

Approximations for the waiting time in (s,nQ)-inventory models for different types of consolidation policies

Citation for published version (APA):

Smits, S. R., & Kok, de, A. G. (2001). *Approximations for the waiting time in (s,nQ)-inventory models for different types of consolidation policies*. (BETA publicatie : working papers; Vol. 67). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2001

Document Version:

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

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.tue.nl/taverne

Take down policy

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

openaccess@tue.nl

providing details and we will investigate your claim.

Approximations for the waiting time in (s,nQ) -inventory models for different types of consolidation policies

S.R. Smits and A.G. de Kok

Faculty of Technology Management
Technische Universiteit Eindhoven
Pav E6, P.O. Box 513,
5600 MB Eindhoven
The Netherlands

Abstract. In many practical situations the coordination of transportation management and inventory management may lead to considerable cost reductions. Transportation management includes the application of different types of shipment consolidation policies. Shipment consolidation takes into account the logistics strategy of combining two or more shipment orders to optimize transportation. When the shipment consolidation policy changes, the shipment lead time changes as well and if the lead time changes, the inventory policy needs to be re-evaluated, since changing lead times affect customer service. In this paper the lead time comprises two elements: waiting time due to order consolidation and the shipment time. The lead time is an important parameter for inventory management. We derive approximations for the lead time behaviour in (s,nQ) models where the items are consolidated according to different types of consolidation policies.

1 Introduction

The coordination of inventory management and transportation management is crucial for an efficient management of the supply chain. Transportation management includes the application of different types of shipment consolidation policies. The consolidation policy coordinates shipment processes of different item orders for the same (intermediate) destination, and this can lead to a reduction in transportation costs. Higginson and Bookbinder (1994) distinguish between two types of consolidation policies: the time policy and the quantity policy. The time policy dispatches orders when a shipping date is expired. The shipping date is usually set through consideration of service levels. Higginson and Bookbinder (1995a) give some normative approaches to set the shipping date. The quantity policy dispatches orders when a fixed quantity is consolidated. Higginson and Bookbinder (1995b) use a Markov chain model to determine the optimal consolidation policy given an (s,S) inventory policy generating shipment orders.

Another line of research is the joint replenishment or coordinated replenishment problem. Goyal and Satir (1989) present an early review of all

models, starting from a simple deterministic problem. In the joint replenishment literature we observe two types of control policy; viz. the continuous review can-order policy (s_i, c_i, S_i) and the periodic review order-up-to policy (R_i, S_i) . In the continuous can-order policy (s_i, c_i, S_i) , when the inventory position of an item i reaches the must-order point s_i , a replenishment is triggered as to raise the item's inventory position to order-up-to level S_i . Meanwhile, any other item in the group with an inventory position at or below its can-order point c_i is included in the replenishment as to raise the inventory position up to S_i . See, e.g., Liu and Yuan (2000) or Federgruen, Groenevelt and Tijms (1984). In the periodic review (R_i, S_i) policy, the inventory position of item i is inspected with intervals R_i and the review moments are coordinated in order to consolidate orders of individual items, see Viswanathan (1998).

In this paper, we analyze shipment consolidation policies under the assumption of compound renewal customer demand. The compound renewal demand process enables accurate modeling of real-world demand processes. In the literature discussed above order-up-to-policies are employed for inventory management. In practice, it is often more appropriate to employ (s, nQ) -policies, which take into account restrictions imposed by material handling units such as pallets and boxes. The focus in this paper is to model the interaction between shipment consolidation processes and inventory management policies. In general this involves multiple items or stock keeping units (sku's) and multiple stock locations. It is easy to see that a building block for the analysis of the interaction between shipment consolidation and inventory management, is the analysis of a line haul between two stocking locations, for example between a warehouse and a retailer. We assume for ease of reference that the warehouse holds stocks of multiple items. The same items are held by the retailer who sells to customers. The inventories at the retailer are controlled according to an (s, nQ) -policy. The retailer has to satisfy fill rate requirements for all items. The reorder level, which ensures the required fill rate, depends on the lead time of orders from the warehouse to the retailer. The lead time of an order comprises the waiting time for truck departure and the transportation time. The waiting time for truck departure is the main subject of this paper. The contribution of this paper to literature is twofold. First of all, we analyze shipment consolidation policies under the assumption of compound renewal demand, where this demand represents the customer orders at the retailer. We present an overall analysis of this problem integrating shipment consolidation and inventory management taking into account material handling restrictions. The latter is dealt with through the use of (s, nQ) -policies.

In this paper, we use the method of Whitt (1982) to superpose renewal processes with mixed-Erlang distributed inter-renewal times. Notice that the superposed process is not a compound renewal process. Yet, our analysis reveals that in line with the research of Whitt (1982), assuming that the

superposed process is renewal yields good approximations for performance characteristics, c.f. Smits, de Kok and, van Laarhoven (2000).

The sequel of this paper is organized as follows: In section 2, we describe in detail the model and we derive approximations for the waiting time. In section 3, we test the approximations through extensive computer simulations and in the last section, we give some conclusions and indicate a few thoughts for further research.

2 Model description

The model considers a line haul between a warehouse and a retailer. At both locations stocks of all items are kept. At the retailer the customer demand for each item arrives according to a compound renewal process, i.e., customer orders for an item arrive according to a renewal process and the demand per customer has some arbitrary distribution function. Demands of different customers for an item are independent and identically distributed. We furthermore assume that the compound renewal demand processes for different items are independent of each other. Additional constraints of the model are the item fill rate constraints at the retailer. This implies that at the retailer for each item a target fill rate is given. The fill rate is the fraction of demand directly delivered from shelf. The inventories at the retailer are controlled by (s, nQ) -policies. It operates as follows: as soon as the inventory position, which is expressed as the physical inventory plus the stock on order minus the backorders, drops below reorder level s an amount nQ is ordered such that the inventory position is raised to a value between s and $s + Q$. Q is called the batchsize, n is an integer. The demand that cannot be met immediately is backordered. We assume that the warehouse always has enough stock to fulfill the replenishment orders towards the retailer. To be able to calculate the reorder levels of the different items, the lead time towards the retailer is needed. In our model the lead time comprises the waiting time due to transport consolidation and the handling and transportation time. In this paper, we derive approximations for the waiting time due to consolidation for the time policy and the quantity policy. Figure 1 gives a graphical representation of the model. Below a list of the used notation is introduced.

