Erasmus University Rotterdam

Erasmus School of Economics

Utilising wait time and fill rate service measures to optimise inventory allocation within multi-echelon distribution systems

Master Thesis Econometrics and Management Science

Operations Research and Quantitative Logistics

April 5, 2024

Author: Xander Maaskant BSc

Abstract

Optimising inventory allocation within multi-echelon distribution systems through fill rate service measures can lead to excessive stock for slow-moving items. This paper addresses this issue by proposing two Various-Service Level (VSL) models implementing both wait time and fill rate service measures. Compared to the baseline model, the VSL models significantly decrease the total stock and cost by a minimum of 31.7% and 5.1%, respectively. Results obtained through simulation indicated an adequate performance across the models. Further testing indicated the challenging scenarios, the importance of adequately selecting order quantities to optimally determine the reorder points and the sensitivity to large demand sizes. In general, the VSL models proved the possibilities to include flexible wait time targets within multi-echelon inventory optimisation.

Internship supervisor:

MSc R. Zoun

Supervisor:

Prof. dr. ir. R. Dekker

Second assessor:

Dr. O. Vicil

Contents

1 Introduction

Effective inventory management poses a significant challenge, especially within multi-echelon distribution systems. The objective of inventory management is ensuring product availability while simultaneously minimising costs. Achieving this objective is considerably more difficult in a multi-echelon setting, as one must consider all locations within the distribution chain during optimisation. These locations may experience direct as well as indirect demand, but, more importantly, they can vastly differ in demand frequency and sizes. In addition, all locations typically have a service measure to suffice, further complicating the optimisation process.

The current paper focuses on a two-echelon distribution system, which consists of a central warehouse or distribution centre (DC) supplying all local warehouses (depots). Within this distribution system depots experience direct demand, whilst only the DC experiences indirect demand. To ensure effective inventory management, all depots must satisfy a predetermined service measure. Prior research, exemplified by Axsäter (2003a) and Geelen et al. (2019), has developed models to optimise inventory allocation within such distribution systems. However, these models primarily rely on fill rate targets as their main service measure. Such a fill rate service measure can apply to the entire distribution system or be location and item specific. These type of models excel when applied to fast-moving items and in scenarios where depots have similar high demand frequencies, as they prefer to stock locally. However, slow-moving items can differ significantly in demand frequency and may require an alternative approach to avoid excessive holding costs and prolonged shelf occupancy.

This research results from an internship at Gordian Logistic Experts, a consultancy firm specialised in the field of supply chain and (spare part) inventory management. Currently at Gordian Logistic Experts, the implemented model (the *vNext* model) uses a fill rate service measure to optimise the inventory allocation within a distribution system. Since each depot must satisfy its fill rate target, the *vNext* model always allocates stock at a depot even if the demand frequency of an item is considerably low and therefore slow-moving. This results in excessive inventory within the distribution system. Hence, we will address the following research question: How can the inventory, especially of slow-moving items, be reduced within a multi-echelon distribution system whilst providing a proper and flexible solution?

At present, no proper solution exists and the implemented solution at Gordian Logistic Experts is very 'ad hoc'. Namely, reorder points at depots with slow-moving items are manually set to minus one, whilst a fill rate target of 65% is maintained for the DC (Gordian Logistic Experts, 2021). For the current research, we assume that depots' replenishment orders from the DC have constant and reliable inter-lead times. Therefore, we propose to incorporate wait time as a service measure at depots with slow-moving items. The wait time service measure permits customers to wait at a depot for a maximum period of time before being served. Thus, allowing the inventory within the distribution system to be reduced as the depots with a wait time service measure are not required to have stock on hand. The introduced models within this paper incorporate both fill rate and wait time as service measures and optimise the reorder point at all locations within the distribution system. The models distinguish between fast-moving and slow-moving items and assign a service measure accordingly. Moreover, it permits placing more inventory centrally rather than locally, which allows it to reduce the total cost within the entire multi-echelon distribution system.

The current research provides practical relevance, as it is not uncommon to encounter widely varying demand frequencies among different depots within a distribution system. Depots may vary in the number of assets, due to a difference in installed base, or even in the types of assets they handle. All of which affect the demand at a depot and making effective inventory management a challenging task. The proposed models offer a flexible alternative, incorporating multiple service measures and diverging from the existing literature that predominantly relies on fill rate service measures. Furthermore, the current paper provides valuable insights into the application of wait time service measures in a multi-echelon distribution system, extending its relevance beyond spare part inventory management.

Scientifically, the present research addresses the gap in existing literature by offering a solution to reduce the inventory in distribution systems concerning items with heavily diverging demand frequencies, especially those considered to be slow-moving. Furthermore, the existing literature mainly optimises the inventory allocation of multi-echelon distribution systems using a fill rate service measure. Approximation methods of the wait time at depots in relation to the central reorder point are proposed in the existing literature. However, unlike existing literature, the current paper combines fill rate and wait time service measures to present an alternative solution to optimise the inventory allocation within a multi-echelon distribution system.

In the subsequent sections, we first provide a detailed problem description, delineate the scope of our research and elaborate on our approach in Section 2. Thereafter, we provide a comprehensive literature review in Section 3, elaborating on the currently existing solutions in scientific literature. Furthermore, it provides an overview of different wait time approximations and discusses their important differences. Section 4 introduces the (current) *vNext* model and the (proposed) Various-Service Level models together with methods to compare their performances. Afterwards, Section 5 briefly discusses the provided data for this research. The results are analysed in Section 6, displaying the reduction in inventory allocation through the proposed models whilst also elaborating on their challenges. Subsequently, Section 7 concludes the research with a summary of our key findings. Finally, Section 8 presents a discussion regarding our most useful insights and foundations for future research.

2 Problem description and approach

This section provides a detailed problem description. First, we illustrate the problem at hand and elaborate why the current approach is inadequate. Following that, we describe our research questions and proposed solution. After which, the assumptions that delineate our scope are defined. Finally, we discuss our problem approach, through answering the research questions.

2.1 Problem definition

Gordian Logistic Experts specialises in supply chain and inventory management. They optimise inventory strategies for their clients, including multi-echelon distribution systems. More specifically, their multi-echelon clients usually deal with two-echelon distribution systems. Figure 1 provides a visualisation of such two-echelon distribution system. It displays the flow of items within the network from outside supplier to customer. The first echelon consists of the local warehouses (depots), which experience the direct demand of arriving customers. The second echelon only contains the central warehouse (DC), which experiences indirect demand and supplies the depots. Within this system, a client of Gordian Logistic Experts has full ownership of all warehouses in both echelons. Therefore, the goal is to optimise the inventory strategy of all warehouses within both echelons.

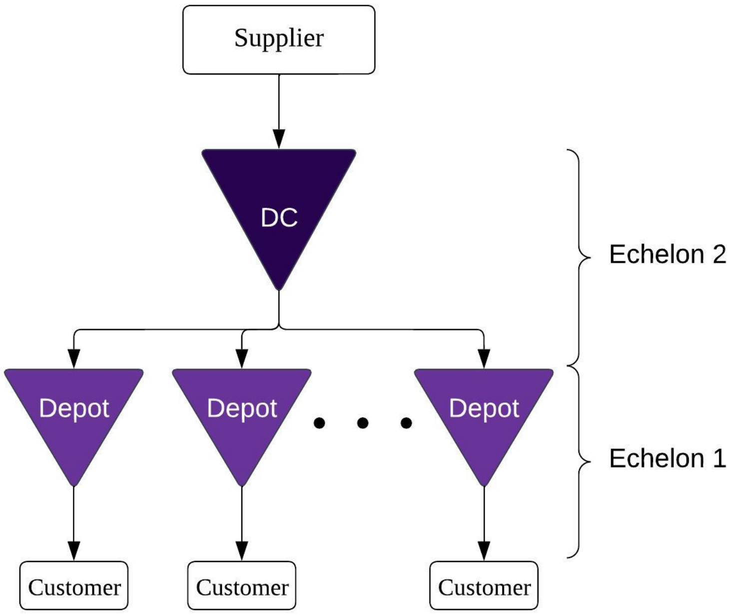

Figure 1: Visualisation of a two-level distribution system

The chosen inventory strategy at every warehouse is a general (R, nQ) strategy, where R represents the reorder point and Q the order quantity at the warehouse. The present model (the *vNext* model) uses fill rate as its service measure to determine the optimal reorder levels at all warehouses (Gordian Logistic Experts, 2021). This approach performs well when customer arrivals are frequent with little variance in demand sizes. However, it is common for depots to experience different demand frequencies, especially when dealing with spare items. When customer arrivals for an item become too infrequent at a depot, i.e. for slow-moving items, the *vNext* model allocates extravagant amounts of inventory at these local warehouses. Leading to excessive stock within the distribution system, which causes unnecessary costs for the client.

Depots experience excessive stock as the fill rate service measure needs a positive target, causing a depot to always need a positive inventory in order to suffice its set target. In other words, if there is never stock on hand available at a depot, the fill rate service measure can not be attained. Therefore, the reorder point will be non-negative, which causes there to be stock on hand continually. Depots with slow-moving items thus experience redundant stock, since stock on hand is obligatory to suffice their fill rate service measure. Therefore, the present model poses the problem of excessive stock within the distribution system due to its chosen service measure.

The currently implemented solution to solve for the excessive stock within the distribution system is very 'ad hoc'. Namely, reorder points of slow-moving depots are manually set to zero and a fill rate target of 65% is applied to DC. Through this 'ad hoc' solution, Gordian Logistic Experts is able to decrease the stock on hand within the distribution system and, therefore, its associated cost. However, as no proper optimisation method is used, this approach does not guarantee an optimal inventory strategy or proper solution. Therefore, our research answers the broad question: "How can the inventory, especially of slow-moving items, be reduced within a multi-echelon distribution system whilst providing a proper and flexible solution?"

2.2 Proposed solution

We propose to use wait time as an alternative service measure at depots with slow-moving items. The target on the wait time determines the time a customer is allowed to wait at a particular depot before being served. Utilising wait time as a service measure naturally facilitates the scenario wherein depots are not inherently required to maintain stock on hand. These depots can instead be served from the central warehouse. Therefore, shifting the inventory within the distribution system to the central warehouse and reducing the stock locally at depots with low demand.

We propose to use a probabilistic wait time. Thus, the wait time service measure has to be met for a predetermined percentile of the customers at a specific depot. This approach allows us to differentiate between the different methods of serving a customer, i.e. through stock on hand or back orders. Moreover, it offers more flexibility compared to a fixed maximum wait time if the central warehouse is out of stock. Therefore, it dismisses the consideration of determining how to weigh the wait times of specific customers or scenarios used in the expected wait time.

In this research, we propose two different models: a Simple Various-Service Level (SVSL) model and an Advanced Various-Service Level (AVSL) model. Both models focus on incorporating the wait time service measure for depots with slow-moving items and reducing the local stock. The SVSL model forces depots with slow-moving items to have a reorder point equal to minus one. Thus, these depots will only replenish upon a customer arrival and mainly serve customers through back orders. Therefore, the target on their wait time service measures must be greater than the inter-lead time between the DC and the depot. The AVSL model allows depots with slow-moving items to have reorder points that are greater than or equal to minus one. Thus, also allowing them to be non-negative. This model is therefore capable of having targets on the wait time service measure that are smaller than or equal to the inter-lead time between the DC and the depot.

Both introduced models incorporate fill rate and wait time as their service measures. A fill rate service measure will be used at depots with fast-moving items, whilst a wait time service measure will be used at depots with slow-moving items. Through these models, we provide a proper and flexible solution to the problem, whilst reducing the inventory within the distribution system. The following sub-research questions aid us in determining the necessary factors and important characteristics of the introduced models:

1. Which type of wait time approximation is most appropriate?
2. What additional assumptions need to be considered for the proposed models?
3. How do the performances of the proposed models compare to the *vNext* model?
4. What parameters significantly influence the solutions and performances of the proposed models?

2.3 Problem assumptions

For the comprehensibility of the problem, we delineate the scope of our research. We define and specify the following assumptions to provide the outline. First of all, let us indicate the assumptions on the entire two-echelon distribution system.

Assumption 1. The two-echelon distribution system is divergent with a single central warehouse.

Assumption 2. The distribution system operates under continuous review policies.

Assumption 3. All warehouses within the distribution system have unlimited storage.

Assumption 4. The distribution system only handles nonperishable items.

Assumption 5. The use of emergency shipments within the distribution system is prohibited.

Assumption 6. The use of lateral transshipments within the distribution system is prohibited.

Assumption 7. The demand within the distribution system is stationary.

Assumption 1 outlines the the structure of the distribution system. The distribution system consists of two echelons, where the first echelon consists of many depots and the second echelon contains a single central warehouse, as depicted in Figure 1. Hence, it is divergent as the single central warehouse supplies all lower echelon depots. Through *Assumption* 2 we define the type of review policies used. The use of a continuous review policy indicates that order levels are followed continuously and replenishment orders may occur at any time, as described by Axsäter (1993). *Assumption* 3 limits the scope of the research to explicitly optimise the total inventory within the distribution system rather than optimising the storage of the items. To ensure that items can only be removed from inventory through customer demand, we introduce *Assumption* 4. Prohibiting emergency shipments (*Assumption* 5) and lateral transshipments (*Assumption* 6), allows us to properly evaluate the performance of the distribution system. Moreover, *Assumption* 5 also ensures that no additional supply enters the distribution system, whilst *Assumption* 6 also ensures that no stock moves between locations within the same echelon, therefore the DC is the only supplier for the depots. *Assumption* 7 entails that no changes in demand size and frequency occur within the distribution system over time.

Finally, we state assumptions with regard to the functioning of the warehouses.

Assumption 8. All warehouses follow an $(R,\, nQ)$ policy, where $R \in \mathbb{Z}$ and $Q \in \mathbb{N}^+$.

Assumption 9. All warehouses operate on a first-come, first-serve basis.

Assumption 10. Partially fulfilled orders are prohibited within the distribution system and unsatisfied demand is back ordered with the available stock on hand being reserved.

Assumption 11. All warehouses place replenishment orders when their inventory position is at or below their reorder point.

As per the problem description, we use *Assumption* 8 to indicate the inventory policy under which all warehouses operate. Furthermore, warehouses only handle complete items, therefore all values must be integer based. More specifically, the order quantity Q is a strictly positive integer. *Assumption* 9 describes how arrivals at all warehouses are handled. Moreover, *Assumption* 10 indicates that an arrival at a warehouses must be completely fulfilled before continuing to the succeeding arrival. Hence, if the demand of an arrival cannot be completely fulfilled, the remainder is back ordered with the currently available stock on hand being reserved for the initial arriving customer. This also coincides with *Assumption* 8, which allows for warehouses to work (almost) completely with back orders, since the reorder points are also allowed to be non-positive. Finally, *Assumption* 11 dictates exactly when replenishment orders are placed.

2.4 Problem approach

Let us now outline our approach to answer the primary research question: How can the inventory within a multi-echelon distribution system be reduced whilst providing a proper and flexible solution? To better address this broad question, we have formulated four sub-research questions, see Section 2.2. These sub-questions aid us in defining a proper solution and offer valuable insights into the essential factors and item characteristics inherent to the proposed models in the current paper.

To address the first sub-research question, we conduct a literature review on existing literature to select a wait time approximation method that best fits the scope of the problem at hand. Section 3 provides an overview of the considered wait time approximations and elaborates on the advantages and disadvantages per method. Subsequently, Section 4.2 elaborates on which approximation method is chosen, considering the current problem and the existing *vNext* model. Throughout Section 4, we introduce additional assumptions, thus addressing the second sub-research question. These assumptions are primarily derived from the insights and recommendations of prior research. Moreover, additional assumptions must be formulated based on the constraints and limitations inherent to the selected wait time approximation method.

Moreover, in Section 4, we present the equations and algorithms for all models, along with a discussion on the implementation and performance measurement. Gordian Logistic Experts provided data, as detailed in Section 5, which we will utilise as a case study to access and evaluate the performance of the different models. Section 6 describes and depicts the obtained results for all models, using the *vNext* model's performance as a baseline for comparing the performances of the proposed models.

To assess the performances, two distinct analyses are conducted. First, the inventory allocation is investigated across all models, illustrating the potential stock and cost savings of the proposed models. Secondly, customer arrivals are simulated within the distribution system to evaluate the inventory allocation performance of the different models over a prolonged time period. The simulations incorporate the historical demand patterns within the provided data to imitate realistic circumstances within the distribution system. During both performance measures, we evaluate two distinct items individually. Subsequently, the overall performance of the entire distribution system is analysed. Through the assessment of the distinct items, we aim to provide a more in-depth analysis and a better insights into the choices made in allocating inventory by the different models. Combining these analyses allows us to address the third sub-research question.

Continuing from the performance measurement, we address the fourth and final sub-research question in Section 6.3 by examining the robustness of the proposed models. The tests assess the robustness of the entire inventory allocation process, emphasising on the overall impact of the modifications made rather than focusing specifically on the precise effects of changes in a single parameter. Various item characteristics will be modified to observe their influence on the performance of the inventory allocation and simulation for the proposed models. Identifying the parameters that significantly impact the performance offers subsequent research a foundation to address possible issues that may arise in the proposed models.

The main research question of in the current paper is addressed throughout Sections 4 to 7. In these section two distinct models are proposed, both with unique purposes and limitations. The SVSL model adopts a more straightforward approach, offering a proper solution to the problem by adequately approximating the wait time and implementing it as a service measure in a multi-echelon distribution system. The AVSL model provides both a proper and flexible solution. It enables a more pragmatic use of wait time service measures by lifting restrictions on the possible targets of the wait time service measure. Both models utilise both fill rate and wait time as service measures within a multi-echelon distribution system, providing an alternative solution to these type of inventory management problems. Furthermore, we conduct an experiment on a simplified scenario to further justify our decision regarding the scope of the proposed models. Finally, Section 7 concludes our findings, and discusses the strengths and crucial parameters of the proposed models.

3 Literature review

This section discusses the relevant literature regarding multi-echelon distribution systems and wait time approximation methods. A lot of different type of inventory management systems and strategies have been developed in the past. de Kok et al. (2018) provided a review and typology study on different stochastic multi-echelon inventory systems. Moreover, they indicated a detailed classification of the research conducted up until 2017. Through their research, they were able to outline major classes within the study of multi-echelon inventory models and classify papers through their different dimensions. These dimensions specify which type of problem and model is being researched combined with their selected assumptions. Therefore, we will review papers within our scope and relevant to the problem at hand, as per Section 2.

An important aspect of inventory management systems is the selection of a review policy. As per Section 2.3, one of the assumptions of the current research is to use a continuous review policy for the distribution system. By opting for this policy, we avoid the use of the so-called *balance assumption*. As explained by Doğru et al. (2009), the *balance assumption* indicates that the supply within an inventory system is exactly equal to the demand of the inventory system over a period of time. In other words, the assumption relaxes the constraint on the inventory positions at local warehouses (depots) after replenishment needing to be greater than or equal to the inventory positions prior to the replenishment. Therefore, optimal inventory allocation within the distribution system is always achievable, and the inventory positions of downstream depots become insignificant.

Axsäter (2003b) stated that, in general, the *balance assumption* provides a good approximation of inventory levels. However, several studies have investigated the effects of the balance assumption on inventory management models. For instance, Axsäter et al. (2002) indicated that the use of this assumption could cause considerable errors when demand characteristics deviate significantly at depots, as is common in practice. Moreover, according to Doğru et al. (2009) and Gallego et al. (2007), the *balance assumption* is unsuitable in many scenarios. In particular, if the lead times to the central warehouse (DC) are lengthy and the lead times to the downstream depots are short, the scenario also encountered in the current paper, see Section 5.

The choice of order policy is another crucial dimension de Kok et al. (2018) consider. Current literature distinguishes between different policies. Chew and Tang (1995) proposed an approach using an (s,S)-policy. This policy orders up until a given level S, i.e. it allows for non-fixed order quantities. Besides allowing for varying order quantities, the (s,S)-policy also frequently applies the *balance assumption*, as done by Diks and De Kok (1998). Most of these policies ignore the restriction of a Minimum Order Quantity (MOQ). In order to account for this, Kiesmüller et al. (2011) proposed a so-called base-stock policy or S-policy. This policy orders up to S whenever the inventory level drops below s. If the order size is smaller than the MOQ, the order size is set to the MOQ. However, models incorporating this policy mainly use periodic review. Therefore, as mentioned in Section 2.3, the current paper operates under the assumption of using (R,nQ)-policies. These policies depend on a fixed order quantity and can incorporate MOQs. Moreover, as discussed in Section 4, it is assumed that the order quantities are known and given.

