.00	1.93	5.31	0.1095	0.0607	0.0337	0.0187	0.0104

Table 13 Estimating the optimal base-Stock level of a positively correlated process

	S^*	TC	Error (%)	$E[X^+]$	Error (%)	$E[X^-]$	Error (%)	$Pr[X < 0]$	Error (%)
MAP/M/1	10	10.7850		0.9586		5.9919		0.1435	
MÂP/M/1	9	10.7973	0	1.1273	18	5.1606	− 14	0.1687	18
PH/M/1	7	11.3881	6	1.5591	63	3.5924	− 40	0.2334	63

6.2.2 Performance of fitted MAP in setting the base-stock level of a negatively correlated inter-arrival times

We generate negatively correlated inter-event data from the output process of a production line with three stations and no buffers in between. The service time at each station is uncertain and follows acyclic PH distribution with the moments given in Table 14. The generator matrix of this production line consists of 1920 states. We fit a MAP with with 13 states to simulated data from this production line. The statistical properties of the real production line, simulated data, and MÂP is given in Table 15. Simulated data almost captures the statistical properties of the production line. MÂP captures the first three moments and first-lag autocorrelation of the simulated data accurately. However, the higher lags differ substantially from that of the simulated data.

The performance measures of systems with real process, MÂP, and exact marginal distribution is shown in Table 16. The optimal base-stock level of the system is 4. Estimating the optimal base-stock level by using MÂP, and exact marginal distribution results in a base-stock level of 5. Controlling the system with this base-stock level results in 2% higher cost. Exact marginal distribution gives as good estimation of the base-stock level as MÂP due to low coefficient of variation of the real MAP, and zig-zag autocorrelation structure of the higher lags of MÂP. If the coefficient of variation of the arrival process increases, performance

Table 14 Moment of the processing time for a production line with three stations

Station	1	2	3
$E[T]$	1.0	1.0	0.8
$E[T^2]$	1.1	1.2	0.85

Table 15 Real, simulated, and fitted statistical descriptors of a negatively correlated process with coefficient of variation less than one

	Moments			Autocorrelation (lag)				
	1	2	3	1	2	3	4	5
Real	1.00	1.29	1.99	− 0.3033	− 0.0500	− 0.0049	− 0.0004	0.0000
Simulation	1.00	1.30	2.01	− 0.3015	− 0.0473	− 0.0042	− 0.0045	− 0.0059
Horváth (2013)	1.00	1.30	2.01	− 0.3015	0.1962	− 0.1277	0.0831	− 0.0541

Table 16 Estimating the optimal base-stock level of a negatively correlated process

	S^*	TC	Error (%)	$E[X^+]$	Error (%)	$E[X^-]$	Error (%)	$Pr[X<0]$	Error (%)
MAP/M/1	4	4.2646		0.4417		2.0559		0.1535	
MÂP/M/1	5	4.3439	2	0.2883	− 35	2.9024	41	0.1001	− 35
PH/M/1	5	4.3439	2	0.2883	− 35	2.9024	41	0.1001	− 35

of the exact marginal distribution in estimating the optimal base-stock level will decrease.

7 Summary of findings

In this section, we summarize our findings based on the numerical experiments reported in the preceding sections:

1. In systems with negatively correlated arrival or service process, ignoring auto-correlation can lead to overestimation of the optimal base-stock level. In other words, renewal approximation of a negatively correlated process overestimates the optimal base stock level, and consequently the total cost of the system.
2. The overestimation of optimal base-stock level in negatively correlated process is due to impact of negative autocorrelation on the stationary queue length probability distribution.
3. In systems with positively correlated arrival or service process, ignoring auto-correlation can lead to underestimation of the optimal base-stock level. In other words, renewal approximation of a positively correlated process underestimates the optimal base stock level, and consequently the total cost of the system.
4. The underestimation of optimal base-stock level in positively correlated process is due to impact of positive autocorrelation on the stationary queue length probability distribution.
5. The queue length distribution of a system with positively correlated arrivals stochastically dominates the same process with less magnitude of autocorrelation. In other words, positive correlation increases the probability of having higher queue length in the system as opposed to negative autocorrelation which increases the probability of having lower queue lengths.
6. Capturing the inter-departure time distribution more accurately while ignoring the autocorrelations does not necessarily improve the accuracy of the results. Modeling the correlated arrival process with a PH distribution does not necessarily result in more accurate estimation of the optimal control parameter. For instance, modeling a positively correlated arrival with a coefficient of variation less than one, with an exponential distribution may result in better estimates of the base-stock level than PH distribution.