Parameters and variables

L_d driving time

Z waiting time due to shipment consolidation

L^* lead time

T time between two truck departures. In the time policy T is deterministic and in the quantity policy T is stochastic

A_i time between two subsequent arrivals of item i
at the retailer

D_i	demand size of item order i (in volume) at the retailer
P_{2i}	fill rate at the retailer
$D_i(L^*)$	demand for item i during the lead time
s_i	reorder level of item i (in volume)
X_i^+	physical inventory level at an arbitrary point in time
Q_{max}	predetermined consolidation quantity (in volume)
$\Delta(t)$	consolidated quantity at time t , $\Delta(T)$ is the shipped quantity
Q_i	batchsize of item i (in volume)
O_i	order process of the retailer towards the warehouse for item i
R_i	time between order placements at the warehouse for item i
O^*	aggregate order process at the warehouse
$O_{\neq i}^*$	aggregate order process of all item except i at the warehouse
R^*	aggregate process of the time between two order placements at the warehouse
$R_{\neq i}^*$	aggregate process of the time between two order placements for all items except i at the warehouse
V	rest part of the split order in the quantity policy with partial shipments
U	undershoot process
$N(T)$	is defined as the number of arrivals in $(0, T]$
W_i	is defined as $O_{\neq i}^* + V + O_i$
$N(X)$	is defined as the number of arrivals between the arrival of an arbitrary customer and the departure of the truck
$Y(t)$	inventory position at moment t

Functions and Operators

$E[Y]$	expectation of the random variable Y
$\sigma^2(Y)$	variance of the random variable Y
$E[Y^2]$	second moment of the random variable Y
c_Y	coefficient of variation of the random variable Y
$(y)^+$	$\max(0, y)$
$P\{A\}$	probability of event A
$F^{n*}(t)$	n -fold convolution of $F_y(t)$
$F_y(t)$	pdf of random variable Y
$M(t)$	renewal function, $M(t) = \sum_{n=0}^{\infty} F^{n*}(t)$ associated with pdf $F_y(t)$

Time policy

In case consolidation employs a time policy the trucks depart at fixed time intervals T (for example, every week). Figure 2, gives a schematic representation of the time policy. All replenishment orders arriving within one time interval T are consolidated and shipped together to the retailer. We define $\Delta(t)$ as the collected quantity at moment t and $\Delta(T)$ as the consolidated

amount that is shipped. We want to find expressions for the time between the arrival of an arbitrary order and the departure of the truck.

Quantity policy

In case consolidation employs a quantity policy, we distinguish between two alternatives; partial shipments and full shipments/flexible truck capacity. In both cases we assume that at time 0 a truck leaves the warehouse.

i) *Partial shipments*

The orders are consolidated until the collected quantity $\Delta(t)$, is larger than or equal to a predetermined quantity Q_{max} . The last order $O_{N(T)}$ is split such that $\Delta(T)$ is equal to Q_{max} , where T is the time of the first truck departure after time 0. The consolidated quantity $\Delta(T)$ is shipped to the retailer and the consolidation process starts all over again with as starting quantity the remaining part of the order V . So $O_{N(T)} - V$ leaves directly and the remaining part V leaves with the next truck. Figure 3a) gives a schematic representation of the process. In order to determine the probability distribution function of $\Delta(t)$ we apply the following proposition.

Proposition 1. *The consolidation process of a quantity policy with partial shipments under a compound renewal demand process is equivalent to the inventory position process under an $(0, nQ_{max})$ control policy and compound renewal customer demand.*

Explanation. In the $(0, nQ_{max})$ inventory policy with compound renewal customer demand, the inventory position $Y(t)$ decreases at the arrival of a customer order. When the inventory position drops below 0 an amount Q_{max} is ordered. The amount by which the inventory position drops below 0 is called the undershoot and is denoted by U . In the quantity consolidation policy with compound renewal replenishment orders, the collected quantity increases at the arrival of a replenishment order. When the collected quantity exceeds Q_{max} , an amount $\Delta(T) = Q_{max}$ is shipped towards the retail warehouse. In the consolidation process the cumulative order between two shipments is $\Delta(T)$. The inventory position has the same course as the consolidated quantity at a moment in time. The undershoot process U in the $(0, nQ_{max})$ inventory policy is similar to the split order process V in the shipment consolidation process (see Figure 3).

ii) *Full shipments/flexible truck capacity*

The orders are consolidated until the collected quantity is larger than or equal to a predetermined quantity Q_{max} . The consolidated quantity $\Delta(T)$ is equal to the entire collected quantity in $(0, T]$ and hence $\Delta(T) \geq Q_{max}$. The consolidated quantity $\Delta(T)$ is shipped to the retailer and the consolidation process starts all over again. Figure 4 a) gives a schematic representation of the process. The probability density function of $\Delta(t)$ can be derived from the following proposition.

Proposition 2. *The quantity consolidation process with full shipments/flexible truck capacity under compound renewal demand is equivalent to the inventory position process under an $(0,S)$ control policy and compound renewal customer demand where $S = Q_{max}$.*

Explanation. In the $(0,S)$ inventory policy, where $S = Q_{max}$, with compound renewal customer demand, the inventory position decreases at the arrival of a customer order. When the inventory position drops below 0, an amount $Q_{max} + U$ is ordered such that the inventory position is raised up to Q_{max} . In the quantity consolidation policy with compound renewal replenishment orders, the collected quantity increases at the arrival of a replenishment order. When the collected quantity exceeds Q_{max} , an amount $\Delta(T) = Q_{max} + V$ is shipped towards the retailer. The inventory position has the same course as the consolidated quantity at every moment in time (see Figure 4).