Within the current literature, models exist to optimise for a multi-echelon distribution system using fill rate targets as its service measure. For example, Axsäter (2003a) developed such a model. However, in their optimisation, they make use of holding and shortage costs, which usually need to be approximated and can introduce errors when estimated incorrectly. Moreover, Forsberg (1997) has shown in their research that exactly determining the holding and shortage cost of an (R,Q)-policy drastically increases the computation time, especially when the order quantities change. Therefore, during a previous internship at Gordian Logistic Experts, Geelen et al. (2019) introduced a model to optimise over a multi-echelon distribution system without the need for holding and shortage costs. Their model is mainly based on the works of Grob and Bley (2018) and Axsäter (2003a). The present multi-echelon model of Gordian Logistic Experts (2021) (*vNext* model) is based on the research of Geelen et al. (2019).

Building on the *vNext* model, the proposed models in the current paper will also use wait time service measures in combination with fill rate service measures. In order to use wait time targets, the wait time has to be approximated. Grob and Bley (2018) provided an overview of four different wait time approximation methods.

The first kind of approach is the so-called METRIC type of approximation. It was first introduced by Sherbrooke (1968) and is based on Little's law. The approach from Axsäter (2003a) was selected to represent this class of approximations. This approach approximates the mean and variance of the wait time under the assumption of normally distributed demand. The wait time is considered to be identical across all depots, not considering their individual contribution to the overall demand at the DC.

Secondly, Kiesmüller et al. (2011) provided a different approach. They approximate the first two moments of the wait time using the stationary interval method from Whitt (1982). Their approach is only applicable to non-negative central reorder points. Moreover, as noted by Grob and Bley (2018) and Berling and Farvid (2014), their approach leads to negative values for the variance in some cases.

Berling and Farvid (2014) provided a third approach that also computes the first two moments of the wait time using a stochastic demand rate. This rate considers the excess demand of other depots at the DC and, therefore, approximates the moments for the wait time per depot. It considers three different cases depending on the central reorder point and the local order quantity. Grob and Bley (2018) do note that the approach of Berling and Farvid (2014) fails when a local order quantity is greater than the sum of the central reorder point and central order quantity. In that case, it produces a negative expected value for the wait time. However, in practice it is improbable for this scenario to occur.

The final approach to approximate the wait time considered, is proposed by Grob and Bley (2018) themselves. They used the negative binomial distribution to approximate the aggregate lead time demand. After which, they compute the first two moments of the wait time identically to the approach provided by Kiesmüller et al. (2011). Their approach solved most issues encountered by Kiesmüller et al. (2011). However, they did assume that the actual order size is always equal to the order quantity. Moreover, their approach is still not applicable for negative central reorder points.

Using one of the wait time approximation methods described by Grob and Bley (2018), we approximate the customer wait time at a depot and utilise it to determine its optimal reorder point. Hence, our research integrates wait time and fill rate service measures, using the present multi-echelon model of Gordian Logistic Experts (2021) as a framework. Thereafter, we propose two new models that optimise all reorder points within a distribution system, utilising both wait time and fill rate service measures within the same distribution system. Moreover, we propose adjustments to the current model to decrease its computation time compared to the original introduced method by Geelen et al. (2019).

4 Methodology

This section defines all models and describes the methods utilised within the current research. We will start by specifying the existing *vNext* model, which uses fill rate as its service measure (Gordian Logistic Experts, 2021). Secondly, we will introduce the Simple and Advanced Various-Service Measure models. These new models implement both fill rate and wait time service measures. Furthermore, we discuss the implementation of the models together with their required decisions. Finally, we provide the methods used to evaluate the performances of the different models.

Before continuing to with describing and defining the models for this research, let us elaborate on the general notation used throughout the remainder of the paper. For every item $i = 1, \ldots, N$, the goal is to optimise for the reorder points R_{ij}, where $j = 0, 1, \ldots, M_i$. Here j represents a warehouse. In particular, it represents the central warehouse (DC) if $j = 0$ and a local warehouse (depot) if $j = 1, \ldots, M_i$. Note that the number of depots differs per item, since not every item is handled by all warehouses (except for the DC, needless to say). For easier notation, we will introduce the set $J_i = \{1, \ldots, M_i\}$ and $J_i^+ = J_i \cup \{0\}$. For the general notation we drop the index i, as we optimise the reorder points per item for all warehouses. Hence, the set J_i^+ indicates that these variables only correspond to item i at all warehouses j. Therefore, the general notation is provided below and all subsequent sections utilise similar notation.

R_j	The reorder point at warehouse $j \in J_i^+$.
Q_j	The order quantity of warehouse $j \in J_i^+$.
L_j	The lead time for warehouse $j \in J_i^+$.
IL_j	The inventory level at warehouse $j \in J_i^+$.
IP_j	The inventory position at warehouse $j \in J_i^+$.

Section 2 provided the scope of the current paper and introduced the necessary assumptions for the problem at hand. We will now introduce the set of assumptions that apply to and hold for the models depicted in the current paper. Any deviations from these assumptions will be clearly indicated and elaborated on in the subsequent sections.

Assumption 12. All models optimise for a single item, no consolidation between different items.

Assumption 13. The DC lead times are independently and identically distributed random variables with known mean and variance.

Assumption 14. The inter-lead times from the DC to the depots are deterministic.

Assumption 15. Single-item batching is allowed.

Assumption 16. The order quantity Q is known for each item at every warehouse.

Assumption 17. Customers arrive according to independent compound Poisson processes across all depots.

Assumption 18. Demand sizes are strictly positive and integer valued.

As stated at the beginning of the current section, the goal is to optimise the reorder points for all warehouses per item. Hence, *Assumption* 12 indicates that the models will optimise for the reorder points on item level. Furthermore, it also states that consolidation between different items is not possible, thus providing item specific solutions. *Assumption* 13 indicates the stochasticity of the DC lead times and assumes that the mean and variance are known. On the other hand, *Assumption* 14 denotes that the inter-lead times are known and constant. *Assumption* 15 states that single-item batching is allowed since all locations operate on an (R_j, nQ_j) policy. Thus indicating that n multiples of the order quantity Q_j can be batched and ordered at once. Moreover, as per *Assumption* 16, we assume that the order quantity is known for all items at all locations as the scope of this research focuses on optimising the reorder points rather than the order quantities.

Finally, *Assumptions* 17 & 18 describe the arrival process of the customers and indicate that demand sizes are strictly positive integers. Moreover, the former indicates that all compound Poisson processes are independent across the depots. Hence, each individual compound Poisson process depicts the arrival process only for a specific depot. As customers will arrive according to a Poisson process, their inter-arrival times are exponentially distributed. Furthermore, as stated by Axsäter (2015), it can be assumed that demand sizes are strictly positive without a loss of generality, as the latter assumption indicates. Lastly, assuming a Poisson arrival process permits the use of the Poisson Arrivals See Time Averages (or PASTA) property (Wolff, 1982), indicating that customers always observe the distribution system in steady state.

4.1 The *vNext* model

The *vNext* model (Gordian Logistic Experts, 2021) is the currently existing model. It determines the optimal reorder points for a two-level distribution system using fill rate service measures. The model is mainly based on the thesis of Geelen et al. (2019) at Gordian Logistic Experts. This section outlines the theoretical framework of the model and the adjustments made compared to Geelen et al. (2019). Section 4.4 discusses the implementation of the *vNext* model, including the alterations made to the theoretical framework for implementation purposes and to decrease computation time.

Let us first introduce the notation for this section:

μ_j	The expected demand at depot $j \in J_i$ during one time period.
σ_j	The standard deviation of the demand at depot $j \in J_i$ during one time period.
q	The greatest common divisor of the local order quantities; $q = gcd(Q_1, \dots, Q_{M_i})$.
$D_j(l)$	The stochastic demand at depot $j \in J_i$ over a time period of length l.
$D_0(L_0)$	The stochastic lead time demand of the DC.
L_j^{eff}	The effective lead time at depot $j \in J_i$.
K_j	The stochastic customer demand sizes at depot $j \in J_i^+$.
$\beta_j(R_j)$	The fill rate at warehouse $j \in J_i^+$ given R_j.
β_j^T	The fill rate target at depot $j \in J_i$.
Δ	The delay experienced by a depot at the DC, it indicates the experienced wait time for a replenishment order at the DC for a depot. The delay is assumed to be identical across all depots $j \in J_i$.
$TS(R_0)$	The total stock within the distribution system given R_0; $TS(R_0) = \sum_{j \in J_i^+} R_j$.

The *vNext* model can be divided into three steps. The coming subsections will discuss these in order. The three main steps are:

I. Compute the distribution of the central lead time demand $D_0(L_0)$.

II. Compute bounds on all the reorder points R_j for $j \in J_i^+$.

III. Compute the optimal central reorder point R_0^* minimising the total stock $TS(R_0)$.

As mentioned in *Assumption* 16, the current paper assumes that Q_j for $j \in J_i^+$ is known. The computation and optimisation of these values are not within the scope of the current paper. Do note that the chosen order quantities can affect the optimisation of the reorder points and, therefore, should be selected carefully.

4.1.1 The central lead time demand

The first step of the *vNext* model is to compute the distribution central lead time demand $D_0(L_0)$. The formulas used for this step are mainly based on the research of Grob and Bley (2018), as explained by Geelen et al. (2019). However, alterations are made in comparison with the original formulas, and are highlighted throughout this section.

We ultimately want to compute the central lead time demand in this step. To do so, we make use of the formulas proposed by Grob and Bley (2018). Let us, therefore, first introduce $\delta_j(k)$. It represents the probability of depot $j \in J_i$ placing at most k orders of size Q_j over the central lead time L_0, for $k \in \mathbb{N}$. As indicated by *Assumption* 13, L_0 is considered to be a random variable with probability density function $pdf_{L_0}(l)$. Hence, we can condition on this probability by setting $L_0 = l$, indicating the stochasticity of this variable. The conditioned probability is given by

$$\delta_j(k, L_0 = l) = \sum_{x=1}^{Q_j} \frac{1}{Q_j} Pr(D_j(l) \leq kQ_j + x - 1), \tag{1}$$

as described by Grob and Bley (2018). In this equation the PASTA property is used such that all arriving customers observe the distribution system in a stationary position. Using this property, Axsäter (2015) described the uniformity property of the inventory position as $IP_j \sim \text{Unif}(R_j + 1, R_j + Q_j)$, or, as used in (1), $IP_j - R_j \sim \text{Unif}(1, Q_j)$. Hence, due to the PASTA property, the probability of the IP_j being any state within the given domain is $\frac{1}{Q_j}$. Furthermore, as k tends to infinity, the probability $\delta_j(k)$ tends to one. Therefore, if k gets sufficiently large, its contribution will become negligible.

Integrating over the lead time and incorporating pdf_{L_0}, allows us to relax the conditioning on L_0:

$$\delta_j(k) = \int_{l=0}^{\infty} \left(\sum_{x=1}^{Q_j} \frac{1}{Q_j} Pr(D_j(l) \leq kQ_j + x - 1) \cdot pdf_{L_0}(l) \right) dl. \tag{2}$$

Using $\delta_j(k)$, Grob and Bley (2018) indicate that the probability of depot j placing exactly k orders of size Q_j over L_0, given as $s_j^{ord}(k)$, can be computed as:

$$s_j^{ord}(k) = \begin{cases} \delta_j(0), & \text{if } k = 0 \\ \delta_j(k) - \delta_j(k-1), & \text{if } k > 0,\ k \in \mathbb{N} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Note that this probability only uses $k \in \mathbb{N}$, as the order quantities and demand sizes are both non-negative integers (*Assumption* 8 & 18). Therefore, all multiples of order quantities must be integer valued as well.

Given (2) and (3) and using the independence of the orders from the depots (due to the independent arrival processes, see *Assumption* 17), we can compute the mean and variance of the central lead time demand as

$$\mathbb{E}[D_0(L_0)] = \sum_{j=1}^{M} \mu_j \mathbb{E}[L_0], \tag{4}$$

$$Var[D_0(L_0)] = \sum_{j=1}^{M} \left[\sum_{k=0}^{\infty} (\mu_j \mathbb{E}[L_0] - kQ_j)^2 s_j^{ord}(k) \right], \tag{5}$$

(Grob and Bley, 2018). The mean equals the summation of the average demand over the central lead time per depot. The variance evaluates the squared deviance between the average demand over the central lead time and the specific order of size kQ_j, which is multiplied by the probability of the depot placing exactly k orders of size Q_j. Orders with a small deviance or with lower probabilities, i.e. uncommon order sizes, add little value to the variance. On the contrary, large deviating orders from the average that are common, i.e. the orders with higher probability, add more significant value to the variance.

The mean and variance can be used to fit a standard distribution for $D_0(L_0)$. Section 4.4 describes which standard distribution will be used and how it will be estimated.

4.1.2 Bounds on the reorder points

In this step, we compute the bounds on the central and local reorder points. Moreover, we provide an algorithm that optimises for the lowest local reorder point given a specific central reorder point and fill rate target. As stated by *Assumption* 8, reorder points must be integer valued. Furthermore, to ensure a positive inventory position which is needed for a positive fill rate, it is essential for the inequality $R > -Q$ to hold, as explained by Axsäter (2015). This property must hold for all warehouses as they need a positive fill rate to oblige by their fill rate service measures.

Besides this strict lower bound on the reorder point, we will bound the reorder level from above and further tighten it from below through defining R_j^{min} and R_j^{max} for all $j \in J_i^+$. Let us however first define how to compute the fill rate of a depot after which we will elaborate on how to determine R_j^{min} and R_j^{max}. The computation of the fill rate is very similar for the depots and the DC, the differences will be discussed after describing the computations.

Computing local fill rates

The fill rate β_j for a depot $j \in J_i$ is affected by the central reorder point R_0. Hence, for the computation of the local fill rates it is assumed that R_0 is known and given. To compute $\beta_j(R_j)$, we first need the effective lead time of a depot. To determine the effective lead time we assume that the lead time of a depot and the delay experienced by the depot are mutually independent. The delay indicates the experienced wait time for a replenishment order at the DC for a depot and is assumed to be identical across all depots $j \in J_i$. We approximate the delay using the METRIC-type approximation methods described by Axsäter (2003a), see Appendix A. Note that the approximation depends on the given R_0. As described by Geelen et al. (2019), the mean and variance of the effective lead time for a depot can be computed as:

$$\mathbb{E}[L_j^{eff}] = \mathbb{E}[L_j] + \mathbb{E}[\Delta], \tag{6}$$

$$Var[L_j^{eff}] = Var[L_j] + Var[\Delta]. \tag{7}$$

Using the effective lead time, Grob and Bley (2018) propose the following equations to determine the mean and variance of the effective lead time demand:

$$\mathbb{E}[D_j(L_j^{eff})] = \mu_j \mathbb{E}[L_j^{eff}], \tag{8}$$

$$Var[D_j(L_j^{eff})] = \sigma_j^2 \mathbb{E}[L_j^{eff}] + \mu_j^2 Var[L_j^{eff}]. \tag{9}$$

These are used to fit a standard distribution to the demand over the effective lead time, see Section 4.4.

Next, we need $Pr[IL_j = p]$, i.e. the probability of the inventory level at depot j equals p. To compute this probability, Grob and Bley (2018) uses the properties and equations provided by Axsäter (2015) for the inventory level and the inventory position based on the PASTA property of the Poisson process (*Assumption* 17). Identically to (1), the uniformity property of the inventory position indicates that $IP_j \sim \text{Unif}(R_j + 1, R_j + Q_j)$. Using this property together with the distribution of the demand over the effective lead time, Grob and Bley (2018) derive the probability of the inventory level at depot j equal to p, as:

$$\begin{aligned}
Pr[IL_j = p] &= Pr[IP_j - D_j(L_j^{eff}) = p] \\
&= \sum_{x=0}^{\infty} Pr[IP_j - D_j(L_j^{eff}) = p \,|\, IP_j = x] \cdot Pr[IP_j = x] \\
&= \sum_{x=\max\{R_j+1,p\}}^{R_j+Q_j} Pr[D_j(L_j^{eff}) = x - p \,|\, IP_j = x] \cdot Pr[IP_j = x]
\end{aligned}$$

$$Pr[IL_j = p] = \frac{1}{Q_j} \sum_{x=\max\{R_j+1,p\}}^{R_j+Q_j} Pr[D_j(L_j^{eff}) = x - p]. \tag{10}$$

Note that, contrary to Grob and Bley (2018), for x we sum over the domain $[\max\{R_j + 1,\, p\},\, R_j + Q_j]$ as also proposed by Axsäter (2015). *Assumption* 18 states that the demand sizes are strictly positive, therefore the demand over a certain time frame must be non-negative. Hence, the probability $Pr(D_j(L_j^{eff}) = x - p)$ is only defined for $x - p \geq 0$. Therefore, $x \geq p$ which is implied by the maximum over the two different starting values for x.

As stated in *Assumption* 10, we do not allow for partially fulfilled orders. Therefore, all customers arriving at depot j are only fulfilled if the inventory level p is greater or equal to the incoming customer demand size k. Let us denote the probability of customers having a demand size of k as $pdf_{K_j}(k)$. It is assumed that these probabilities are known. As a result, we can compute the order fill rate

$$\begin{aligned}
\beta_j(R_j) &= \sum_{k=1}^{R_j+Q_j} \sum_{p=k}^{R_j+Q_j} pdf_{K_j}(k)\, Pr(IL_j = p) \\
&= \sum_{k=1}^{R_j+Q_j} \sum_{p=k}^{R_j+Q_j} pdf_{K_j}(k) \frac{1}{Q_j} \sum_{x=\max\{R_j+1,p\}}^{R_j+Q_j} Pr(D_j(L_j^{eff}) = x - p).
\end{aligned} \tag{11}$$

Determining the local bounds

All depots $j \in J_i$ must satisfy their specified fill rate target β_j^T. For a depot j we can determine its optimal R_j such that it suffices the target β_j^T for a given R_0, through Algorithm 1. R_0 influences the central wait time, affecting a depot's effective lead time, see (6) and (7). The optimal R_j is the smallest R_j such that it suffices the target β_j^T. Moreover, the fill rate β_j is increasing in R_j. Therefore, if the fill rate is too low for R_j^{min} and too high for R_j^{max}, we can determine the optimal R_j through bisection. Hence, Algorithm 1 uses a bisection-based method.

The algorithm uses R_j^{min} and R_j^{max} as starting values. Thus, we let $R_j^{min} = -Q_j$ and $R_j^{max} = V \cdot Q_j$, where V is sufficiently large such that $\beta_j(V \cdot Q_j) \geq \beta_j^T$ always holds. Geelen et al. (2019) also computes the optimal R_j for a given R_0. Contrary to their approach, Algorithm 1 does not iterate over all possible values of R_j between R_j^{min} and R_j^{max}, thus requiring less computations.

