Boxma, Onno; Perry, David; Stadje, Wolfgang

Article — Published Version

Perishable inventories with random input: a unifying survey with extensions

Annals of Operations Research

Suggested Citation: Boxma, Onno; Perry, David; Stadje, Wolfgang (2023) : Perishable inventories with random input: a unifying survey with extensions, Annals of Operations Research, ISSN 1572-9338, Springer US, New York, Vol. 332, Iss. 1, pp. 1069-1105, https://doi.org/10.1007/s10479-023-05317-2

This Version is available at:
https://hdl.handle.net/10419/317736

Annals of Operations Research (2024) 332:1069–1105
https://doi.org/10.1007/s10479-023-05317-2

ORIGINAL - SURVEY OR EXPOSITION

Perishable inventories with random input: a unifying survey with extensions

Onno Boxma[1] · David Perry[2] · Wolfgang Stadje[3]

Accepted: 24 February 2023 / Published online: 30 May 2023

Abstract
This paper is devoted to the theory of perishable inventory systems. In such systems items arrive and stay 'on the shelf' until they are either taken by a demand or become outdated. Our aim is to survey, extend and enrich the probabilistic analysis of a large class of such systems. A unifying principle is to consider the so-called virtual outdating process $\mathbf{V}$, where $V(t)$ is the time that would pass from t until the next outdating if no new demands arrived after t. The steady-state density of $\mathbf{V}$ is determined for a wide range of perishable inventory systems. Key performance measures like the rate of outdatings, the rate of unsatisfied demands and the distribution of the number of items on the shelf are subsequently expressed in that density. Some of the main ingredients of our analysis are level crossing theory and the observation that the $\mathbf{V}$ process can be interpreted as the workload process of a specific single server queueing system.

Keywords Perishable inventories · Level crossings methodology · Satisfied demand conservation law · Laplace Transform · Steady state analysis · Busy period

1 Introduction

Background and motivation The theory of *perishable inventory systems* (PIS) deals with one of the classical topics of stochastic operations research: items of a certain type arrive at a collecting point from where they are taken away by incoming demands. If an item stays too long it can become unusable due to random deterioration or a predetermined maximum

✉ Onno Boxma
o.j.boxma@tue.nl

David Perry
davidper@hit.ac.il

Wolfgang Stadje
wstadje@uos.de

[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[2] Holon Institute of Technology, P.O. Box 305, Holon 5810201, Israel

[3] Institute of Mathematics, University of Osnabrück, 49069 Osnabrück, Germany

expiration time. The arrival intervals of items as well as those of demands may be random or deterministic, in either case a controller may try to govern them, and arrivals may also occur in batches. The standard real-world example is of course a storage place for commodities, but other applications include blood banks, spot markets for special goods, distribution sites for transplantation organs, or peer-to-peer lending agencies in the internet, where many individual lenders deposit money for a limited period, which can then be borrowed by debtors.

Our aim in this paper is to survey, extend and considerably supplement and enrich the probabilistic analysis of a large class of these PIS. Over the last four decades inventory models of this basic structure were treated in hundreds of articles, textbooks, and monographs. In many examples the system input is generated by replenishment orders of a controller who decides about the timing of the orders and their sizes. The great majority of papers studies optimization problems for this kind of systems, looking for optimal ordering policies; see for example the monograph by Nahmias (2012) and his earlier review (Nahmias, 1982). However, there is a second line of research dealing with PIS with random input (without involvement of a controller) and focusing on their stochastic analysis. That research is surveyed and expanded in this paper.

The survey (Karaesmen et al., 2011) contains a comprehensive section about the papers on PIS with random input that were published until 2011. The authors of the present paper were involved in many of the studies surveyed in Karaesmen et al. (2011) and since 2011 have extended the earlier results in several directions, contributing, jointly with various co-authors, more than 20 publications. This survey provides a unifying presentation of the published material and also develops several new model variants. Our approach also enables us to derive additional results for models studied in the past. In view of space constraints, the presentation of some of the new model variants necessarily is rather concise. We plan to elaborate on these discussions in an extended version of the present paper, accessible as a Eurandom report via https://www.eurandom.tue.nl/pre-prints/.

Even after having restricted ourselves to PIS with random input, there are topics that we largely ignore in order to present a methodologically coherent overview; for those topics we refer to the valuable survey of Krishnamoorthy et al. (2020). An interesting line of research that will not be touched upon in the following sections is that of perishable inventory systems *with common lifetime*. In this type of PIS, which was first studied in Lian et al. (2005), it is assumed that all items of a batch of the same age fail at the same time. The topic was taken up by Chakravarthy (2010) and subsequently by Krishnamoorthy and collaborators (see e.g. Krishnamoorthy et al., 2016); several more references can be found in Section 6.2.5.1 of Krishnamoorthy et al. (2020), and a very recent one is Shajin et al. (2022).

Another interesting line of research that we do not discuss in the following sections is that of queueing/inventory systems with positive service times: If a demand arrives and an item is in stock, it may still require a non-negligible amount of time to take that item. Such systems are extensively surveyed by Krishnamoorthy et al. (2020). They sometimes give rise to a product form; often an asymptotic product form for the joint distribution of the number of waiting customers and the number of items in stock, or (cf. Otten et al., 2015 that considers a queueing system in a random environment) the joint distribution of the number of customers and the environment state—where the environment represents the inventory-replenishment structure. Some references in Table 6.1 and Subsection 6.2.1.1 of Krishnamoorthy et al. (2020) on such product form results in queueing-inventory systems also allow perishability; and again see (Shajin et al., 2022).

Finally, we also ignore PIS with retrials of demands. We refer to Section 6.2.4 of Krishnamoorthy et al. (2020) for references to queueing-inventory systems with retrials, and to Ko (2020) for a recent study on this topic with perishability of items.

Motivation The original motivation to consider PIS with external random input (without ordering policies) was formed by *blood bank systems* in which a random stream of blood donations serves as input and the output consists of satisfied demands, taking away blood units, and outdated blood units. Note that the maximum shelf life (sojourn time) of every blood unit is determined by some external authorities; that is, every country or province has its own health regulations regarding the expiration dates of the blood units. *Organ transplantation sites* (Boxma et al., 2011a) constitute another application area. Organs are removed from the bodies of just deceased people. In some countries this is possible either after receiving permission from the family or because the deceased had given permission while alive; in others it is mandated by law that every citizen is a potential donor. Both the (usually deterministic) shelf lifetimes of the organs and the random "impatience times" (lifetimes) of the waiting patients are finite. A *spot market* for perishable goods constitutes another PIS with random input; the flower market in The Netherlands may serve as a colorful example.

Methodology Our main focus is always on the analysis of the stationary (long-run) behavior of the PIS in question, often leading to closed-form expressions for the most relevant performance measures and functionals, or their transforms. The obtained explicit formulas can subsequently be used for numerical optimization of an objective (cost or profit) function. Cost functions usually involve the holding costs of items, which makes it important to study the distribution of the number of items in the system. Cost functions will also typically take into account penalties for unsatisfied demands and for outdatings of items.

Let us denote the above-mentioned number of items that are present in the system ("on the shelf") at any time t by $K(t)$. In general $\mathbf{K} = \{K(t), t \geq 0\}$ is not a Markov process, since at any given time t_0 the distribution of $(K(t))_{t \geq t_0}$ usually depends on the evolution of the process before t_0 and not just on $K(t_0)$—indeed, the age of the items is important. One could try to use supplementary variables to retain the Markov property, but the resulting process would become too complex even for quite simple PIS. To overcome this difficulty, we now introduce a one-dimensional process that turns out to be Markovian in many PIS.

Let $\mathbf{A} = \{A(t) : t \in [0, \infty)\}$ where $A(t)$ is the age of the oldest item on the shelf or, if the shelf is empty at time t, $A(t)$ denotes a "negative age", defined to be minus the time it takes until the next arrival at the shelf after time t. For simplicity assume that items expire at age 1. Then set $V(t) = 1 - A(t)$; cf. Fig. 1. A little reflection shows that $V(t)$ is the time that would pass from t until the next outdating if no new demands arrived after t. This "virtual" process $\mathbf{V} = \{V(t), t \geq 0\}$, the so-called Virtual Outdating Process (VOT), is closely related to $\mathbf{K}$ provided that any newly arriving demand is always satisfied by the oldest item present, if at all. Indeed, the shelf is empty if and only if the age of the oldest item is negative ($A(t) < 0$, so $V(t) > 1$), and the number of items on the shelf equals n if and only if $n - 1$ items have arrived during the age time interval of the oldest item.

For all PIS models in this survey $\mathbf{V}$ turns out to be a key process. Under certain Poisson-type assumptions $\mathbf{V}$ is a Markov process and, when the process is stationary, a regenerative process. When its steady-state distribution exists, it is characterized by an integral equation of the *Pollaczek–Khintchine* type. This integral equation, a Volterra integral equation of the second type, is for all $x \geq 0$ given by

$$r(x) f(x) = \int_0^{k(x)} \mu(w)[1 - G(L(x) - L(w))] f(w) \mathrm{d}w + f(0)[1 - G(L(x))]. \quad (1)$$

In (1) the function $f(\cdot)$ is the steady-state density of $\mathbf{V}$ and the functions $r(\cdot)$, $L(\cdot)$, $G(\cdot)$, $\mu(\cdot)$ and $k(\cdot)$ are specific in every model variant. This equation can be obtained by deriving the Kolmogorov forward equation for the $\mathbf{V}$ process. Typically one can also obtain it via

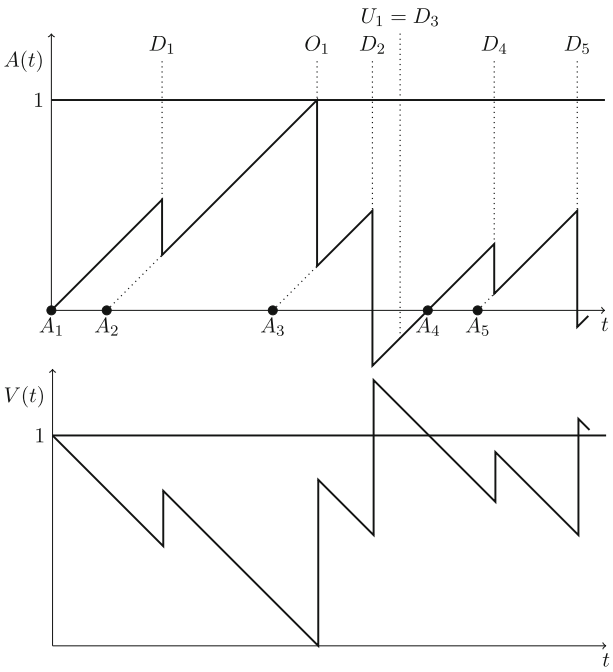

Fig. 1 A typical sample path of the age process **A** (top panel) and of the VOT process **V** (bottom panel). A_n (D_n) denotes the arrival time of the nth item (demand); O_1 denotes the first outdating, and U_1 the first unsatisfied demand; it coincides with D_3

application of the Level Crossing Theory (LCT) (Brill, 2008; Doshi, 1992). LCT observes and exploits the fact that, for the process **V** in steady state, the long-run average number of downcrossings of any level x per time unit is equal to the long-run average number of upcrossings of that level per time unit.

For several model variants we shall show that the lefthand side of (1) equals the rate of downcrossings of level x, and the righthand side the rate of upcrossings of level x. In particular, we show this in some detail when proving Theorem 3. Solving the integral equation yields the steady-state density $f(\cdot)$ (or its Laplace transform). In several cases, we shall use that result to also determine the steady-state distribution of the number of items on the shelf.

An important observation, suggested by Fig. 1 for the **V** process, is that this process can be interpreted as the workload process of a specific single server queueing system. Here the jumps upward indicate service requirements of arriving customers, and the slope downward reflects the fact that a server is working (in the figure: at a constant speed of one unit of work per time unit). In addition, customers have a patience of length one; they do not enter the system if their waiting time would be larger than one (corresponding to unsatisfied demands). This observation allows us to make use of methods which have been developed, and results which have been obtained, for single server queues. It should be noticed that, in most queueing systems, when the system becomes empty it stays empty until the next arrival (an idle period); the graph for **V** can be viewed as representing the workload in a queue after the idle periods have been removed and the busy periods have been glued together.

Structure of the paper In Sect. 2 we present a model description for a large class of PIS, and we also introduce some preliminary results, including a conservation law for the rate of the

satisfied demands that is valid for all model variants. In Sect. 3 we introduce a class of PIS models in which the arrival processes of items and of demands are Poisson processes with rates that depend on the current value of the VOT process. This model is studied in detail in Sects. 3 and 4, along with several variants. In Sect. 5 we globally discuss PIS models in which the arrival process of items, or of demands, forms a renewal process. Section 6 contains a detailed analysis of three such models. Finally, Sect. 7 briefly mentions some model variants and problems which in our view are of considerable interest, but for which we lack the space to discuss them at length. This section also contains some open problems.

2 Model description and preliminaries

We consider a perishable inventory system with infinite storage capacity, in which input and demands are both random. Items arrive at the system according to a point process $\mathbf{N}_I := \{N_I(t), t \geq 0\}$. Each item has the same deterministic usable lifetime, which w.l.o.g. is assumed to be one time unit. Demands arrive according to a point process $\mathbf{N}_D := \{N_D(t), t \geq 0\}$. Upon arrival, a demand removes the oldest item in storage ('on the shelf'), or leaves unsatisfied if the system is empty (but in Sect. 6.2 we shall also study a model in which demands are willing to wait). An item that has not been taken within one time unit of arrival becomes outdated and must be scrapped. The arrival processes of items and demands are assumed to be independent. We assume for simplicity that the system is empty at time 0.

The process of outdated items is denoted by $\mathbf{N}_O := \{N_O(t), t \geq 0\}$, and the process of unsatisfied demands is denoted by $\mathbf{N}_U := \{N_U(t), t \geq 0\}$. $\mathbf{N}_O$ is a filtered process of $\mathbf{N}_I$ and $\mathbf{N}_U$ is a filtered process of $\mathbf{N}_D$.

As will be seen in Theorem 2 below, if both $\mathbf{N}_I$ and $\mathbf{N}_D$ are Poisson processes, then both $\mathbf{N}_O$ and $\mathbf{N}_U$ are renewal processes. If $\mathbf{N}_I$ is a renewal process but not Poisson, and $\mathbf{N}_D$ is a Poisson process, then $\mathbf{N}_O$ still is a renewal process but $\mathbf{N}_U$ need not be a renewal process.

Let us assume that the following long-run arrival rates exist:

$$\begin{aligned}
\lambda &\stackrel{def}{=} \lim_{t\to\infty} \frac{N_I(t)}{t} = \lim_{t\to\infty} \frac{EN_I(t)}{t},\\
\mu &\stackrel{def}{=} \lim_{t\to\infty} \frac{N_D(t)}{t} = \lim_{t\to\infty} \frac{EN_D(t)}{t},\\
\lambda^* &\stackrel{def}{=} \lim_{t\to\infty} \frac{N_O(t)}{t} = \lim_{t\to\infty} \frac{EN_O(t)}{t},\\
\mu^* &\stackrel{def}{=} \lim_{t\to\infty} \frac{N_U(t)}{t} = \lim_{t\to\infty} \frac{EN_U(t)}{t}.
\end{aligned}$$

The limits in the middle are almost-sure limits. λ and μ are the arrival rates of items and of demands, respectively, while λ^* and μ^* denote the outdating rate and the rate of unsatisfied demands. In all PIS considered in this paper these rates will be seen to exist due to the underlying regenerative structures.

Now let $K(t)$ be the number of items in the system at time t. Clearly, $K(t)$ is equal to the number of items that have arrived up to t minus the number of items that have left until then, which in turn is the sum of the number of outdatings and that of satisfied demands up to t. Hence,

$$K(t) = N_I(t) - [N_O(t) + N_D(t) - N_U(t)]. \tag{2}$$

Dividing both sides of (2) by t and taking the limit as $t \to \infty$ yields the following conservation law.

Theorem 1 (Satisfied demand conservation law) *Consider a PIS for which* $\mathbf{N}_I$ *and* $\mathbf{N}_D$ *are arbitrary counting processes and* $\lambda, \lambda^*, \mu, \mu^*$ *exist. Then*

$$\lambda - \lambda^* = \mu - \mu^*. \tag{3}$$

The conservation law is intuitively clear: the left side represents the rate of items that do not become outdated, while the right side represents the rate of satisfied demands. The conservation law is very useful, e.g., when one out of $\mathbf{N}_I$ or $\mathbf{N}_D$ is a Poisson process and the second is a renewal process. Then, one process out of $\mathbf{N}_O$ or $\mathbf{N}_U$ is a renewal process while the second process is not, but at least its rate can be found by (3). The conservation law shows that the higher the outdating rate, the lower the rate of unsatisfied demands. Both a high outdating rate and a high unsatisfied demands rate have their drawback; the conservation law reflects the trade-off between the goals of efficient resource usage and customer satisfaction.