To be able to calculate the waiting time between the placement of an arbitrary order and the departure of the truck, we must derive expressions for the arrival process of orders to be consolidated. To calculate this arrival process, we derive approximations for the order process of the different items from the retailer towards the warehouse. These approximations are described in Appendix A. After that, we superpose the order processes of the different items. The approximations for the superposed process are described in Appendix B. The superposed compound renewal process constitutes the demand for the consolidation process. With this superposed process we can derive expressions for the first two moments of the waiting time due to shipment consolidation. Notice that the superposed process is not a compound renewal process. Yet, our analysis reveals that assuming that the superposed process is a compound renewal process yields good approximations for the performance characteristics. From the first two moments of the waiting time due to shipment consolidation and the first two moments of the transportation time we compute the first two moments of the lead time of an arbitrary order for an item. Using the analysis in Smits et al. (2000) we can compute the reorder levels that yield the required fill rate level and the associated average net stocks.

2.1 Waiting time due to consolidation

In this section, we derive expressions for the waiting time due to shipment consolidation. Again we distinguish between two types of consolidation policies, the time policy and the quantity policy. Due to the compound renewal demand process we cannot hope for exact results of the waiting time distribution. Our generic approach is to derive expressions for the first two moments of the waiting time and fit a tractable distribution to these first two moments.

Time policy

We assume that the orders arrive at the warehouse according to a compound renewal process. In appendix A and B approximations of the superposed order arrival process are given. This process is independent of the truck departure process, which is a renewal process with deterministic inter-renewal times. Under stationarity it holds that the waiting time until truck departure of an arbitrary order is uniformly distributed on $(0, T)$. The proof of this statement can be found in appendix D. Thus we find

$$E[Z] = \frac{T}{2} \quad (1)$$

$$E[Z^2] = \frac{T^2}{3} \quad (2)$$

Quantity policy

The derivation of the waiting time distribution under the various types of quantity policies is much more complicated than in case of the time policies. Exact results are only available for special cases. Therefore we have to resort to the derivation of approximations for the first two moments of the waiting time distribution.

We have defined $N(X)$ as the number of arrivals between the placement of an arbitrary order and the departure of the truck. For the first moment we obtain

$$E[N(X)] = \sum_{n=1}^{\infty} n(F^{(n-1)*}(Q_{max} - X) - F^{(n)*}(Q_{max} - X)) \quad (3)$$

$$E[N(X)] = \int_0^{Q_{max}} M(Q_{max} - x) dF_X(x) \quad (4)$$

To be able to evaluate this for the different consolidation policies, we have to find an expression for $F_X(x)$ for the two different policies.

i) Partial shipment

We can use Proposition 1 and the fact that the inventory position of an (s, nQ) inventory policy is uniformly distributed between $(s, s+Q)$. Therefore we can conclude that $\Delta(t)$ is uniformly distributed between $(0, Q_{max})$, this gives

$$P\{X \leq x\} = \frac{x}{Q_{max}}.$$

ii) Full shipment/flexible truck capacity

We can use Proposition 2 and the fact that the inventory position of a (s,S) inventory policy is a renewal process. Therefore we can conclude that $\Delta(t)$ is a renewal process, this gives

$$P\{X \leq x\} = \frac{M(x)}{M(Q_{max})}.$$

When O_i is Poisson distributed, we could precisely calculate $N(X)$ for the first two cases, but in practice the coefficient of variation of the replenishment orders are lower than 1. Another difficulty is the calculation of the waiting time from $N(X)$, since we have to take into account that a part of the last replenishment V may not leave directly in the partial shipment case. We observe that it is difficult to find an exact expression for the waiting time. However in practice, it may be possible to standardize the volume of the batchsizes for the different sku's to pallets or boxes and the volume of the truck to a multiple of this volume unit. In this case the consolidated quantity $\Delta(T)$ is exactly equal to the predetermined quantity Q_{max} and we can compute exact derivations for the waiting time.

a. Equal batchsizes for all items

In this subsection we derive the first two moments of the waiting time when the volume of the batchsizes of the different items are equal to some Q . The volume of the predetermined shipped quantity Q_{max} is assumed to be a multiple of Q . In this case $\Delta(t)$ has a discrete distribution. The consolidation process starts at the arrival of the first batchsize Q , after this a second batch arrives and the consolidated quantity is $2Q$, then $3Q$ until the predetermined shipped quantity is reached. In the steady state, $\Delta(t)$ is uniformly distributed over $[0, Q, 2Q, \dots, (\frac{Q_{max}}{Q} - 1)Q]$. The different consolidated quantities have the same probability namely $\frac{Q}{Q_{max}}$. It easily follows that

$$E[N(X)] = \sum_{k=0}^{(\frac{Q_{max}}{Q}-1)} k \frac{Q}{Q_{max}} = \frac{1}{2} \left(\frac{Q_{max}}{Q} - 1 \right) \frac{Q_{max}}{Q} \frac{Q}{Q_{max}} = \frac{1}{2} \left(\frac{Q_{max}}{Q} - 1 \right) \quad (5)$$

$$E[N(X)^2] = \sum_{k=0}^{(\frac{Q_{max}}{Q}-1)} k^2 \frac{Q}{Q_{max}} = \frac{1}{6} \left(2 \frac{Q_{max}^2}{Q^2} - 3 \frac{Q_{max}}{Q} + 1 \right) \quad (6)$$

$$E[Z] = E[N(X)]E[R^*] \quad (7)$$

$$E[Z^2] = E[N(X)]\sigma^2(R^*) + E[N(X)^2]E[R^*]^2 \quad (8)$$

b. *Heuristic for non-equal batchsizes*

In the derivations of the waiting time for the quantity policy with non-equal batchsizes we encounter two difficulties: the waiting time is dependent on the quantity consolidation policy (partial shipments or full shipments/flexible truck capacity) and the waiting time may be different for different items. If a batchsize is very large compared to other ones then it is likely that the waiting time for this batchsize is smaller than for the other ones.