Algorithm 1 Find optimal R_j for a warehouse $j \in J_i$ given R_0

Input β_j^T, R_0, R_j^{min} and R_j^{max}

Output R_j^*

function FINDREORDERPOINT(β_j^T, R_0, R_j^{min} and R_j^{max})
  Determine $\mathbb{E}[\Delta]$ and $Var[\Delta]$ ▷ Axsäter (2003a)
  $R_j^{low} = R_j^{min}$
  $R_j^{top} = R_j^{max}$
  $R_j^{mid} = \left\lceil \frac{R_j^{low}+R_j^{top}}{2} \right\rceil$

  Compute $\beta_j(R_j^{low})$ ▷ Equation (11)
  if $\beta_j(R_j^{low}) >= \beta_j^T$ **then**
    $R_j^* = R_j^{low}$
    return R_j^*
  end if

  while $R_j^{mid} \neq R_j^{low}$ and $R_j^{mid} \neq R_j^{top}$ **do**
    Compute $\beta_j(R_j^{mid})$ ▷ Equation (11)
    if $\beta_j(R_j^{mid}) >= \beta_j^T$ **then**
      $R_j^{top} = R_j^{mid}$
    else
      $R_j^{low} = R_j^{mid}$
    end if
    $R_j^{mid} = \left\lceil \frac{R_j^{low}+R_j^{top}}{2} \right\rceil$
  end while

  $R_j^* = R_j^{top}$
  return R_j^*

22: **end function**

Algorithm 1 first computes the expected value and variance of the central wait time, which depends on R_0. If the fill rate target is satisfied for R_j^{min} then it is the optimal value. Otherwise, the algorithm computes the fill rate for the intermediate reorder point R_j^{mid}. The current upper bound R_j^{top} is decreased to R_j^{mid} if the fill rate target β_j^T is satisfied for R_j^{mid}. Otherwise the lower bound R_j^{low} is increased to R_j^{mid}. This continues until either the current lower or upper bound is equal to the new intermediate reorder point, i.e. the difference between the lower and upper bound is exactly one. The algorithm then always returns R_j^{top}, which is the lowest reorder point to satisfy the specified fill rate target β_j^T and thus optimal.

As stated earlier, a reorder point for a depot $j \in J_i$ is affected by the central reorder point R_0. However, as depicted above, R_0 does not directly affect the local fill rate. Instead it influences the central wait time. Therefore, we can identify which cases bound the central wait time that in turn bounds the local reorder points. Hence, the lower and upper bound on the reorder points of a depot j can be described as:

R_j^{min*}: Achieved when the DC is always able to supply the depot j on time implicating the central wait time is zero. Depot j holds the minimum amount of stock on hand to satisfy it's demand requests.

R_j^{max*}: Achieved when the DC restocks very rarely, i.e. $R_0 = -Q_0$, implicating the central wait time is maximal. Depot j holds maximal amount of stock on hand to satisfy it's demand requests.

The starting values R_j^{min} and $R_j max$ of Algorithm 1 can thus be improved upon. The description above determines the scenarios in which R_j^{min} and R_j^{max} occur. Therefore, to compute R_j^{min}, we set $\mathbb{E}[\Delta]$ and $Var[\Delta]$ equal zero in Algorithm 1, the corresponding R_0 is irrelevant in this case. The resulting reorder point is R_j^{min*}. To compute R_j^{max} we set R_0 to its theoretical minimum, as stated by Axsäter (2015), of $-Q_0$. Hence, $\mathbb{E}[\Delta]$ and $Var[\Delta]$ attain their maximum values. The resulting reorder point is R_j^{max*}. We can now replace the initial R_j^{min} and R_j^{max} by the lower bound R_j^{min*} and the upper bound R_j^{max*}, respectively, for all depots $j \in J_i$.

Computing the central fill rate

To compute the fill rate of the central warehouse ($\beta_0(R_0)$), we use the distribution of the central lead time demand, determined in Section 4.1.1, as the distribution of the demand in (11). Moreover, as stated by Geelen et al. (2019), the distribution of the order size at the DC is given as:

$$pdf_{K_0}(k) = \frac{\sum_{j=1|Q_j=k}^{M_i} \left(\sum_{z=0}^{\infty} s_j^{ord}(z) * z\right)}{\sum_{j=1}^{M_i} \left(\sum_{z=0}^{\infty} s_j^{ord}(z) * z\right)}. \tag{12}$$

Therefore, the fill rate of the central warehouse can be computed as:

$$\beta_0(R_0) = \sum_{k=1}^{R_0+Q_0} \sum_{p=k}^{R_0+Q_0} pdf_{K_0}(k) \frac{1}{Q_0} \sum_{x=\max\{R_0+1,p\}}^{R_0+Q_0} Pr(D_0(L_0) = x - p). \tag{13}$$

Note that contrary to the local fill rates, the central fill rate simply uses L_0 instead of the effective lead time as the delays encountered for the DC are encompassed by the central lead time. Moreover, as noted by Geelen et al. (2019), the probability in (12) only holds for an (R, Q) policy but not an (R, nQ) policy as assumed in *Assumption* 8. However, to adjust (12) such that it fits an (R, Q) policy requires additional complexity. Therefore, since the central fill rate will only be applied to determine the initial bounds on the central reorder point, R_0^{min} and R_0^{max} see below, the benefit of the additional accuracy does not outweigh the additional complexity needed to adjust (12).

Furthermore, the arrival process at the DC is not naturally a Poisson process as the replenishment orders are generated through the arrival processes of the depots. Therefore, we assume that the central arrival process is a compound Poisson process in (13), which allows us to use the PASTA property of the Poisson process (*Assumption* 17) for (13) to compute the probabilities of the central inventory level and inventory position. Furthermore, we expect the arrival process at the DC to converge to a Poisson process when the number of independent depots with distinct order quantities is increasing.

Finally, *Assumption* 16 indicates that all order quantities are known and given. However, as also indicated by Grob and Bley (2018), for a proper approximation of $\beta_0(R_0)$ we need $q = gcd(Q_0, \dots, Q_{M_i}) = 1$ to hold. Therefore, if $q > 1$, we need to divide all demand related variables, i.e. μ_j, σ_j^2 and Q_j for all $j \in J_i$, by q in (1) - (5).

Determining the central bounds

For the DC, we need β_0^{min} and β_0^{max} to determine the bounds, these are assumed to be given and indicate the limits on the fill rate for the DC. Moreover, it should hold that $0 < \beta_0^{min} < \beta_0^{max} \leq 1$. With these values, we can describe and compute these bounds as

R_0^{min}: Compute the smallest R_0 for which the central fill rate $\beta_0(R_0) \geq \beta_0^{min}$.

R_0^{max}: Compute the smallest R_0 for which the central fill rate $\beta_0(R_0) \geq \beta_0^{max}$.

In the case of $R_0 = R_0^{min}$ the distribution system encounters the maximum central delay Δ given β_0^{min}. On the other hand, if $R_0 = R_0^{max}$ the distribution system encounters the minimum Δ given β_0^{max}. Note that as β_0 goes towards zero, the $\mathbb{E}[\Delta]$ increases towards its limit. Additionally, as β_0 goes towards one, the $\mathbb{E}[\Delta]$ decreases towards zero.

To compute the bounds R_0^{min} and R_0^{max}, one could start from $-Q_0$ and increase R_0 by one until the fill rate is greater or equal to the provided target value for the lower and upper bound. However, we can slightly alter Algorithm 1 to determine these bounds in similar fashion to R_j^{min} and R_j^{max}. The central wait time computation can be omitted. We use the given β_0^{min} and β_0^{max} as previously described. Finally, the fill rate computation as in (11) will be replaced by (13). Through these alterations, Algorithm 1 allows us to find all the lower and upper bounds for the DC and thus every warehouse $j \in J_i^+$.

4.1.3 Computing optimal reorder points

The final step is to compute the optimal central reorder point R_0^* such that the total stock $TS(R_0)$ is minimised. The total stock given a central reorder point is defined as $TS(R_0) = \sum_{j \in J_i^+} R_j$. Geelen et al.

(2019) determines the optimal R_0 for which the sum over the reorder points of all depots, $\sum_{j \in J_i} R_j$, is minimised. However, if R_0 is not taken into consideration this summation will lead to the summation over all lower bounds, i.e. $\sum_{j \in J_i} R_j^{min}$, as this is the minimum sum over the local reorder points and is achieved at R_0^{max}. Therefore, we need to minimise for $TS(R_0)$ to determine R_0^*.

Let us first introduce Algorithm 2 through which we can determine the total stock at R_0.

Algorithm 2 Find $TS(R_0)$ for a given R_0

Input R_0
Output $TS(R_0)$
function GETTOTALSTOCK(R_0)
  $TS(R_0) = R_0$
  for all $j \in J_i$ **do**
    $R_j^* =$ FINDREORDERPOINT(β_j^T, R_0, $R_{j^*}^{min}$ and $R_{j^*}^{max}$) ▷ Algorithm 1
    $TS(R_0) = TS(R_0) + R_j^*$
  end for
  return $TS(R_0)$
end function

Using the previous algorithm, we can determine the optimal central reorder point R_0^* and $TS(R_0^*)$ through Algorithm 3. In addition, the algorithm also optimises the local reorder points through the computation of $TS(R_0^*)$. The original idea of Geelen et al. (2019) is to iterate over all distinct values for R_0 between the computed bounds. At each iteration we compute $TS(R_0)$ and select the R_0 for which $TS(R_0)$ achieves its minima. However, iterating over all possibilities is computationally very heavy. Therefore, we adjust the proposed optimisation approach by Geelen et al. (2019) and introduce a type of bisection method to limit the number of computations. For this to be feasible, it is assumed that the relation between R_0 and $TS(R_0)$ is unimodal, which is made plausible by Geelen et al. (2019) as they indicated that this relation is generally convex. Hence, Algorithm 3 is a bisection-based algorithm similar to Algorithm 1, and uses a step size equal to half of the current search domain.

Algorithm 3 Find optimal R_0 and $TS(R_0)$

$R_0^{LB} =$ FINDREORDERPOINT(β_0^{min}) ▷ Algorithm 1
$R_0^{UB} =$ FINDREORDERPOINT(β_0^{max})
$TS^{LB} =$ GETTOTALSTOCK(R_0^{LB}) ▷ Algorithm 2
$TS^{UB} =$ GETTOTALSTOCK(R_0^{UB})

while $R_0^{UB} - R_0^{LB} > 1$ **do**
  $R_0^{mid} = \left\lceil \frac{R_0^{LB} + R_0^{UB}}{2} \right\rceil$
  $TS^{mid} =$ GETTOTALSTOCK(R_0^{mid}) ▷ Algorithm 2
  if $TS^{UB} - TS^{mid} \geq TS^{LB} - TS^{mid}$ **then**
    $R_0^{UB} = R_0^{mid}$
    $TS^{UB} = TS^{mid}$
  else

$R_0^{LB} = R_0^{mid}$
$TS^{LB} = TS^{mid}$
end if
end while

if $TS^{LB} < TS^{UB}$ **then**
return TS^{LB} and R_0^{LB}
else
return TS^{UB} and R_0^{UB}
end if

Note that the provided solution will be near-optimal if the objective function $TS(R_0)$ is not strictly convex but only unimodal.

4.2 The wait time service measure

Before introducing the proposed Various-Service Level (VSL) models, let us first provide the notation and additional assumptions needed for the wait time service measure. We define $W_i \subseteq J_i$ as the set of depots with wait time service measures for an item i. Note that it is possible for W_i to equal J_i. In that case, all depots use a wait time service measure. The additional notation used for the subsequent sections is provided below.

WT_j^T The wait time target at depot $j \in W_i$, $WT_j^T > 0$. It represents the total time a customer is allowed wait at a depot.

P_j^T The probability target on the wait time at a depot $j \in W_i$. It indicates the probability of a customer having a shorter wait time than WT_j^T, or the percentile of customers or demand requests that must satisfy the wait time service measure.

Δ_j The stochastic delay at the DC experienced by depot $j \in W_i$.

ρ_j The probability a customer encounters wait time at depot $j \in W_i$.

Note that contrary to Section 4.1, here the stochastic delay Δ_j is depot specific and will be estimated per depot. The delay for a depot j denotes the wait time incurred for a replenishment order by depot j at the DC.

In addition to the assumptions made in Section 2 and Section 4, we introduce the following additional assumptions for the wait time service measure of theVSL models.

Assumption 19. The set $W_i \subseteq J_i$ is known and given.

Assumption 20. The central lead time L_0 is constant.

Assumption 19 indicates that the depots using wait time service measures are known. Finally, in Section 4, *Assumption* 13 indicates that the DC lead times are stochastic. However, for the wait time service measure, we require these to be constant. Therefore, *Assumption* 20 is made and overrules *Assumption* 13. The necessity of this assumption will be further elaborated on in Section 4.2.1.

The VSL models use fill rate and wait time service measures. The wait time depots $j \in W_i$, are allowed to have negative reorder points. Having negative reorder points allows depots to service their customers through back orders. Hence, this lowers the total stock within the distribution system whilst maintaining the set service level targets.

However, by allowing negative reorder negative reorder points for wait time depots, we need to consider the following two reorder point cases:

1. $-1 \leq R_j$,
2. $-Q_j \leq R_j < -1$.

To outline the major difference between the reorder point cases we will look at what happens to the customer wait time when the customer demand sizes K_j is less and equal to the inventory position IP_j at a depot $j \in W_i$, $K_j \leq IP_j$, and when $K_j > IP_j$.

Let us first examine when $K_j \leq IP_j$ holds, in other words when the demand size of the arriving customer order is less or equal to the inventory position at a depot j. Per *Assumption* 18, $K_j > 0$ must hold and K_j is integer valued. Moreover, per definition of the inventory position, we know that $IP_j = IL_j + OO_j$, where OO_j are the outstanding replenishment orders at depot j. Combining these gives $1 \leq K_j \leq IL_j + OO_j$. The best case is when $K_j \leq IL_j$, the customer is served from stock on hand and experiences zero wait time. Otherwise, the customer will be served using outstanding orders, which in the worst case are ordered by the arrival of the previous customer. Hence, the customer experiences a wait time of at most $\Delta_j + L_j$. For both the reorder point cases if $K_j \leq IP_j$ holds, the wait time is at most $\Delta_j + L_j$.

Now, we will investigate when $K_j > IP_j$ holds, in other words when the demand size of the arriving customer is greater than the inventory position at a depot j. First, let us rewrite the inequality using *Assumption* 18, the definition of the inventory position and the fact that K_j and IP_j are always integer valued:

$$\begin{aligned} & K_j > IP_j \\ \Leftrightarrow\, & K_j - IP_j > 0 \\ \Leftrightarrow\, & K_j - IP_j \geq 1 \\ \Leftrightarrow\, & IP_j - K_j \leq -1. \end{aligned} \tag{14}$$

Using (14) and $-1 \leq R_j$, the first case of the reorder point cases, we get the inequality $IP_j - K_j \leq R_j$. This implicates that a reorder is always triggered if $K_j > IP_j$ and $-1 \leq R_j$ hold simultaneously. Therefore, the wait time of a customer is still at most $\Delta_j + L_j$ as we always order on or below the reorder point, *Assumption* 11.

However, in the second reorder point case where we have $-Q_j \leq R_j < -1$, the inequality in (14) implicates that not every arriving customers triggers a reorder. Therefore, customers will have to wait upon more arriving customers before a reorder will be triggered to serve them. Hence, the wait time now consists of and can be at most $\Delta_j + L_j$ plus the additional time of more customers arriving. This is problematic since the wait time service measure is only intended for depots with slow-moving items, thus customers may have to wait an extensive amount of time before other customers arrive to trigger a reorder. Therefore, we restrict the reorder points of depots $j \in W_i$ to $-1 \leq R_j$. Section 4.3.3 will provide a simpler setting to allow for an experiment on the second case, to demonstrate how cumbersome allowing for the second reorder point case is for the problem at hand.

In this section we start by explaining the computation of the individual wait time. Thereafter, we will introduce the wait time condition depots need to suffice by in order to satisfy their wait time service measure.

4.2.1 Approximating the wait time per depot

To use wait time as a service measure, we need to approximate the wait time for all depots $j \in W_i$. As shown prior, the customers' wait time is at most $\Delta_j + L_j$, where the lead time L_j is constant as per *Assumption* 14. The delay Δ_j is a random variable and will be approximated in the current section. The approximation is used to approximate the customer wait time of a depot $j \in W_i$.

Section 3 discussed four different methods to approximate the wait time. We opt to apply the approach from Berling and Farvid (2014) to approximate the delay. Their idea revolves around finding the total wait time for a batch of Q_j units placed by a depot $j \in J_i$. This approximation provides the best fit and integration considering the problem at hand, Section 2, in combination with the already existing model, Section 4.1. As previously mentioned, their approach approximates the first two moments of the delay at the depot level and is able to incorporate negative central reorder points. This approximation differs from the wait time approximation in Section 4.1.2, which uses the approach from Axsäter (2003a) and approximates a general central wait time identical across all depots. To determine the delay per depot, Berling and Farvid (2014) assume a constant central lead time L_0, similarly we use *Assumption* 20 as previously introduced.

The additional notation needed for this section is:

$\zeta_{0,j}$	The demand at the DC during the lead time L_0 in excess of the current order of Q_j placed by depot $j \in W_i$.
$f_j(\zeta_{0,j})$	The probability density function of $\zeta_{0,j}$.
$\mu_{\zeta_{0,j},p}$	The expected demand for depot p at the DC over time period L_0, i.e. the individual contribution of depot p to $\mu_{\zeta_{0,j}}$.
$\sigma^2_{\zeta_{0,j},p}$	The variance of the demand for depot p at the DC over time period L_0, i.e. the individual contribution of depot p to $\sigma^2_{\zeta_{0,j}}$.
τ	The time between the arrival of the final unit(s) to complete the current order placed by depot j and the placing of the current order by depot j. Differs for each depot $j \in W_i$.
$\mu_{0,j}$	The expected demand at the DC over time period τ expressed in units per time unit of L_0.

First of all, let us determine the expected value and variance of $\zeta_{0,j}$ according to the formulas in

Berling and Farvid (2014):

$$\mathbb{E}[\zeta_{0,j}] = \sum_{p=1}^{M} \mu_{\zeta_{0,j},p}, \tag{15}$$

$$Var[\zeta_{0,j}] = \sum_{p=1}^{M} \sigma^2_{\zeta_{0,j},p}. \tag{16}$$

Here, if $p \neq j$, we can determine $\mu_{\zeta_{0,j},p}$ as:

$$\mu_{\zeta_{0,j},p} = \mu_p L_0, \tag{17}$$

and if $p = j$, then

$$\mu_{\zeta_{0,j},j} = \sum_{k=0}^{\infty} kQ_j s_j^{ord}(k), \tag{18}$$

holds for $\mu_{\zeta_{0,j},j}$. Moreover, we can estimate $\sigma^2_{\zeta_{0,j},p}$ for all $p \in J_i$ as:

$$\sigma^2_{\zeta_{0,j},p} = \sum_{k=0}^{\infty} (\mu_{\zeta_{0,j},p} - kQ_p)^2 s_p^{ord}(k). \tag{19}$$

Combining these equations, and writing out (15) and (16) completely, gives us:

$$\mathbb{E}[\zeta_{0,j}] = \sum_{p=1,\, p\neq j}^{M} \mu_p L_0 + \sum_{k=0}^{\infty} kQ_j s_j^{ord}(k), \tag{20}$$

$$Var[\zeta_{0,j}] = \sum_{p=1,\, p\neq j}^{M} \left[\sum_{k=0}^{\infty} (\mu_p L_0 - kQ_p)^2 s_p^{ord}(k) \right] + \sum_{k=0}^{\infty} \left[\left(\left(\sum_{x=0}^{\infty} xQ_j s_j^{ord}(x) \right) - kQ_j \right)^2 s_j^{ord}(k) \right]. \tag{21}$$

Note that for the $\mathbb{E}[\zeta_{0,j}]$, only the term for depot j, $\mu_{\zeta_{0,j},j}$ changes compared to (4). Moreover, the formula for the variance $Var[\zeta_{0,j}]$ is also similar to (5), again except from the term of depot j, $\sigma^2_{\zeta_{0,j},j}$.

The term for depot j in (20) and (21) is altered to properly represent the demand at the DC, excluding the current order batch of depot j. For all other depots, the terms are identical to those used in (4) and (5), respectively. Note that these functions are denoted using the notation introduced by Grob and Bley (2018), yet they correctly represent the proposed formulas from Berling and Farvid (2014). Appendix B provides a complete overview regarding the difference in notation.

Berling and Farvid (2014) do indicate that, if Q_j is small or tends to one ($Q_j \in \mathbb{N}$), it is appropriate to use $\delta_j(k)$ for depot j as expressed in (1) in the computation of $\zeta_{0,j}$. However, if Q_j is relatively large an adjusted $\delta_j(k)$ should be used for depot j, since the demand at the DC is more influenced by the current ordered batch Q_j. This is due to the likelihood of depot j ordering additional batches during L_0 is decreasing in k. The adjusted $\delta_j(k)$ for depot j is given as:

$$\delta_j(k, L_0 = l) = Pr(D_j(l) \leq (k+1)Q_j). \tag{22}$$

Note for depot $p \neq j$ in (20) and (21), we always use the unadjusted $\delta_j(k)$ as given in (1).