The conservation law (3) can be extended to the case in which either arrivals occur as batches or every demand is for a batch of items. In the latter case a demand can be either partially or fully satisfied. To see the generalization, suppose that the demand batch sizes are i.i.d. (independent, identically distributed) random variables with mean χ. Then the balance equation (1) becomes

$$\lambda - \lambda^* = \mu\chi - \mu^*. \tag{4}$$

Applications of (3) and (4) occur in later sections.

Remark 1 Keeping in mind the definition of the age process that was introduced in Sect. 1, it is readily seen that the following alternative representation of the number of items $K(t)$ also holds:

$$K(t) = 1 + N_I(t) - N_I(t - A(t)), \quad \forall\, t \geq 0. \tag{5}$$

Note that $N_I(t) - N_I(t - A(t))$ is the number of items arriving during the age of the oldest item at t and that $\mathbf{N}_I$ is right continuous. This identity is useful for determining the steady-state mean number of items on the shelf.

The case in which both $\mathbf{N}_I$ and $\mathbf{N}_D$ are Poisson processes is called the **basic** PIS. Recalling our remark above on the connection to queueing systems, one can view $\mathbf{V}$ for the basic PIS as the workload process of a certain $M/M/1 + D$ queue in which customers do not wait more than $D = 1$ time unit for their service and the idle periods are deleted. The following theorem explores the probabilistic structure of the basic PIS, also allowing $\mathbf{N}_I$ to be a renewal process.

Theorem 2 *If for a PIS* $\mathbf{N}_I$ *is a renewal process and* $\mathbf{N}_D$ *is a Poisson process, the number of items process* $\mathbf{K}$*, the VOT process* $\mathbf{V}$ *and the age process* $\mathbf{A}$ *are regenerative with the sequence of outdating times as regeneration points, and* $\mathbf{N}_O$ *is a renewal process. For the basic PIS* $\mathbf{N}_U$ *is also a renewal process.*

Proof Consider the system at a time T (> 1) of an outdating. Looking forward in time from T, the times until the next arrival of an item and until the next demand arrival do not depend on the prior history of the PIS and are independent of each other, the first having the distribution of an item interarrival time conditioned to be greater than 1 and the latter being exponentially distributed with parameter μ, and they are also independent of T. Furthermore, the items that arrived in $[T - 1, T)$ are still on the shelf, and their number is independent of the past, its distribution being equal to that of $M = \max\{n : X_1 + \ldots + X_n < 1\}$, where the X_i are i.i.d. item interarrival times (for the basic PIS this number is Poisson(λ)-distributed).

Altogether this implies that T is a regeneration time for the processes $\mathbf{K}$, $\mathbf{V}$ and $\mathbf{A}$, and the times between successive outdatings are i.i.d. Hence $\mathbf{N}_O$ is a renewal process.

At any arrival time of an unsatisfied demand the shelf is empty, so that for the basic PIS its future evolution depends only on the Poisson processes of future item arrivals and demand arrivals, which are independent of each other and, by their memoryless property, also independent of all past events. Thus $\mathbf{N}_U$ is a renewal process for the basic PIS. □

In the present paper we shall devote much attention to the following generalization that was instigated in Nahmias et al. (2004b). Let the item arrival rate and the demand arrival rate depend on the Virtual Outdating Time $V(t)$, in the following way:

Given that $V(t) = x$, the probability of one item (respectively demand) arrival in the time interval $[t, t + \Delta t)$ equals $\lambda(x)\Delta t + o(\Delta t)$ (respectively $\mu(x)\Delta t + o(\Delta t)$), for $\Delta t \downarrow 0$; the probability of two or more such arrivals is $o(\Delta t)$.

It should be observed that one might be able to improve the performance of the system by adapting $\lambda(x)$ and $\mu(x)$, which may be done by varying the purchase price and the selling price. A controller might wish to choose these rates in order to stay away from VOT level 0 (too many outdatings) or level 1 (too many unsatisfied demands). Control based on the number of items on the shelf might also seem natural, but it has the disadvantage that $\mathbf{K}$ is not a Markov process.

In Sect. 3 we shall study PIS with such Markovian Arrival Processes which we refer to as the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ model.

3 The $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ model

For the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS model let us consider the VOT process $\mathbf{V}$. Recall that items arrive with instantaneous rate $\lambda(x)$ when $V(t) = x$, demands arrive with instantaneous rate $\mu(x)$ when $V(t) = x$, demands arriving at an empty shelf leave unsatisfied, and any item that has not been taken within one time unit after arrival becomes outdated and must be scrapped. We note already here that for the calculation of the steady-state density of $\mathbf{V}$ we may set $\mu(x) = 0$ for $x > 1$, because a demand arriving at some time t when $V(t) > 1$ leaves unsatisfied and thus does not influence the virtual outdating time at t.

Also for this PIS the process $\mathbf{V}$ is regenerative with outdating times as cycle beginnings. The proof of Theorem 2 works almost verbatim, since at any outdating time T the items on the shelf are the ones that arrived in $[T - 1, T)$, the evolution of $\mathbf{V}$ after T only depends on their ages and on item and demand arrivals after T.

The mean cycle length is finite. To see this, note that since the probability of no demand arrivals in $[T, T + 1)$ is obviously positive, there is a positive probability, say $p > 0$, that the next outdating occurs before $T + 1$, so that the time until the next outdating is bounded by a geometric random variable with parameter p. Hence, the mean cycle length is smaller than $1/p$. It follows that $\mathbf{V}$ is stable.

Alternatively to the VOT process of the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS model, one could also speak of the workload process of a very specific $M/M/1$-type queueing system with state-dependent customer arrival rate $\mu(x)$ and state-dependent service rate $\lambda(x)$, in which a customer arriving when the current workload is w, say, causes the workload to jump upward to a new level that has distribution function

$$x \mapsto 1 - \mathrm{e}^{-[L(x)-L(w)]}, \quad x > w,$$

where $L(x) := \int_0^x \lambda(y)\mathrm{d}y$. There are no idle periods in this queue, because whenever the workload process hits zero, it immediately jumps upward to a new level that has distribution function $x \mapsto 1 - \mathrm{e}^{-L(x)}$, $x > 0$. We denote the workload (or virtual waiting time, VWT) process of this queueing system by $\tilde{\boldsymbol{V}}$. In the case $\mu(x) = 0$ for $x > 1$ there are no customer arrivals as long as the workload stays above 1. Then the queue workload process has the same law as the VOT process. This identity of distributions also occurs if we modify the queueing system by adding the feature that customers who upon their arrival encounter a workload larger than one do not enter the queue. View the bottom panel of Fig. 1 to see the equivalence between the PIS sample path and the familiar workload sample path in a single server queue.

For the queueing system we do not need the restriction $\mu \equiv 0$ on $(1, \infty)$. However, we make the following

Assumption (i) $\mu(x)$ is bounded and $\lim_{x\to\infty} \mu(x) = 0$,
Assumption (ii) $0 < a \leq \lambda(x) \leq b < \infty$ for some $a, b > 0$.

Then the arrival rate is for large x always smaller than the service rate minus $a/2$ (this ensures that $\lambda(x) - \mu(x)$ is bounded away from zero for large x, which will be needed in the proof of Theorem 4), and the marginal workload added at a customer arrival time is stochastically bounded by an $\exp(a)$ random variable. Hence the VWT process is a regenerative process, with the time periods between successive downcrossings of level 0 as cycles having finite mean cycle length. This implies that the VWT process $\tilde{\boldsymbol{V}}$ is stable.

Next we derive an integral equation for the steady-state density $\tilde{f}(\cdot)$ of $\tilde{\boldsymbol{V}}$, and use it to determine $\tilde{f}(\cdot)$ in closed form. Let $M(x) := \int_0^x \mu(w)\mathrm{d}w$.

Theorem 3 *Let $\tilde{f}(\cdot)$ be the steady-state density of the VWT process $\tilde{\boldsymbol{V}}$. Then*

$$\tilde{f}(x) = \int_0^x \mu(w)\mathrm{e}^{-[L(x)-L(w)]}\tilde{f}(w)dw + \tilde{f}(0)\mathrm{e}^{-L(x)}, \quad x \geq 0. \tag{6}$$

Proof We use level crossing theory (LCT) to derive the *Pollaczek-Khintchine* type equation (6). According to LCT, the rate of downcrossing any level x equals the rate of upcrossing that level. It is readily seen that the downcrossing rate equals $\tilde{f}(x)$. We now show that the righthand side of (6) is equal to the corresponding upcrossing rate. Given that the state is $w \in (0, x)$, a jump occurs with instantaneous rate $\mu(w)$, and it upcrosses level x if and only if there were no item arrivals (in the PIS), while the VWT decreased from x to w (which happens with probability $\exp -\{L(x) - L(w)\}$), where the same $\tilde{f}(\cdot)$ appears both in the lefthand side and in the righthand side by PASTA. Level x can also be upcrossed just after $\tilde{\boldsymbol{V}}$ hits level 0. In this case the probability to jump above x is $\exp\{-L(x)\}$. By LCT the rate of hitting level 0 is $\tilde{f}(0)$. The theorem is proved by deconditioning with respect to the position of $\tilde{\boldsymbol{V}}$ just before the jump. □

Remark 2 (i) The application of PASTA (Poisson Arrivals See Time Averages) is here somewhat delicate, because the arrival rate is state-dependent. However, by taking $\bar{\mu} := \sup_{x\geq 0}\mu(x) < \infty$ we can rewrite the integral in (6) as $\bar{\mu}\int_0^x \frac{\mu(w)}{\bar{\mu}}\mathrm{e}^{-[L(x)-L(w)]}\tilde{f}(w)\mathrm{d}w$, thus we can assume that arrivals occur according to a Poisson process with constant rate $\bar{\mu}$, while an arrival at t is admitted to the system with probability $\mu(w)/\bar{\mu}$ when $\tilde{V}(t) = w$.

(ii) Balance equations for the workload in single server queues with state-dependent arrival rate, service requirement rate and also service speed are discussed in detail in Section 3.2 of Bekker (2005). He uses Kolmogorov forward equations to derive such balance/integral equations. Apart from the technical issue of having deleted the idle periods, the above

theorem follows from his Formula (3.2). See also Bekker et al. (2004) for the case of state-dependent arrival rate and service speed.

Theorem 4 *The steady-state density $\tilde{f}(\cdot)$ of the VWT process $\tilde{\mathbf{V}}$ in the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ model above is given by*

$$\tilde{f}(x) = \frac{\mathrm{e}^{-[L(x)-M(x)]}}{\int_0^\infty \mathrm{e}^{-[L(y)-M(y)]}\mathrm{d}y}, \quad x \geq 0. \tag{7}$$

Proof Multiplying both sides of (6) by $\mathrm{e}^{L(x)}$ and introducing $g(x) := \tilde{f}(x)e^{L(x)}$, we obain the integral equation $g(x) = \int_0^x \mu(w)g(w)\mathrm{d}w + g(0)$. Differentiation yields a first-order differential equation, and we readily find that

$$\tilde{f}(x) = \tilde{f}(0)\mathrm{e}^{-[L(x)-M(x)]}. \tag{8}$$

The normalizing condition $\int_0^\infty \tilde{f}(x)\mathrm{d}x = 1$ gives (7). Note that, by Assumptions (i) and (ii), we have $\int_0^\infty \mathrm{e}^{-[L(y)-M(y)]}\mathrm{d}y < \infty$. □

Now let us return to the VOT process of the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS. Since demands arriving while $V(t) > 1$ have no bearing on the future evolution of $\mathbf{V}$ we may replace the demand rate function $\mu(x)$ by $\tilde{\mu}(x) = \mu(x)1_{(0,1)}(x)$. Recall that this means that no demands enter the PIS as long as the shelf contains no items. (In the queueing interpretation in this case, customers who arrive and see a workload larger than one, i.e., would have a waiting time larger than one, do not join the queue—this behavior is called balking in queueing terminology.)

To derive the steady-state density of the VOT process, the only adaptation in the balance equation in Theorem 3 is that we have to replace the upper integration value x by $x \wedge 1$:

$$f(x) = \int_0^{x\wedge 1} \mu(w)e^{-[L(x)-L(w)]}f(w)dw + f(0)e^{-L(x)}, \quad x \geq 0. \tag{9}$$

This yields

Corollary 1 *The steady-state density of the VOT process of the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS is given by*

$$f(x) = \begin{cases} ce^{-[L(x)-M(x)]}, & 0 \leq x \leq 1 \\ ce^{-[L(x)-M(1)]}, & x > 1 \end{cases}, \tag{10}$$

where

$$c = \left[\int_0^1 e^{-[L(x)-M(x)]}\mathrm{d}x + e^{M(1)}\int_1^\infty e^{-L(x)}\mathrm{d}x\right]^{-1}.$$

This result immediately gives us the following *key performance measures:*

(a) The long-run arrival rate of items is given by $\lambda = \int_0^\infty \lambda(x)f(x)\mathrm{d}x$.
(b) The long-run outdating rate equals $f(0) = c$.
(c) The long-run arrival rate of demands is given by $\mu = \int_0^\infty \mu(x)f(x)\mathrm{d}x$.
(d) The long-run rate of unsatisfied demands equals $\int_1^\infty \mu(x)f(x)\mathrm{d}x$, since all demands that find the shelf empty (i.e., the VOT above 1) depart unsatisfied.

Remark 3 The time between two successive outdatings of items is a busy period in the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}/1$ queue with customer balking when the waiting time exceeds 1. The lengths of successive busy periods, i.e., the times between successive downcrossings of level 0, are i.i.d.; the outdating times form a renewal process.

Remark 4 Consider a dual PIS (we speak of duality of type 1; later we also introduce a duality of type 2) in which every *demand* is willing to wait exactly one unit of time and an *item* that arrives and finds the waiting line in the PIS empty of waiting demands leaves immediately. The abandonment of items now is of the balking type, whereas the abandonment of the demands now is of the reneging type: each demand is admitted to the system, but it has a constant patience of length one and leaves unsatisfied when that patience runs out. A little reflection shows the following: (i) the instants of jumps are the times of item arrivals into the shelf, (ii) the jump sizes are the times between successive demand arrivals, (iii) a downcrossing of level 1 is a time of a first demand arrival into an empty waiting line, (iv) a downcrossing of level 0 is a time of a demand leaving unsatisfied, and (v) the abandonments of items occur when they arrive and find that $\mathbf{V} > 1$. From the above, we immediately conclude that in the dual model the times between unsatisfied demands have the same law as the times between outdatings in the primal PIS where $\lambda(x)$ and $\mu(x)$ are reversed.

We now consider a few special choices for $\lambda(x)$ and $\mu(x)$. In all examples, $f(\cdot)$ is the density of the VOT process, as given in Corollary 1.

Example 1 *The case* $\lambda(x) = \mu(x)$. In this case the constant c of the corollary becomes

$$c = \left[1 + e^{L(1)} \int_1^\infty e^{-L(x)} \mathrm{d}x\right]^{-1}.$$

The density f is constant on $(0, 1)$. The steady-state distribution is a mixture with weights c and $1 - c$ of the uniform distribution on $(0, 1)$ and the distribution on $(1, \infty)$ having density $x \mapsto e^{-L(x)} / \int_1^\infty e^{-L(u)} \mathrm{d}u$, $x > 1$.

Example 2 *The basic PIS* This special case was already treated in Kaspi and Perry (1983). From (10) we obtain for $\lambda \neq \mu$:

$$f(x) = \begin{cases} \dfrac{\lambda(\lambda - \mu)e^{-(\lambda-\mu)x}}{\lambda - \mu e^{-(\lambda-\mu)}}, & 0 \leq x \leq 1, \\ \dfrac{\lambda(\lambda - \mu)e^{-(\lambda x-\mu)}}{\lambda - \mu e^{-(\lambda-\mu)}}, & x > 1. \end{cases} \tag{11}$$

In the case $\lambda = \mu$ it follows from Example 1 that we get a mixture of the uniform distribution on $(0, 1)$ and the distribution of $1 + Y$ where Y is $\exp(\lambda)$-distributed. Once we know the steady-state density $f(\cdot)$, the rates λ^* and μ^* are easily determined. The outdating rate is

$$\lambda^* = f(0) = \frac{\lambda(\lambda - \mu)}{\lambda - \mu e^{-(\lambda-\mu)}},$$

and by the conservation law

$$\mu^* = \mu - \lambda + \frac{\lambda(\lambda - \mu)}{\lambda - \mu e^{-(\lambda-\mu)}}.$$

The rate of unsatisfied demands μ^* can also be obtained in an alternative way. By PASTA a demand is unsatisfied whenever it arrives at an empty system, which implies that $\mu^* = \mu \int_1^\infty f(x)\mathrm{d}x$.