We observe that it is not efficient to have a high probability of having more than two batches of the same item in one truck. When there is high probability of having two orders in one truck, we can increase the batchsize without increasing the inventory level which leads to the same inventory costs but may lead to lower transportation and handling costs. To calculate the waiting time of batchsize i , we assume two types of order processes. The order process of item i and the order process of all other items except i . We use the formulas in appendix A and B to calculate the aggregate order process of all products except item i . We define $E[R_{\neq i}^*]$ and $E[R_{\neq i}^{*2}]$ as the first two moment of the inter-arrival times of all other items except item i and we define $E[O_{\neq i}^*]$ and $E[O_{\neq i}^{*2}]$ as the first two moments of the order size of all items except i .

i) *Partial shipments*

The last order is split such that the consolidated quantity is equal to the predetermined quantity. The part of the last order which is split is denoted by V . Since this consolidation policy is equivalent to the $(0, nQ_{max})$ inventory policy, V is equivalent to the undershoot process in the $(0, nQ_{max})$ inventory model.

For the first two moments of the undershoot process (Appendix B formula's 33 and 34), we use the asymptotic results for the first two moments of the forward recurrence time distribution. Therefore the first two moments of V are

$$E[V] \simeq \frac{E[O^{*2}]}{2E[O^*]} \quad (9)$$

$$E[V^2] \simeq \frac{E[O^{*3}]}{3E[O^*]} \quad (10)$$

Now let us define n_i as the amount of products orders i that are consolidated in an arbitrary consolidated shipment. We neglect the probability that $n_i > 1$, since we deduced previously that it is not cost efficient to have a high probability of having more than 2 orders of the same item in one truck. Figure 5 gives a schematic representation of the truck consolidation process. We define $W_i = O_{\neq i}^* + V + O_i$. Given that $n_i = 1$, the probability that the number of arrivals of order process $\neq i$ is equal to k , is as follows:

$$P\{n_{\neq i} = k | n_i = 1\} = P\left\{\sum_{i=1}^k W_i \leq Q_{max}\right\} - P\left\{\sum_{i=1}^{k+1} W_i \leq Q_{max}\right\} \quad (11)$$

If O_i is discrete then we can evaluate the formula above by assuming that $\sum_{i=1}^k O_{\neq i}^* + V$ is mixed Erlang distributed, else we assume that $\sum_{i=1}^k W_i$ is mixed Erlang distributed. Formula (11) gives correct results when k is large and when $O_{\neq i}^*$ and V are exponentially distributed. We assume that the probability that the item i order is the first one is equal to the probability that the item i order is the second one, the third or the last one, given that $n_{\neq i} = k$ and $n_i = 1$. We define $N(X)_i$ as the number of arrivals between the placement of an arbitrary order i and the departure of the truck.

$$E[N(X)_i] \simeq \sum_{k=0}^{\infty} P\{n_{\neq i} = k | n_i = 1\} \left(\frac{1}{k+1} \sum_{s=1}^k s + \frac{1}{2} \right) \quad (12)$$

$$E[N(X)_i^2] \simeq \sum_{k=0}^{\infty} P\{n_{\neq i} = k | n_i = 1\} \left(\frac{1}{k+1} \sum_{s=1}^k s^2 + \frac{1}{4} \right) \quad (13)$$

The first two moments for the waiting time are

$$E[Z_i] \simeq E[N(X)_i] E[R_{\neq i}^*] \quad (14)$$

$$E[Z_i^2] \simeq E[N(X)_i] \sigma^2(R_{\neq i}^*) + E[N(X)_i^2] E[R_{\neq i}^*]^2 \quad (15)$$

We can derive similar expressions for the full shipment/flexible truck capacity.

3 Simulations and results

In this section, we test the approximations found for the first two moments of the waiting time. The testing is done with the help of discrete event simulations. The simulations start all with the same seed and stop after 40 000 arrivals of item orders, which ensures accuracy of the simulation results. We assume that the inter-arrival time between customer orders at the retailer is mixed Erlang distributed. Furthermore we assume that the customer order sizes are mixed Erlang distributed.

3.1 Input for the simulations

For the time policy, we ran 84 different simulations to test the approximations of the waiting time distribution. In the derivations of the aggregate order process some approximations are made to estimate the second moment of the time between two replenishment orders, the approximations perform less if the number of superpositions is smaller than 16. For this reason we varied in the simulations the number of items between the 16 and 32. This is in line with practice where the number of different items is usually larger than 16. All items are identical with $E[A_i] = 1$ and $E[D_i] = 100$, because the derived

approximations for the waiting time in time policy are independent of the arrival process of orders to be consolidated. We varied $c_{A_i}^2$ and $c_{D_i}^2$ between 0.4, 1 and 1.6. The batchsizes at the retailer were varied between 500, 1000, 1500, ..., 5000. T is varied between 1, 3 and 6 and P_{2i} between 90 %, 95 % and 99 %. L_d is varied between 2 and 8.

For the quantity policy with equal batchsizes we performed 84 different simulations. The number of different items are varied between 16 and 32. For all items $E[A_i] = 1$ and $c_{A_i}^2$ is varied between 0.2, 1 and 2. In the quantity policy the waiting time due to shipment consolidation is dependent on the arrival process of orders to be consolidated, therefore we assume for the different items different demand processes. We used the fact that in practice, a few items have a large demand and many items have a small demand. We distinguish between two types of items large demand and small demand. 66 % of the items have a small demand. $E[D_i] = 100$ for the large demand and $E[D_i] = 10$ for the small demand. $c_{D_i}^2$ is varied between 0.2, 1 and 2. L_d is varied between 2, 4, 8 and 16. The batchsize is varied between 500 and 2000. Q_{max} is varied between 2000 and 4000 and P_{2i} is varied between 90 %, 95 % and 99 %.