Similar to the lead time demand $D_0(L_0)$, the mean and variance can be used to fit a standard distribution for $\zeta_{0,j}$. Section 4.4 describes which standard distribution will be used and how it will be estimated. Both will assume the same standard distribution for proper comparison and computation of the fill rate and wait time service measures.

Using the distribution of $\zeta_{0,j}$, we can compute the expected value and variance of the wait time at depot j as done in Berling and Farvid (2014). They implicated three separate cases which depend on the relation between the central reorder point R_0 and the relationship to the order quantity Q_j of a depot j:

Case 1: $\quad Q_j \leq R_0$

$$\mathbb{E}[\Delta_j] = \frac{L_0}{Q_0} \sum_{x=R_0+1}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right) f_j(\zeta_{0,j}) \right), \tag{23}$$

$$\mathbb{E}[\Delta_j^2] = \frac{L_0^2}{Q_0} \sum_{x=R_0+1}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right)^2 f_j(\zeta_{0,j}) \right), \tag{24}$$

Case 2: $\quad 0 \leq R_0 < Q_j$

$$\mathbb{E}[\Delta_j] = \frac{L_0}{Q_0} \left(Q_j - R_0 - 1 + \sum_{x=Q_j}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right) f_j(\zeta_{0,j}) \right) \right), \tag{25}$$

$$\mathbb{E}[\Delta_j^2] = \frac{L_0^2}{Q_0} \left(Q_j - R_0 - 1 + \sum_{x=Q_j}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right)^2 f_j(\zeta_{0,j}) \right) \right), \tag{26}$$

Case 3: $\quad R_0 < 0$

$$\mathbb{E}[\Delta_j] = \frac{L_0}{Q_0} \left(\sum_{y=R_0+1}^{-1} \left(1 + \frac{y-R_0}{\mu_{0,j}}\right) + Q_j + \sum_{x=Q_j}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right) f_j(\zeta_{0,j}) \right) \right), \tag{27}$$

$$\mathbb{E}[\Delta_j^2] = \frac{L_0^2}{Q_0} \left(\sum_{y=R_0+1}^{-1} \left(1 + \frac{y-R_0}{\mu_{0,j}}\right)^2 + Q_j + \sum_{x=Q_j}^{R_0+Q_0} \left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} \left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right)^2 f_j(\zeta_{0,j}) \right) \right). \tag{28}$$

Note that (25) and (26) differ from the original formulation provided by Berling and Farvid (2014). We improved these equations and they now contain the correct probability for the central inventory position IP_0, following a $\text{Unif}(R_0+1, R_0+Q_0)$ distribution. Appendix C discusses the slight error in the original formula and derives the correct probability as utilised above.

The three different cases compute all compute the expected value and second moment of the delay depending on the relation between the central reorder point and local order quantity. In each case, the current order of Q_j units plays a distinct part in when the order can be shipped and when a replenishment order is triggered by the DC. Hence, the different cases take into consideration the placing of the current order, the size of the orders, the size of the additional demand at the DC and the lead time of the DC to compute the expected value and second moment of the delay.

Only in (27) and (28) $\mu_{0,j}$ is used, because τ is only positively defined for $R_0 < 0$. Berling and Farvid (2014) state the following equation to calculate $\mu_{0,j}$:

$$\mu_{0,j} = \left(\frac{\mu_{0,j}^j}{\tau} + \sum_{p=1,\, p\neq j}^{M} \mu_p \right) \cdot L_0. \tag{29}$$

Here, $\mu_{0,j}^j$ is the expected demand of depot j over the time period τ. Hence, the contribution of depot j to the total expected demand $\mu_{0,j}$ is $\frac{\mu_{0,j}^j}{\tau} \cdot L_0$. Berling and Farvid (2014) indicate that finding $\mu_{0,j}^j$ and τ is not straightforward, thus they provided an approximation. τ can be approximated as $\frac{-R_0/2}{\sum_{j\in J_i} \mu_j}$. They based this approximation on the average inventory position when a shortage occurs at the DC that is small enough to not trigger an additional replenishment order at the DC. They then divide this by the mean demand rate at the DC. Using this approximation for τ, we can compute $\mu_{0,j}^{j}$. Note that $\mu_{0,j}^{j} = \sum_{k=0}^{\infty} kQ_j s_j^{ord}(k)$, it therefore coincides with the latter part of (20). Therefore, we can compute it in similar fashion, but we use τ as the time period instead of L_0 in (2).

With the equations (23) - (28), we can determine the first and second moment of the delay at the DC for any combination of R_0 and Q_j. The variance of the delay experienced by a depot j is determined through:

$$Var[\Delta_j] = \mathbb{E}[\Delta_j^2] - \mathbb{E}[\Delta_j]^2. \tag{30}$$

The delay distribution $F[\Delta_j]$ can be fitted to a standard distribution using the expected value and variance. Note that $F[\Delta_j]$ corresponds with the CDF of the delay at depot j. Thus, $F_{\Delta_j}(x)$ corresponds with the probability $Pr[\Delta_j \leq x]$, where $Pr[\Delta_j \leq x] = 0$ if $x < 0$ holds as the delay non-negatively defined. We can assume a standard distribution depending on how conservative we want the delays to be. Heavy-tailed distributions, specifically heavy right-tailed distributions, are more conservative as they place more weight on the possibility of longer delays and ensure that extreme values are captured when determining the wait time at higher probability levels. Section 4.4 specifies which distributions will be used and elaborates on how these distributions can be fitted to the expected value and variance of the delay.

4.2.2 Computing the probability of waiting and defining the wait time condition

The goal of the wait time service measure is to serve P_j^T percent of the customers within the set wait time target WT_j^T at depot $j \in W_i$. Therefore, we define a condition that must hold for a depot to suffice its target. It is straightforward that R_0 affects the wait time. However, R_j affects the probability of waiting. Therefore local and central reorder point both need to be taken into account since they affect different parts of the condition.

Before introducing the wait time condition, we need to compute ρ_j, the probability a customer encounters wait time at depot $j \in W_i$. A customer experiences wait time when it arrives at a depot j with a greater demand size K_j than the current inventory level IL_j. Therefore, we first need the probability of the customer demand being greater than the current inventory level of depot j.

To compute the cumulative probability of the inventory level, we need to integrate over (10) until the selected value. However, since we only allow for discrete quantities, this is similar to summing over all possible discrete values:

$$\begin{aligned} Pr[IL_j \leq y] &= \frac{1}{Q_j} \int_{-\infty}^{y} \sum_{x=R_j+1}^{R_j+Q_j} Pr[D_j(L_j^{eff}) = x - p] \ dp \\ &= \frac{1}{Q_j} \sum_{p=-\infty}^{y} \sum_{x=R_j+1}^{R_j+Q_j} Pr[D_j(L_j^{eff}) = x - p]. \end{aligned} \tag{31}$$

Note that similarly to (10), the demand over the effective lead time is used here as described in (8) and (9). However, in Section 4.1.2, the effective lead time is defined using the METRIC-type approximation of the general wait time, (6) and (7). Now in (31), we need to redefine the effective lead time and use the delay experienced by depot $j \in W_i$, as approximated in (23) - (28) and (30). The effective lead time is now defined as:

$$\mathbb{E}[L_j^{eff}] = \mathbb{E}[L_j] + \mathbb{E}[\Delta_j], \tag{32}$$

$$Var[L_j^{eff}] = Var[L_j] + Var[\Delta_j]. \tag{33}$$

With the inventory level distribution, we can determine the probability of a customer demand size being greater than the current inventory level. Similar to (11), we denote the probability of a customer with demand size k as $pdf_{K_j}(k)$. It is assumed that these probabilities are known. Furthermore, we introduce the set $\kappa_j = \{k \in \mathbb{N}^+ \,|\, pdf_{K_j}(k) > 0\}$, denoting the set of all customer demand sizes at depot $j \in W_i$. Using these two probabilities, we can determine ρ_j as:

$$\rho_j = \sum_{k \in \kappa_j} pdf_{K_j}(k)\, Pr[IL_j \leq k-1]. \tag{34}$$

Note that we need the cumulative probability of the inventory level to be less than or equal to the customer demand size minus one $(k-1)$ as we need to compute the probability of a customer with demand size k experiencing wait time.

Having determined ρ_j, we introduce a sufficient wait time condition for a depot j in order to suffice its wait time service measure in (35). The left-hand side of the condition adds distinct probabilities of a customer being served within the wait time target. The first term denotes the probability of a customer having to wait, ρ_j, times the probability that the delay of the depot j is less than the wait time target minus the inter-lead time. In the second term, $1-\rho_j$ indicates the probability of a customer being served through stock on hand, i.e. experiences zero wait time. A customer with zero wait time always suffices the wait time target. Hence, this probability $(1-\rho_j)$ is actually multiplied by one. The addition of the two terms need to be greater than or equal to the probability target P_j^T.

$$\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1-\rho_j) \geq P_j^T. \tag{35}$$

In general this condition is too strict as it assumes that every customer that incurs wait time triggers a replenishment order at a depot j. In case customers experience wait time, but do not trigger a replenishment order as there is already a large enough outstanding order, the probability of sufficing the wait time target is greater than the first term in (35). To further elaborate on the approximation and the inaccuracy that is paired with it, we need to consider the following cases: $WT_j^T < L_j$ and $WT_j^T \geq L_j$.

If $WT_j^T < L_j$ holds, then $F_{\Delta_j}(WT_j^T - L_j) = 0$, see Section 4.2.1, therefore (35) reduces to the simple condition of $(1-\rho_j) \geq P_j^T$. This indicates that the probability of a customer being served through stock on hand must be greater than or equal to the probability target. Since the condition assumes that a waiting customer triggers a replenishment order, it is unfeasible for this customer to be served within WT_j^T if $WT_j^T < L_j$ holds. Although the condition disregards the exact value of WT_j^T, it can matter for the fulfilment of the wait time target.

By disregarding WT_j^T if $WT_j^T < L_j$ holds, the condition is too strict in general. Hence, it is only a sufficient condition. Yet, given the problem of the current paper, see Section 2, the use of wait time is only applied to slow-moving items. For small order sizes Q_j and large inter-arrival times the likelihood of customers that incur wait time but do not trigger a replenishment order is small, benefiting the accuracy of this condition. Furthermore, as further elaborated on in Section 4.3.2, this case only occurs in the AVSL model where the condition is applied to optimise the local reorder point R_j. Due to the integrality of R_j, see *Assumption* 8, large increases of ρ_j are expected when the reorder level is increased. Hence, the approximation is expected to have an insignificant effect during optimisation.

For the second case, if $WT_j^T \geq L_j$ holds, the condition only considers two types of customers. A customer that incurs no wait time as it is served through stock on hand, or a customer that incurs wait time and triggers a replenishment order. A customer is served through stock on hand with a probability of $(1 - \rho_j)$ and satisfies the wait time target WT_j^T with a probability of 1. Furthermore, a customer incurs wait time with a probability of ρ_j and satisfies the wait time target WT_j^T with a probability of $F_{\Delta_j}(WT_j^T - L_j)$. However, a customer can incur wait time but not trigger a replenishment order when Q_j is large enough. The condition assigns all wait time incurring customers the identical probability, namely that of a wait time incurring customer triggering a replenishment order. Hence, all customers that incur wait time have a probability of at least $F_{\Delta_j}(WT_j^T - L_j)$ to be served within the wait time target WT_j^T. The condition therefore overestimates their wait time and is therefore only a sufficient condition.

The overestimation is expected to have no significant effect on the optimisation of the reorder points in Section 4.3, as we expect the amount of customers that incur wait time and do not trigger a replenishment order, to be very limited. Similarly to the first case, the likelihood of these customers decreases as Q_j gets smaller and the inter-arrival times get larger, thus reducing the effect of this approximation on the condition.

As stated previously, both the local and central reorder point affect the condition. The local reorder point affects ρ_j. A decrease in R_j leads more customers are served through back orders and therefore encounter wait time. An increase in R_j gives the opposite effect. The central reorder point affects the delay Δ_j. A decrease in R_0 leads to an increase in delay experienced by the depot j, increasing R_0 lowers the delay. Note that R_0 does also affects ρ_j through Δ_j and the effective lead time. A decrease in R_0, reduces ρ_j and vice versa.

4.3 Various-Service Level models

The current paper proposes two different models. In both models, the first two steps remain identical to the *vNext* model, as described in Section 4.1. After these steps, we distinguish between the Simple and Advanced Various-Service Level models (the SVSL and AVSL models).

Given the restriction on the reorder points in 4.2, the main difference between the SVSL and the AVSL model is the allowed values for the reorder points of the wait time depots. In the SVSL model, we only allow the reorder point of depot $j \in W_i$, to be equal to -1. This ensures that a depot reorders upon customer arrival and has minimal stock on hand, serving customer demand mainly through back orders. The central reorder point is expected to increase to guarantee the wait time service measures. By choosing a set reorder point at these depots, the wait time targets must be greater than the inter-lead time between the DC and the depot to always ensure a feasible solution.

For the AVSL model, we allow the reorder point of the depot $j \in W_i$, to be greater than -1 and thus also to be non-negative. By allowing an increase in local stock, targets on the wait time are enabled to be less than or equal to the inter-lead time between the DC and the depot. Thus, providing more options for the wait time service measure and resulting in a more flexible model.

In this section we first introduce the SVSL model. Afterwards, we describe the optimisation procedure of the AVSL model. Finally, we will sketch a simpler scenario in which we allow for reorder points below -1.

4.3.1 The Simple Various-Service Level model

We will now outline the SVSL model. For this model, we assume that the reorder point $R_j = -1$ for all $j \in W_i$, such that customers are mainly served through back orders. Compared to the *vNext* model in Section 4.1, the SVSL model consists of the same steps, with the only addition being the third step below:

I. Compute the distribution of the central lead time demand $D_0(L_0)$.

II. Compute bounds on all the reorder points R_j for $j \in J_i^+$.

III. Compute a new lower bound on the central reorder point R_0, such that all depots $j \in W_i$ satisfy their wait time targets.

IV. Compute the optimal central reorder point R_0^* minimising the total stock $TS(R_0)$.

Since we want to ensure that $R_j = -1$ is valid for all depots $j \in W_i$, we need to make an assumption for the imposed wait time target:

Assumption 21. The wait time target WT_j^T must be greater than the inter-lead time L_j between the DC and the depot $j \in W_i$.

Section 4.3 provided a brief explanation of the model and elaborated on why this assumption must be made. Moreover, the wait time condition in (35) indicates the necessity of this assumption. Since depot $j \in W_i$ has a constant reorder point of -1, a wait time target smaller than or equal to the lead time always results in $F_{\Delta_j}(WT_j^T - L_j) = 0$. Therefore, it could prove $R_j = -1$ to be invalid as the depot j is not able to suffice the condition due to insufficient customers being served by stock on hand.

The first two steps remain identical to Sections 4.1.1 and 4.1.2, respectively. The third step of the SVSL model determines the new lower bound R_0^{min*}, this is the lowest central reorder point that satisfies the wait time targets of all depots $j \in W_i$. The existing bounds R_0^{min} and R_0^{max} are determined in step **II.** through satisfying a predetermined fill rate target at the DC, in Section 4.1.2. However, the lower bound here does not necessarily provide a feasible solution to all the depots $j \in W_i$ given the wait time condition in (35).

We first introduce Algorithm 4. This algorithm determines the lowest central reorder point $R_0^{min,j}$ such that the wait time target of depot $j \in W_i$ is satisfied at the P_j^T percentile of the distribution F_{Δ_j}. Since the delay is decreasing in R_0, we use a bisection-based approach in this algorithm to limit the necessary computations.

Algorithm 4 Find the minimum R_0 that satisfies the target at depot $j \in W_i$

Input WT_j^T, P_j^T, R_0^{min} and R_0^{max}

Output $R_0^{min,j}$

function FINDREORDERPOINTSIMPLE(WT_j^T, P_j^T, R_0^{min} and R_0^{max})

$R_0^{low} = R_0^{min}$

$R_0^{top} = R_0^{max}$

$R_0^{mid} = \left\lceil \frac{R_0^{low} + R_0^{top}}{2} \right\rceil$

Determine $\mathbb{E}[\Delta_j]$, $Var[\Delta_j]$ and $F[\Delta_j]$ for R_0^{low} ▷ Equations (23) - (28), (30)

Determine ρ_j for R_0^{low} ▷ Equation (34)

if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)

$R_0^{min,j} = R_0^{low}$

return $R_0^{min,j}$

end if

while $R_0^{mid} \neq R_0^{low}$ and $R_0^{mid} \neq R_0^{top}$ **do**

Determine $\mathbb{E}[\Delta_j]$, $Var[\Delta_j]$ and $F[\Delta_j]$ for R_0^{mid} ▷ Equations (23) - (28), (30)

Determine ρ_j for R_0^{mid} ▷ Equation (34)

if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)

$R_0^{top} = R_0^{mid}$

else

$R_0^{low} = R_0^{mid}$

end if

$R_0^{mid} = \left\lceil \frac{R_0^{low} + R_0^{top}}{2} \right\rceil$

end while

$R_0^{min,j} = R_0^{top}$

return $R_0^{min,j}$

end function

Note that if R_0^{low} is already sufficient, we can return this value without further computations.

We can now determine the new R_0^{min*}, which is done through Algorithm 5. This algorithm iterates through all depots $j \in W_i$ and identifies the depot j^* with the largest gap between between P_j^T and $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j)$. Whereafter, we use Algorithm 4 to determine R_0^{min,j^*}. We then iterate through all remaining depots to check if this reorder point also satisfies the wait time of the remaining depots $j \in W_i \setminus \{j^*\}$. If it does we determine that R_0^{min,j^*} is the new lower bound R_0^{min*}. Otherwise, we continue and determine a new depot j^* and repeat this process until we all wait time targets are satisfied or until the upper bound R_0^{max} is reached.

Algorithm 5 Find the R_0^{min*} such that all depots $j \in W_i$ satisfy their targets

Input R_0^{min}, R_0^{max}, and WT_j^T, P_j^T $\forall j \in W_i$

Output R_0^{min*}

function FINDMINIMUMREORDERPOINTWT(R_0^{min}, R_0^{max}, and WT_j^T, P_j^T $\forall j \in W_i$)

Set $j^* = \text{NA}$

Set $\gamma = -\infty$

for all $j \in W_i$ **do**

Determine $\mathbb{E}[\Delta_j]$, $Var[\Delta_j]$ and $F[\Delta_j]$ for R_0^{min} ▷ Equations (23) - (28), (30)

Determine ρ_j for R_0^{min} ▷ Equation (34)

if $P_j^T - (\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j)) > \gamma$ **then**

$j^* = j$

$\gamma = P_j^T - (\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j))$

end if

end for

while $\gamma > 0$ and $R_0^{min} < R_0^{max}$ **do**

$R_0^{min} =$ FINDREORDERPOINTSIMPLE($WT_{j^*}^T, P_{j^*}^T, R_0^{min}$ and R_0^{max}) ▷ Algorithm 4

Set $\gamma = -\infty$

for all $j \in W_i \setminus \{j^*\}$ **do**

Determine $\mathbb{E}[\Delta_j]$, $Var[\Delta_j]$ and $F[\Delta_j]$ for R_0^{min} ▷ Equations (23) - (28), (30)

Determine ρ_j for R_0^{min} ▷ Equation (34)

if $P_j^T - (\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j)) > \gamma$ **then**

$j^* = j$

$\gamma = P_j^T - (\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j))$

end if

end for

end while

$R_0^{min*} = \min(R_0^{min}, R_0^{max})$

return R_0^{min*}

end function

The algorithm above returns the new lower bound on the central reorder point R_0^{min*} that satisfies the wait time targets of all depots $j \in W_i$. Note that it is possible that the returned reorder point R_0^{min*} does not suffice all wait time targets due to R_0^{max} being too low. If this is the case, one should return to step **II.** and increase β_0^{max} in Section 4.1.2, such that R_0^{max} is always big enough to satisfy all the wait time targets. With R_0^{min*}, we can continue with step **IV.**, this step is identical to the step explained in Section 4.1.3.