7. MAP fitting methods allows setting the base-stock level in a way that performs at least as the performance obtained by using the exact marginal distribution. Employing more sophisticate fitting methods can result in more accurate estimated optimal base-stock levels.

8 Conclusions

In this paper, our objective is controlling a production system by using shop-floor inter-event data such as the inter-arrival, inter-departure, and service completion times. Although, empirical studies show that inter-event times of a production system are correlated, most of the analytical studies for the analysis and control of production systems ignore correlation. We use Markovian arrival processeses (MAP) to model shop-floor inter-event data because of their ability in estimating any inter-event process arbitrarily close enough (Asmussen and Koole 1993). We present a comprehensive review of MAPs and the existing MAP fitting methods to build MAPs from inter-event shop-floor data for the manufacturing systems literature.

We analyze the impact of autocorrelation on the optimal base-stock level by using the structural properties of the MAPs. Our analysis show that ignoring autocorrelation in modeling inter-arrival and service times of production-inventory systems can lead to misleading results.

We present the effectiveness of the existing MAP fitting methods as they are used to build MAP models from the inter-event time data and then used to determine the base-stock levels. Our analysis show that even using the simplest MAP fitting method gives at least as good estimation of the optimal base-stock level as the exact marginal distribution. Employing more sophisticate fitting methods can result in more accurate estimated optimal base-stock levels.

We conjecture that queue length distribution of a system with positively correlated arrivals, stochastically dominates queue length distribution of a processes with the same distribution an lesser magnitude of autocorrelation. In addition, the queue length distribution of a system with negatively correlated arrivals, is stochastically dominated by the queue length distribution of a process with the same distribution an lesser magnitude of autocorrelation. The analytical demonstration of this statement requires an investigation regarding the behavior of geometric matrix R (or its determinant) as a function of the autocorrelation or covariance of the arrival or service processes of the system.

We conclude by stating that MAP is suitable process to model shop-floor inter-event data. There exist numerous MAP fitting methods which can be employed to fit a MAP into the shop-floor inter-event data. MAP can capture the correlation in inter-arrival and service times which affect the optimal base-stock level of the system, significantly. Our findings on the effects of autocorrelation on the system can be summarized as follows:

1. The autocorrelation structure should be taken into account while designing a production system.

2. Positively (Negatively) correlated systems need a larger (smaller) buffer size in comparison with the correlated processes with lower magnitude of first-lag auto-correlation. In other words, the expected inventory in front of a service station increases (decreases) as a function of the positive (negative) autocorrelation.
3. Positive autocorrelation may increase the average cycle time of the products in the system.
4. The arrival of products into the system can be modulated in order to control the correlation in inter-arrival times. For instance, if low buffer levels is needed in front of a line creating negative autocorrelation may decrease the optimal buffer level.

We developed an estimation-then-optimization framework for controlling a production line in which production decision can be taken by employing full statistical properties of the shop-floor inter-event data. A possible research direction is to adopt model free methods instead of estimation-then-optimization approach in setting the base-stock level. Model-free methods decrease the model misspecification errors at a cost of computational complexity. Integrating MAPs into a model-free method may reduce their computational costs and result in more efficient algorithms in estimating the optimal base-stock levels. More information from shop-floor such as the state of the machines, and quality of the products can be modeled by using model-free methods.

Our results are presented here for single-product problems. Our framework can be extended to a multi-product problem by adopting marked MAP. Buchholz et al (2010), Horváth and Okamura (2013) propose methods to fit a marked MAP into data. These methods can be adopted in controlling a production line with multiple products by using shop-floor inter-event data. Further research directions include performance of the base-stock policy in controlling a system with correlated arrival or service processes.

References

Adan I, Kulkarni V (2003) Single-server queue with Markov-dependent inter-arrival and service times. Queueing Syst 45(2):113–134

Altiok T, Melamed B (2001) The case for modeling correlation in manufacturing systems. IIE Trans 33(9):779–791

Asmussen S, Koole G (1993) Marked point processes as limits of Markovian arrival streams. J Appl Probab 30(2):365–372

Bauerle N (1997) Monotonicity results for mr/gi/1 queues. J Appl Probab 34(2):514–524

Bause F, Buchholz P, Kriege J (2010) Profido: the processes fitting toolkit dortmund. In: 2010 Seventh international conference on the quantitative evaluation of systems, pp 87–96

Biller B, Nelson BL (2003) Modeling and generating multivariate time-series input processes using a vector autoregressive technique. ACM Trans Model Comput Simul 13(3):211–237

Bini DA, Meini B, Steffé S, Van Houdt B (2006) Structured markov chains solver: algorithms. In: Proceeding from the 2006 workshop on tools for solving structured markov chains, SMCtools '06. ACM, New York