For the quantity policy with non-equal batchsize, we performed 14 different simulations. In the simulations we assumed 16 items with four different types of demand. Type j demand is defined as D_j . The L_d is varied between 2 and 8 and the fill rate is varied between 90 %, 95 % and 99 %. The input for the cases are given Table 1.

Table 1: Input for the quantity policy simulations with non-equal batch-sizes

Case	$E[A]$	c_A^2	$E[D_1]$	$E[D_2]$	$E[D_3]$	$E[D_4]$	c_D^2	$E[Q]$	Q_{max}
1	1	1	100	150	200	250	1	900	4000
2	1	1	50	70	90	110	1	900	4000
3	1	0.4	50	70	90	110	1	900	4000
4	1	1.6	50	70	90	110	1	900	4000
5	1	1	50	70	90	110	0.4	900	4000
6	1	1	50	70	90	110	1.6	900	4000
7	1	1	50	70	90	110	1	300	8000
8	1	1	50	70	90	110	1	600	8000
9	1	1	50	70	90	110	1	900	8000
10	1	0.4	50	70	90	110	1	900	8000
11	1	1.6	50	70	90	110	1	900	8000
12	1	1	50	70	90	110	0.4	900	8000
13	1	1	50	70	90	110	1.6	900	8000
14	1	1	50	70	90	110	1	2000	8000

where $E[Q]$ is the average batchsize over all different items.

3.2 Results

We used the approximations derived previously in this paper to calculate the first two moments of the waiting time due to shipment consolidation. We fit a mixed Erlang distribution to these first two moments. Using the expressions in Appendix C we computed the reorder levels that ensure the required service levels and the resulting average physical inventory levels. To test our approximations we compare the error in the first two moments of the waiting times, the fill rate and physical inventory level. The error in the first two moments of the waiting time and the average physical inventory level is expressed in a percentage error. The error between the target P_{2i} and the one obtained with the simulation is expressed in absolute percentage error. For every simulation we calculate the relative absolute deviation as follows:

$$RAD_i = |P_{2i} - P_{2i}^{target}| * 100 \quad (16)$$

The percentage error in the average inventory (PEAI) is calculated as follows:

$$PEAI_i = \frac{|E[X_i^+] - E[X_i^{+*}]|}{E[X_i^{+*}]} * 100 \quad (17)$$

where $E[X_i^{+*}]$ is the calculated average inventory and $E[X_i^+]$ is the simulated average inventory. To be able to draw meaningful conclusions, we define acceptable margins for the RAD_i values and the $PEAI_i$ values. To construct a realistic margin, we look at the error in the probability of having backlog ($1 - P_{2i}$). In Table 2 shows good and acceptable values for the fill rate.

Table 2: Good and acceptable fill rates.

P_{2i}^{target}	Good P_{2i}			Acceptable P_{2i}		
	Min	Max	RAD_i values	Min	Max	RAD_i values
90 %	89 %	91 %	1	88 %	92 %	2
95 %	94.5 %	95.5 %	0.5	94 %	96 %	1
99 %	98.9 %	99.1 %	0.1	98.8 %	99.2 %	0.2

For the $PEAI_i$ the good margin is 2.5 % and the acceptable margin is 5%.

Time policy

In this section, we discuss the results of the simulations with time policy. For 84 simulations, the percentage error in the $E[Z]$ is between 0.5 % and 1.5 %, the percentage error in the second moment of the waiting time is between 2 % and 3.5 %. Table 3 summarizes the results.

Table 3: Summary of the results for the "time" policy

P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
90 %	0.38	0.93	100 %	100 %	1.03	6.45	83 %	95 %
95 %	0.52	1.17	93 %	98 %	0.79	5.06	90 %	100 %
99 %	0.16	0.36	64 %	78 %	0.57	3.75	90 %	100%

Where RAD_i^m and $PEAI_i^m$ are respectively the maximum RAD and $PEAI$, GP_2 and AP_2 are the percentage within respectively the good and the acceptable range for the error in the fill rate and GI and AI are the percentages within the good and acceptable range for the errors in the average inventory. Hence we conclude that our approach yields excellent results for the time policy.

Quantity policy, equal batchsizes

The results for the situation with batchsizes that are equal in volume are summarized in Table 4.

Table 4: Summary of the results for the quantity policy with equal batch-sizes

P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
90 %	0.38	1.13	92 %	100 %	1.90	7.46	80 %	99 %
95 %	0.30	0.77	81 %	100 %	1.47	6.53	80 %	99 %
99 %	0.20	0.45	20 %	60 %	1.22	4.32	90 %	99 %

The error in $E[Z]$ is between 0.7 % and 4.2 % and $\sigma^2(Z)$ is between 3.4% and 7.2 %. The peak values are caused by a high probability of having two orders of the same item in one truck. If there is a high probability of having two orders of the same item in one truck then the demand during the lead time can be no longer approximated by a mixed Erlang distribution because the distribution function of the demand during the lead time has more than one peak. The results were not satisfying in cases of extreme large or low coefficients of variation of the inter-arrival times and order sizes, which is in line with observations about two-moment approximations in general in Tijms (1994). Generally, when the number of items increase, the coefficients of variation will converge to 1 and the results will show better outcome.

Quantity policy, non-equal batchsizes

In section 2.1, we derived approximations for the waiting time due to shipment consolidation for the partial shipment non-equal batchsize quantity policy. In a similar way, we derived expressions for the waiting time due to shipment consolidation for the full shipment/flexible truck capacity policy and for the full shipment/fixed truck capacity. In the full shipment/fixed truck capacity the orders are consolidated until the collected quantity is larger than or equal to a predetermined quantity Q_{max} . In this policy, $\Delta(T) = Q_{max} - O_{N(T)} + V$ is directly shipped to the retailer and the last order $O_{N(T)}$ will be shipped with the next truck to the retailer.