4.3.2 The Advanced Various-Service Level model

The idea for the AVSL model is to allow a more flexible selection of the wait time targets WT_j^T for all depots $j \in W_i$. We want to also allow for targets to be below or equal to the inter-lead time between the DC and the depot. Therefore, we drop *Assumption* 21 of the SVSL model. To enable these targets, we have to allow for non-negative reorder points. Moreover, we need to be able to determine the minimum local reorder point R_j that suffices the wait time target for a given R_0 for all depots $j \in W_i$.

The steps in the AVSL model are identical to those of the *vNext* model (Section 4.1) with only an adjustment to Algorithm 2. Since we now need to optimise for depots with different service levels. Similar to Algorithm 1, we now need to find the optimal reorder points for the wait time depots given R_0. Let us introduce Algorithm 6 that finds these reorder points for all $j \in W_i$. Given a central reorder point, it determines the wait time distribution and finds the minimum reorder point R_j^* that satisfies the wait time service measure at depot j. The algorithm uses R_j^{min} and R_j^{max} as starting values. Here, R_j^{min} is always equal to -1, as explained in Section 4.3. Let R_j^{max} be $V \cdot Q_j$, V must be large enough such that $V \cdot Q_j$ always suffices the target in (35). These initial values are, respectively, the lower and upper bound in this bisection-based method. These bounds need to be used instead of the computed bounds from Section 4.1.2, as those are based on the fill rate. Note that this bisection-based approach is similar to Algorithm 4, we again use it since ρ_j is decreasing in R_j and it decreases the necessary computations.

Algorithm 6 Find best R_j for a warehouse $j \in W_i$

Input R_0, WT_j^T, P_j^T, R_j^{min} and R_j^{max}

Output R_j^*

function FINDREORDERPOINTWTADVANCED(R_0, WT_j^T, P_j^T, R_j^{min} and R_j^{max})
    Determine $\mathbb{E}[\Delta_j]$, $Var[\Delta_j]$ and $F[\Delta_j]$ ▷ Equations (23) - (28), (30)
    $R_j^{low} = R_j^{min}$
    $R_j^{top} = R_j^{max}$
    $R_j^{mid} = \left\lceil \frac{R_j^{low} + R_j^{top}}{2} \right\rceil$

    Determine ρ_j for R_j^{low} ▷ Equation (34)
    if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)
        return R_j^{low}
    end if

    while $R_j^{mid} \neq R_j^{low}$ and $R_j^{mid} \neq R_j^{top}$ **do**
        Determine ρ_j for R_j^{mid} ▷ Equation (34)
        if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)
            $R_j^{top} = R_j^{mid}$
        else
            $R_j^{low} = R_j^{mid}$
        end if

$R_j^{mid} = \left\lceil \frac{R_j^{low} + R_j^{top}}{2} \right\rceil$
end while

$R_j^* = R_j^{top}$
return R_j^*
end function

Note that if the lower bound already suffices, we can return it and no further computations are necessary. Moreover, much like in Section 4.1.2, we can tighten the upper bound in similar fashion, namely:

R_j^{max*}: Achieved when the DC restocks very rarely, i.e. $R_0 = -Q_0$, implicating that the delay experienced by depot j is maximal. Depot j holds maximal amount of stock on hand to to comply by its wait time target.

Hence, the upper bound in Algorithm 6 can be improved upon. Following the description above, we let $R_0 = -Q_0$ be the starting value of the algorithm. The $\mathbb{E}[\Delta_j]$ and $Var[\Delta_j]$ will attain their maximum value, increasing the local reorder point such that (35) is still satisfied. This results in the outcome of the algorithm being R_j^{max*}, the tightened upper bound.

Let us now redefine Algorithm 2. Algorithm 7 computes the total stock $TS(R_0)$ for a given central reorder point R_0. It determines the minimum reorder point R_j^* for a depot j based on its assigned service measure through Algorithms 1 or 6.

Algorithm 7 Find $TS(R_0)$ for a given R_0

Input R_0
Output $TS(R_0)$

function GETTOTALSTOCK(R_0)
$TS(R_0) = R_0$
for all $j \in J_i$ **do**
if $j \in W_i$ **then**
$R_j^* =$ FINDREORDERPOINTWTADVANCED(R_0, WT_j^T, P_j^T, R_j^{min} and R_j^{max}) ▷ Algorithm 6
else
$R_j^* =$ FINDREORDERPOINT(β_j^T, R_0, R_j^{min} and R_j^{max}) ▷ Algorithm 1
end if

if $R_j^* < R_j^{min}$ **then**
$R_j^* = R_j^{min}$
end if
$TS(R_0) = TS(R_0) + R_j^*$
end for
return $TS(R_0)$
end function

This algorithm is used in Algorithm 3 to find all optimal reorder points. The final algorithm of the *vNext* model, therefore, remains unchanged and replaces Algorithm 2 with the algorithm above.

4.3.3 Allowing local reorder points below -1

As stated previously, the VSL models only allow for reorder points that are greater than or equal to minus one. Section 4.2 elaborated on why this restriction is chosen. However, in this section we will discuss an experiment in which we include the following boundary $-Q_j \leq R_j < -1$, the second reorder point case, within the solution space of depots $j \in W_i$. To do so we will make an important additional assumption:

Assumption 22. All customers arrive with a demand size of 1. In other words, the customer demand size K_j is constant and equal to 1 in the compound Poisson process for all $j \in W_i$.

This assumption simplifies the compound Poisson process to a standard Poisson process for the customer arrivals in the distribution system. However, it does allows us to establish why including the second reorder point case does not yield any benefit.

For the first reorder point case, $-1 \leq R_j$ for $j \in W_i$, we will use the computations and formulas from the AVSL model. Given *Assumption* 22 and if $-1 \leq R_j$ holds, ρ_j in (34) can be rewritten as:

$$\rho_j = Pr[IL_j \leq 0] \tag{36}$$

which is possible since $\kappa_j = 1$ and $pdf_{K_j}(1) = 1$. Therefore, this remains of the summation in (34). The left hand side multiplies the probability of a customer having to wait, if $IL_j \leq 0$, by the largest wait time it can experience at most. If the above inequality holds, the wait time target for a depot $j \in W_i$ is satisfied if $-1 \leq R_j$.

For the second reorder point case we need to introduce some additional formulas, but extend on the AVSL model. For the remainder of this paper we will refer to this extension of the AVSL model as the Experiment Various-Service Level model, the EVSL model. Let us first introduce the additional notation needed for this section:

C_j^k The stochastic customer inter-arrival time at depot $j \in W_i$, given that the customer must wait on k other customers.

ω_j The average additional wait time a customer encounters at depot $j \in W_i$. The average time over all customers until enough customers to arrive to prompt a replenishment order.

Now, if $-Q_j \leq R_j < -1$ holds, a customer encounters wait time if $IL_j \leq 0$. However, a customer experiences additional wait time ω_j if $IP_j \leq 0$ since it has to wait upon other customers to arrive in order for a replenishment order to be triggered. This poses the following two cases:

1. $IP_j > 0$, a customer experiences a wait time of at most $\Delta_j + L_j$.
2. $IP_j \leq 0$, a customer experiences a wait time of at most $\Delta_j + L_j + \omega_j$.

Therefore, we alter the condition in (35), such that now the following must hold for a depot $j \in W_i$ to suffice its wait time target when $-Q_j \leq R_j < -1$ holds:

$$\rho_j \cdot \left[Pr[IP_j > 0] \cdot F_{\Delta_j}(WT_j^T - L_j) + Pr[IP_j \leq 0] \cdot F_{\Delta_j}(WT_j^T - L_j - \omega_j)\right] + (1 - \rho_j) \geq P_j^T. \tag{37}$$

Note that we use the ρ_j from (36). Moreover, compared to (35), the term multiplied by ρ_j is split into two parts as per the two cases described above.

To approximate the additional wait time ω_j, let us first define $F[C_j^k]$ as the CDF of the inter-arrival times of customers at depot j, with known mean and variance. Thus, $F_{C_j^k}[x]$ corresponds with the probability $Pr[C_j^k \leq x]$. Since all customers arrive according to a Poisson process (*Assumptions* 17 & 22), the inter-arrival times between customers are exponentially distributed. Therefore, C_j^k follows an Erlang k-distribution. The rate of the inter-arrival times, λ, is assumed to be known.

Next we know that $IL_j \leq 0$ and $IP_j \leq 0$ both hold, as otherwise a customer would not experience additional wait time. Therefore, the probability $Pr[IP_j = x \,|\, IP_j \leq 0]$ is the probability that the inventory position is in a non-positive state. Using Bayes' Theorem, it can be rewritten as:

$$\begin{aligned} Pr[IP_j = x \,|\, IP_j \leq 0] &= \frac{Pr[IP_j \leq 0 \,|\, IP_j = x]\, Pr[IP_j = x]}{Pr[IP_j \leq 0]} \\ &= \frac{1 \cdot \frac{1}{Q_j}}{\frac{-R_j}{Q_j}} \\ Pr[IP_j = x \,|\, IP_j \leq 0] &= \frac{1}{-R_j}, \end{aligned} \tag{38}$$

making use of the discrete uniform distribution of the inventory position. Moreover, $Pr[IP_j \leq 0 \,|\, IP_j = x] = 1$ for $x \in [R_j + 1, 0]$. Since, it is given that x is non-positive this always holds and the outcome of this probability can be used in the equation above.

Utilising the distribution of the inter-arrival times and (38), we can now approximate ω_j as:

$$\begin{aligned} \omega_j &= Pr[IP_j = x \,|\, IP_j \leq 0] \sum_{x=R_j+1}^{0} \mathbb{E}[C_j^x] \\ &= \frac{1}{-R_j} \cdot \sum_{x=1}^{|R_j+1|} \mathbb{E}[C_j^x]. \end{aligned} \tag{39}$$

The sum in (39) computes on how many additional customers arrivals a customer has to wait per inventory position. If a customer arrives when $IP = 0$, then it has to wait for $|R_j + 1|$ additional customers. On the other hand, arriving at $IP = |R_j + 1|$ triggers a replenishment order immediately, thus the customer does not have to wait for any additional customers. Moreover, we use the mean of the k^{th}-customer inter-arrival time as this suffices to approximate the average additional wait time of a customer for this experiment.

Finally, we need to replace Algorithm 6 used in Algorithm 7 to allow for the second reorder point case. Therefore, we introduce Algorithm 8, it incorporates both the reorder point cases with its corresponding equations and replaces Algorithm 6

Algorithm 8 Find best R_j for a warehouse $j \in W_i$

Input R_0, WT_j^T, P_j^T and V

Output R_j^*

```
1: function FINDREORDERPOINTWTEXPERIMENT(R_0, WT_j^T, P_j^T and V)
2:     Determine E[Δ_j], Var[Δ_j] and F[Δ_j]          ▷ Equations (23) - (28), (30)
3:     R_j^low = −Q_j
```

$R_j^{top} = V \cdot Q_j$

$R_j^{mid} = \left\lceil \frac{R_j^{low} + R_j^{top}}{2} \right\rceil$

Determine ρ_j for R_j^{low} ▷ Equation (36)

if $-1 \leq R_j^{low}$ **then**

  if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)

    $R_j^* = R_j^{low}$

  end if

else

  Determine ω_j for R_j^{low} ▷ Equation (39)

  if $\rho_j \cdot \left[Pr[IP_j > 0] \cdot F_{\Delta_j}(WT_j^T - L_j) + Pr[IP_j \leq 0] \cdot F_{\Delta_j}(WT_j^T - L_j - \omega_j)\right] + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (37)

    $R_j^* = R_j^{low}$

  end if

end if

if $R_j^* = R_j^{low}$ **then**

  return R_j^*

end if

while $R_j^{mid} \neq R_j^{low}$ and $R_j^{mid} \neq R_j^{top}$ **do**

  Boolean bool = *False*

  Determine ρ_j for R_j^{mid} ▷ Equation (36)

  if $-1 \leq R_j^{mid}$ **then**

    if $\rho_j \cdot F_{\Delta_j}(WT_j^T - L_j) + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (35)

      bool = *True*

    end if

  else

    Determine ω_j for R_j^{low} ▷ Equation (39)

    if $\rho_j \cdot \left[Pr[IP_j > 0] \cdot F_{\Delta_j}(WT_j^T - L_j) + Pr[IP_j \leq 0] \cdot F_{\Delta_j}(WT_j^T - L_j - \omega_j)\right] + (1 - \rho_j) \geq P_j^T$ **then** ▷ Equation (37)

      bool = *True*

    end if

  end if

  if bool **then**

    $R_j^{top} = R_j^{mid}$

  else

$R_j^{low} = R_j^{mid}$

end if

$R_j^{mid} = \left\lceil \frac{R_j^{low} + R_j^{top}}{2} \right\rceil$

end while

$R_j^* = R_j^{top}$

return R_j^*

end function

Note that the starting values for the algorithm are $-Q_j$ and $V \cdot Q_j$. Again, V must be great enough such that $V \cdot Q_j$ always satisfies the target in (37). In the algorithm above we separate the different reorder point cases. Moreover, the structure is adjusted slightly compared to Algorithm 6 for readability, since we have to consider both cases every time we want to evaluate the wait time for a specific R_j.

4.4 Implementation of the models

This section discusses the implementation of the models. It first elaborates on the choice of distributions and to fit these to the computed moments. Furthermore, we elaborate on the discretisation of the distributions. Next, the associated choices for the implementation are disclosed. Finally, the stochasticity of the central lead time is discussed.

4.4.1 Choice and fitting of distributions

As mentioned in the preceding sections, distributions are needed for the probabilities of the order sizes, demand over time, delay and inter-arrival times. Let us now explain which distributions we use and how these are fitted to the known or computed first two moments.

First of all, we will make use of the empirical distribution for the demand sizes, thus making use of historical data. The probabilities of the demand sizes K_j for $j \in J_i$ are assumed to be known and given. Therefore, it follows an empirical distribution in (11) and (34). To determine the probability density function (PDF) and cumulative density function (CDF), we use the definition given by Dekking et al. (2005). It is assumed that the historical data points $(X_1, \dots, X_n)$ are all independently and identically distributed random variables. We can then identify their PDF ($\hat{f}_n(x)$) and CDF ($\hat{F}_n(x)$) as:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i = x} \tag{40}$$

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x}. \tag{41}$$

Here, $\mathbb{1}$ represents the indicator function and equals either one or zero depending on the state of the defined event.

Furthermore, a $Gamma(k, \theta)$ distribution is assumed for the total demand over time, $D_j(t)$ (where t is either L_0 or L_j^{eff}) for all $j \in J_i^+$, used in (1), (2), (10), (11) and (31). Hence, also the central lead time demand, $D_0(L_0)$ follows a $Gamma(k, \theta)$ distribution in (13). In addition, the delay, Δ_j, and the demand in excess at the DC, $\zeta_{0,j}$, also both assume a $Gamma(k, \theta)$ distribution in (35) and (37), and

(23) - (28), respectively. For the *Gamma* distribution, $k > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. These parameters are estimated exactly using the mean ($\mathbb{E}[X]$) and variance ($Var[X]$):

$$k = \frac{\mathbb{E}^2[X]}{Var[X]} \tag{42}$$

$$\theta = \frac{Var[X]}{\mathbb{E}[X]}. \tag{43}$$

Note that C_j^k in Section 4.3.3 is $Erlang(k, \lambda)$ distributed, which is a special case of the $Gamma(k, \theta)$ distribution, in (39). However, for C_j^k it holds that k is the number of customers upon which the current customer has to wait and λ is the rate at which customers arrive at the depot j. Since all customers arrive with a demand size of one in Section 4.3.3, the rate can be estimated by dividing the total time period by the total demand experienced at depot j.

Finally, we consider the $LogNormal(\mu, \sigma^2)$ distribution for the delay, Δ_j, to compare the performance of the two distributions. Further elaborated on in Section 4.5. For the *LogNormal* distribution, μ and σ are the mean and standard deviation, respectively. We can determine these parameters as:

$$\mu = \ln\left(\frac{\mathbb{E}[X]}{\sqrt{1 + \frac{Var[X]}{\mathbb{E}^2[X]}}}\right) \tag{44}$$

$$\sigma^2 = \ln\left(1 + \frac{Var[X]}{\mathbb{E}^2[X]}\right). \tag{45}$$

Note that when $Var[X] = 0$ in (42) and (44), there is a division by zero. As this leads to an improper and unattainable implementation, we set $Var[X] = 10^{-10}$ if this occurs. This leads to a significantly small variance, thus not affecting the distribution but allowing us to use the distribution by preventing division by zero.

4.4.2 Discretisation of the distributions

As stated in *Assumption* 18, demand sizes are integer valued. However, as mentioned previously, we assume a continuous *Gamma* distribution for all demand related variables, $D_j(t)$ and $\zeta_{0,j}$. The domain is non-negative per definition for both. However, acquiring the exact probability of an integer valued demand size through the PDF is not consistent. Therefore, we need to discretise the *Gamma* distribution.

We use the CDF of the *Gamma* distribution can be used to discretise the distribution. For all equations in which we require the probability of demand to exactly equal a certain value, e.g. as in (10) and (23), the CDF will be used in implementation instead of the PDF. However, the CDF computes the probability $Pr[X \leq x]$ instead of the probability $Pr[X = x]$, as done by the PDF. Therefore, to compute the probability around a given integer x, we use the following substitution:

$$Pr[X = x] = Pr[X \leq x + (1 - \alpha)] - Pr[X \leq x - \alpha]. \tag{46}$$

Here α, with $0 \leq \alpha \leq 1$, determines how conservative the substitution will be. If $\alpha = 0$, the least conservative option, the entire probability value between x and $x + 1$ will be counted towards the probability of $Pr[X = x]$. On the other hand, if $\alpha = 1$, the most conservative option, the entire probability value between $x - 1$ and x will be counted towards the probability of $Pr[X = x]$.

In our approach, we opt for $\alpha = 0.6$ to be slightly more conservative. However, using this α, we still allow the probability of a random variable being equal to zero to exist. This is necessary to attain minimal reorder points at the depots, since there is a positive probability of the demand over a certain time period being equal to zero.

Furthermore, implementing the exact probabilities using the CDF, as described above, also solves for the cases in which the PDF observes a significant peak between two integer values. In such cases the PDF would assign probabilities that are far too low to yield any adequate results in the necessary equations. Therefore, this implementation discretises and assigns proper probabilities to the integer valued random variables. Appendix D provides an in-depth example of this implementation and elaborates more on why it would be problematic if peaks occur in between integer values in the PDF.

Finally, as mentioned in Section 3 and noted by Grob and Bley (2018), the wait time approximation of Berling and Farvid (2014) can provide a negative expected value if $Q_j > R_0 + Q_0$ holds for $j \in J_i$, which is the second case. As explained in Section 4.2.1, we improved the formulation of Berling and Farvid (2014) in (25). Therefore, this phenomenon should not occur anymore.

4.4.3 Implementation choices

In order to successfully implement the equations and models, we need to solve some issues that may arise from them. The main reason for these choices are to decrease the computation time.

Let us start on how we cope with the infinite sums in Sections 4.1 & 4.3. These sums are computationally heavy and contain parts that are obsolete, as they do not add alter the result of the sum significantly beyond a certain point. Consequently, we want to decrease the computations needed whilst upholding the integrity of the equations.

We encounter two different types of infinite sums. The first type considers sums with a limit of one at infinity, i.e. cumulative probabilities. In (5), (12) and (20) - (28), such a term is contained in the form of a probability, more specifically, the terms $s_j^{ord}(k)$ and $f_j(\zeta_{0,j})$. For equations with these terms, we sum until the following inequalities hold respectively:

$$\sum_{k=0}^{\infty} s_j^{ord}(k) > 1 - \epsilon, \tag{47}$$

$$\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty} F_j(\zeta_{0,j}) > 1 - \epsilon. \tag{48}$$

Here $F_j(\zeta_{0,j})$ is the CDF of $\zeta_{0,j}$, and ϵ is a specified tolerance level; $\epsilon = 10^{-6}$ is used. With these inequalities, we can limit the necessary computations without introducing large errors.