One could also derive the distribution of the time between two successive outdatings by observing that this time is also the busy period in the $M/M/1$ queue with arrival rate μ and service rate λ, in which customers do not enter the system if they have to wait more than one time unit; this busy period has been studied in Kaspi and Perry (1983). In the dual model introduced in Remark 4, the same law holds with λ and μ reversed.

We next derive the generating function of the steady-state number of items on the shelf. Use that

$$K(t) = \begin{cases} 0, & V(t) > 1, \\ n, & V(t) \le 1 \text{ and } n-1 \text{ arrivals during the age of the oldest item.} \end{cases}$$

Define K and V as the number of items on the shelf and the VOT in steady state. We have $\lim_{t\to\infty} \mathbb{P}(V(t) > 1) = \mathbb{P}(V > 1)$ and (by dominated convergence) $\lim_{t\to\infty} \mathbb{E}\, z^{K(t)} = \mathbb{E}\, z^K$, $|z| < 1$, so that

$$\mathbb{E}\, z^K = \mathbb{P}(V > 1) + z \int_0^1 e^{-\lambda(1-w)(1-z)} f(w) \mathrm{d}w, \tag{12}$$

and by substituting (11) we obtain (for $\lambda \neq \mu$)

$$\begin{aligned} \mathbb{E}\, z^K &= \frac{(\lambda-\mu)e^{-(\lambda-\mu)}}{\lambda-\mu e^{-(\lambda-\mu)}} + z\int_0^1 e^{-\lambda(1-w)(1-z)} \cdot \frac{\lambda(\lambda-\mu)e^{-(\lambda-\mu)w}}{\lambda-\mu e^{-(\lambda-\mu)}} \mathrm{d}w \\ &= \frac{(\lambda-\mu)e^{-(\lambda-\mu)}}{\lambda-\mu e^{-(\lambda-\mu)}} \, \frac{\lambda z \mathrm{e}^{\lambda z-\mu} - \mu}{\lambda z - \mu}. \end{aligned} \tag{13}$$

Formula (13) is not contained in Kaspi and Perry (1983).

In many inventory applications, younger items on the shelf are worth more than older items. In Nahmias et al. (2004a) the value of the inventory in steady state is determined from an actuarial point of view. A function $R(x)$ is introduced that denotes the value of an item of age x. Let Z denote the steady-state total value of all items in the system. If $V > 1$ then the total value is zero. Otherwise, if $V = w \in (0, 1)$, let N denote the number of arrivals during the age $1-w$ of the oldest item, and denote their ages by $1-w-T_N, \ldots, 1-w-T_1$ with T_i the time between the arrival of the oldest item and the ith arrival. Then $Z = R(1-w) + \sum_{j=1}^{N} R(1-w-T_j)$. Since the item arrival process is Poisson, we can use a familiar property of the Poisson process: the arrival times of the N items are independent and uniformly distributed on $(0, 1-w)$. Hence

$$\mathbb{E}[\mathrm{e}^{-\alpha Z} | V = w] = \mathbb{E}[\mathrm{e}^{-\alpha R(1-w)} G^N(\alpha, 1-w)],$$

where

$$G(\alpha, u) := \int_0^u \mathrm{e}^{-\alpha R(y)} \frac{\mathrm{d}y}{u}.$$

A straightforward calculation now yields, with $f(w)$ being given in (11):

$$\begin{aligned} &\mathbb{E}[\mathrm{e}^{-\alpha Z}] \\ &= \mathbb{P}(V > 1) + \int_0^1 \mathrm{e}^{-\alpha R(1-w)} \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda(1-w)} \frac{(\lambda(1-w))^n}{n!} G^n(\alpha, 1-w)\, f(w) \mathrm{d}w \\ &= \frac{(\lambda-\mu)\mathrm{e}^{\mu-\lambda}}{\lambda-\mu\mathrm{e}^{\mu-\lambda}} + \int_0^1 \mathrm{e}^{-\alpha R(1-w)-\lambda(1-w)(1-G(\alpha,1-w))}\, f(w) \mathrm{d}w. \end{aligned} \tag{14}$$

Example 3 (*See* Nahmias et al., 2004b) Let $\lambda(x) = \lambda_0(x \wedge 1)^a$ and $\mu(x) = \mu_0[1-(1-x)^+]^b$, for some positive numbers λ_0, μ_0, a and b. In this case we have for $0 \le x \le 1$,

$$L(x) = \frac{\lambda_0 x^{a+1}}{a+1}, \qquad M(x) = \frac{\mu_0[1-(1-x)^{b+1}]}{b+1},$$

so that (with c following by normalization)

$$f(x) = \begin{cases} c \exp\left\{ \dfrac{\mu_0[1-(1-x)^{b+1}]}{b+1} - \dfrac{\lambda_0 x^{a+1}}{a+1} \right\}, & 0 \le x \le 1, \\ c \exp\left\{ \dfrac{\mu_0}{b+1} - \dfrac{\lambda_0}{a+1} - \lambda_0(x-1) \right\}, & x > 1. \end{cases}$$

Example 4 Let $\lambda(x) = \lambda_0$ and $\mu(x) = \frac{\mu_0}{1+bx}$, for some constant $b > 0$. Then $L(x) = \lambda_0 x$ and $M(x) = \frac{\mu_0}{b} \ln(1+bx)$, so that

$$f(x) = \begin{cases} c(1+bx)^{\mu_0/b} \mathrm{e}^{-\lambda_0 x}, & 0 \le x \le 1, \\ c(1+b)^{\mu_0/b} \mathrm{e}^{-\lambda_0 x}, & x > 1. \end{cases}$$

Example 5 Divide $(0, 1)$ in N intervals, and for the kth interval take $\lambda(x) = \lambda_k$ and $\mu(x) = \mu_k, k = 1, \ldots, N$. The expression for $f(\cdot)$ in (10) is easily evaluated. This might be relevant in the case of blood donations in a country or large city that is divided in a number of sections. In each section there are independent Poisson streams of blood donations and blood demands. By systematically adding and deleting sections a controller can adjust the effective arrival rates of blood items and of demands, to reduce the rates of outdated blood doses and of unsatisfied demand.

4 Applications

In this section we consider several variants of the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS model.

4.1 Variant 1: Hysteretic control

This application is taken from Perry and Posner (1990).

Model variant We consider a $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS model in which $\lambda(x) \equiv \lambda$, and in which the $\mu(x)$ function has the following special feature. There are two switchover levels a and b, such that $0 < a < b < 1$, and there are two possible demand rates: μ_L and $\mu_H > \mu_L$. Whenever a downcrossing of a occurs, the demand rate switches to μ_L. It keeps that value until level b is upcrossed; it then switches back to μ_H; etc. When one defines a cycle C as the period between two successive downcrossings of a, then each cycle consists of (first) a subcycle C_L with demand rate μ_L and then a subcycle C_H with demand rate μ_H. See Fig. 2. The VOT process $\mathbf{V}$ clearly is a regenerative process with regeneration epochs the successive downcrossing epochs of a. The difficulty in analyzing such a so-called *hysteretic policy* is that it is not a priori clear whether the demand rate is μ_L or μ_H when the VOT process takes a value between a and b.

Motivation A reason for using this policy could be that it will have fewer (possibly expensive) demand rate switches than if one would have no hysteresis but different demand rates in the three intervals $(0, a)$, (a, b) and $(b, 1)$. A possible application is found, e.g., in bloodbanks where a controller might wish to alternately include and exclude particular regions in submitting blood demands (and possibly also blood donations), in a hysteretic way.

The VOT process We construct two artificial regenerative processes from the VOT $\mathbf{V}$ such that in each cycle $\mathbf{V}$ is split into two separate VOT processes, $\mathbf{V}_L$ and $\mathbf{V}_H$. $\mathbf{V}_L$ ($\mathbf{V}_H$) is generated by deleting the C_H (C_L) periods from C and gluing together the C_L (C_H) periods.

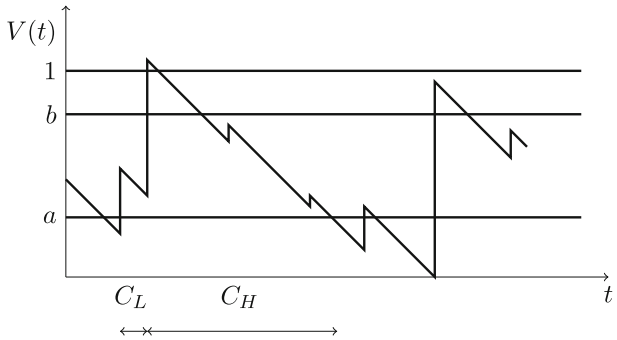

Fig. 2 Hysteresis with two switchover levels. The demand rate is μ_L in C_L, an interval from a downcrossing of a until the next upcrossing of b. The demand rate is μ_H in C_H, an interval from such an upcrossing until the next downcrossing of a

Key performance measures We derive the steady-state proper (conditional) densities $f_L(\cdot)$ and $f_H(\cdot)$ by using LCT, and by weighing them appropriately we obtain $f(\cdot)$:

$$f(x) = \frac{EC_L}{EC_L + EC_H} f_L(x) + \frac{EC_H}{EC_L + EC_H} f_H(x).$$

In the next lemma we first relate EC_L and EC_H to $f_H(a)$ and $f_L(\cdot)$, and then we derive integral equations for $f_L(x)$ and $f_H(x)$ which can be solved in a straightforward manner (first solving $f_L(x)$, then $f_H(x)$) and finally $f(x)$ follows using normalization; we leave the details to the reader (see also Perry and Posner, 1990). Define the constant ω by

$$\omega := \mu_L \int_0^b \mathrm{e}^{-\lambda(b-w)} f_L(w)\mathrm{d}w + \mathrm{e}^{-\lambda b} f_L(0). \tag{15}$$

Lemma 1

$$\frac{EC_L}{EC_L + EC_H} = \frac{f_H(a)}{f_H(a) + \omega}. \tag{16}$$

$$f_L(x) = \begin{cases} \mu_L \int_0^x \mathrm{e}^{-\lambda(x-w)} f_L(w)\mathrm{d}w + \mathrm{e}^{-\lambda x} f_L(0), & 0 \le x \le a, \\ \mu_L \int_0^x \mathrm{e}^{-\lambda(x-w)} f_L(w)\mathrm{d}w + \mathrm{e}^{-\lambda x} f_L(0) - \omega, & a < x \le b, \end{cases} \tag{17}$$

and

$$f_H(x) = \begin{cases} \mu_H \int_a^x \mathrm{e}^{-\lambda(x-w)} f_H(w)\mathrm{d}w + f_H(a), & a \le x \le b, \\ \mu_H \int_a^{x\wedge 1} \mathrm{e}^{-\lambda(x-w)} f_H(w)\mathrm{d}w + f_H(a)\mathrm{e}^{-\lambda(x-b)}, & b < x. \end{cases} \tag{18}$$

Proof We first prove (16). By LCT, ω is the rate of upcrossings of level b by $\mathbf{V}_L$. However, level b is upcrossed only once at the end of the cycle C_L. Thus, by LCT $\omega = \frac{1}{EC_L}$. Similarly, level a is downcrossed only once at the end of the cycle C_H, so that $f_H(a) = \frac{1}{EC_H}$. By renewal theory, $\mathbf{V}$ is a regenerative process whose successive cycles are composed of a C_L period followed by a C_H period. Thus, $\frac{EC_L}{EC_L+EC_H}$ and $\frac{EC_H}{EC_L+EC_H}$ are the steady state probabilities of the C_L and the C_H periods, respectively. Formula (16) follows.

The first relation in (17) is a simple level crossing identity, but the second one is more subtle. Notice that for $x \in (a, b]$, with probability 1, the number of upcrossings in every cycle of $\mathbf{V}_L$ minus the number of downcrossings equals 1, and that level b is upcrossed only once—at the end of C_L. By LCT the rate of downcrossings of level x is $f_L(x)$ and in steady state, the rate of upcrossings must be equal to the rate of upcrossings of level x minus ω,

which is the rate of upcrossings of level b. In terms of rates, we claim that during the L periods for every $a < x \leq b$,

$$\{\text{downcrossing rate of } x\} = \{\text{upcrossing rate of } x\} - \{\text{upcrossing rate of } b\}.$$

Note that as an intuitive consequence, the steady-state density $f_L(b) = 0$ and by substituting $x = b$ this is what we get.

Next consider (18). During the C_H periods there can be two types of upcrossings of level $x \geq a$. Firstly there are Poisson(μ_H) jumps; those jumps start at some level $w \geq a$. Secondly, at the start of each C_H period there is exactly one jump above level b, that causes the transition from C_L to C_H period. The latter event has rate $f_H(a)$.

For the region $x > 1$ the first integral runs until 1, since jumps are not admitted if they arrive to find $\mathbf{V}_H$ above level 1. □

4.2 Variant 2: Obsolescence

This section is mainly based on Perry and Stadje (2000b) (see also Perry, 1985); it extends the former paper by allowing state-dependent $\mu(\cdot)$ and by considering the joint distribution of the number of obsolescent and non-obsolescent items.

Model variant In this subsection we consider the $MPIS_{\mu(x)}/MPIS_{\lambda(x)}$ PIS model, where we restrict ourselves to $\lambda(x) \equiv \lambda$ while adding the following feature. So far, items that were not taken by a demand stayed on the shelf until reaching their fixed expiration age of 1. In the present subsection we also allow the possibility that an item becomes obsolete, i.e., is no longer of use for any demand, *before* the fixed unit expiration time. We assume that obsolescence depends on the age, occurring at rate $s(x)$ if $V(t) = x$, for $0 < x < 1$. That is, if the lifetime distribution of an item is $H(\cdot)$ (ignoring the fact that an item that reaches age one is scrapped), then $s(x) = \frac{\mathrm{d}H(x)}{\mathrm{d}x}/(1 - H(x))$.

Motivation In many medical and food applications of PIS, an item can deteriorate while on the shelf.

The VOT process By LCT we derive the following integral equation, of Pollaczek-Khintchine type, for $f(\cdot)$:

$$f(x) = \begin{cases} \int_0^x [\mu(w) + s(1-w)]\mathrm{e}^{-\lambda \int_w^x [1-H(1-y)]dy} f(w)\mathrm{d}w \\ \quad + f(0)\mathrm{e}^{-\lambda \int_0^x [1-H(1-y)]\mathrm{d}y}, & 0 \leq x < 1, \\ c\mathrm{e}^{-\lambda(x-1)}, & x \geq 1, \end{cases} \tag{19}$$

where the constant c can be found by the continuity condition $f(1-) = f(1+)$ and $f(0)$ can be found by the normalizing condition $\int_0^\infty f(x)\mathrm{d}x = 1$.

To derive (19), suppose that $\mathbf{V}$ is at level $w \in (0, 1]$ at some time t, i.e. the oldest item in the system is of age $1 - w$.