For the three possible consolidation policies the errors in the $E[Z]$ are between 2.58 % and 7.44 % and in $\sigma^2(Z)$ are between 4.90 % and 16.10 %. The results for the fill rate and the average inventory are as follows:

Table 5: Summary of the results for the non-equal batchsizes

Quantity policy	P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
1	90 %	0.74	2.22	79 %	97 %	2.16	4.96	64 %	98 %
1	95 %	0.32	1.64	68 %	100 %	2.56	5.46	64 %	98 %
1	99 %	0.15	0.46	46 %	93 %	1.86	4.05	64 %	98 %
2	90 %	0.93	1.99	86 %	97 %	3.22	5.84	52 %	91 %
2	95 %	0.48	1.49	71 %	96 %	2.87	5.29	58 %	91 %
2	99 %	0.15	0.48	61 %	86 %	2.33	5.14	58 %	94 %
3	90 %	1.17	2.20	50 %	86 %	3.65	6.36	0.33	87 %
3	95 %	0.67	1.93	46 %	75 %	3.56	6.02	0.39	87 %
3	99 %	0.19	0.93	43 %	86 %	3.03	6.34	0.33	87 %

In Table 5, quantity policy 1 refers to the partial shipment policy, quantity policy 2 refers to the full shipment /flexible truck capacity policy and quantity policy 3 refers to the full shipment/fixed truck capacity policy.

In quantity policy 2, the fill rates obtained with the simulations were higher than the target fill rates and in quantity policy 3, the fill rates obtained with the simulations were lower than the target fill rates. The errors increase when the coefficients of variation of demand are high (1.6) or low (0.4) due to approximations made in the superposition of mixed Erlang distributions. For example, for the cases 2 to 6, for reference see Table 1, with $P_{2i} = 90\%$ and $L_d = 2$, Table 6 shows the differences in $E[RAD_i]$ for different coefficient of variations.

Table 6: Differences in $E[RAD_i]$ for different coefficient of variations.

c_A^2	1	0.4	1.6	1	1
c_D^2	1	1	1	0.4	1.6
$E[RAD_i]$	0.19	0.60	0.42	0.70	0.74

The errors in the fill rate increase when the average batchsize is large compared to the truck size. Table 7 shows these results, the truck size was kept constant at 8000 units and the target fill rate was 90 %.

Table 7: Differences in $E[RAD_i]$ for different number of orders per truck

$\frac{Q_{max}}{E[Q]}$	27	13	9	4
$E[RAD_i]$	0.06	0.1	0.42	1.18

The heuristic performs well as long as the order sizes are not too large compared to the truck size and c_A^2 and c_D^2 are not extremely large or low.

4 Conclusions and further research

In this paper we studied the interactions between shipment consolidation policies and inventory management policies. We argued that the lead time of orders from a retailer to a warehouse is influenced by the shipment consolidation policy used at the warehouse. In order to get a deeper insight into this interaction we derived approximations for the waiting time distribution of retailer replenishment orders due to consolidation at the warehouse for different consolidation policies used in practice. An extensive numerical study was conducted to understand and test the different approximations. The study revealed that the approximations performed well. The only problems occurred when the coefficient of variations were very low or very high. Those errors were usually due to approximations made in the superposition of the inter-arrival times. We know from previous research (Smits, de Kok, and van Laarhoven 2000) that if the number of customers and items is large and the demand between customers varies a lot the errors will diminish.

A next step in this research will be to find expressions for the waiting time in a multi-echelon setting and to find close to optimal values for the batchsizes and T in the time policy and close to optimal values for the batchsizes and Q_{max} in the quantity policy. These extensions should enable to develop models for the integrated design of transportation and supply networks that incorporate the operational characteristics of the processes under considerations, such as stochastic demand and stochastic lead times.

APPENDIX

A Replenishment process

In this section we present a procedure that translates the demand process of an item at the warehouse to a replenishment process of the item towards the central warehouse. These expressions have been derived by Pyke, De Kok and Baganha (1996). Assume O_i is the order process of product i . The first two moments of O_i are derived as follows:

$$E[O_i] = \frac{Q_i E[D_i]}{\int_0^{Q_i} P\{D_i \geq x\} dx} \quad (18)$$

$$E[O_i^2] = Q_i^2 \sum_{z=0}^{\infty} (2z+1) P\{U_i \geq zQ_i\} \quad (19)$$

where

$$P\{U_i \geq zQ_i\} = \frac{\int_0^{Q_i} P\{D_i \geq zQ_i + x\} dx}{\int_0^{Q_i} P\{D_i \geq x\} dx} \quad (20)$$

The time between the placement of two orders is defined as R_i . It is evaluated as follows:

$$R_i = \sum_{j=1}^{N_i} A_{ij} \quad (21)$$

where N_i is defined as the number of arrivals during an arbitrary replenishment cycle at the central warehouse and A_{ij} as the j^{th} inter-arrival time during this replenishment cycle at the central warehouse. There a replenishment cycle is defined as the time that elapses between two consecutive replenishment orders generated by the retailer for product i . Then the first two moments of R_i can be calculated as follows:

$$E[R_i] = E[N_i]E[A_i] \quad (22)$$

$$E[R_i^2] = E[N_i]\sigma^2(A_i) + E[N_i^2]E^2[A_i] \quad (23)$$

Due to flow conservation the following relation holds for $E[N_i]$:

$$E[N_i] = \frac{E[O_i]}{E[D_i]} \quad (24)$$

To determine an expression for $E[N_i^2]$, we apply the following approximation which is accurate when $\frac{Q_i}{E[D_i]}$ is not too small ($\frac{Q_i}{E[D_i]} > 1$):

$$E[N_i^2] \simeq \left(\frac{Q_i^2}{E^2[D_i]} + c_{D_i}^2 \frac{Q_i}{E[D_i]} + \frac{E^2[D_i^2]}{2E^4[D_i]} - \frac{E[D_i^3]}{3E^3[D_i]} \right) \frac{E[O_i]}{Q_i} \quad (25)$$