Finally, there is one more infinite sum in (31). Here we sum over x from negative infinity to a given value y. In the implementation, we reverse this sum. Keeping the latter in mind, let T_x be the total value of the sum from y until x. Thereafter, we sum until it holds that $x < -100$ and $T_x - T_{x+1} < \epsilon$. Note that it has to hold that $T_{x+1} \leq T_x$ by reversing the summation.

4.4.4 The stochasticity of the central lead time

In Section 4.1 the central lead time L_0 is considered to be stochastic. However, this can greatly affect the computations of the *vNext* model. If for example the variance is very large, a depot might need a higher reorder point to satisfy its fill rate target as it needs to account for the additional demand due to the variance in central lead time. The VSL models, in Section 4.3, consider a constant central lead time. Therefore, depots in these models do not need encounter these case of large variance.

To make a fairer comparison between the models during performance measurement, see Section 4.5, we determine which items have an acceptable level of variance in the central lead time. These are determined through the coefficient of variation. If the coefficient of variation of the central lead time is less than or equal to a half, i.e. $\frac{\sqrt{Var[L_0]}}{\mathbb{E}[L_0]} \leq 0.5$, we consider it to be reliable and use the mean and variance as given for its stochastic process. Therefore, in equations were the central lead time is considered to be constant but also reliable, we use $\mathbb{E}[L_0]$ as the constant value. Otherwise, if $\frac{\sqrt{Var[L_0]}}{\mathbb{E}[L_0]} > 0.5$ holds, the variance is too large and the central lead time is considered to be unreliable. Therefore, we will use a constant central lead time constant across all models for that specific item. Section 5 indicates which constant value will be used in case of an unreliable central lead time.

4.5 Measuring the performance

This section elaborates on how the performance and robustness of the various models will be measured and compared. The performance of the models is assessed in two ways. First of all, all models determine their optimal reorder points, after which the inventory allocation is evaluated per model. Secondly, we simulate customer demand to assess the performance of the distribution systems with the models' inventory allocation and evaluate their ability to achieve the service measures. Finally, the robustness of the new models is tested by adjusting specific parameters and assessing how these affect the performance of the models.

Throughout the analysis we will distinguish between assessments on distribution system level and item level. Assessing the distribution system will be done by comparing the models' overall performance between one and another over all items. Whilst for the assessment on item level, we use two distinct items, see Section 5, and evaluate the models' performance. This provides a more comprehensive and in-depth analysis of the performance of the different models.

4.5.1 Comparison of inventory

As previously stated, we will assess the models on their inventory allocation within the distribution system. To assess the inventory allocation, we will first analyse the two individual items and their inventory allocation per model. Moreover, the analysis delves into the optimisation range and domain per model, specifically exploring the relationship between the central reorder point and the total stock, determined by summing over all reorder points within the distribution system. This relation is expected to lead to a convex optimisation domain for the *vNext* model, as also elaborated upon by Geelen et al. (2019). We expect the AVSL models to have a similar optimisation domain. However, the SVSL is expected to have a much smaller optimisation range, as it determines a new lower bound on the central reorder point. This lower bound is expected to be significantly higher than the original lower bound of the *vNext* model since the reorder points at the wait time depots are constant, thus the central reorder point must increase to satisfy the service measures at these depots.

Secondly, the models are assessed on their performance on the distribution system level, or across all items see Section 5. Each model is evaluated on the total stock, determined by summing over all the reorder points within the distribution system. Additionally, we assess the total inventory cost per model, which is computed by multiplying the positive reorder points for an item by its respective price. In this instance, only depots with a positive reorder point are used as these are the only depots that add to the inventory of the distribution system. Note that the order quantities are given and therefore assumed to be constant, as per *Assumption* 16. Hence, left out during evaluation of the total stock and the total inventory cost, as they do not affect the performance across the different models.

Finally, since the SVSL is bound to wait time targets that exceed the DC-depot inter-lead time, a comparable target will be selected for the AVSL model to provide the fairest comparison between the two models. The selected target for the AVSL model will still show the possible flexibility it allows for. The results of the in-depth analyses on item level are shown through bar plots and line graphs, and the results on distribution system level are displayed in tables.

4.5.2 Simulation

Secondly, we will measure the models' overall performance through simulation. We evaluate the performances by measuring the actual service levels and compare these to the set service measures. The general layout and set-up of the simulation is described below. However, the exact parameters of the simulation will be provided in Section 6.2.

For the simulation, we create a distribution system per item. Within the distribution system, customers arrive according to a compound Poisson process as per *Assumption* 17. Hence, the inter-arrival times at each depot follow an exponential distribution with a parameter λ_j, the average inter-arrival time at depot $j \in J$. We estimate λ_j by dividing the entire time span of the historical data by the total number of customers that arrived at depot j, further elaborated on in Section 5. Customer demand is simulated per depot according to the historical data, as the demand sizes K_j in (11) and (34) use an empirical distribution. We expect the means and variances of the demand size to decrease in the simulation, due to the choice of distribution.

For a fair comparison we will use identical random number sequences for each model to simulate the customer arrivals and demand sizes. However, it is possible for an item that different depots have the same λ parameter for the inter-arrival time distribution, i.e. they have identical demand frequencies, further displayed in Section 5.1. In such cases it is important that depots with the same demand frequencies do not get assigned identical random number sequences, as these depots otherwise always experience demand requests at the same time. It is natural this occurs at some point within the simulation, however to consistently experience demand requests simultaneously across different depots is unrealistic and overload the distribution system. Therefore, to prevent such an occurrence we assign a unique seed to each depot, which is a constant value multiplied by the depot's assigned number. This ensures identically generated random number sequences across the different models without overloading the distribution system.

To verify the validity of the simulation, the simulated average inter-arrival times of customers, demand frequency and demand sizes will be tested against the actual values from the historical data. A Wilcoxon Signed-Rank test (Wilcoxon, 1945) will be performed between each pair of actual and simulated variables. The test is a non-parametric test, therefore does not require us to assume a distribution over the variables. The Wilcoxon Signed-Rank test tests the location between two dependent samples, the simulated values depend on the actual values as the individual variables are simulated according to a distribution based on the historical data. As stated by Rey and Neuhäuser (2011), we can define the null and alternative hypotheses of this test, respectively, as:

H_0: The distribution of differences between the simulated and actual values are symmetric around zero, i.e. the two samples do not differ in location.

H_1: The distribution of differences between the simulated and actual values are not symmetric around zero, i.e. the two samples differ in location.

If we accept the null hypothesis, the simulated values do not differ significantly from the historical data. The simulation therefore accurately replicates the realistic setting based on the historical data.

Next, let us elaborate on the layout of the simulation. The simulation has a set time frame and is simulated per day. On a simulated day, first all replenishment orders arrive at all warehouses j for $j \in J_i^+$. After which, we check the depots $j \in J_i$, in random order, if they have a customer arrival. It is possible for a depot j to have multiple customer arrivals on the same day. A depot can only place a replenishment order per day, which is placed after observing all customer arrivals, i.e. at the end of the day. After all depots are checked and have placed their replenishment orders if necessary, the DC is checked. The DC sends out the demand requests on first-come first-serve basis, *Assumption* 9, which corresponds to the checking order of the depots for that day. The DC can only place a supplier replenishment order if they receive an order from a depot. This process corresponds to our assumption of a continuous review policy, see *Assumption* 2.

Furthermore, replenishment orders are only placed when the inventory position of a warehouse $j \in J_i^+$ is less than or equal to its reorder point, i.e. $IP_j \leq R_j$ for $j \in J_i^+$, corresponding to *Assumption* 11. Replenishment orders are of size nQ_j, *Assumption* 8, such that $IP_j + nQ_j > R_j$ for $j \in J_i^+$. The order size is thus equal to n times the order quantity, such that the new inventory position is greater than the reorder point.

The DC-depot inter-lead times L_j are constant, *Assumption* 14, and provided in Section 5. Moreover, the central lead time L_0 follows a *Gamma* distribution if reliable, see Section 4.4.4, which is estimated according to Section 4.4.1. We opted for the *Gamma* distribution as the exponential distribution has a coefficient of variation of one, which is too large to be reliable. Otherwise if the central lead time is unreliable, it is considered to be constant. Section 5 elaborates on the constant value used.

The simulated distribution systems do not use a warm start through simulating an extensive time period before measuring the performance. The simulation focuses on replicating a realistic situation to its best extent. The models are intended to optimise the inventory within a distribution system given historical data in combination with possible forecast data. Therefore, in practice, the parameters of the models should be updated after every customer arrival or on a scheduled basis. As the demand within the distribution systems are assumed to be stationary, *Assumption* 7, this is unnecessary.

Using a warm start would either require each depot to experience a minimum number of customer arrivals or the entire distribution system to encounter a predetermined number of customers. The first option is unrealistic as customer arrivals can be very infrequent for a wait time depot $j \in W_i$, as will be shown in Section 5. Geelen et al. (2019) did use a warm start and measured the service level for each depot per arriving customer. They showed that these remain stable when enough customers arrived in the distribution system. However, they were generally dealing with more fast-moving items. Since our models are generally intended for more slow-moving items, customer arrivals are expected to be more infrequent. Hence, we expect service levels to fluctuate more and these depots may need an unrealistic amount of time or customer arrivals to attain stable service levels. The second option provides an unfair start to the wait time depot, as fill rate depots experience vastly more customer arrivals. Therefore, the wait time depots might have yet to converge to their attainable service level.

Hence, we opt to let warehouses start with a predetermined amount of stock and use this as the starting set-up, Section 6.2 indicates the specific set-up parameters of the simulation. Moreover, the distribution systems will be simulated for an extensive period of time, such that each depot experiences sufficient customer arrivals to measure its performance.

Finally, the results of the in-depth analysis of the two distinct items will be analysed first. The analysis inspects which depots meet their service measure and why some do not satisfy their service measure. Moreover, it will also investigate the wait time distributions of specific wait time depots. Furthermore, we will analyse the results on distribution level. Hence, the simulation is performed individually for the distinct items and simultaneously for all items. Analyses will be conducted on the depots to determine the number of service measures satisfied and also the type service measures satisfied.

4.5.3 Testing the robustness

After the simulation, the robustness of the AVSL model will be analysed. It is the most flexible model, therefore we want to test its performance under varying scenarios. Moreover, we also want to highlight important parameters that influence the model's performance. These tests will mainly be carried out on the individual items, as given in Section 5 but can be broadened to the entire distribution system. We test the robustness by adjusting different parameters and analysing how the model's proposed solution and performance are affected. Below we describe the various tests and characteristics that are explored. The specific modifications will be provided in Section 6.3. All test results will be compared to the baseline results presented in Sections 6.1 and 6.2.

Firstly, as mentioned in Section 4.4.1, we compare the two different wait time distributions. In general, the gamma distribution is used in all computations regarding the wait time. However, we want to analyse the effects on the wait time when using a more heavy-tailed distribution, i.e. the *LogNormal* distribution as stated in Section 4.4.1. The expectation is that a heavier-tailed distribution causes a more conservative approach. Therefore, leading to higher reorder points locally or, in other words, the model prefers to stock locally.

Secondly, the main strength of the AVSL model is the ability of set wait time targets WT_j^T below the DC-depot inter-lead times L_j. A baseline target is determined in Section 6.1 to compare the VSL models fairly. However, we want to explore the effects on the AVSL model and evaluate its performance when the wait time targets are altered. We expect that the AVSL model's performance will be negatively effected when $WT_j^T < L_j$ holds.

Furthermore, we expect that the order quantities are of great influence on the performance of the AVSL model. These are determined by Gordian Logistic Experts according to the historical demand data, therefore incorporating the demand frequencies and sizes. Keeping the order quantity constant, we want to explore the effect on the AVSL model when the demand frequency and sizes are altered, and the order quantity is no longer valid for the demand. The expectation is that when there is a significant difference in average demand size and order quantity at a depot, especially when the average demand size is greater than the order quantity, the AVSL model can encounter difficulties in predicting the correct central reorder point to account for the increasing orders of a depot.

Finally, we will explore the effect of the order quantity. The demand statistics in Table 1 indicate that the demand sizes generally experience a lot of variances. Therefore, the order quantities might be unable not cover most of the demand requests, possibly negatively impacting the performance as the DC needs to correctly account for such high demand sizes. Therefore, we want to explore the possibility of assigning every depot an order quantity that covers at least 80% of its demand requests. In other words, the order quantity is the 80% quantile of the demand sizes. This percentage is obtained through conferring with Gordian Logistic Experts. Moreover, the equations in Section 4.2 suggest that the DC accounts for kQ_j replenishment orders. Therefore, we want to examine the effects on the AVSL model's performance when using increased order quantities. We expect that using order quantities that cover most demand sizes, provides a better input for the AVSL model and thus positively affects its performance.

5 Data

This section provides a detailed description of the data employed in the current paper. A client of Gordian Logistic Experts, referred to as "Company A", provided the data. Company A is a South-African enterprise specialised in the manufacturing, distributing, and supplying of material handling equipment to various industries. They primarily focus on the selling and distribution of agriculture, forestry, and mining equipment. Company A operates on a global scale, with central warehouses in Africa and Europe. Our paper focuses on the African distribution system, consisting of 25 local warehouses (depots) and a central warehouse (DC). However, it should be noted that the DC also encounters direct demand. A fictitious depot is created to account for this demand, as also proposed in Axsäter et al. (2007). This depot will be treated as a regular, non-prioritised depot to ensure consistency throughout this paper, increasing the total number of depots to 26.

The data consists of 10,000 distinct items handled by the distribution system. Each item has its own data, including demand and supply data per warehouse. The data covers a five-year time frame, from 01-01-2018 to 31-12-2022. Upon further inspection of the data, 51 items have no demand data within the set time frame. Hence, these items are excluded, and the remainder of the research only considers the other 9949 items.

Section 2 indicates that the proposed models in the current paper use distribution systems containing depots handling slow-moving items. Gordian Logistic Experts classifies items per depot based on their demand frequency. Items are therefore classified with regards to all the other items at each distinct depot. We use the classification of Gordian Logistic Experts. However, the rule-of-thumb is that an item is considered to be slow-moving if it has a demand frequency of 1.0 or less. Table 2 discusses the demand statistics in detail and elaborates on the possible divergent items from the rule-of-thumb.

Upon analysing the data, 11 items are classified as pure fast-moving items. Therefore, they do not classify as slow-moving at any depot. Since these items compose less than one percent of the data, they will not significantly affect the results and are therefore not excluded.

As already indicated, items are classified per depot. Therefore, a depot can be considered fast-moving for one item and slow-moving for another. For ease of analysis, and the displaying of the data and results, we divide the complete distribution system into a distribution system per item. This does not alter the data or results in any way, but each item now has its own distribution system. Therefore, we now have a total of 38041 depots. For convenience, we denote a depot with an item classified as fast-moving as FL and a depot with an item classified as slow-moving as WL. Resulting in, 5208 FL depots and 32833 WL depots.

Table 1 presents the general statistics of all items at the DC. The contractual central lead time is the agreed-upon lead time between Company A and its external suppliers. However, as shown, the actual central lead times is also known. The table indicates that the contractual lead times and the actual lead times can differ significantly. Therefore, if the central lead time is considered to be unreliable for a specific item, as per Section 4.4.4, we use the contractual lead time as the constant central lead time across all models. The table further shows that for some items the contractual central lead times are 999 days. Conferring with Gordian Logistic Experts, they indicated that these items do not have a contractual lead time. Therefore, for the remainder of this paper, the contractual central lead time of these items is set to 888 days, the highest actual contractual central lead time. This alteration causes the average contractual central lead time to decrease to 64.479 days.

As per *Assumptions* 14 & 16, the DC-depot inter-lead times as well as the order quantities are constant. The order quantities are predetermined by Gordian Logistic Experts based on the time frame of the data. Their precise methods remain undisclosed, but use the EOQ method (Harris, 1913) and Croston method (Croston, 1972) to determine their optimal order quantities (Gordian Logistic Experts, 2021). However, Gordian Logistic Experts conferred with Company A, and due to their preferences the order quantities are adjusted downwards, especially for the depots with slow-moving items. Company A prefers order quantities of 1 at these depots. The depots' lead times are divided into three groups; one depot has a lead time of 1 day, 21 depots of 3 days and finally four depots have a lead time of 14 days. The average depot lead time is 4.615 days. Note that these refer to the original 26 mentioned depots within the complete distribution system. When we consider the weighted values of the 38041 depots in the distribution systems per item, the average depot lead time is 3.609 days.

Furthermore, Table 1 also indicates the number of items per depot for the complete distribution system. As well as the number of depots per item for the distribution systems per item. Obviously, each item is distributed to at least one depot. The maximum number of depots is 26, which equals the total number of existing depots. The average amount of depots per item is about four. The average amount of FL depots per item is significantly lower than the average amount of WL depots. Following this table, we will refer solely to the distribution systems per item for all further analyses.

Table 1: Statistics of the original data at the DC

	Min.	Mean	Max.
Price (ZAR)	0.001	3574.328	800086.200
Contract central lead time (days)	2	65.326	999
Avg. central lead time (days)	0.000	46.380	888.000
Avg. variance of central lead time	0.000	2012.10	127569.330
Avg. supplies at the DC per item	0.000	4.604	362.000
DC order quantity	1	16.281	5000
Items per depot	311	1462.962	7569
DC-depot inter-lead time (days)	1	4.615	14
Depots per item	1	3.804	26
FL depots per item	0	0.521	26
WL depots per item	0	3.283	24

Note: FL depots handle items classified as fast-moving. WL depots handle items classified as slow-moving.

Table 2 displays the statistics regarding the demand data. This includes the demand frequencies, inter-arrival times and demand sizes. It displays the statistics for all depots combined and per classification of the depots with fast-moving or slow-moving items, FL and WL, respectively. The average time inter-arrival time at a depot is computed by dividing the time frame (in days) by the number of customer arrivals. Moreover, we computed the yearly demand frequency by multiplying the number of customer arrivals by 365 days (the assumed number of days within a year) and then dividing this by the time frame.

The data shows a clear difference in yearly demand frequency between the FL and WL depots. The mean demand frequency is more than 20 times greater for the FL depots than for the WL depots. The inter-arrival times also reflect and further highlight this difference. The mean of the WL depots is almost 1000 days higher than that of the FL depots. Note that there appear to be WL depots with a relatively high demand frequency. As mentioned previously, classification is done per item with regard to all other items at a single depot. Therefore, although these items defer from the rule-of-thumb, they are still considered to be slow-moving with regard to the other items at that depot. As this only concerns a handful of items, we opt to retain all of these items for the remainder of this research.

Finally, the average demand sizes' mean differs slightly between the WL and FL depots. However, the variance of the average demand sizes differs significantly between the two types of depots. The FL depots have a generally much higher variance of demand sizes compared to the WL depots. Moreover, we notice a significant difference in average demand size and average order quantity, especially for the WL depots as their average demand size is double their average order quantity. However, for the FL depots, the opposite holds, their average order quantity is greater than their average demand size.

Table 2: Demand statistics of the data specified per depot type

	Depots	Min.	Mean	Max.
Demand frequency (yearly)	All	0.200	2.708	1770.630
	FL	0.200	15.402	1770.630
	WL	0.200	0.672	48.400
Avg. inter-arrival time (days)	All	0.206	939.732	1826.000
	FL	0.206	78.563	1826.000
	WL	7.541	1077.793	1826.000
Avg. demand size	All	1.000	2.483	1481.000
	FL	1.000	3.152	292.208
	WL	1.000	2.376	1481.000
Var. demand size	All	0.000	35.350	69,938.000
	FL	0.000	72.480	65,179.300
	WL	0.000	25.230	69,938.000
Order quantity	All	1.000	1.407	325.000
	FL	1.000	3.502	325.000
	WL	1.000	1.071	69.000

Note: FL depots handle items classified as fast-moving. WL depots handle items classified as slow-moving.

5.1 Individual items

As stated previously, two distinct individual items are selected for a more in-depth analysis in Section 6. With these items, we aim to provide further insight into the workings of the different models. For the first item we consider a general slow-moving item. The demand across all depots is very infrequent, and the demand sizes vary little to not at all. On the other hand, the second item we consider is classified as a more fast-moving item across some depots, yet it is slow-moving for most depots. The demand sizes vary quite heavily for both different sets of depots. Hence, these two items have been selected in particular to further investigate the performances of the models for these varying types of items.