Considering Fig. 3, it is easily seen that there is an instantaneous upcrossing of level x at time t if and only if the following two events occur:

(i) The oldest item is removed either by an arriving demand or due to obsolescence. The rate of this to happen is $\mu(w) + s(1 - w)$.
(ii) None of the items that have entered the system during the time interval $(t - (x - w) - (1 - x); t - (1 - x)]$ is still present at time t (as otherwise $V(t)$ would still be less than x after the jump, see Fig. 3). Conditioning on the number of arrivals in an interval, the

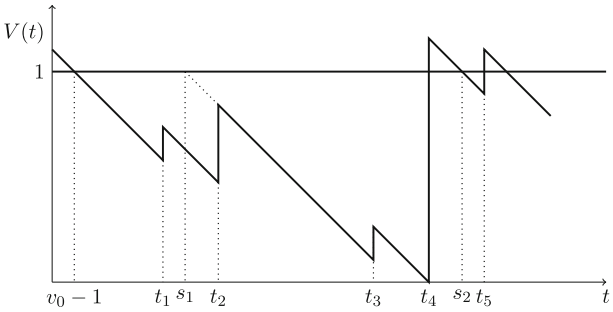

Fig. 3 A typical sample path of the VOT process **V** in the case of random item lifetimes. $V(0) = v_0 > 1$ so that the first item arrives at time $v_0 - 1$. The jumps at times t_1, t_2, t_3, t_5 can be due to a demand or because the life of the oldest item present ends at that time. The item that arrived at s_1 becomes the oldest item after the jump at t_2. At t_4 the oldest item becomes outdated, leaving an empty shelf behind. A new item arrives at s_2, and it is removed at t_5 either by demand or because its life ends

arrival times are independent and uniformly distributed on an interval of length $x - w$, so that the probability that all these items are gone at time t is equal to

$$\sum_{n=0}^{\infty} \mathrm{e}^{-\lambda(x-w)} \frac{[\lambda(x-w)]^n}{n!} \left(\frac{\int_0^{x-w} H(1-w-u)\mathrm{d}u}{x-w} \right)^n = e^{-\lambda \int_w^x [1-H(1-y)]dy}.$$

Key performance measures. Let us first determine $f(\cdot)$ from (19). Introducing $g(x) := f(x) \exp(\lambda \int_0^x [1 - H(1-y)]\mathrm{d}y)$ for $x \in [0, 1)$, the first part of (19) reduces to

$$g(x) = \int_0^x [\mu(w) + s(1-w)]g(w)\mathrm{d}w + g(0), \quad 0 \le x < 1.$$

Differentiation w.r.t. x yields $g'(x) = [\mu(x) + s(1-x)]g(x)$, so

$$g(x) = g(0)\mathrm{e}^{\int_0^x [\mu(y)+s(1-y)]\mathrm{d}y},$$

and hence

$$f(x) = f(0)\mathrm{e}^{\int_0^x [\mu(y)+s(1-y)-\lambda(1-H(1-y))]\mathrm{d}y}, \quad 0 \le x < 1. \tag{20}$$

For $x \ge 1$ it is clear that $f(x) = c\mathrm{e}^{-\lambda(x-1)} = f(1)\mathrm{e}^{-\lambda(x-1)}$, as no jumps can occur in the VOT process for $x > 1$. $f(1)$ is expressed in $f(0)$ via (20), and finally $f(0)$ follows from the normalizing condition.

As in previous model variants, various performance measures can be obtained once $f(\cdot)$ is known. Firstly, the outdating rate equals $f(0)$. Secondly we focus on the rate of unsatisfied demands. Using the conservation law (3) with $\mu = \int_0^\infty \mu(x) f(x)\mathrm{d}x$ and $\lambda^* = f(0)$, the unsatisfied demand rate is found to be $\mu^* = \int_0^\infty \mu(x) f(x)\mathrm{d}x - \lambda + f(0)$. Finally, we determine the (generating function of the) steady-state joint distribution of the number of non-obsolescent items K_{NO} in the system and the number of items K_O that, during the age of the oldest item, have left prematurely due to obsolescence. The reasoning in Sect. 3 leading to (12) is still valid: the system is empty at time t iff $V(t) > 1$, and if it is not empty and $V(t) = w \in (0, 1)$, then the number of arrivals during the age $1 - w$ of the oldest item is Poisson($\lambda(1-w)$). The latter Poisson process is split into two independent Poisson processes, with rates $\lambda \int_0^{1-w} (1 - H(y))\mathrm{d}y$ for the ones that have not become obsolete and $\lambda \int_0^{1-w} H(y)\mathrm{d}y$ for the ones that have become obsolete. Hence we have, with $\mathbb{P}(V > 1) =$

$1 - F(1)$:

$$\mathbb{E}[z_1^{K_{NO}} z_2^{K_O}] = 1 - F(1) + z_1 \int_0^1 \exp[-\lambda(1-z_1)\int_0^{1-w}(1-H(y))\mathrm{d}y - \lambda(1-z_2)\int_0^{1-w} H(y)\mathrm{d}y] f(w)\mathrm{d}w. \tag{21}$$

Notice that $z_1 = z_2 = 1$ gives (12), and that there is a factor z_1 in front of the w-integral corresponding to the item with the oldest age (during whose lifetime the other items have arrived).

We end this subsection by briefly discussing some choices for $H(\cdot)$. For these choices, the above formulas may be somewhat simplified; in particular, $s(x)$ and integrals like $\int_0^{1-w} H(y)\mathrm{d}y$ can be evaluated.

(i) If $H(\cdot)$ is uniform on $(0, a)$ for some $a > 1$, then $s(x) = 1/(a - x)$.
(ii) If $H(x) = 1 - (1 - x)^2$ for $0 \le x \le 1$ (triangular density), then $s(x) = 2/(1 - x)$.
(iii) If $H(x) = 1 - \mathrm{e}^{-\eta x}$, $x > 0$, then $s(x) = \eta$.
(iv) If $H(x) = x/(1 + x)$, $x > 0$ (Pareto), then $s(x) = 1/(1 + x)$.
(v) The case $H(x) = 0$ for $x < 0$, $H(x) = q$ for $0 \le x < a$ and $H(x) = 1$ for $a \le x < b < 1$ is somewhat different. Here the maximum shelf life alternates between two constants a, b, with $0 < a < b < 1$. If an item has reached age a, it is inspected. With probability p it is found to be good and then $b - a$ time units are added to its expiration date. But with probability $q = 1 - p$ it is found to be unfit for issuance and it is removed from the shelf. Observe that no item stays longer than b in the system; we can now take $V(t) = b - A(t)$. We refer to Perry (1999) for a detailed discussion of this case, when $\mu(w) \equiv \mu$. In the case of general $\mu(\cdot)$, the balance equations are readily seen to be the following (notice that there are minor differences with Theorem 1 of Perry (1999), where in a few places a should have been replaced by $b - a$):

$$\begin{aligned} f(x) &= \int_0^x \mu(w)\mathrm{e}^{-\lambda p(x-w)} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda p x}, \quad 0 \le x \le b-a, \\ f(x) &= \int_0^{b-a} \mu(w)\mathrm{e}^{-\lambda p(b-a-w)-\lambda(x-(b-a))} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda p(b-a)-\lambda(x-(b-a))} \\ &\quad + \int_{b-a}^{x\wedge b} \mu(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + q f((b-a)+)\mathrm{e}^{-\lambda(x-(b-a))}, \quad x > b-a. \end{aligned} \tag{22}$$

Notice that $b - a$ is a point of discontinuity for $\mathbf{V}$, since $pf((b-a)+) = f((b-a)-)$. This is intuitive because the proportion between the downcrossing rates of levels $(b-a)-$ and $(b-a)+$ is p. The first equation is trivially solved by first multiplying both sides by $\mathrm{e}^{\lambda p x}$ and then differentiating. The second equation is solved by distinguishing between $x \le b$ and $x > b$. In the latter case, $f(x) = C\mathrm{e}^{-\lambda x}$ for some constant C. In the former case, multiplying both sides by $\mathrm{e}^{\lambda x}$ and differentiating results in a simple first-order differential equation.

4.3 Variant 3: Risk management

An event of unsatisfied demand could have serious consequences, e.g., in organ transplant and blood bank settings. Hence a controller may want to avoid unsatisfied demands, or even risky situations, as much as possible. In this section we briefly discuss four possible strategies

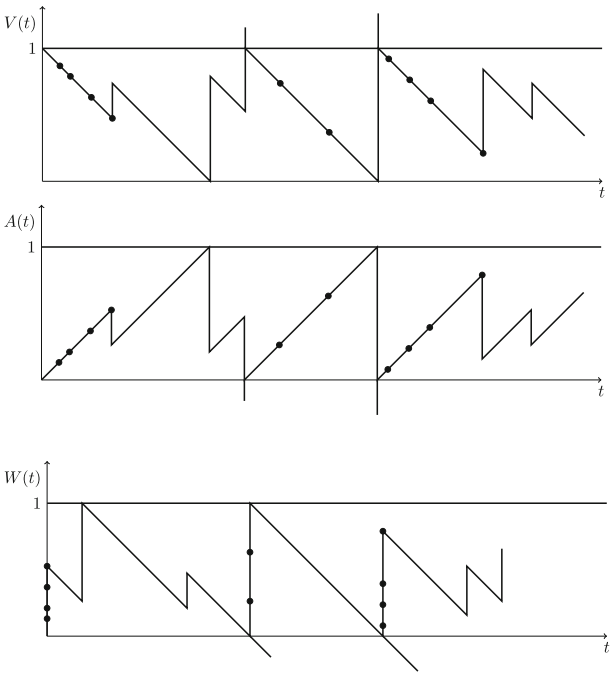

Fig. 4 A typical sample path of the VOT process **V** (top panel), the age process **A** (middle panel) and the transformation into the process **W** (bottom panel), for the case of special deliveries with $n_0 = 4$. The dots correspond to demand arrivals that take one of the instant delivery items

to accomplish this. In each case we assume that item arrivals (regular ones, see below) are Poisson (μ) and demand arrivals are Poisson(λ).

Model variant (i): Outsourcing; cf. Bar-Lev et al. (2005).
In this variant, the possibility of unsatisfied demands is excluded by introducing an alternative source of fresh items that is completely reliable and delivers with zero delay. When the shelf becomes empty, the controller places an order at this source, and it instantaneously delivers a batch of n_0 items. In the **V** process, after each upcrossing of level 1, the next n_0 demands do not cause jumps in **V** as long as the age of these items is less than 1. If some of the last of the n_0 items become outdated at age 1, the next demand does cause an exp(λ) jump. However, if that jump happens to be larger than 1, it is cut off in the **V** process by 1 and again n_0 items are ordered.

The VOT process A typical realization of the **V** process is shown in the top panel of Fig. 4. Note that the demand process is a Poisson process with rate μ, but the jump process of **V** is not a Poisson process. In order to cope with this situation we apply a duality argument, called duality of type 2, in which we first look at the age process **A**, with $A(t) = 1 - V(t)$. The original process ($V(t)$ in the top panel of Fig. 4, with steady-state density $f(\cdot)$) is a regenerative process whose cycle is the time between two downcrossings of level 1. Construct a new process **W**, with steady-state density $f_W(\cdot)$, in the following way. Every trajectory of slope 1 in **A** becomes a jump to the same level in **W** and every negative jump in **A** becomes a trajectory to the same level in **W**. Now, the **A** process is a regenerative process whose cycles are the times between two successive downcrossings of level 0. The sample path of **W** is the

same as that of a finite dam (queueing) model in which the distribution of the first jump in a cycle is different from that of the other jumps. The first jump size is the sum of n_0 exp(λ) distributed random variables, hence Erlang(n_0, λ), but the jump is truncated (if necessary) at 1. All other jumps are exp(λ) distributed. Also, the idle periods are deleted.

Key performance measures For the steady-state density $f(\cdot)$ of $\mathbf{V}$ we have, for all $0 \leq x \leq 1$:

$$f_W(x) = f(1-x) = \lambda \int_0^x \mathrm{e}^{-\mu(x-w)} f_W(w)\mathrm{d}w + f_W(0) \sum_{j=0}^{n_0-1} \frac{\mathrm{e}^{-\mu x}(\mu x)^j}{j!}. \tag{23}$$

By introducing $g(x) := \mathrm{e}^{\mu x} f_W(x)$ and differentiating w.r.t. x, we obtain a simple first-order differential equation which is readily solved (we leave the details to the reader):

$$g'(x) = \lambda g(x) + f_W(0)\mu \sum_{j=1}^{n_0-1} \frac{(\mu x)^{j-1}}{(j-1)!}.$$

Since $f(x) = f_W(1-x)$, we now also have $f(\cdot)$, and one can subsequently obtain other performance measures. In particular, the conservation law becomes

$$\mu = \lambda + n_0 f(1) - f(0)\,\mathbb{E}\,J, \tag{24}$$

where J is the number of items that are outdated when $\mathbf{V}$ reaches level 0, and this yields $\mathbb{E}\,J$. The reasoning behind (24) is the following. Since there are no unsatisfied demands, the satisfied demand rate equals μ. This should equal the item input rate $\lambda + n_0 f(1)$, minus the rate of outdated items.

Some interesting performance measures which were not discussed in Bar-Lev et al. (2005) are: (i) The distribution of the number of items which are on the shelf in steady state. Here there could be several oldest items. (ii) The distribution of the number of items that are outdated when $\mathbf{V}$ reaches level 0. (iii) The distribution of the busy period; it can be obtained via a martingale argument.

Model variant (ii): Urgency Classes; cf. Bar-Lev et al. (2005).
In this variant it is not possible to place additional orders. The incoming demands are classified into different categories of urgency. For simplicity, assume that there are two such categories whose demand arrival times form independent Poisson processes of intensities $\mu_1(w)$ and $\mu_2(w)$, respectively; item arrivals are Poisson(λ). One possible policy is to satisfy high-urgency (type 1) demands whenever possible (i.e., if the system is not empty) and less urgent demands (type 2) only if there are at least $m_0 > 1$ items on the shelf. An undesirable aspect of this policy is that it does not take the lifetime of the oldest item into account. For example, under this policy the oldest item will not be used for a less urgent demand even if its residual lifetime is very short, so its outdating is imminent. To avoid this drawback, we propose the following policy refinement. Fix $\gamma \in (0, 1)$ and an integer $m_0 > 1$. A demand of type 1 is satisfied if and only if the system is not empty; a demand of type 2 is satisfied if and only if there are at least m_0 items in the system or the shelf age of the oldest item is at least $1 - \gamma$. Any demand of type 1 or 2 that is not immediately satisfied is lost. This model was studied in Bar-Lev et al. (2005) for the case $\mu_1(w) \equiv \mu_1$, $\mu_2(w) \equiv \mu_2$.

The VOT process It is readily seen that the VOT density $f(\cdot)$ satisfies the following balance equations:

$$f(x) = \int_0^x (\mu_1(w) + \mu_2(w))\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, \quad 0 \le x < \gamma,$$

$$f(x) = \int_0^x \mu_1(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + \int_0^\gamma \mu_2(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w$$
$$+ \sum_{i=m_0-1}^{\infty} \mathrm{e}^{-\lambda(1-x)} \frac{(\lambda(1-x))^i}{i!} \int_\gamma^x \mu_2(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, \quad \gamma \le x < 1,$$

$$f(x) = \int_0^1 \mu_1(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + \int_0^\gamma \mu_2(w)\mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, \quad x \ge 1.$$

The first equation can be solved easily by multiplying both sides by $\mathrm{e}^{\lambda x}$ and differentiating. The third equation shows that $f(x) = c\mathrm{e}^{-\lambda x}$ with c some constant. The second equation can be formally solved via the technique of Picard iteration; this method is discussed in some detail in Sect. 5. In the, not unrealistic, case that the ratio $\mu_1(w)/\mu_2(w)$ is constant, the equation can be solved more explicitly, using the same approach as for the first equation. Possible extensions which were not treated in Bar-Lev et al. (2005) are: (i) the distribution of the number of items on the shelf; taking lead times for special orders into account.

Model variant (iii): Risk of an empty shelf. In this third model variant (which has not been considered before) we assume that the controller carries out the following policy. When only one item is on the shelf when a demand arrives, the condition of that demand (e.g., a person requiring blood or an organ) is inspected. The demand is diagnosed with probability p as urgent, and then the demand is immediately satisfied. If it is diagnosed as non-urgent, the demand is released unsatisfied.

The VOT process The balance equation in this case is easily seen to be

$$f(x) = \begin{cases} \mu p \left[\int_0^x \mathrm{e}^{-\lambda(1-w)} f(w)\mathrm{d}w + \lambda(1-x)\int_0^x \mathrm{e}^{-\lambda(1-w)} f(w)\mathrm{d}w\right] & 0 \le x \le 1, \\ +\mu \int_0^x [\mathrm{e}^{-\lambda(x-w)} - \mathrm{e}^{-\lambda(1-w)}] f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, & \\ \mu p \int_0^1 \mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, & x > 1. \end{cases}$$

Indeed, for $0 \le x \le 1$ let S be the jump size, which is the generic time between arrivals at the shelf. By conditioning on both $V = w$ and on the number N_I of items seen by the arriving demand we get for $0 \le x \le 1$:

$$f(x) = \mu p \int_0^x \mathbb{P}(S > 1 - w \mid N_I(1-w) = 0)\, \mathbb{P}(N_I(1-w) = 0) f(w)\mathrm{d}w$$
$$+\mu p \int_0^x \mathbb{P}(x - w \le S < 1 - w \mid N_I(1-w) = 1)\, \mathbb{P}(N_I(1-w) = 1) f(w)\mathrm{d}w$$
$$+\mu \int_0^x \sum_{n=2}^{\infty} \mathbb{P}(x - w \le S < 1 - w \mid N_I(1-w) = n)\, \mathbb{P}(N_I(1-w) = n) f(w)\mathrm{d}w.$$

The first conditional probability given $N_I(1-w) = 0$ is equal to 1, since the events $\{N_I(1-w) = 0\}$ and $\{S > 1-w\}$ are equivalent events. In the second line S, given $\{N_I(1-w) = 1\}$, is uniformly distributed on $(0, 1-w)$ and in the third line S given $\{N_I(1-w) = n\}$ (for $n \ge 2$) is stochastically equal to the minimal order statistic taken from a uniform distribution on $(0, 1-w)$. Note that the second and the third lines are separated from each other, since the demand rates are μp and μ respectively. The above equation thus becomes, for $0 \le x \le 1$:

$$f(x) = \mu p \int_0^x \mathrm{e}^{-\lambda(1-w)} f(w)\mathrm{d}w + \mu p \int_0^x \tfrac{1-x}{1-w} \mathrm{e}^{-\lambda(1-w)} \lambda(1-w) f(w)\mathrm{d}w$$
$$+\mu \int_0^x \sum_{n=2}^{\infty} \left(\tfrac{1-x}{1-w}\right)^{n-1} \tfrac{\mathrm{e}^{-\lambda(1-w)}[\lambda(1-w)]^{n-1}}{(n-1)!} f(w)\mathrm{d}w.$$

The proof is completed after some simple algebra and the fact that for $x > 1$ an upcrossing means that an arriving demand sees only one item on the shelf and is satisfied by it.