B Aggregate order process

In this paragraph we will find expressions for the first two moments of the inter-arrival time and the order size of an arbitrary order towards the central warehouse. To do this, we apply the stationary interval method developed by Whitt (1982), to superpose renewal processes. Instead of superposing hyper-exponential and shifted exponential distributions we superpose mixtures of Erlang distributions. In the superposition procedure it is assumed that the superposed process is a renewal process, which is not true. Because when we superpose N renewal processes, the first renewal time of the superposed process should be the minimum of the first renewal times of the N individual renewal processes. The superposition gives exact results when the renewal processes are Poisson distributed. The superposed process converges to the Poisson process when N tends to infinity (Tijms (1994)). The N products are represented by index i . The first two moments of T^* and O^* are calculated as follows: (see De Kok (1996))

$$E[R^*] = \frac{1}{\sum_{i=1}^N \frac{1}{E[R_i]}} \quad (26)$$

$$E[R^{*2}] \simeq 2E[R^*] \int_0^\infty \left(\prod_{i=1}^N \frac{1}{E[R_i]} \right) \left(\prod_{i=1}^N \int_0^\infty (1 - F_{R_i}(y)) dy \right) dx \quad (27)$$

$$E[O^*] = \sum_{i=1}^N \frac{E[R^*]}{E[R_i]} E[O_i] \quad (28)$$

$$E[O^{*2}] = \sum_{i=1}^N \frac{E[R^*]}{E[R_i]} E[O_i^2] \quad (29)$$

C Derivation of the reorder level and physical inventory level

First, we derive some analytical approximations to calculate the reorder levels for a target fill rate. Given the inter-arrival time A_i of each item and its

demand size D_i , the reorder levels can be analytically evaluated. The reorder levels s_i are calculated as follows: (see Janssen (1998) for a proof)

$$P_{2i} \simeq 1 - \frac{E[(D_i(L_i^*) + U_i - s_i)^+] - E[(D_i(L_i^*) + U_i - s_i - Q_i)^+]}{Q_i} \quad (30)$$

It is possible to evaluate the reorder level s_i using the bisection method. L^* is the total lead time, it is expressed as the sum of the waiting time due to consolidation and the transportation time. $D_i(L_i^*)$ is the demand for product i at the warehouse during the lead time L_i^* . The mean and the variance of $D_i(L_i^*)$ are calculated as follows: (For a detailed explanation see De Kok (1991))

$$E[D_i(L_i^*)] \simeq \left(\frac{E[L_i^*]}{E[A]} + \frac{E[A^2]}{2E^2[A]} - 1 \right) E[D_i] \quad (31)$$

$$\begin{aligned} \sigma^2(D_i(L_i^*)) \simeq & \frac{E[L_i^*]}{E[A]} \sigma^2(D_i) + \frac{E[L_i^*]}{E[A]} c_A^2 E^2[D_i] + \sigma^2(L_i^*) \frac{E^2[D_i]}{E^2[A]} \\ & + \frac{(c_A^2 - 1)}{2} \sigma^2(D_i) + \frac{(1 - c_A^4)}{12} E^2[D_i] \end{aligned} \quad (32)$$

Expressions for the first and the second moment of the undershoot are

$$E[U_i] \simeq \frac{E[D_i^2]}{2E[D_i]} \quad (33)$$

$$E[U_i^2] \simeq \frac{E[D_i^3]}{3E[D_i]} \quad (34)$$

For a derivation of these results see Tijms (1994). The average stock on hand is calculated as follows:

$$E[X_i^+] \simeq \frac{1}{2Q_i} (E[(s_i + Q_i - D_i(L_i^*))^{+2}] - E[(s_i - D_i(L_i^*))^{+2}]) \quad (35)$$

For a derivation see Janssen (1998) or De Kok (1991).

D Derivation of Z for the time policy

Theorem 3. *Z is uniformly distributed over the interval (0,T) for time policy.*

Proof. We define \tilde{A} as the residual life time of the inter-arrival time of an arbitrary customer at an arbitrary moment in time. We define W as the residual lifetime of the truck arrival process at the arbitrary moment in time. Since W is the time between an arbitrary moment in time and the departure of the truck, W is uniformly distributed over $(0, T)$, for references see Doob (1953). $k \in N$

$$\tilde{A} + Z = W + kT$$

$$\tilde{A} + Z + T - W = (k + 1)T$$

We define $X = T - W$, it is easy to see that X is uniform distributed between $(0, T)$ and $\tilde{k} = k + 1$