Let us introduce the Item I, the first item. This item is distributed to five depots, all classifying it as slow-moving. Table 3 displays the statistics across all warehouses. Note that there is no demand data for the DC, as it only receives indirect demand and is therefore entirely dependent on the orders from the depots. It holds that the demand is relatively infrequent for all depots, with only WL5 experiencing demand multiple times a year on average. Moreover, as the table shows, the demand sizes vary little to not at all, with only WL5 having any variance in its demand sizes. For WL2 - WL4, no variance can be determined, since only one demand request was encountered for each depot. Order quantities do not equal the average demand sizes, even though they are fairly constant.

Table 3: Descriptive statistics of Item I

Warehouse	Contractual lead time (days)	Order quantity	Avg. time between demand (days)	Demand frequency (per year)	Avg. demand size	Var. demand size
DC	47	5	-	-	-	-
WL1	3	1	912.500	0.400	2.000	0.000
WL2	3	1	1825.000	0.200	2.000	-
WL3	3	1	1825.000	0.200	1.000	-
WL4	14	1	1825.000	0.200	2.000	-
WL5	1	1	121.667	3.000	1.267	0.638

Note: FL depots classify the item as fast-moving, and WL depots as slow-moving.

The statistics for Item II are displayed in Table 4. An immediate noticeable difference compared to the previous item is the number of depots. Item II has a total of 21 depots, of which 17 depots classify the item as slow-moving and four as fast-moving. Therefore, Item II deals with more frequent demand, greater average demand sizes, and heavier varying demand sizes. The difference between the slow-moving and fast-moving depots is very notable. The WL depots have a demand frequency of at most 1.2 items per year. On the other hand, the FL depots have an average demand frequency of at least three items per year. Finally, the table indicates that, similar to Item I, the order quantities do not correspond to the average demand size when they have a low variance .

Table 4: Descriptive statistics of Item II

Warehouse	Contractual lead time (days)	Order quantity	Avg. time between demand (days)	Demand frequency (per year)	Avg. demand size	Var. demand size
DC	22	500	-	-	-	-
WL1	3	1	365.000	1.000	5.600	34.800
WL2	3	1	304.167	1.200	5.000	36.000
WL3	3	1	365.000	1.000	2.000	0.000
WL4	3	1	365.000	1.000	3.200	7.200
WL5	3	1	608.333	0.600	2.000	0.000
WL6	3	2	1825.000	0.200	16.000	-
WL7	3	1	456.250	0.800	28.250	1240.250
WL8	3	2	608.333	0.600	14.000	252.000
WL9	3	1	1825.000	0.200	16.000	-
WL10	3	1	365.000	1.000	10.000	24.000
WL11	3	1	912.500	0.400	12.000	32.000
WL12	3	1	608.333	0.600	8.333	56.333
WL13	3	1	1825.000	0.200	8.000	-
WL14	14	1	365.000	1.000	8.000	22.000
WL15	14	1	365.000	1.000	3.600	6.800
WL16	14	1	1825.000	0.200	2.000	-
WL17	14	3	608.333	0.600	24.000	0.000
FL1	3	3	24.333	15.000	2.200	1.432
FL2	3	1	53.676	6.800	6.676	48.529
FL3	3	1	96.053	3.800	3.474	2.485
FL4	1	5	9.973	36.600	5.792	91.836

Note: FL depots classify the item as fast-moving, and WL depots as slow-moving.

5.2 Alteration for the EVSL model

As elaborated on in Section 4.3.3, we will conduct an experiment in which we allow reorder points below -1. To facilitate these reorder points, we require *Assumption* 22 stating that all wait time depots, $j \in W_i$, have a Poisson arrival process with a constant demand size of one. Therefore, we alter the data as in Table 2 to accommodate for this assumption and only permit a constant demand size of one.

The following alteration is performed. For all WL depots, every customer with a demand size greater than one is split into distinct customers, such that all new customers have a demand size of exactly one. Every new customer arrives on the initial demand date. Note that by altering the data in such manner only the demand frequency, inter-arrival times and demand sizes are modified. All other parameters remain unaffected.

Table 5 displays the modified demand statistics. Comparing these statistics to Table 2, one should note that the statistics regarding the FL depots are unaffected. Moreover, the order quantity statistics are omitted from the table as they remain unmodified. However, the remaining demand statistics behave as expected. The demand frequency increases on average whilst the average inter-arrival time decreases for the WL depots, and therefore a similar change occurs for the averages of all depots combined. Finally, the average demand size of the WL depots is exactly one and the variance of the demand size drops to zero. Logically, this also lowers the average and variance of the demand size for all depots combined.

Table 5: Demand statistics of the modified data specified per depot type for the EVSL model

	Depots	Min.	Mean	Max.
Demand frequency (yearly)	All	0.200	3.615	1770.630
	FL	0.200	15.402	1770.630
	WL	0.200	1.725	489.732
Avg. inter-arrival time (days)	All	0.206	735.392	1826.000
	FL	0.206	78.563	1826.000
	WL	0.745	840.754	1826.000
Avg. demand size	All	1.000	1.297	292.208
	FL	1.000	3.152	292.208
	WL	1.000	1.000	1.000
Var. demand size	All	0.000	10.020	65,179.300
	FL	0.000	72.480	65,179.300
	WL	0.000	0.000	0.000

Note: FL depots handle items classified as fast-moving. WL depots handle items classified as slow-moving.

6 Results

This section discusses the results of our research, we analyse the outcomes of the models as discussed in Section 4. We will first analyse the inventory allocation as done per model. Secondly, a simulation will be conducted, through which we access the models' performances. Thirdly, we will provide results on the robustness of the developed models and determine which parameters are most influential during the optimisation of the inventory allocation. Hence, this section's order corresponds to the analysis's explanation in Section 4.5. Finally, as discussed in Section 4.3.3, we conduct an experiment on allowing reorder points below -1 at depots with a wait time service measure, i.e. the EVSL model. We will analyse the results of the inventory allocation and simulation for this model.

All results are obtained using the Java programming language (Oracle Corporation, 1995). The results are visualised using the R programming language (R Core Team, 2019). Finally, computation times are analysed within this section, therefore it is important to note that these results are obtained on a 2020

MacBook Pro with an 8-core Apple M1 CPU-chip and 16 GB of RAM.

6.1 Inventory allocation results

Before discussing the results, let us elaborate on the targets and parameters for the various models. As mentioned in Section 4, the various models cannot use identical targets due to their intended usage. Therefore, we will assign similar targets across all models for the fairest comparison. All models use a fill rate target of 95% at all fill rate depots, i.e. $\beta_j^T = 0.95$ for all $j \in J_i \setminus W_i$. For the SVSL model, the wait time target at all wait time depots equals the DC-depot inter-lead time plus one day, i.e. $WT_j^T = L_j + 1$ for all $j \in W_i$, as the target must be greater than said lead time, see Section 4.3.1. For the AVSL model, a wait time target of four days is used at all wait time depots, i.e. $WT_j^T = 4.0$. This wait time target is chosen as most depots have a DC-depot inter-lead time of three days, see Section 5. Therefore, this target selection provides the fairest comparison between the SVSL and AVSL model, whilst still displaying the flexibility of the AVSL model as some depots have an inter-lead time of 14 days.

Furthermore, as mentioned in Section 4.2, the wait time service measure is computed as a probabilistic wait time, i.e. it must hold for predetermined percentile of the customers. The SVSL and AVSL models both use a 95% probability target, i.e. $P_j^T = 0.95$, indicating that at least 95% of all customers must be satisfied within the set wait time target at a specific depot. The remaining parameters for the models include $\beta_0^{min} = 0.6$ and $\beta_0^{max} = 0.99$. These values are determined in collaboration with Gordian Logistic Experts. They prefer the DC to have a minimal fill rate of 60% and consider it to be maximal at 99%. Furthermore, Section 4.2.1 indicates that for the VSL models an adjusted $\delta_j(k)$ should be used if Q_j is relatively large. Therefore, we will use the adjusted $\delta_j(k)$ in (22) if $Q_j > 10$.

Depots will use a fill rate service measure when they classify an item as fast-moving, and a wait time service measure is used for depots that classify an item as slow-moving. The assignment for the type of service measure corresponds to the grouping of depots in Section 5. Thus, fill rate depots are denoted as FL and wait time depots as WL, respectively, indicating fast-moving and slow-moving depots. Finally, this section only analyses the inventory allocation across the different models. Therefore, these results provide an indication of the possibilities regarding the inventory reduction, yet they are not indicative for the performances of the models. We access the models' performances in Section 6.2, in which we highlight the importance of effective inventory allocation over simply allocating minimum inventory.

Finally, as elaborated in Section 4, we refer to the total stock (given a specific reorder point) as the sum over all reorder points, i.e. $TS(R_0) = \sum_{j \in J_i^+} R_j$. This definition of the total stock will be used throughout the remainder of the current section and is an important term and concept within our analyses.

6.1.1 Individual items

We first analyse the two individual items presented in Section 5.1. Since all models optimise between a certain lower and upper bound of the central reorder point, as per Section 4.1.2, Figure 2 displays the optimisation range for Item I. The highlighted points are the determined solutions per model. The figure shows that the *vNext* model's solution lies at the lower bound, as its total stock is monotonically increasing in the central reorder point on the entire domain.

The solution of the AVSL model is located in a local minimum and on a section of the domain for which the total stock remains constant, after which the relation increases slightly and experiences a sudden decrease to the global optimum at (5, 2). The AVSL model is able to assign more stock centrally, therefore it can decrease the local stock and provide a better solution. However, due to the bisection-based optimisation approach, the determined solution is non-optimal and is not at the global minimum. The bisection-based approach requires unimodality on the selected domain, this is violated by the AVSL model in general. Moreover, the step size used in our bisection-based approach is generally too big. Combining these causes the AVSL model to achieve non-optimal solutions as it can get stuck in a local minimum, or possibly not achieve a minimum as by leaping over it due to its step size.

The lower bound on the central reorder point R_0 for the SVSL model is greater compared to the other two models, as it must provide a valid solution at all the wait time depots. Its optimal solution therefore lies at this lower bound, and, similarly to *vNext* model, the total stock is monotonically increasing in the central reorder point. Finally, the graph displays that the total stock is lowest for the SVSL model, whilst the *vNext* model allocates the most stock.

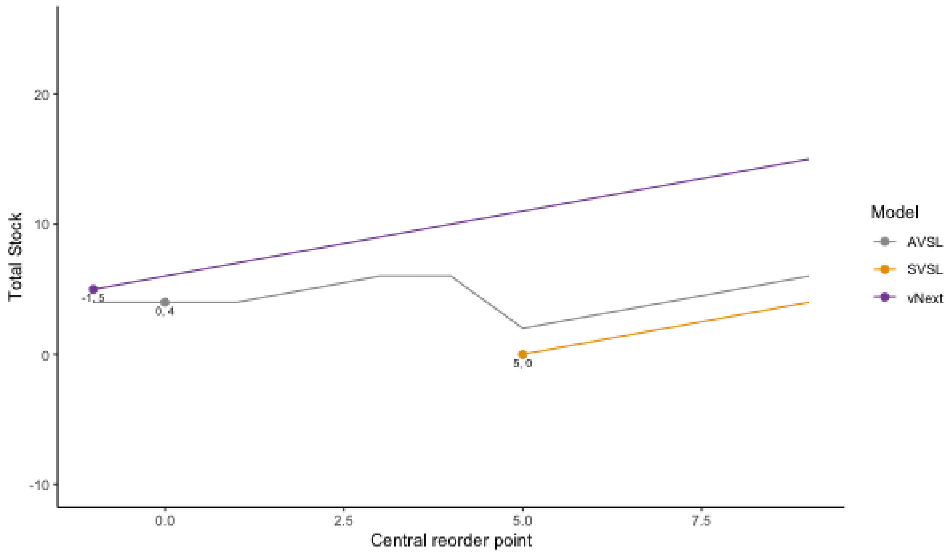

Figure 2: Total stock versus central reorder point per model for Item I

Note: The highlighted points indicate the determined solution per model

Figure 3 shows the reorder points per warehouse of all models for Item I, such that all depots achieve their fill rate and wait time service measures. Note that since all depots are considered to be slow-moving, they all use a wait time service measure in the the the SVSL and AVSL models. Hence, all depots have a negative reorder point, equal to minus one, for the SVSL model.

The *vNext* and AVSL models stock similarly for most of the depots. The AVSL model prefers a larger central reorder point at the DC, whereas the *vNext* model prefers a larger reorder point at WL5 instead. Note that both ensure a positive inventory position across all depots, even though some reorder points are set to zero or even negative one for the DC. The inventory position (IP) will still always be positive, since order quantities are always positive and the IP is defined on $[R_j + 1, R_j + Q_j]$ for all $j \in J_i^+$. Moreover, for the depots with zero or no variance in demand size, their lower bound on the IP is equal to the average demand size, see Table 3. This explains the difference in stock for WL5, as this depot varies in demand size and has a lead time of one day. Therefore, the AVSL model is able to replenish it in time for the wait time service measure. It should be noted that the gap between the allowed wait time and lead time is the largest at this depot. Finally, due to the SVSL model's restrictions on the local reorder points, it must allocate significantly more stock at the DC compared to the other models in order to satisfy all wait time service measures.

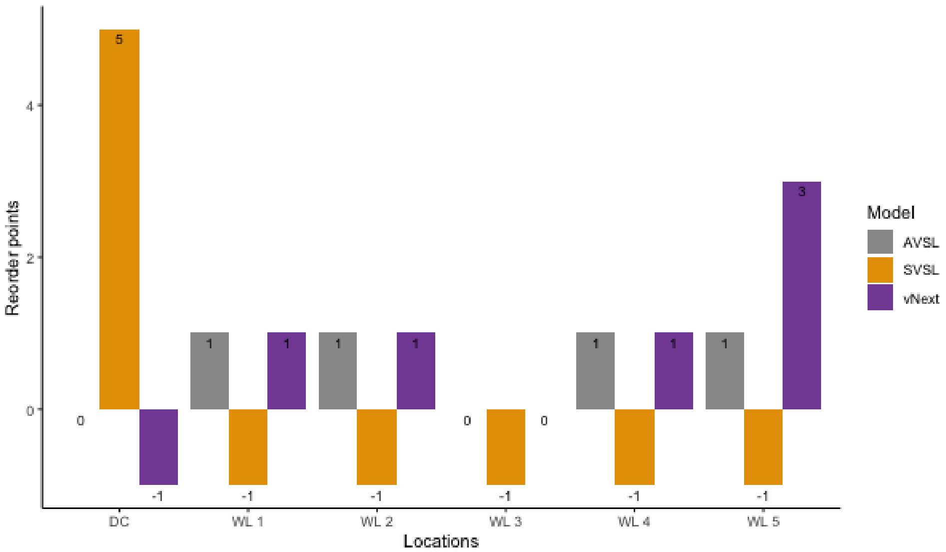

Figure 3: Reorder points at all warehouses per model for Item I

For Item II, Figures 4 and 5 display a similar graph and plot as above. Figure 4 shows a similar result to Figure 2 for Item I. Again, the SVSL model allocates the least amount of stock, while the *vNext* model allocates the most. Moreover, the SVSL model has a smaller optimisation range as its lower bound is in the domain of the *vNext* and AVSL models.

As the order quantity at the DC for Item II is considerably large, the optimisation range is greater than that of Item I. In general, the order quantity at the DC greatly affects the optimisation range. The graph displays the general form of the optimisation domain for the *vNext* and AVSL models, as the optimisation range is extensive. As expected, the relation between the central reorder point and the total stock is (almost) convex for the *vNext* model, see Section 4.1.3. For the AVSL model, it was expected to have a similar shape and thus also be convex. However, the graph indicates that the relation between the central reorder point and the total stock is not convex at all, or even unimodal. Therefore, similar to Item I, using a bisection-based optimisation approach to determine the solution does not lead to the global minimum but a non-optimal solution. In general, due to the optimisation range being not unimodal and the step sizing being rather large, the AVSL model is unable to attain the global minimum. Hence, for Item II, the global minimum is actually achieved at (11, 96), instead of the highlighted point at (18, 102).

The graph displays another interesting result for the AVSL model. It was already visible on a smaller scale in Figure 2, but the total stock heavily decreases at certain points and then slightly increases again as the central reorder point increases. However, when the optimisation reaches a central reorder point of 11, the total stock increases monotonically thereafter. Thus, the AVSL model prefers to stock at most 11 items at the DC and increasing the central reorder point further provides no additional benefit. Furthermore, these sudden declines in total stock occur due to the model being able to decrease the local stock significantly for a certain central reorder point. Thereafter, slightly increasing it, does not yield any further decrease in local stock until another specific central reorder point is reached, after which this process repeats itself. This phenomenon occurs due to the integrality of the reorder points, as per *Assumption 8*.

Finally, a final observation can be made from Figures 2 and 4. As mentioned, the AVSL model cannot determine an optimal solution and thus does not arrive at the global minimum for both items. However, its global minimum corresponds to the optimal solution of the SVSL model. This could implicate that the solutions from the SVSL model can indicate where the global minimum of the AVSL model is located. Moreover, both models prefer a non-negative central reorder point, contrary to the *vNext* model which prefers a negative central reorder point.

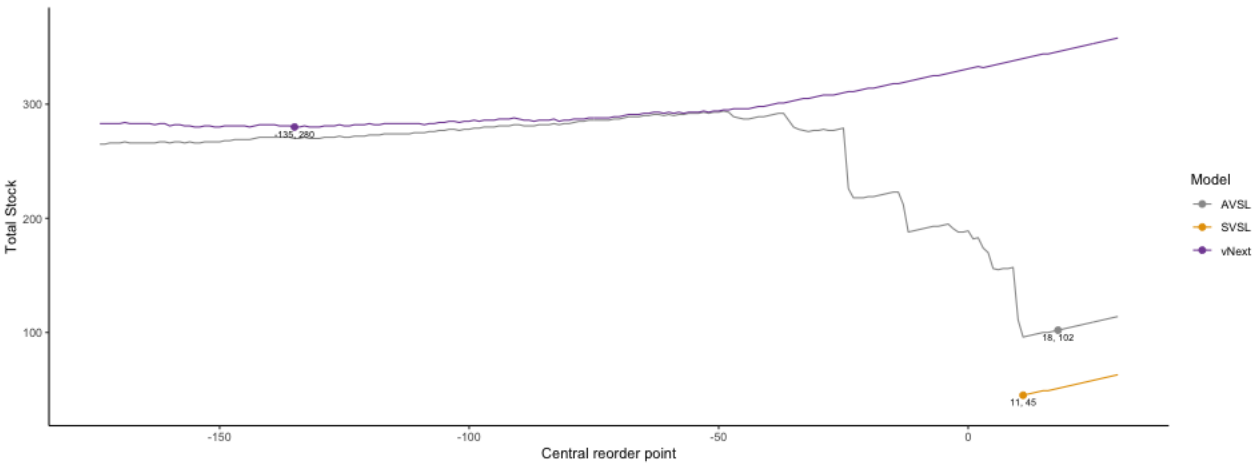

Figure 4: Total stock versus central reorder point per model for Item II

Note: The highlighted points indicate the determined solution per model

Considering the previous graph, Figure 5 shows an unsurprising result. It displays the reorder points for each model per warehouse. It is clearly visible that both the SVSL and AVSL models have a considerable positive central reorder point, allowing for a lower stock allocation locally. On the other hand, the *vNext* model tends to stock more locally due to its fill rate service measures, and is therefore selects a negative central reorder point.

The most interesting depots are FL4, WL14, WL15, WL16 and WL17. The first depot is a fill rate depot. However, due to the SVSL and AVSL stocking more centrally, they are able to allocate significantly less stock to this depot compared to the *vNext* model. For WL14, WL15, WL16 and WL17, it is interesting that the AVSL and *vNext* models both allocate stock locally. This is due to their DC-depot inter-lead time of 14 days, see Table 4, and their wait time target of 4 days. Therefore, the AVSL must allocate stock to these depots, similar to the *vNext* model, in order to satisfy their wait time service measures. For the remainder of the wait time depots, the AVSL and SVSL do not allocate any stock contrary to the *vNext* model. Therefore, especially at WL6 - WL13, the allocation differs significantly.