Breuer L (2002) An EM algorithm for batch Markovian arrival processeses and its comparison to a simpler estimation procedure. Ann Oper Res 112(1):123–138

Breuer L, Kume A (2010) An EM algorithm for Markovian arrival processeses observed at discrete times. In: Müller-Clostermann B, Echtle K, Rathgeb EP (eds) Measurement, modelling, and evaluation of computing systems and dependability and fault tolerance. Springer, Heidelberg, pp 242–258

Brickner C, Indrawan D, Williams D, Chakravarthy SR (2010) Simulation of a stochastic model for a service system. In: Proceedings of the 2010 winter simulation conference, pp 1636–1647

Buchholz P (2003) An EM-algorithm for MAP fitting from real traffic data. In: Kemper P, Sanders WH (eds) Computer performance evaluation. Modelling techniques and tools, Springer, Heidelberg, pp 218–236

Buchholz P, Kriege J (2009) A heuristic approach for fitting MAPs to moments and joint moments. In: 2009 Sixth international conference on the quantitative evaluation of systems, pp 53–62

Buchholz P, Kriege J (2017) Fitting correlated arrival and service times and related queueing performance. Queueing Syst 85(3):337–359

Buchholz P, Panchenko A (2004) A two-step EM algorithm for MAP fitting. In: Aykanat C, Dayar T, Korpeoglu I (eds) Computer and information sciences-ISCIS 2004. Springer, Heidelberg, pp 217–227

Buchholz P, Kemper P, Kriege J (2010) Multi-class Markovian arrival processeses and their parameter fitting. Perform Eval 67(11):1092–1106

Buchholz P, Kriege J, Felko I (2014) Input modeling with phase-type distributions and Markov models: theory and applications. Springer briefs in mathematics, Springer, Berlin

Casale G, Zhang EZ, Smirni E (2010) Trace data characterization and fitting for Markov modeling. Perform Eval 67(2):61–79

Civelek I, Biller B, SchellerWolf A (2009) The impact of dependence on queuing systems. Working paper, Carnegie Mellon University, Pittsburgh

Dahl TA, Willemain TR (2001) The effect of long-memory arrivals on queue performance. Oper Res Lett 29(3):123–127

Distefano S, Trivedi KS (2013) Non-Markovian state-space models in dependability evaluation. Qual Reliab Eng Int 29(2):225–239

Dizbin N, Tan B (2017) Production and inventory control with correlated demand. In: Stochastic models of manufacturing and service operations conference, Acaya, Italy

Dizbin NM (2016) Modeling and analysis of autocorrelation in manufacturing systems. Master's thesis, Graduate School of Business, Koç University, Istanbul, Turkey

Gerhardt I, Nelson B (2009) On capturing dependence in point processes: matching moments and other techniques. Technical Report, Northwestern University

Hadidi N (1985) Further results on queues with partial correlation. Oper Res 33(1):203–209

He QM (2013) Fundamentals of matrix-analytic methods. Springer, Berlin

Hendricks KB, McClain JO (1993) The output processes of serial production lines of general machines with finite buffers. Manag Sci 39(10):1194–1201

Horváth G (2013) Matching marginal moments and lag autocorrelations with MAPs. In: Proceedings of the 7th international conference on performance evaluation methodologies and tools, ValueTools '13. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, pp 59–68

Horváth G, Okamura H (2013) A fast EM algorithm for fitting marked Markovian arrival processeses with a new special structure. In: Balsamo MS, Knottenbelt WJ, Marin A (eds) Comput Perform Eng. Springer, Heidelberg, pp 119–133

Horváth G, Telek M (2017) Butools 2: a rich toolbox for Markovian performance evaluation. In: Proceedings of the 10th EAI international conference on performance evaluation methodologies and tools on 10th EAI international conference on performance evaluation methodologies and tools.

ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, pp 137–142

Horváth G, Buchholz P, Telek M (2005) A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. In: Second international conference on the quantitative evaluation of systems (QEST'05), pp 124–133

Hwang GU, Sohraby K (2003) On the exact analysis of a discrete-time queueing system with autoregressive inputs. Queueing Syst 43(1):29–41

IEC (2015) Factory of the future. IEC-International Electrotechnical Commission White Paper

Inman RR (1999) Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. Prod Oper Manag 8(4):409–432

Jemai Z, Karaesmen F (2005) The influence of demand variability on the performance of a make-to-stock queue. Eur J Oper Res 164(1):195–205