Remark 5 A weakness of the policy is that it does not take the age of the oldest item on the shelf into account. Suppose that, when a demand arrives, the age of the oldest item is close to 1. If the demand is not satisfied by the oldest item, the item will become obsolete very soon anyway. Thus, it would be reasonable to issue the item regardless of the demand's condition. Accordingly, it is natural to fix a certain switchover level, say a, such that if the age of the item is greater than $1 - a$ (alternatively, $V < a$), the demand will be satisfied even if it is the only item present on the shelf. We distinguish between two cases: if just before a moment of demand arrival, there is one item on the shelf the demand is satisfied by the item with probability p_1 and the shelf becomes empty. But if just before a moment of demand arrival, there are at least two items on the shelf the demand is satisfied by the oldest item with probability p_2 and immediately after the issuance only one item is left on the shelf. It is not hard to derive the integral equation for density $f(\cdot)$ for this adaptation.

Finally we refer to Balcioglu et al. (2008) for a risk management study of a basic PIS with a demand rate that is either high or low depending on the value of $\mathbf{V}$.

Model variant (iv): Optional shelf life In some cases, it might be disastrous to have an unsatisfied demand. Keeping the blood bank example in mind, one can imagine that there are situations in which it is opportune to slightly extend the fixed maximum shelf life time when the alternative—an unsatisfied demand—is likely to have worse consequences. Accordingly, we propose a model variant in which the system controller is allowed to lengthen the expiration date of items under certain conditions. Our easy-to-apply control policy consists of the following simple rules.

(1) When the oldest item on the shelf is the *only* item on the shelf and reaches age 1, an additional amount of time a is added to its life time.
(2) If in the next a time units a demand arrives, before a fresh item, then the item with extended life time satisfies this demand.
(3) If in the next a time units a fresh item arrives on the shelf, before a demand, then the item with extended life time is removed from the shelf.
(4) If in the next a time units no demand and no fresh item arrive, then the item with extended life time is removed from the shelf.

It is readily verified that this gives rise to the following balance equations for the VOT, when we assume Poisson(λ) item arrivals and Poisson(μ) demand arrivals; see also Fig. 5.

$$\begin{aligned} f(x) &= (\lambda+\mu)F(x) + f(0), \quad 0 \le x < a, \\ f(x) &= (\lambda+\mu)F(a) + f(0) + \mu \int_a^x \mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w \\ &\quad + f(a+)[\mathrm{e}^{-\lambda(x-a)} - \mathrm{e}^{-\lambda}], \quad a \le x < 1+a, \\ f(x) &= \mu F(a)\mathrm{e}^{-\lambda(x-a-1)} + f(0)\mathrm{e}^{-\lambda(x-a-1)} \\ &\quad + \mu \int_a^{1+a} \mathrm{e}^{-\lambda(x-w)} f(w)\mathrm{d}w, \quad x \ge 1+a. \end{aligned} \tag{25}$$

The three terms on the right in the first line of (25) correspond respectively to the following events, when the oldest item has age $\in (1, 1+a]$: (i) a fresh item arrives (cf. t_4 in the figure), (ii) a demand arrives (cf. t_5); and (iii) outdating of the oldest item occurs (cf. t_6). The term in the third line also deserves to be mentioned. It represents the event in which a

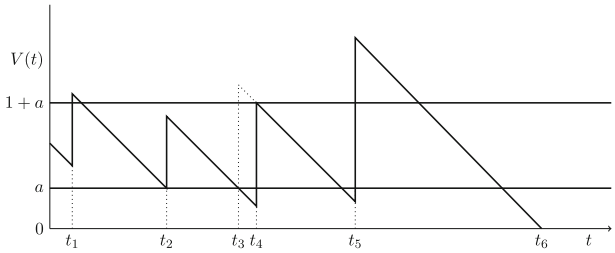

Fig. 5 A typical sample path of the VOT process **V** in the case of optional shelf life a. t_1 is the time of a demand arrival; it is satisfied by the oldest item, whose age is less than 1. t_2 is a time at which a is reached from above. The age of the oldest item reaches 1 and it is removed because it is not the only item on the shelf. t_3 is another time at which a is reached from above; now it is downcrossed, because the oldest item is the only item present. t_4 marks the arrival of a fresh item, that instantaneously replaces the oldest (and outdated) item. t_5 is the arrival time of a demand that is satisfied by an item with age > 1; subsequently the shelf becomes empty. t_6 is the time of outdating of an item with age $1 + a$

is reached from above but not downcrossed; the oldest item reaches age 1 and is removed because there is a younger item on the shelf (cf. t_2 in the figure). In the fourth line we have a term $\mu F(a)\mathrm{e}^{-\lambda(x-a-1)}$ instead of an integral from 0 to a, because any jump from below a will exceed level $a + 1$ and then continue for an exp(λ) distributed amount, regardless of the precise level it jumped from. Similarly for the $f(0)$ term. The other terms are self-explanatory and/or the same as in the case $a = 0$.

Finally we observe that it is straightforward to determine $f(\cdot)$ from (25) (in particular, $f(x)$ is proportional to $\mathrm{e}^{(\lambda+\mu)x}$ for $x \in [0, a)$ and to $\mathrm{e}^{-\lambda x}$ for $x \geq 1 + a$); we leave the details to the reader.

Remark 6 It should be noticed that the density $f(\cdot)$ has two points of discontinuity: (i) $f(a+) = f(a-)\mathrm{e}^{\lambda}$ (cf. t_2) and (ii) $f((a+1)+) - f((a+1)-) = -\lambda F(a)$ (cf. t_4).

The conservation law For the present model variant, we have

$$\lambda - f(0) - [f(a+) - f(a-)] - [f((a+1)-) - f((a+1)+)] = \mu F(a+1).$$

Both sides represent the rate of the satisfied demands. This is obvious for the righthand side. The lefthand side is the arrival rate λ minus the rate of the three outdating components: $f(0)$ for items that become outdated at age $1+a$; $f(a+)-f(a-) = f(a+)[1-\mathrm{e}^{-\lambda}]$ for items that are removed at age 1 and left a non-empty shelf; and $f((a+1)-) - f((a+1)+) = \lambda F(a)$ for items that are removed at an age between 1 and $1 + a$ due to the arrival of a fresh item.

Model variant (v): Secondary products In some real-world applications items are not scrapped when reaching their formal, preset expiration age, but are taken for secondary use. In a blood bank, expired blood portions that are no longer suitable for transfusions could be processed to other medical products; similarly, in a PIS for certain foodstuff, e.g. fruits, products that are not any more considered to be marketable could be transformed into others, e.g. juice or jam.

A first attempt at modeling this situation is a PIS with two storage places, say the upper shelf and the lower shelf for the primary and the secondary items, respectively. Let the maximal shelf life of the primary product again be 1. When an item reaches this age it is immediately transferred to the lower shelf where its maximal shelf age is b. There are two independent Poisson arrival streams of demands with rates μ_0 and μ_1, respectively. This and more refined models were analyzed in Perry and Stadje (2000a).

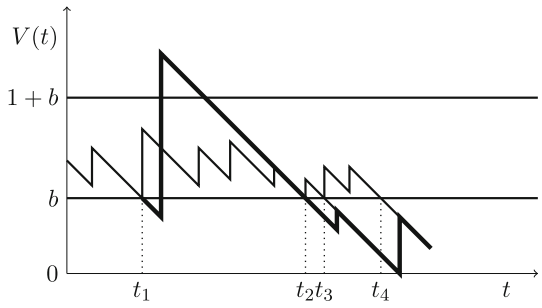

Fig. 6 A typical sample path of the two VOT processes in the case of primary and secondary products

In this PIS there are two coupled VOT processes, one for the upper shelf and one for the lower shelf. A typical sample path is shown in Fig. 6.

The times t_i are the times of a removal in the upper VOT (times of outdatings, hence transfers from the upper to the lower shelf), and the intervals between these times are also the jump sizes in the lower (bold) VOT. Note that when the lower shelf is empty, there is only one common VOT for both the upper and the lower shelves.

It is easy to conclude that the upper VOT has the same law as the workload process in the $M/M/1+D$ $(D=1)$ queue in which the idle periods are deleted (this is the VOT of the basic model), while the lower VOT behaves like the workload process in the $M/G/1+D$ $(D=b)$ queue in which the idle periods are deleted and G denotes the distribution of inter-outdating times of the upper shelf. The marginal laws of both systems can be determined; the analysis of the $M/M/1+D$ case is described in the basic PIS above and that of the $M/G/1+D$ case in Sect. 5 below.

Another interpretation of this PIS with two shelves is an inventory subject to two types of demands: *normal demands* and *urgent demands*. The urgent demands have priority over the normal ones and require young items. Accordingly, the controller prepares an infrastructure of two shelves such that the items on the upper shelf are young and are supposed to satisfy only the urgent demands, while the items on the lower shelf are older and are supposed to satisfy only the normal demands. Whenever an item on the upper shelf reaches its maximum admissible age, it is moved to the lower shelf. In case a normal demand arrives at an empty lower shelf, it leaves unsatisfied even if the upper shelf is not empty. (The rationale behind this policy is that the controller wants to keep young items for potential future arrivals of urgent demands.) This model clearly coincides mathematically with the PIS above for primary and secondary products.

5 PIS Models with renewal arrival processes

This section is devoted to the $MPIS_{\mu(x)}/G$ PIS, a model in which the item arrival process is a renewal process, while the demand arrival process is a Poisson process with rate $\mu(x)$ when the age is $1-x$. Let $G(\cdot)$ denote the distribution of the i.i.d. item interarrival times. As before, we assume that each item has a usable lifetime of one time unit and that, upon arrival, a demand removes the oldest item on the shelf—leaving unsatisfied if the shelf is empty. An item that has not been taken within one time unit of arrival becomes outdated. Finally, the arrival processes of items and demands are again assumed to be independent.

This model is studied in Kaspi and Perry (1984), for the case $\mu(w) \equiv \mu$. In the present section we first derive an integral equation for the steady-state density $f(\cdot)$ of the VOT

process $\mathbf{V}$ for general $\mu(\cdot)$, and we subsequently outline its solution. Thereafter, we restrict ourselves to $\mu(w) \equiv \mu$, in which case the solution of the integral equation becomes more explicit. We also express the distribution of the number of items on the shelf into $f(\cdot)$. Three natural applications of $MPIS/G$ and M/G PIS models will be discussed in Sect. 6.

Again consider the age process $\mathbf{A}$ and, in particular, the VOT process $\mathbf{V}$, where $V(t) = 1 - A(t)$. Since we allow no patience of demands, the steady-state density $f(\cdot)$ of $\mathbf{V}$ always exists. To determine it, we can again take $\mu(w) = 0$ for $w > 1$, because a demand arriving at some time t when $V(t) > 1$ leaves unsatisfied and thus does not influence the virtual outdating time at t. A level crossing argument readily yields that $f(\cdot)$ satisfies the following integral equation:

$$f(x) = \int_0^{x\wedge 1} \mu(w)(1 - G(x - w))f(w)\mathrm{d}w + f(0)(1 - G(x)), \quad x \geq 0. \tag{26}$$

Notice that the case $G(x) = 1 - \mathrm{e}^{-\lambda x}$ was treated in Corollary 1 (in fact, we there allowed $\lambda(\cdot)$); for that exponential case, a straightforward solution procedure is to multiply both sides by $\mathrm{e}^{\lambda x}$, after which differentiation results in a simple first-order differential equation. That approach breaks down for general $G(\cdot)$. However, there is a standard (albeit somewhat formal) solution procedure, Picard iteration, for such Volterra integral equations of the second kind (see, e.g., Chapter I of Mikhlin (1957)). We now outline that procedure. Let $K(x, w) := \mu(w)(1 - G(x - w))$ for $0 \leq w \leq x$. Then (26) becomes: $f(x) = \int_0^x K(x, w)f(w)\mathrm{d}w + cK(x, 0)$, where $c := f(0)/\mu(0)$. Iteration yields:

$$\begin{aligned} f(x) = {} & cK(x, 0) + c\int_0^x K(x, w)K(w, 0)\mathrm{d}w \\ & + c\int_0^x K(x, w)\int_0^w K(w, y)K(y, 0)\mathrm{d}y\mathrm{d}w + \cdots \end{aligned} \tag{27}$$

Introducing $K_1(x, w) := K(x, w)$ and $K_n(x, w) := \int_w^x K_{n-1}(x, z)K(z, y)\mathrm{d}z$ for $n = 2, 3, \ldots$, one can verify that $f(x)$ is given by the following convergent sum:

$$f(x) = c\sum_{n=1}^{\infty} K_n(x, 0), \quad x \geq 0. \tag{28}$$

Now take $\mu(w) \equiv \mu$ for $0 \leq w \leq 1$, and $\mu(w) = 0$ otherwise. Kaspi and Perry (1984) exploit the fact that the VOT process now coincides with the above-mentioned workload process in an $M/G/1$ queue with restricted accessibility (an $M/G/1 + D$ queue) and deleted idle periods. Using a result of Daley (1964) for the so-called finite dam, they find the distribution $F(x) = \int_0^x f(w)\mathrm{d}w$. With $1/\lambda$ the mean of $G(\cdot)$ and $n*$ denoting an n-fold convolution, their Formula (3.21) states that

$$\begin{aligned} F(x) &= \frac{\frac{\lambda}{\mu}\sum_{n=0}^{\infty}\int_{0+}^{x}\frac{[-\mu(x-u)]^n}{n!}\mathrm{e}^{\mu(x-u)}\mathrm{d}G^{n*}(u)}{\sum_{n=0}^{\infty}\int_{0-}^{1}\frac{[-\mu(1-u)]^n}{n!}\mathrm{e}^{\mu(1-u)}\mathrm{d}G^{n*}(u)}, \quad 0 < x \leq 1, \\ F(x) &= \mu\int_{y=0}^{x}\int_{u=0}^{1}(1 - G(y - u))f(u)\mathrm{d}u\mathrm{d}y, \quad x > 1. \end{aligned} \tag{29}$$

When $G(\cdot) \sim \exp(\lambda)$, this expression is readily seen to simplify to (11).

Repeating an argument that was already used for that exponential $G(\cdot)$ case, we can also find the distribution of the steady-state number of items on the shelf:

$$\begin{aligned}\mathbb{P}(K=n) &= \int_0^1 [G^{(n-1)*}(1-w) - G^{n*}(1-w)]f(w)\mathrm{d}w, \quad n=1,2,\ldots,\\ \mathbb{P}(K=0) &= 1-F(1).\end{aligned} \tag{30}$$

Remark 7 It would also be interesting to study PIS models in which the item arrival process is Poisson but the *demand* arrival process is not Poisson but a renewal process. The LCT approach now breaks down: jumps do not occur according to a Poisson process, so PASTA does not hold and the VOT at a jump epoch is not the same as the steady-state VOT. We briefly sketch an approach that one can follow in this case. Just like in Sect. 4.3, see in particular Fig. 4, one could construct a new process $\mathbf{W}$ from $\mathbf{A}$ by replacing upward trajectories with slope 1 by upward jumps to the same level, and downward jumps by downward trajectories of slope -1 to the same level. This so-called duality of type 2 results in an artificial $MPIS/G$ process $\mathbf{W}$ with the same steady-state law as $\mathbf{V}$. The balance equations for that $MPIS/G$ model can be derived using LCT.