$$\begin{aligned}
P\{Z \leq z\} &= P\{X + \tilde{A} \in (\tilde{k}T - z, \tilde{k}T), \tilde{k} \in N\} \\
&= \sum_{\tilde{k}=1}^{\infty} \frac{1}{T} \int_0^T P\{L_s \in (\tilde{k}T - z - x, \tilde{k}T - x)\} dx \\
&= \sum_{\tilde{k}=1}^{\infty} \frac{1}{T} \int_0^T \left[\int_{\tilde{k}T - z - x}^{\infty} dF_{\tilde{A}}(\tilde{a}) - \int_{\tilde{k}T - x}^{\infty} dF_{\tilde{A}}(\tilde{a}) \right] dx \\
&= \frac{1}{T} \left[\int_0^{T-z} \int_{T-z-x}^{\infty} dF_{\tilde{A}}(\tilde{a}) dx + \int_{T-z}^T \int_0^{\infty} dF_{\tilde{A}}(\tilde{a}) dx - \int_0^T \int_{T-x}^{\infty} dF_{\tilde{A}}(\tilde{a}) dx \right] \\
&\quad + \sum_{\tilde{k}=2}^{\infty} \frac{1}{T} \left[\int_{\tilde{k}T - z}^{\infty} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) + \int_{((\tilde{k}-1)T - z)}^{\tilde{k}T - z} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) \right. \\
&\quad \left. - \int_{\tilde{k}T}^{\infty} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) - \int_{((\tilde{k}-1)T)}^{\tilde{k}T} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) \right] \\
&= \frac{1}{T} \left[\int_0^{T-z} \int_{T-z-\tilde{a}}^{T-z} dx dF_{\tilde{A}}(\tilde{a}) + \int_{T-z}^T \int_0^{T-z} dx dF_{\tilde{A}}(\tilde{a}) - \int_0^T \int_0^T dx dF_{\tilde{A}}(\tilde{a}) \right. \\
&\quad \left. + z - \int_0^T \int_{T-\tilde{a}}^T dx dF_{\tilde{A}}(\tilde{a}) \right] + \sum_{\tilde{k}=2}^{\infty} \left[\int_{(\tilde{k}T - z)}^{\infty} dF_{\tilde{A}}(\tilde{a}) - \int_{\tilde{k}T}^{\infty} dF_{\tilde{A}}(\tilde{a}) \right. \\
&\quad \left. + \int_{(\tilde{k}-1)T - z}^{\tilde{k}T - z} \left(\frac{\tilde{a} - z - (\tilde{k}-1)T}{T} \right) dF_{\tilde{A}}(\tilde{a}) - \int_{(\tilde{k}-1)T}^{\tilde{k}T} \frac{\tilde{a} - (\tilde{k}-1)T}{T} dF_{\tilde{A}}(\tilde{a}) \right] \\
&= \frac{z}{T} + \int_{T-z}^{\infty} \frac{T-z}{T} dF_{\tilde{A}}(\tilde{a}) - \int_{T-z}^T \frac{\tilde{a}}{T} dF_{\tilde{A}}(\tilde{a}) - \int_T^{\infty} dF_{\tilde{A}}(\tilde{a}) \\
&\quad + \sum_{\tilde{k}=2}^{\infty} \left[\int_{(\tilde{k}-1)T - z}^{\tilde{k}T - z} \frac{z}{T} dF_{\tilde{A}}(\tilde{a}) + \int_{(\tilde{k}-1)T - z}^{(\tilde{k}-1)T} \frac{\tilde{a} - (\tilde{k}-1)T}{T} dF_{\tilde{A}}(\tilde{a}) \right. \\
&\quad \left. - \int_{\tilde{k}T}^{\tilde{k}T} \frac{\tilde{a} - \tilde{k}T}{T} dF_{\tilde{A}}(\tilde{a}) \right] \\
&= \frac{z}{T} + \int_{T-z}^{\infty} \frac{T-z}{T} dF_{\tilde{A}}(\tilde{a}) - \int_{T-z}^T \frac{\tilde{a}}{T} dF_{\tilde{A}}(\tilde{a}) - \int_T^{\infty} dF_{\tilde{A}}(\tilde{a}) \\
&\quad + \int_{T-z}^{\infty} \frac{z}{T} dF_{\tilde{A}}(\tilde{a}) + \int_{T-z}^T \frac{\tilde{a} - T}{T} dF_{\tilde{A}}(\tilde{a}) \\
&= \frac{z}{T}
\end{aligned}$$

References

- 1 **Doob, J.L. (1953)** *Stochastic processes* . Wiley London
- 2 **Federgruen, A., H. Groenevelt and H.C. Tijms (1984)** *Coordinated replenishments in a multi-item inventory system with compound Poisson demands* . Management Science 30, p 344-357
- 3 **Goyal, S.K. and A.T. Satir, (1989)** *Joint replenishment inventory control: deterministic and stochastic models* . European Journal of Operational research 38, p 3-13
- 4 **Higginson, J.K. and J.H. Bookbinder, (1994)** *Policy recommendations for a shipment consolidation program*. Journal of Business Logistics 15, p 87-112
- 5 **Higginson, J.K. and J.H. Bookbinder, (1995a)** *Probabilistic models of freight consolidation systems*. Working paper, Department of Management Science, University of Waterloo, 1995
- 6 **Higginson, J.K. and J.H. Bookbinder, (1995b)** *Markovian Decision Processes in Shipment Consolidation*. Transportation Science 29, nb. 3
- 7 **Janssen, F.B.S.L.P. (1998)** *Inventory management systems*. Ph.D. thesis, Center for economic research, Tilburg University, The Netherlands
- 8 **Kok, A.G. de (1991)** *Basics of Inventory Management*. FEW working papers 520-525, Tilburg University, The Netherlands
- 9 **Kok, A.G. de (1996)** *Analysis of (s,nQ) -installation stock policies in divergent multi-echelon networks*. Research report TUE/TM/LBS/96-09, Eindhoven University of Technology, The Netherlands.
- 10 **Liu, L. and X. Yuan, (2000)** *Coordinating replenishments in inventory systems with correlated demands*. European Journal of Operations Research 123, p 490-503
- 11 **Pyke, D.A., A.G. de Kok and M. Baganha (1996)** *The under-shoot distribution in (s,nQ) -models* . Working paper TUE/TM/LBS/96-05, Eindhoven University of Technology, The Netherlands.
- 12 **Smits, S.R., A.G. De Kok and P.J.M. Van Laarhoven (2000)** *Analysis of Divergent N -echelon (s,nQ) -policies Under Compound Renewal Demand*. Submitted for publication
- 13 **Tijms, H.C. (1994)** *Stochastic models: an algorithmic approach*. John Wiley & Sons.
- 14 **Viswanathan, V., (2000)** *Periodic review (s,S) policies for joint replenishment inventory systems* . Working paper, Nanyang Business School, Nanyang tech. University, Singapore 1996

- 15 **Whitt, W. (1982)** *Approximating a point process by a renewal process, I: Two basic methods* . Operations Research 30, p 125-147.