Klemm A, Lindemann C, Lohmann M (2002) Traffic modeling of IP networks using the batch Markovian arrival processes. In: Field T, Harrison PG, Bradley J, Harder U (eds) Computer performance evaluation: modelling techniques and tools. Springer, Heidelberg, pp 92–110

Kriege J, Buchholz P (2010) An empirical comparison of MAP fitting algorithms. In: Müller-Clostermann B, Echtle K, Rathgeb EP (eds) Measurement, modelling, and evaluation of computing systems and dependability and fault tolerance. Springer, Heidelberg, pp 259–273

Kriege J, Buchholz P (2014) PH and MAP fitting with aggregated traffic traces. In: Fischbach K, Krieger UR (eds) Measurement, modelling, and evaluation of computing systems and dependability and fault tolerance. Springer, Cham, pp 1–15

Lakatos L, Szeidl L, Telek M (2013) Introduction to queuing system with telecommunication applications. Springer, Berlin

Livny M, Melamed B, Tsiolis AK (1993) The impact of autocorrelation on queuing systems. Manag Sci 39(3):322–339

Neuts MF (1975) Probability distributions of phase type. University of Louvain, Louvain, Liber Amicorum Prof Emeritus H Florin

Neuts MF (1979) A versatile Markovian point process. J Appl Probab 16(04):764–779

Okamura H, Dohi T (2009) Faster maximum likelihood estimation algorithms for Markovian arrival processeses. In: 2009 Sixth international conference on the quantitative evaluation of systems, pp 73–82

Okamura H, Dohi T, Trivedi KS (2009) Markovian arrival processes parameter estimation with group data. IEEE ACM Trans Netw 17(4):1326–1339

Okamura H, Kishikawa H, Dohi T (2013) Application of deterministic annealing EM algorithm to MAP/PH parameter estimation. Telecommun Syst 54(1):79–90

Patuwo B (1989) The effects of the dependency in the Markov renewal arrival process on the various performance measures of exponential server queues. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA

Patuwo B, Disney RL, McNickle DC (1993) The effect of correlation arrivals on queues. IIE Trans 25(3):105–110

Pereira DC, del Rio Vilas D, Monteil NR, Prado RR, del Valle AG (2012) Autocorrelation effects in manufacturing systems performance: a simulation analysis. In: Proceedings of the 2012 winter simulation conference (WSC), pp 1–12

Resnick S, Samorodnitsky G (1997) Performance decay in a single server exponential queueing model with long range dependence. Oper Res 45(2):235–243

Runnenburg J (1961) An example illustrating the possibilities of renewal theory and waiting-time theory for Markov-dependent arrival intervals. Proc Ser A Kon Neder Akad Weten 64:560–576

Runnenburg JT (1962) Some numerical results on waiting-time distributions for dependent arrival-intervals. Stat Neerl 16(4):337–347

Schomig AK, Mittler M (1995) Autocorrelation of cycle times in semiconductor manufacturing systems. Winter Simul Conf Proc 1995:865–872

Szekli R, Disney RL, Hur S (1994a) MR/GI/1 queues with positively correlated arrival stream. J Appl Probab 31(2):497–514

Szekli R, Disney RL, Hur S (1994b) On performance comparison of MR/GI/1 queues. Queueing Syst 17(3):451–470

Takahashi K, Nakamura N (1998) The effect of autocorrelated demand in JIT production systems. Int J Prod Res 36(5):1159–1176

Tan B, Lagershausen S (2017) On the output dynamics of production systems subject to blocking. IISE Trans 49(3):268–284

Telek M, Horváth G (2007) A minimal representation of Markov arrival processes and a moments matching method. Perform Eval 64(9):1153–1168

Xu SH (1999) Structural analysis of a queueing system with multiclasses of correlated arrivals and blocking. Operations Research 47(2):264–276

Nima Manafzadeh Dizbin is a Ph.D. student at Koç University, Istanbul, Turkey. He received a B.S. degree in Industrial Engineering from the Sharif University of Technology and a master's degree in Business Theory from Koç University. His research interests are in data-driven decision making, data-driven performance evaluation, and control of manufacturing systems, stochastic modeling of operations.

Barış Tan is a Professor of Operations Management and Industrial Engineering at Koç University, Istanbul, Turkey. He received a B.S. degree in Electrical and Electronics Engineering from Bogaziçi University. He received a master's degree in Industrial and Systems Engineering, a Manufacturing Systems Engineering degree, and a Ph.D. degree in Operations Research from the University of Florida. His main expertise is in design and control of production systems, analytical and numerical modeling techniques, supply chain management, and operations research. His current research interests are in stochastic modeling of operations, data-driven performance evaluation, and control of manufacturing systems, and cooperation in supply chain management.