6 Renewal arrivals: Three variants

In this section we discuss three PIS models that may be viewed as special cases of $MPIS/G$ models. In Sect. 6.1 we consider PIS models with batch arrivals of either items or demands. Sect. 6.2 is devoted to the case in which demands are willing to wait. In Sect. 6.3 we take a closer look at intervals between successive outdatings and intervals between successive unsatisfied demands when item arrivals follow a renewal process.

6.1 Batch arrivals

In this subsection we briefly discuss three different cases in which items and/or demands arrive in batches.

Case 1: Poisson(λ) item arrivals and Poisson($\mu(x)$) demand arrivals; demands arrive in batches (cf. Kaspi & Perry, 1984*).* Let θ_n be the probability that the demand batch size equals $n, n=1,2,\ldots,$ with generating function $J(\cdot)$. This case basically is a special version of the M/G model studied in Sect. 5, with the jump sizes being a random sum of exp(λ) random variables with distribution $G(x) = \int_0^x \sum_{n=1}^{\infty} \theta_n \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}\mathrm{d}t$. There are two exceptions to this: (i) if the jump size is above level 1 then the overflow above 1 is always just one exp(λ) phase; and (ii) the jump from zero is also exp(λ), since such a jump is not due to a demand. In passing we observe that $G(\cdot)$ is a phase-type distribution, with LST $J(\frac{\lambda}{\lambda+\alpha})$. We also observe that a batch demand can alternatively be viewed as a single demand for a random number of items.

The balance equations for the density $f(\cdot)$ of the VOT process are given by

$$\begin{aligned}f(x) &= \int_0^x \mu(w)[1-G(x-w)]f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, \quad 0\le x\le 1,\\ f(x) &= \int_0^1 \mu(w)[1-G(1-w)]\mathrm{e}^{-\lambda(x-1)}f(w)\mathrm{d}w + f(0)\mathrm{e}^{-\lambda x}, \quad x>1.\end{aligned} \tag{31}$$

One could solve the first equation using the Picard iteration outlined in Sect. 5, while the second equation immediately translates into $f(x) = ce^{-\lambda x}$ for some constant c. If $\theta_n = (1-a)a^{n-1}$ for $n = 1, 2, \ldots$ (i.e., geometric batch sizes), then $1 - G(x) = e^{-(1-a)x}$, and it is straightforward to determine $f(\cdot)$ more explicitly.

Remark 8 If items arrive according to a renewal process with distribution $S(\cdot)$ and demands arrive in batches according to a Poisson process, then the above phase-type $G(\cdot)$ is replaced by a more general $G(\cdot)$ with density $g(\cdot)$. Formula (31) becomes

$$f(x) = \int_0^x [1 - G(x-w)]f(w)\mathrm{d}w + f(0)[1 - S(x)], \quad 0 \le x \le 1,$$

$$f(x) = \int_0^1 \int_{y=w}^1 \sum_{n=0}^{\infty} g^{n*}(y-w)[1 - S(x-y)]\mathrm{d}y f(w)\mathrm{d}w + f(0)[1 - S(x)], \quad x > 1.$$

Here g^{n*} denotes the n-fold convolution of the density $g(\cdot)$, and $g^{0*}(\cdot) = 1$. The resulting integral equation can again be solved via Picard iteration.

Case 2: Poisson(λ) item arrivals and Poisson(μ) demand arrivals; items arrive in batches. Successive item batch sizes $J_1, J_2, \ldots$ are i.i.d., with generating function $J(\cdot)$. In Fig. 7 we display the VOT process $\mathbf{V}$ and the age process $\mathbf{A}$ in the top and middle panel. All the dots in those two panels occur at times of satisfied demands—except for the dot at $A(t) = 1$. In the figure we have $J_1 = 4$, $J_2 = 2$, $J_3 = 6$ and $J_4 \ge 3$; observe that J_4 may have been larger than 3, because the VOT jumps up from 0 at an outdating, and if $J_4 > 3$ then more than 3 items are simultaneously outdated. The jump process is not a Poisson process, since only the last item in the batch is accompanied by a jump.

Just like in Sect. 4.3 we apply the duality of type 2: every negative jump in $A(t)$ becomes a trajectory with decreasing slope of rate 1 in $W(t)$ and every trajectory with increasing slope of rate 1 in $A(t)$ becomes a positive jump in $W(t)$ (cf. Fig. 7). LCT implies that the processes $\mathbf{A}$ and $\mathbf{W}$ have the same steady-state law, since by the above construction the numbers of up- and downcrossings of every level $x > 0$ in both processes are the same for every realization. The process $\mathbf{W}$ describes a finite dam model with Poisson arrivals of rate λ and the jump sizes have a phase-type distribution $G(\cdot)$ with LST $J(\frac{\mu}{\alpha+\mu})$. That is, the LST of the jump size is the generating function of a random sum of exp(μ) random variables. Also note that the emptiness period in $\mathbf{W}$ is exactly the time during which the shelf is empty in $\mathbf{A}$. We can now apply LCT to the $\mathbf{W}$ process and thus get the following balance equation for its steady-state density $f_W(\cdot)$:

$$f_W(x) = \lambda \int_0^x [1 - G(x-w)]f_W(w)\mathrm{d}w + \pi e^{-\lambda x}, \quad 0 \le x \le 1, \tag{32}$$

where the probability of an empty dam

$$\pi = \frac{1/\lambda}{1/\lambda + 1/f_W(0)} = \frac{f_W(0)}{f_W(0) + \lambda}.$$

In principle (32) can be solved via Picard iteration. One additional equation is provided by the normalizing condition:

$$\int_0^1 f_W(x)\mathrm{d}x = 1 - \pi.$$

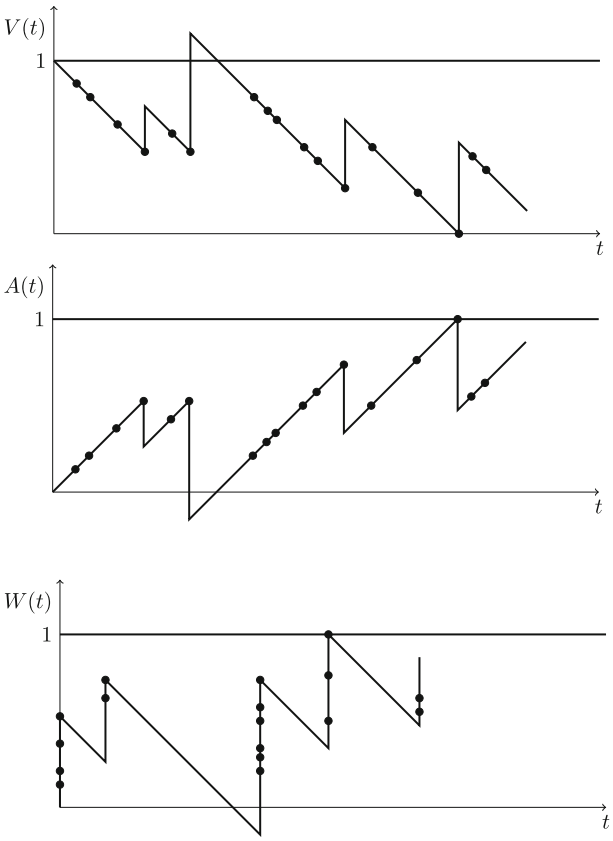

Fig. 7 A typical sample path of the VOT process **V** (top panel), the age process **A** (middle panel) and the transformation into the process **W** (bottom panel), for the case of batch arrivals of items

Finally, by the duality construction of $W(t)$ from $V(t)$ we get the density $f(\cdot)$ of **V** from $f_W(\cdot)$: $f(x) = f_W(1-x)$ for $0 \le x \le 1$ and $f(x) = \pi e^{-\lambda(x-1)}$ for $x > 1$ (recall that the overflow above level 1 in **V** is exp(λ)).

Case 3: Poisson arrivals of item batches and Poisson arrivals of demand batches. We assume independence of the arrival processes and of the various batch sizes. Furthermore, item batch sizes are geometric(γ_I) distributed, and demand batch sizes geometric(γ_D); i.e., a generic item batch size B_I has distribution $\mathbb{P}(B_I = n) = \gamma_I(1-\gamma_I)^{n-1}$, $n = 1, 2, \ldots$, and similarly for a generic demand batch size. This case was discussed in Goh et al. (1993); they focus on the busy period, and not on the steady-state analysis of **V**. A generic item batch size B_I is smaller than or equal to a generic demand batch size B_D with probability ρ, where it is readily verified that $\rho = \frac{\gamma_I}{\gamma_I+\gamma_D-\gamma_I\gamma_D}$. Hence a demand arrival epoch is with probability $1-\rho$ not accompanied by a jump. By the memoryless property of the geometric distribution, the residual size of the partially taken item batch again is geometric(γ_I) distributed, and so forth. Similarly, if the demand batch size is strictly larger than the item batch size, then the batch demand is only partially satisfied and the residual size of the demand batch again is geometric(γ_D) distributed, and so forth. A conclusion from the above is that we again almost have the basic PIS, in which now jumps upward form a Poisson process with rate $\mu\rho$ and

the jump sizes are independent and $\exp(\lambda(1-\rho))$ distributed. The two exceptions are: (i) a jump size from 0 (after an outdating) is $\exp(\lambda)$ distributed and (ii) the overflow is also $\exp(\lambda)$. Accordingly, we get the balance equations

$$f(x) = \begin{cases} \int_0^x \mu\rho e^{-\lambda(1-\rho)(x-w)} f(w)dw + f(0)e^{-\lambda x}, & 0 \le x \le 1, \\ \int_0^1 \mu\rho e^{-\lambda(1-\rho)(1-w)-\lambda(x-1)} f(w)dw + f(0)e^{-\lambda x}, & x > 1. \end{cases}$$

$f(x)$ can be easily determined from these equations (again multiplying both sides of the first equation by $e^{\lambda x}$ and then differentiating).

Conservation law of satisfied demand We have

$$\frac{\lambda - f(0)}{\gamma_I} = \frac{\mu F(1) - f(1)(1-\gamma_D)}{\gamma_D}.$$

Indeed, the lefthand side represents the rate of not outdated items. The righthand side is the rate of all the demands that arrive when the VOT is below level 1, with one correction: An upcrossing of level 1 means that a residual amount of the arriving demand batch (which is still geometric(γ_D)) is unsatisfied, *except* if B_D exactly equals B_I. The probability of the latter event is $\rho(1-\gamma_D)$. Hence $\frac{\mathbb{P}(B_D > B_I)}{\mathbb{P}(B_D \ge B_I)} = 1 - \gamma_D$.

6.2 Demands that are willing to wait

Model variant Perishable items arrive at the shelf according to a renewal process with interrenewal time distribution $G(\cdot)$, having mean $1/\lambda$. Demands for items arrive according to a Poisson process with rate μ, independent of the item arrival times. A demand that upon its arrival finds the shelf of items not empty is satisfied immediately by the oldest item present. Demands that arrive at an empty shelf join the line of waiting demands; newly arriving items are assigned on the spot to waiting demands on a first-come-first-served basis. It should be observed that the main difference with all previously discussed models is that here *demands are willing to wait*.

Each demand possesses its own *random patience time*. Denoting by P_n the patience time of the nth arriving demand, we assume that $P_1, P_2, \ldots$ is a sequence of i.i.d. positive random variables which are independent of the arrival times of items and demands. P_1 has distribution $H(\cdot)$, with mean $1/\eta$. If the waiting time of the nth demand exceeds its patience, then it *abandons* the waiting line without receiving an item. The shelf lifetime of the stored items, i.e., their maximum usage time, is (as before) set to 1. Thus, each item is stored until it either satisfies some demand or, after one time unit on the shelf, is outdated (and then scrapped).

Motivation This type of model occurs, e.g., when persons demand an organ, or a portion of blood. In both cases, the demanded item can only be stored for a limited amount of time. The organ transplantation problem and the blood transfusion process have been captured in various stochastic models. For further references see (Zenios et al., 2000) for an excellent introduction to the modeling of live organ transplantations by means of a waiting list, see (Perry & Stadje, 1999) for another paper on PIS with demands that are willing to wait, and see the paper (Boxma et al., 2011a) on which the present subsection is based. The model that we discuss here captures the essential aspects of the organ transplantation process, while

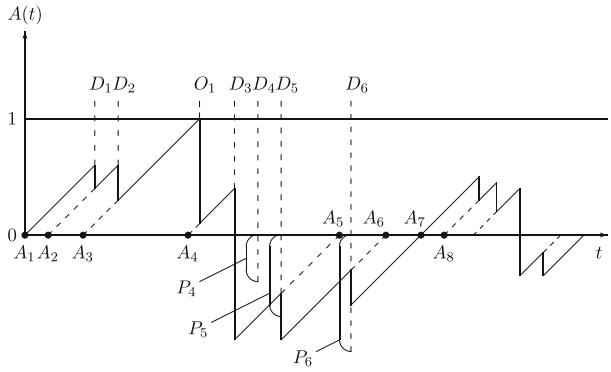

Figure 1a

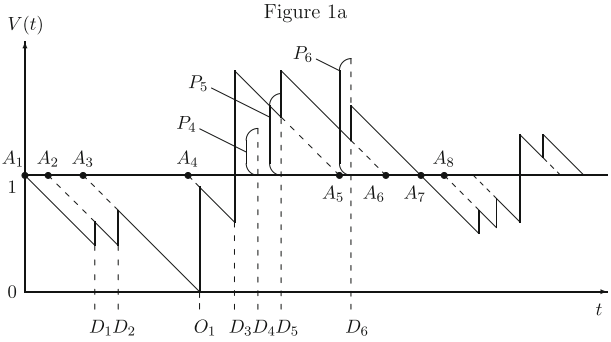

Figure 1b

Fig. 8 a typical sample path of $A(t)$ and the corresponding sample path of the VOT. A_n (D_n) denotes the arrival time of the nth item (demand); P_n denotes the patience time of the nth demand, and O_n the time of the nth outdating

ignoring some aspects which are relevant in the blood transfusion process (like the fact that not all types of blood are of use for a patient).

The VOT process Again let $A(t)$ denote the age of the oldest item on the shelf at time t, and let $V(t) = 1 - A(t)$. The VOT process $\mathbf{V}$ again is a Markov process. It can also be interpreted as the workload process in an $M/G/1+G$ queue—a queue with Poisson(μ) arrivals, service requirements with distribution $G(\cdot)$ and patience time $1 + P_n$. If the idle periods in such a queue are deleted and the busy periods are glued together, a workload process results which has the same law as $\mathbf{V}$. See Fig. 8 for a graphical representation of the age and the VOT process.

Once again applying LCT, it follows that the stationary density $f(\cdot)$ of $\mathbf{V}$ satisfies the integral equations

$$f(x) = \begin{cases} \mu \int_0^x [1 - G(x-w)] f(w)\, \mathrm{d}w + f(0)[1 - G(x)], & 0 < x \le 1, \\ \mu \int_0^1 [1 - G(x-w)] f(w)\, \mathrm{d}w + f(0)[1 - G(x)] & \\ \quad + \mu \int_1^x [1 - G(x-w)][1 - H(w-1)] f(w)\, \mathrm{d}w, & x > 1. \end{cases} \tag{33}$$

If the arrival times of items form a Poisson(λ) process, then the VOT is the workload process in an $M/M/1+G$ queue with deleted idle periods. Solving for $f(\cdot)$ in (33) with $1-G(x)$ being replaced by $\mathrm{e}^{-\lambda x}$, we obtain (see also Section IV of Baccelli et al. (1981))

$$f(x) = \begin{cases} k_0 e^{-(\lambda-\mu)x}, & 0 < x \le 1, \\ k_1 \exp\left\{-[\lambda x - \mu \int_1^x (1 - H(z-1))\mathrm{d}z]\right\}, & x > 1, \end{cases} \tag{34}$$

for certain constants k_0 and k_1. To find k_0 and k_1 note that $f(x)$ is continuous at 1. We get $k_0 = k_1 e^{\mu}$ and k_0 can be easily calculated via the normalizing condition for $f(\cdot)$:

$$k_0 = \left[\int_0^1 e^{-(\lambda-\mu)x} \mathrm{d}x + e^{\mu} \int_1^{\infty} \exp\left\{-[\lambda x - \mu \int_1^x (1 - H(z-1))\mathrm{d}z]\right\} \mathrm{d}x \right]^{-1}.$$

The workload density $f(\cdot)$ for *general* $G(\cdot)$ and $H(\cdot)$ can be obtained from (33) in the following way: (i) solve the integral equation (33) in the interval [0, 1] via Picard iteration (in terms of an infinite series of convolutions and the constant $f(0)$); (ii) insert this solution in the equation for $x \in (1, \infty)$, which can then also be solved in terms of an infinite series of convolutions in which the first series occurs as under the integral sign; and (iii) determine $f(0)$ from the normalization condition $\int_0^{\infty} f(x)\ \mathrm{d}x = 1$. We refer to Section 4 of Boxma et al. (2011a) for a different approach. There $\mathbf{V}$ is decomposed into two processes, which are constructed by deleting the time periods in which $\mathbf{V} > 1$ respectively $\mathbf{V} \le 1$. The first process is then related to a so-called finite dam, and the second process, decreased by one, represents the workload in an $M/G/1+G$ queue with deleted idle periods in which the first service time of a busy period has a different distribution. The densities of those two processes are subsequently determined.

Key performance measures The rate of item outdatings is given by $\lambda^* = f(0)$. The rate of unsatisfied demands equals $\mu^* = \mu - \lambda + f(0)$. By LCT, the rate of item arrivals at an empty system equals $f(1)$.

We next focus on the steady-state number of items K on the shelf. This number is zero when $\mathbf{V} > 1$, and otherwise it equals one plus the number of item arrivals during the age of the oldest item. Hence

$$\begin{aligned} \mathbb{E}\, z^K &= \int_1^{\infty} f(x)\ \mathrm{d}x + \int_0^1 \sum_{n=1}^{\infty} z^n\, \mathbb{P}(n-1 \text{ arrivals in } 1-x) f(x)\ \mathrm{d}x \\ &= \int_1^{\infty} f(x)\ \mathrm{d}x + \int_0^1 \sum_{n=1}^{\infty} z^n (G^{(n-1)*}(1-x) - G^{n*}(1-x)) f(x)\ \mathrm{d}x. \end{aligned} \tag{35}$$

When items arrive according to a Poisson(λ) process, the sum over n becomes $z\mathrm{e}^{-\lambda(1-z)(1-x)}$. We refer to Boxma et al. (2011a) for a study of the steady-state waiting time of demands, of the long-run fraction of time the shelf is empty, and of the outdating process. The number of waiting demands has been studied in the setting of the $M/G/1+G$ queue, cf. Boxma et al. (2011b).

6.3 Outdating and unsatisfied demands in the *M/G* PIS model

Model variant In this subsection we again have Poisson(μ) demand arrivals while item arrivals form a renewal process with renewal distribution $G(\cdot)$. As before, we assume that each item has a usable lifetime of one time unit and that, upon arrival, a demand removes the oldest item on the shelf—leaving unsatisfied if the shelf is empty. An item that has not

been taken within one time unit of arrival becomes outdated. Finally, the arrival processes of items and demands are again assumed to be independent.

We already know how to obtain the density $f(\cdot)$ of the VOT process, but in this subsection we shall exploit its knowledge only in a few places. We shall mainly focus on the following three performance measures, and on some useful techniques for analyzing them: The time between two successive outdatings, the time between two successive unsatisfied demands, and the shelf emptiness period.

The distribution of the time between two successive outdatings. As observed before, the time between successive outdatings is a busy period in the $M/G/1+D$ queue. Its distribution has been derived in Perry et al. (2000). As the analysis in that paper, and the end result, are very complicated, we here present a different approach that is applicable when $G(\cdot)$ has a phase-type distribution or, more specifically, a mixture of Erlang distributions with the same mean for all exponential phases. This class of distributions is known to lie dense in the class of all probability distributions of nonnegative random variables (cf. Section III.4 of Asmussen (2003)). To explain the approach, we restrict ourselves here even further to the case that $G(\cdot)$ is an Erlang distribution with two exponential phases: $G(x) = \int_0^x \lambda^2 t \mathrm{e}^{-\lambda t} \mathrm{d}t$ (note that the mean item interarrival time now is $2/\lambda$). A key quantity in the analysis is the stopping time

$$\tau = \min\{t : V(t) = 0 \text{ or } V(t) \geq 1\},$$

when starting in some state x. We use the abbreviation $\mathbb{E}_x = \mathbb{E}(\cdot \mid V(0) = x)$. The conditional joint LST $\mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)-\beta\tau} | V(\tau) > 1\right)$ of the overflow and the time of the overflow given that an overflow occurred is not easy to obtain for general $G(\cdot)$ due to the dependence between $V(\tau)$ and τ. However, when the item interarrival times are Erlang distributed, $V(\tau)$ and τ are conditionally independent given the number of exponential phases of the overflow above level 1. We now first show how it can be obtained in the Erlang$(2, \lambda)$ case. Defining the events I ={level 1 is upcrossed by the second phase of the jump} and II ={level 1 is upcrossed by the first phase of the jump}, we have for $Q = I, II$:

$$\mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)-\beta\tau} | V(\tau) > 1, Q\right) = \mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)} | V(\tau) > 1, Q\right) \mathbb{E}_x\left(\mathrm{e}^{-\beta\tau} | V(\tau) > 1, Q\right), \tag{36}$$

and

$$\begin{aligned}\mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)-\beta\tau} 1_{\{V(\tau)>1\}}\right) &= \mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)-\beta\tau} 1_{\{V(\tau)>1, I\}}\right) + \mathbb{E}_x\left(\mathrm{e}^{-\alpha V(\tau)-\beta\tau} 1_{\{V(\tau)>1, II\}}\right) \\ &= \mathrm{e}^{-\alpha} \frac{\lambda}{\lambda+\alpha} \mathbb{E}\left(\mathrm{e}^{-\beta\tau} 1_I\right) + \mathrm{e}^{-\alpha} \left(\frac{\lambda}{\lambda+\alpha}\right)^2 \mathbb{E}\left(\mathrm{e}^{-\beta\tau} 1_{II}\right).\end{aligned} \tag{37}$$

Then, for $0 < x \leq 1$ we get

$$\mathbb{E}_x\, \mathrm{e}^{-\alpha V(\tau)-\beta\tau} = \phi_0(\beta; x) + \mathrm{e}^{-\alpha} \tfrac{\lambda}{\lambda+\alpha} \phi_I(\beta; x) + \mathrm{e}^{-\alpha} \left(\tfrac{\lambda}{\lambda+\alpha}\right)^2 \phi_{II}(\beta; x),$$

where $\phi_0(\beta; x)$, $\phi_I(\beta; x)$ and $\phi_{II}(\beta; x)$ are the partial LSTs of τ such that

$$\begin{aligned}&\phi_0(\beta; x) := \mathbb{E}_x\, \mathrm{e}^{-\alpha V(\tau)-\beta\tau} 1_{\{V(\tau)=0\}} = \mathbb{E}_x\, \mathrm{e}^{-\beta\tau} 1_{\{V(\tau)=0\}}, \\ &\phi_I(\beta; x) := \mathbb{E}_x\, \mathrm{e}^{-\beta\tau} 1_I, \\ &\phi_{II}(\beta; x) := \mathbb{E}_x\, \mathrm{e}^{-\beta\tau} 1_{II}.\end{aligned} \tag{38}$$

Note that $\phi_0(0; x)$, $\phi_I(0; x)$ and $\phi_{II}(0; x)$ are the probabilities of the event $\{V(\tau) = 0\}$, the event I and the event II, respectively. To find $\phi_0(\beta; x)$, $\phi_I(\beta; x)$ and $\phi_{II}(\beta; x)$ consider the following process $M(s)$ which is a Kella-Whitt martingale (cf. Kella and Whitt 1992):

$$M(s) := \left[\alpha - \mu\left(1 - \left(\frac{\lambda}{\lambda+\alpha}\right)^2\right) - \beta\right] \int_0^s \mathrm{e}^{-\alpha V(t)-\beta t}\mathrm{d}t + \mathrm{e}^{-\alpha x} - \mathrm{e}^{-\alpha V(s)-\beta s}.$$

Now use the optional sampling theorem with stopping time τ to obtain $\mathbb{E}\, M(\tau) = 0$, i.e., the following fundamental identity:

$$\left[\alpha - \mu\left(1 - \left(\frac{\lambda}{\lambda+\alpha}\right)^2\right) - \beta\right] \mathbb{E}_x \int_0^\tau \mathrm{e}^{-\alpha V(t)-\beta t}\mathrm{d}t = -\mathrm{e}^{-\alpha x}$$
$$+\phi_0(\beta; x) + \mathrm{e}^{-\alpha}\frac{\lambda}{\lambda+\alpha}\phi_I(\beta; x) + \mathrm{e}^{-\alpha}\left(\frac{\lambda}{\lambda+\alpha}\right)^2 \phi_{II}(\beta; x). \tag{39}$$

The term between square brackets in the lefthand side has three zeroes (which actually are real), while the $\mathbb{E}_x$ term must be finite for finite α; hence we get three linear equations for the three unknowns $\phi_0(\beta; x)$, $\phi_I(\beta; x)$ and $\phi_{II}(\beta; x)$.

Remark 9 Briefly consider the case that $G(\cdot)$ is a mixture of Erlang distributions with the same mean for all exponential phases. Its LST is given by $b(\alpha) = \sum_{k=1}^n p_k(\frac{\lambda}{\lambda+\alpha})^k$, with all $p_k > 0$ and summing to one. We then have to distinguish between n instead of two events, corresponding to the number of phases of overshoots above 1. The Kella-Whitt martingale now yields a generalization of (39) with $n+1$ unknown functions of x in the righthand side, while the term between square brackets in the lefthand side is replaced by $\alpha - \mu(1-b(\alpha)) - \beta$. This is a familiar term in the study of the transient behavior of the $M/G/1$ queue, cf. p. 259 and p. 548 of Cohen (1982). The term has $n+1$ zeroes in our case; Rouché's theorem can be used to prove that one of them lies in the righthalf plane, but the zeroes are not necessarily real. It is also not a priori clear how to prove that the resulting $n+1$ linear equations for the $n+1$ unknowns are independent.

We are now ready to obtain the LST of T, the time between two successive outdatings. We have

$$\mathbb{E}\,\mathrm{e}^{-\beta T} = \int_0^1 \lambda^2 x e^{-\lambda x}\, \mathbb{E}_x\, \mathrm{e}^{-\beta T}\mathrm{d}x + \mathbb{E}_1\, \mathrm{e}^{-\beta T}\left[\lambda \mathrm{e}^{-\lambda}\frac{\lambda}{\lambda+\beta} + \mathrm{e}^{-\lambda}\left(\frac{\lambda}{\lambda+\beta}\right)^2\right]. \tag{40}$$

To solve for $\mathbb{E}\,\mathrm{e}^{-\beta T}$ we first have to find $\mathbb{E}_x\, \mathrm{e}^{-\beta T}$ for $0 < x \leq 1$. Observe that the factor $\lambda \mathrm{e}^{-\lambda}$ is the probability that the first phase of the item arrival interval does not exceed 1, but the sum of the two phases does exceed 1, and that the factor $\mathrm{e}^{-\lambda}$ is the probability that the first phase of the item arrival interval exceeds 1. We have

$$\mathbb{E}_x\, e^{-\beta T} = \phi_0(\beta; x) + \phi_I(\beta; x)\frac{\lambda}{\lambda+\beta}\,\mathbb{E}_1\, e^{-\beta T} + \phi_{II}(\beta; x)\left(\frac{\lambda}{\lambda+\beta}\right)^2 \mathbb{E}_1\, e^{-\beta T},$$

and by substituting $x = 1$ we obtain the LST of the time between two successive outdatings:

$$\mathbb{E}_1\, e^{-\beta T} = \frac{\phi_0(\beta; 1)}{1 - \phi_I(\beta; 1)\frac{\lambda}{\lambda+\beta} - \phi_{II}(\beta; 1)\left(\frac{\lambda}{\lambda+\beta}\right)^2}.$$

The time between two successive unsatisfied demands Consider an upcrossing of level 1 of the VOT process $\mathbf{V}$. Because of the Erlang(2, λ) item arrival intervals, such an upcrossing is either with one or with two exp(λ) phases. If the next demand occurs before an arrival of an item on the shelf, it is unsatisfied. The unsatisfied demand process is not a renewal process. However, we can find the LST of the time U between two successive unsatisfied demands

by distinguishing whether an upcrossing of level 1 occurs with one or two exp(λ) phases. Let $\Psi^{(i)}(\beta)$ denote the LST of the time from such an upcrossing until the next unsatisfied demand, if that upcrossing is with i phases, $i = 1, 2$. We have, with $\mathbb{E}_1\, \mathrm{e}^{-\beta U}$ the conditional LST of the remaining length of U, from the moment that level 1 is downcrossed:

$$\Psi^{(1)}(\beta) = \frac{\mu}{\mu+\lambda+\beta} + \frac{\lambda}{\mu+\lambda+\beta}\,\mathbb{E}_1\, \mathrm{e}^{-\beta U},$$

and

$$\Psi^{(2)}(\beta) = \frac{\mu}{\mu+\lambda+\beta} + \frac{\lambda}{\mu+\lambda+\beta}\,\Psi^{(1)}(\beta).$$

To compute $\mathbb{E}_1\, \mathrm{e}^{-\beta U}$, and more generally $\mathbb{E}_x\, \mathrm{e}^{-\beta U}$, the LST of the time until the next unsatisfied demand when starting from level x, we distinguish between the three possibilities that stopping time τ first occurs via an upcrossing of level 1 with one phase, or with two phases, or that it occurs by reaching level zero:

$$\begin{aligned}\mathbb{E}_x\, \mathrm{e}^{-\beta U} &= \phi_I(\beta; x)\Psi^{(1)}(\beta) + \phi_{II}(\beta; x)\Psi^{(2)}(\beta) \\ &\quad + \phi_0(\beta; x)\left[\int_0^1 \lambda^2 y \mathrm{e}^{-\lambda y}\,\mathbb{E}_y\, \mathrm{e}^{-\beta U} dy + \lambda \mathrm{e}^{-\lambda}\Psi^{(1)}(\beta) + \mathrm{e}^{-\lambda}\Psi^{(2)}(\beta)\right].\end{aligned} \tag{41}$$

To solve for $\mathbb{E}_x\, e^{-\beta U}$ multiply both sides of (41) by $\lambda^2 x \mathrm{e}^{-\lambda x}$, integrate and introduce

$$\Omega(\beta) := \int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\,\mathbb{E}_x\, \mathrm{e}^{-\beta U} \mathrm{d}x.$$

Then we get

$$\begin{aligned}\Omega(\beta) &= \Psi^{(1)}(\beta)\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_I(\beta; x)\mathrm{d}x \\ &\quad + \Psi^{(2)}(\beta)\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_{II}(\beta; x)\mathrm{d}x \\ &\quad + [\Omega(\beta) + \lambda \mathrm{e}^{-\lambda}\Psi^{(1)}(\beta) + \mathrm{e}^{-\lambda}\Psi^{(2)}(\beta)]\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_0(\beta; x)\mathrm{d}x,\end{aligned}$$

so that

$$\begin{aligned}\Omega(\beta) &= \frac{\Psi^{(1)}(\beta)\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_I(\beta; x)\mathrm{d}x + \Psi^{(2)}(\beta)\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_{II}(\beta; x)\mathrm{d}x}{1 - \int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_0(\beta; x)\mathrm{d}x} \\ &\quad + \frac{[\lambda \mathrm{e}^{-\lambda}\Psi^{(1)}(\beta) + \mathrm{e}^{-\lambda}\Psi^{(2)}(\beta)]\int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_0(\beta; x)\mathrm{d}x}{1 - \int_0^1 \lambda^2 x \mathrm{e}^{-\lambda x}\phi_0(\beta; x)\mathrm{d}x}.\end{aligned}$$

Now observe that this expression for $\Omega(\beta)$ still contains (hidden in $\Psi^{(1)}(\beta)$) the unknown $\mathbb{E}_1\, \mathrm{e}^{-\beta U}$. By taking $x = 1$ in (41) we obtain a second linear equation between $\Omega(\beta)$ and $\mathbb{E}_1\, \mathrm{e}^{-\beta U}$, and thus both functions can be determined.

Remark 10 We emphasize that the time periods between unsatisfied demands are neither independent nor identically distributed, but they are conditionally independent given the number of phases of the overshoot above level 1. It is also of relevance to determine the probability ζ that an overshoot above level 1 is one phase. ζ can be determined by observing

that $1/f(1)$ is the mean time between successive unsatisfied demands, and that hence the fraction of time above level 1 can be written as

$$1 - F(1) = f(1)\left[\frac{\zeta}{\lambda} + \frac{2(1-\zeta)}{\lambda}\right] = f(1)\frac{2-\zeta}{\lambda}.$$

Hence

$$\zeta = \frac{\lambda[\lambda - 2f(1)]}{f(1)}\left(\frac{1}{2\mu} - \frac{1}{\lambda}\right).$$

Incidentally, that factor $(2-\zeta)/\lambda$ is not only the expected overshoot above level 1, but of course it is also the expected length of the emptiness period.

The emptiness period: The general case Earlier we already emphasized the analogy between the $M/G/1+D$ queue and the PIS model with Poisson demand arrival process and with a renewal arrival process of items. In particular, the workload process of the latter queue, with the idle periods deleted, agrees with the VOT process of the PIS. However, different quantities are of interest in the queue and PIS setting. For example, the amount of overflow above level 1 is not of that much relevance in the $M/G/1+D$ queue, but it represents the important emptiness period of the shelf in the PIS. We study it below. For this, we need the steady-state density $f(\cdot)$ of the VOT $\mathbf{V}$. In Sect. 5 we have discussed how $f(\cdot)$ can be obtained by solving (26) via Picard iteration. Using this knowledge about $f(\cdot)$ we obtain the distribution $B(\cdot)$ of the overshoot above level 1. The following lemma has been introduced in Boxma et al. (2011a), with a different motivation, and for the case of constant $\mu(\cdot)$.

Lemma 2

$$B(x) = 1 - \frac{f(x+1)}{f(1)}, \quad x \geq 0,$$

where $f(\cdot)$ is the solution of (26).

Proof On the one hand, $b_e(x) := \frac{f(x+1)}{1-F(1)}$ for $x > 0$ is the conditional steady-state density of $\mathbf{V}$ given that the shelf is empty. On the other hand, by deleting the time periods in which $\mathbf{V} \leq 1$ and gluing together the time periods in which $\mathbf{V} > 1$ we see that above level 1 the behavior of the VOT $\mathbf{V}$ is stochastically equal to that of the equilibrium forward recurrence time associated with $B(\cdot)$, so that $b_e(x) = \frac{1-B(x)}{\int_0^\infty [1-B(y)]dy}$. From the argument above, $\int_0^\infty [1-B(y)]dy = \frac{1-F(1)}{f(1)}$. Thus, by equating $\frac{f(x+1)}{1-F(1)}$ with $\frac{1-B(x)}{\int_0^\infty [1-B(y)]dy}$, the lemma follows. □

We end this section by briefly mentioning Cohen's approach (Cohen, 1976, Chapter III) to the workload distribution in the $M/G/1+D$ queue, with buffer size $D=1$. Note that this immediately translates into results for the M/G PIS model. Cohen derives $\xi(x) := \mathbb{P}_x(V(\tau)=0) = 1 - \mathbb{P}_x(V(\tau) \geq 1)$, for the case of general $G(\cdot)$, as well as the steady-state workload distribution. Note that $\xi(x)$ is the probability that the shelf becomes empty before an outdating when the starting state is x. This probability is of importance from an *operational research* point of view, since outdating (perishability) and emptying the shelf involve costs of an opposite type that should be optimally balanced. Cohen introduces $\tilde{G}(x) := \rho \int_0^x (1 - G(y))\mathrm{d}y$, with $\rho := \lambda\, \mathbb{E}\, G$, and a parameter δ which is zero if $\rho \leq 1$ and which otherwise is the unique positive zero of $\int_0^\infty \mathrm{e}^{-xy}\mathrm{d}\tilde{G}(y) - 1$. He then proves that, with $\hat{G}(x) := \int_0^x \mathrm{e}^{-\delta y}\mathrm{d}\tilde{G}(y)$ and with

$$\tilde{V}(x) := \int_{0-}^{x} \mathrm{e}^{\delta y}\mathrm{d}\sum_{n=0}^{\infty} \hat{G}^{n*}(y), \quad x \geq 0, \tag{42}$$

one has the following expression for the steady-state distribution of the workload $V^{(1)}$ in the finite dam $M/G/1+D$ with buffer of size $D=1$:

$$\mathbb{P}(V^{(1)} < x) = \frac{\tilde{V}(x)}{\tilde{V}(1)}, \quad 0 < x \leq 1. \tag{43}$$

See also (29) for another representation, up to a multiplicative constant. For $\rho < 1$ the steady-state workload distribution of the ordinary $M/G/1$ queue exists, and $\tilde{V}(x)$ is proportional to that distribution, with proportionality factor $1/(1-\rho)$; and it is a well-known result that the steady-state distribution of the finite dam $(M/G/1+D)$ for $\rho < 1$ is proportional to the steady-state workload distribution in the infinite dam $(M/G/1)$; see, e.g., Hooghiemstra (1987) for an elegant sample-path proof.

Cohen (1976) subsequently obtains the following expression for the probability $\xi(x)$ that the workload process hits level 0 before hitting level 1, when starting from level x:

$$\xi(x) = \frac{\tilde{V}(1-x)}{\tilde{V}(1)}, \quad 0 \leq x < 1. \tag{44}$$

Such exit probability results have later been obtained in much greater generality for Lévy processes, typically expressing these exit probabilities in terms of so-called scale functions (cf. Section 8.2 of Kyprianou, 2006).

7 Conclusion and suggestions for further research

In this paper we have surveyed, extended and enriched the probabilistic analysis of a large class of perishable inventory systems. We have emphasized that a unifying principle is to consider the so-called virtual outdating process $\mathbf{V}$, where $V(t)$ equals one minus the age of the oldest item on the shelf at time t. The steady-state density of $\mathbf{V}$ was shown to be the main vehcle to obtain key performance measures like the rate of outdatings, the rate of unsatisfied demands and the distribution of the number of items on the shelf.

Through the years, we have devoted a significant part of our research efforts towards the probabilistic analysis of perishable inventory systems, and we hope to inspire others to also study them at length. There are many interesting methods and fascinating open problems in this area. Moreover, perishable inventory systems have huge societal relevance, and there is an abundance of practically relevant variants of the basic PIS that we have described in Sect. 2. Some such variants and generalizations have been described and analyzed in the present paper, but neither did we have the space to discuss them exhaustively, nor were we able to treat all the major variants. Below we mention a few more interesting PIS problems.

- *FIFO versus LIFO.* In the PIS literature, it is commonly assumed that items are issued First-In-First-Out (FIFO). One exception is Parlar et al. (2011), where the Last-In-First-Out (LIFO) issuance policy is studied for the basic PIS. Under LIFO, the shelf sojourn time of an item is shown to be distributed as the minimum of 1 and the busy period of an $M/M/1$ queue with arrival rate λ and service rate μ. This result is used to derive several other performance measures. Subsequently FIFO and LIFO are compared according to some cost criterion. Interestingly, while FIFO performs better in most cases, LIFO is better when the holding costs of items are high. It would be interesting to study other issuance policies like a random selection policy. For the secondary products/emergency demands model discussed at the end of Sect. 4.3, it could be natural to assume that in the upper shelf (items for emergency demands) the issuance policy is LIFO.

- *Disasters.* In Perry and Stadje (2001), the basic PIS is studied with the following additional feature: at Poisson epochs, all items become obsolete (e.g., because of a power failure). The steady-state density of the VOT is derived using LCT and solving a second-order homogeneous differential equation. In an extended version of this paper, we intend to treat this model for the case of renewal arrivals of items.
  In Perry and Stadje (2001) the system is also studied in heavy traffic. Under those conditions, the system is only instantaneously empty. Between disasters, the VOT evolves like reflected Brownian motion on [0, 1]. At disasters, it restarts at 1. Using the theory of reflected Brownian motion (cf. Chapter 5 of Harrison (1985)), several cost functionals are determined.
- *Heavy traffic.* Another heavy-traffic study of a PIS is performed in Perry (1997). It proposes a diffusion approximation for a basic PIS with the additional feature of hysteresis (cf. also Sect. 4.1). A reflected Brownian motion between barriers 0 and 1 is obtained, with the special feature that the drift becomes γ_L when the process downcrosses a level a, and becomes γ_H the first time that the process subsequently upcrosses some level $b > a$. The stationary law of the process is analysed by using a martingale, and the total expected discounted costs are evaluated. This heavy-traffic approach seems to have potential for a wider class of PIS, and could be explored further.
- *Two systems with one-way substitution.* Liu et al. (2022) study two PIS that are correlated through a so-called one-way substitution of demands. If the shelf of PIS II is empty when it receives a type-II demand, then that demand is redirected to PIS I. However, if the shelf of PIS I is empty when it receives a type-I demand, then that demand cannot be redirected. This problem is inspired by blood banks, in which persons of a particular blood type can or cannot use blood of another type. The mathematical analysis of PIS II is straightforward, but that of PIS I gives rise to a modulated Poisson demand process in a non-Markovian environment, for which Liu et al. develop an approximation method. The study of correlated systems of PIS still is almost unexplored territory.
- *Lead times.* In many real-life inventory systems, there is an item ordering policy, under which one or more items are ordered when the number of items in stock decreases to a certain level; and typically there is a lead time involved in such a replenishment order. We refer to Berk and Gürler (2008) for the study of a PIS with the so-called (Q, r) replenishment policy with lead times.
- *Positive service times.* As mentioned in Sect. 1, there is an interesting line of research regarding PIS with positive service times: if a demand arrives and finds an item in stock, it takes a positive amount of time to take that item. It would be interesting to explore whether the VOT approach can shed light on such PIS models.

Finally, we would like to point out that the transient behavior of PIS has hardly received attention so far. Also, it would be useful to have sharp approximations and bounds for key performance measures in cases for which it is too hard to obtain explicit expressions for the steady-state density of the VOT.

Acknowledgements The authors are indebted to Professor Krishnamoorthy for his support and encouragement, and to the four referees for their valuable comments. The research of Onno Boxma is supported by the NWO Gravitation Programme NETWORKS (Grant Number 024.002.003). The research of David Perry is partly supported by a grant of the Israel Science Foundation (Grant Number 3274/19).

Declarations

Conflict of interest All the authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Asmussen, S. (2003). *Applied probability and queues*. Springer.

Baccelli, F., & Hébuterne, G. (1981). On queues with impatient customers. In F. J. Kylstra (Ed.), *Proceedings of the Performance'81*. Elsevier.

Balcioglu, B., Kopach, R., & Carter, M. (2008). Tutorial on constructing a red blood cell inventory management system with two demand rates. *European Journal of Operational Research, 185*, 1051–1059.

Bar-Lev, S. K., Perry, D., & Stadje, W. (2005). Control policies for inventory systems with perishable items: Outsourcing and urgency classes. *Probability in the Engineering and Informational Sciences, 19*, 309–326.

Bekker, R. (2005). *Queues with state-dependent rates*. PhD Thesis, Eindhoven University of Technology.

Bekker, R., Borst, S. C., Boxma, O. J., & Kella, O. (2004). Queues with workload-dependent arrival and service rates. *Queueing Systems, 46*, 537–556.

Berk, E., & Gürler, Ü. (2008). The (Q, r) inventory model for perishables with positive lead times and non-negligible ordering costs. *Operations Research, 56*, 1238–1246.

Boxma, O. J., David, I., Perry, D., & Stadje, W. (2011a). A new look on organ transplantation models and double matching queues. *Probability in the Engineering and Informational Sciences, 25*, 135–156.

Boxma, O. J., Perry, D., & Stadje, W. (2011b). The $M/G/1 + G$ queue revisited. *Queueing Systems, 67*, 207–220.

Brill, P. (2008). *Level crossing methods in stochastic models*. Springer.

Chakravarthy, S. R. (2010). An inventory system with Markovian demands, phase type distributions for perishability and replenishment. *OPSEARCH, 47*, 266–283.

Cohen, J. W. (1976). *On regenerative processes in queueing theory*. Springer.

Cohen, J. W. (1982). *The single server queue* (2nd ed.). North-Holland.

Daley, D. J. (1964). Single server queueing system with uniformly limited queueing times. *Journal of the Australian Mathematical Society, 4*, 489–505.

Doshi, B. (1992). Level crossing analysis of queues. In U. N. Bhat, & I. V. Basawa (Eds.), *Queueing and related models. Oxford computer science statistics* (vol. 9, pp. 3–33). Clarendon Press.

Goh, C.-H., Greenberg, B. S., & Matsuo, H. (1993). Perishable inventory systems with batch demand and arrivals. *Operations Research Letters, 13*, 1–8.

Harrison, J. M. (1985). *Brownian motion and stochastic flow systems*. Wiley.

Hooghiemstra, G. (1987). A path construction for the virtual waiting time of an $M/G/1$ queue. *Statistica Neerlandica, 41*, 175–182.

Karaesmen, I., Scheller-Wolf, A., & Deniz, B. (2011). Managing perishable and aging inventories: Review and future research directions. In: K. Kempf, A. Keskinocak, & P. Uzsoy (Eds.), *Handbook of production planning* (vol. 1, pp. 393–436). Kluwer.

Kaspi, H., & Perry, D. (1983). Inventory system of perishable commodities. *Advances in Applied Probability, 15*, 674–685.

Kaspi, H., & Perry, D. (1984). Inventory system of perishable commodities with renewal input and Poisson output. *Advances in Applied Probability, 16*, 402–421.

Kella, O., & Whitt, W. (1992). Useful martingales for stochastic storage processes with Lévy input. *Journal of Applied Probability, 29*, 396–403.

Ko, S.-S. (2020). A nonhomogeneous quasi-birth-death process approach for an (s, S) policy for a perishable inventory system with retrial demands. *Journal of Industrial and Management Optimization, 16*, 1415–1433.

Krishnamoorthy, A., Shajin, D., & Lakshmy, B. (2016). On a queueing-inventory with reservation, cancellation, common lifetime and retrial. *Annals of Operations Research, 247*, 365–389.
Krishnamoorthy, A., Shajin, D., & Narayanan, V. C. (2020). Inventory with positive service time: A survey. In V. Anisimov & N. Limnios (Eds.), *Queueing theory 2: Advanced trends* (pp. 201–237). Wiley.
Kyprianou, A. E. (2006). *Introductory lectures on fluctuations of Lévy processes with applications*. Springer.
Lian, Z., Liu, L., & Neuts, M. F. (2005). A discrete-time model for common lifetime inventory systems. *Mathematics of Operations Research, 30*, 718–732.
Liu, L., Adan, I., & Perry, D. (2022). Two perishable inventory systems with one-way substitution. *Annals of Operations Research, 317*, 107–128.
Mikhlin, S. G. (1957). *Integral equations*. Pergamon Press.
Nahmias, S. (1982). Perishable inventory theory: A review. *Operations Research, 30*, 680–708.
Nahmias, S. (2012). *Perishable inventory systems*. Springer.
Nahmias, S., Perry, D., & Stadje, W. (2004a). Actuarial valuation of perishable inventory systems. *Probability in the Engineering and Informational Sciences, 18*, 219–232.
Nahmias, S., Perry, D., & Stadje, W. (2004b). Perishable inventory systems with variable input and demand rates. *Mathematical Methods of Operations Research, 60*, 155–162.
Otten, S., Krenzler, R., Daduna, H., & Kruse, K. (2015). Queues in a random environment. *Queueing Systems, 80*, 127–153.
Parlar, M., Perry, D., & Stadje, W. (2011). FIFO versus LIFO issuing policies for stochastic perishable inventory systems. *Methodology and Computing in Applied Probability, 13*, 405–417.
Perry, D. (1985). Inventory system of perishable commodities with random life-time. *Advances in Applied Probability, 17*, 234–236.
Perry, D. (1997). A double band control policy of a Brownian perishable inventory system. *Probability in the Engineering and Informational Sciences, 11*, 361–373.
Perry, D. (1999). Analysis of a sampling control scheme for a perishable inventory system. *Operations Research, 47*, 966–973.
Perry, D., & Posner, M. J. M. (1990). Control of input and demand rates in inventory systems of perishable commodities. *Naval Research Logistics, 37*, 85–97.
Perry, D., & Stadje, W. (1999). Perishable inventory systems with impatient demands. *Mathematical Methods of Operations Research, 50*, 77–90.
Perry, D., & Stadje, W. (2000a). An inventory system for perishable items with by-products. *Mathematical Methods of Operations Research, 51*, 287–300.
Perry, D., & Stadje, W. (2000b). Inventory systems for goods with censored random lifetimes. *Operations Research Letters, 27*, 21–27.
Perry, D., & Stadje, W. (2001). Disasters in Markovian inventory systems for perishable items. *Advances in Applied Probability, 33*, 61–75.
Perry, D., Stadje, W., & Zacks, S. (2000). Busy period analysis for $M/G/1$ and $G/M/1$-type queues with restricted accessibility. *Operations Research Letters, 27*, 163–174.
Shajin, D., Krishnamoorthy, A., & Manikandan, R. (2022). On a queueing-inventory system with common life time and Markovian lead time process. *Operational Research, 22*, 651–684.
Zenios, S. A., Chertow, G. M., & Wein, L. M. (2000). Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research, 48*, 549–569.