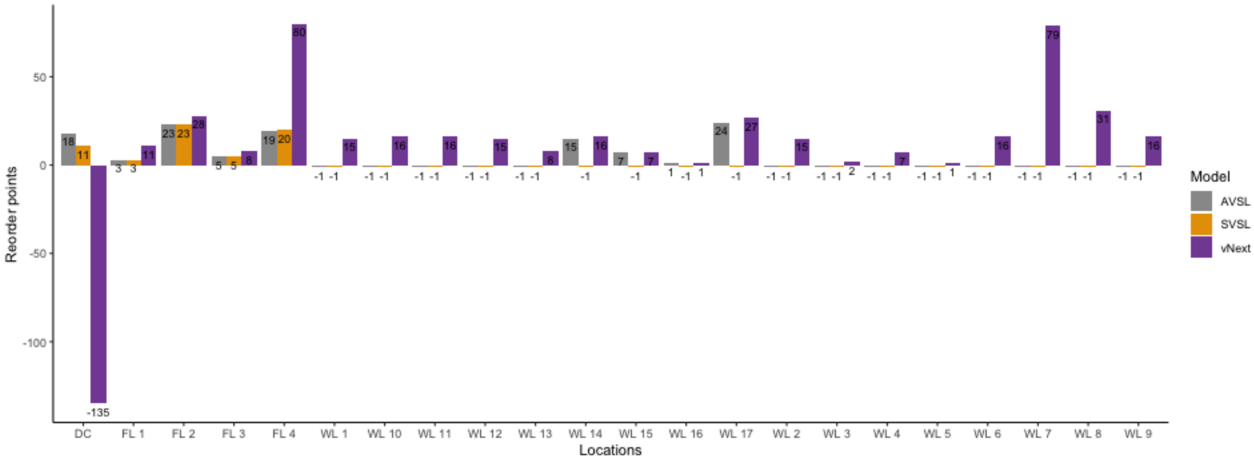

Figure 5: Reorder points at all warehouses per model for Item II

6.1.2 All items

Let us now analyse the inventory allocation results of all items for each model. Tables 6, 7 and 8 present the inventory allocation for the *vNext*, SVSL and AVSL models, respectively, obtained through the bisection-based approaches utilised by all models as described in Section 4. In these tables the minimum (Min.), average (Avg.) and maximum (Max.) are given per item, whilst the final column (Total) is the summation over all items. The total stock TS is computed similarly to Section 4 and is therefore equal to the summation over the reorder points. The total cost on the other hand is computed by multiplying the price by $R_j + Q_j$, as this provides the fairest comparison.

The central reorder points and total stock behave as expected upon analysing the results. The *vNext* models prefers stock at the depots, whilst the SVSL and AVSL models prefer to stock centrally. Therefore, the average central reorder point is significantly lower for the *vNext* model. The average central reorder point is largest for the SVSL model due to having fixed reorder points of -1 at the wait time depots; as expected, the AVSL model finds an intermediate solution. Identical results are observed for the total central reorder point. However, the difference between the *vNext* model and the VSL models is astonishingly immense. The total central reorder points for the AVSL model increased by 388.6% and for the SVSL they increased by 1116.4% compared to the *vNext* model.

As a result of the central reorder points, we observe the opposite result for the total stock within the distribution system. Here it holds that the *vNext* allocates the most stock while the SVSL model allocates the least. Again, the AVSL determined intermediate solutions in between the other models. However, it is considerably closer to the total stock of the SVSL model, as it only differs 2.8%. The decrease in the total stock compared to that of the *vNext* model is 33.6% for the SVSL model and 31.7% for the AVSL model. Therefore, both models exhibit great potential as they decrease the total stock within the distribution system significantly.

Following the results above, the *vNext* model's inventory allocation is the most expensive. However, the AVSL model's inventory allocation is the cheapest. It seems that even though the AVSL model used more stock than the SVSL model, it distributes the allocation of the more expensive items better; thus having a lower total cost compared to the SVSL model. Due to the SVSL model's fixed local reorder points, the central reorder point sometimes has to increase significantly to accommodate the wait time service measures. Alternatively, the AVSL model is allowed to allocate stock at the depots. Therefore, the central reorder point does not increase as much, resulting in a lower total cost. Comparing the total cost to the *vNext* model, the SVSL model decreases the total cost by 5.1%, whilst the AVSL model decreases it by 7.0%.

Finally, the results do indicate that the decrease of total stock and total cost for the AVSL model are paired with a significant increase in computation time. On average, the AVSL model takes about 0.1 seconds more to compute the inventory allocation per item versus the *vNext* model. The difference between the SVSL model and AVSL model is not significant. The total computation time of the AVSL model is about 18.3 minutes, which is an increase of 275.9% in comparison with the *vNext* model, but only 12.2% more than the SVSL model. Thus, the AVSL model is able to produce the cheapest inventory allocation of the three models, but it takes much more computation time.

Table 6: Inventory allocation results of all items for the *vNext* model

	Min.	Avg.	Max.	Total
Central reorder point	-1988	0.713	1642	7128
Total stock	-1163	13.860	2934	138,524
Inventory cost (ZAR)	0.00	33,238.00	4,646,336.00	332,209,936.00
Computation time (s)	0.000	0.029	18.441	292.657

Table 7: Inventory allocation results of all items for the SVSL model

	Min.	Avg.	Max.	Total
Central reorder point	-224	8.675	2747	86,706
Total stock	-57	9.209	3403	92,040
Inventory cost (ZAR)	0.00	31,534.00	4,795,590.00	315,177,437.00
Computation time (s)	0.000	0.098	162.967	980.746

Table 8: Inventory allocation results of all items for the AVSL model

	Min.	Avg.	Max.	Total
Central reorder point	-1988	3.484	1646	34,827
Total stock	-1184	9.470	2932	94,655
Inventory cost (ZAR)	0.00	30,902.00	4,646,336.00	308,812,736.00
Computation time (s)	0.000	0.110	165.535	1100.189

6.2 Simulation results

This section depicts the simulation results of the various models and accesses their performance. The general outline of the simulation is given in Section 4.5.2. For convenience, the simulation utilises the identical service measures from the inventory allocation results (Section 6.1). The results of individual items will be analysed first, after which the overall results regarding the performance of the models are evaluated.

Before discussing the results, let us elaborate on the simulation parameters. As described in Section 4.5.2, we use a starting set-up instead of a warm start. Therefore, we assign a starting inventory at each warehouse. For a depot $j \in J_i$, its starting inventory equals zero if $R_j < 0$ and is equal to $R_j + Q_j$ otherwise. Moreover, the starting inventory at the DC is always equal to $R_0 + Q_0$. These values allow for a fair start of the simulation.

Furthermore, as elaborated on in Section 4.5.2, customers arrive according to a Poisson process. Hence, their inter-arrival times follow an exponential distribution based on the average time between demands in days, as given in Section 5. The demand sizes are based on the historical data and therefore follow an empirical distribution. The supply lead times of the DC, L_0, are *Gamma* distributed if the historical data is reliable, see Section 4.4.4. Otherwise, L_0 is considered to be constant and equals the contractual lead time, see Section 5.

The simulation has a duration of 1000 years and is simulated on a daily basis. This duration provides each depot with sufficient customer arrivals to measure its performance. To verify the starting set-up of the simulation will divide the simulation into five equal periods of 200 years to compare the results per period. This is done to verify that the performance remains similar across the different periods. Moreover, the starting period, the first 200 years, is compared to the total duration, 1000 years, as it is possible for the service measures to be more influenced by individual customer arrivals in the first 200 years. This phenomenon is more likely to occur at the wait time depots due to their general lower demand frequencies. Therefore, to verify these results we again use a Wilcoxon-Signed-Rank Sum test (Wilcoxon, 1945), as previously elaborated on in Section 4.5.2 for the simulation parameters.

6.2.1 Individual items

Let us first analyse the results for Items I and II. Note that all tables display the achieved service level per depot in percentages. A depot satisfies its fill rate or wait time service measure if it achieves a service level of at least 95%, as described in Section 6.1. Moreover, as discussed in Section 5, Tables 3 and 4 display the statistics of both items and will be referred back to during the analysis.

The simulation results of Item I are displayed in Tables 9, 10 and 11 for the *vNext*, SVSL and AVSL models, respectively. Note that, Table 12 provides the results of the DC per model. Moreover, as indicated in Section 4.5.2, all models experience identical demand requests. The results indicate that all depots satisfy their service measure for the *vNext* and SVSL models. For the AVSL model, however, WL5 misses its target by 1.7%. Note that this is the only depot with a demand frequency greater than one per year and a demand size variance greater than zero. Therefore, it seems that the AVSL model is less able to cope with these parameters.

Another noticeable result is that the AVSL model is able to achieve its service measure at WL4. This is the only depot where $L_j > WT_j^T$ holds, or in other words its lead time is larger than the set wait time target. The AVSL model is able to achieve this as it is allowed to assign non-negative reorder points, as described in Section 2.2 and 4.3.2.

Finally, the wait time service levels of the SVSL model are greater than those of the AVSL model, in general. Yet, the local reorder points are lower as they must be equal to minus one for the SVSL model. Due to the higher central reorder point, the SVSL model is able to replenish the depots consistently, as will be shown in Figures 6 and 7, therefore the depots are able to serve their customers more consistently within their wait time target. Leading to a higher achieved wait time service levels for the SVSL model.

Table 9: Simulation results of *vNext* model for Item I, $R_0 = -1$

Depot	Reorder point	Order Quantity	Fill rate target (%)	Achieved service level (%)	Total customers	Total demand
WL1	1	1	95.0	97.8	447	894
WL2	1	1	95.0	98.7	224	448
WL3	0	1	95.0	99.5	187	187
WL4	1	1	95.0	98.3	177	354
WL5	3	1	95.0	96.9	3001	3838

Table 10: Simulation results of SVSL model for Item I, $R_0 = 5$

Depot	Reorder point	Order Quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	4	99.8	447	894
WL2	-1	1	4	100.0	224	448
WL3	-1	1	4	100.0	187	187
WL4	-1	1	15	100.0	177	354
WL5	-1	1	2	99.8	3001	3838

Table 11: Simulation results of AVSL model for Item I, $R_0 = 0$

Depot	Reorder point	Order Quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	1	1	4	99.1	447	894
WL2	1	1	4	99.6	224	448
WL3	0	1	4	100.0	187	187
WL4	1	1	4	98.9	177	354
WL5	1	1	4	**93.3**	3001	3838

Subsequently, Table 12 displays the simulation results for the DC of Item I per model. Note that the total replenishment orders represent the fulfilled orders by the DC within the simulation. Therefore, this indicates the demand requests at the DC. Moreover, the total replenishment size is the total demand requested by the depots, from Tables 9, 10 and 11, at the DC. The total amount of orders is identical across all models. However, there is a significant difference in fill rate between the *vNext* model's DC and the DC of the VSL models. The fill rate is about 20% lower compared to the AVSL model and more than 40% lower compared to the SVSL model. Moreover, it is also lower than the set minimum central fill rate β_0^{min} of 60%. This is likely due to the DC of the *vNext* model having a negative reorder point and the positive demand size variance at WL5.

Table 12: Simulation results DC for Item I, $Q_0 = 5$

Model	Reorder point	Achieved fill rate (%)	Total replenishment orders	Total replenishment size
vNext	-1	58.4	4024	5721
SVSL	5	99.8	4024	5721
AVSL	0	78.3	4024	5721

Finally, for Item I, Figures 6 and 7 display the wait time density distributions per location for the SVSL and AVSL models. Note that we exclude the *vNext* model since all depots achieve their fill rate service measure and thus, in general, customers did not experience any wait time. Furthermore, for the DC, the wait time indicates the time depots have to wait before their replenishment order is sent out, whilst for the depots it indicates the wait time experienced by the customers.

Within each density distribution, three lines are drawn that represent the 25%, 50% and 75% quantiles. However, it holds for all distributions that these lines overlap and are concentrated at the peaks, indicating that the densest part of the distribution is within the 75% quantile. This is consistent with the previous results, as all depots achieved a service level of at least 93.3%. The peaks of the density distributions are consistent with their wait time targets for the SVSL model. For the AVSL model, the peaks are concentrated at zero since they mostly have non-negative reorder points. Therefore, they satisfy most customers with the stock on hand.

The dots within the plots indicate the individual observations of the wait time for a customer. Note that darker dots indicate a larger amount of observations. Since the depots of the SVSL model all satisfied their service measure, most observations overlap. Only for the depots with a service level of less than 100%, some observations deviate from the peak. For the AVSL model however, we observe far more deviations. We especially encounter more observations above the wait time target of four days. This is due to a lower central reorder point and therefore depots frequently have to wait upon their replenishment orders, which is indicated by the wait time at the DC.

A right tail of the density distribution only exists for the DC of the AVSL model. We determined that the lead time data of the DC is unreliable, see Section 4.4.4, therefore the DC used a constant lead time of 47 days, i.e. $L_0 = 47$, in both the inventory allocation and simulation. The wait time distribution explains the lower fill rate at the DC for the AVSL model. The wait times at the DC also correlate with the increased wait times at WL5. Together with Figure 2, these results indicate that this allocation for the AVSL model is sub-optimal. Increasing the central reorder point and decreasing the reorder point at WL5, proves to be a better solution as shown by the SVSL model.

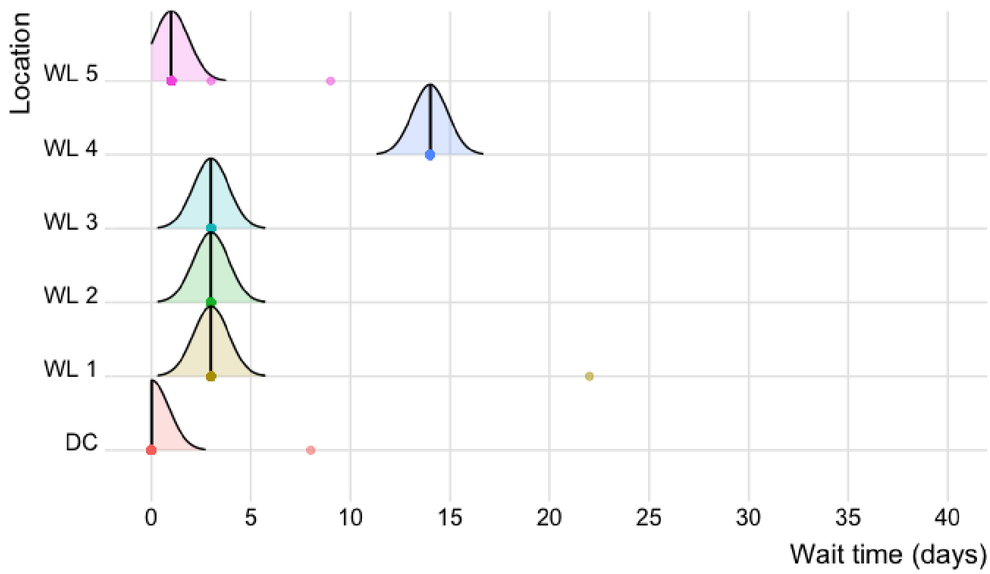

Figure 6: Wait time density distribution per location for the SVSL model of Item I

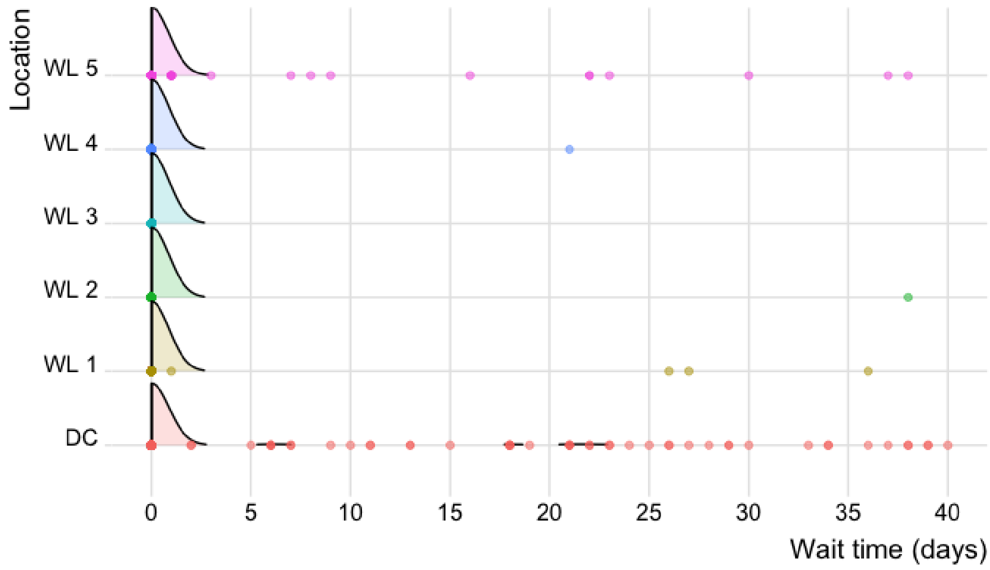

Figure 7: Wait time density distribution per location for the AVSL model of Item I

For Item II, Tables 13, 14 and 15 display the simulation results for the *vNext*, SVSL and AVSL models, respectively. Similarly to Item I, the results of the DC per model are provided in Table 16. Note that this item has both FL and WL depots, i.e. depots with a fill rate service measure and depots with a wait time service measure, respectively. A depot either has a fill rate or wait time service measure. The excluded service measure for a depot is denoted by 'NA'.

The *vNext* and SVSL models struggle with achieving all service measures, and satisfy only 14 out of 21 and 16 out of 21 depots, respectively. Table 13 further indicates that the *vNext* model fails to achieve the service measure of 95% at all of the FL depots. Both models also struggle with depot WL17, implicating that the relatively large DC-depot inter lead time of 14 days and a significant average demand size of 24, see Table 4, challenge the models' capabilities.

For the AVSL model, Table 15 indicates only WL7 and FL1 fail to achieve their service measure. These depots also fail to satisfy their service measure for the SVSL model. It implies that both the VSL models struggle with wait time depots that have a large average demand size and a significant variance in demand size, see Table 4. Contrary to the other two models, the AVSL model is able to satisfy the service measure at WL17, due to the increased central reorder point and the placement of local stock.

Finally, it is interesting that all models fail to achieve the fill rate service measure at FL1. Comparing this depot to the other fill rate depots, again see Table 4, it has relatively low average demand size and variance of the demand size. Therefore, it should be easier for the depot to attain its target. However, the higher relative demand frequency causes complications. FL4 has the highest demand frequency, relatively large average demand size and the greatest variance of the demand size amongst the fill rate depots. Yet, due to only having a DC-depot inter-lead time of one day, it is able to be replenished in time to satisfy its customers for the SVSL and AVSL models.

Table 13: Simulation results of *vNext* model for Item II, $R_0 = -135$

Depot	Reorder point	Order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	15	1	95.0	NA	97.7	988	5458
WL2	15	1	95.0	NA	97.3	1184	5969
WL3	2	1	95.0	NA	**92.5**	1036	2072
WL4	7	1	95.0	NA	98.3	967	3014
WL5	1	1	95.0	NA	95.3	574	1148
WL6	16	2	95.0	NA	99.5	206	3296
WL7	79	1	95.0	NA	97.7	786	21,172
WL8	31	2	95.0	NA	97.0	605	8638
WL9	16	1	95.0	NA	96.9	194	3104
WL10	16	1	95.0	NA	**94.3**	980	9734
WL11	16	1	95.0	NA	98.7	395	4512
WL12	15	1	95.0	NA	96.2	631	5492
WL13	8	1	95.0	NA	99.0	193	1544
WL14	16	1	95.0	NA	95.4	1028	8386
WL15	7	1	95.0	NA	96.2	932	3332
WL16	1	1	95.0	NA	99.0	207	414
WL17	27	3	95.0	NA	**92.3**	544	13,056
FL1	11	3	95.0	NA	**92.4**	15,025	33,087
FL2	28	1	95.0	NA	**93.6**	6758	45,011
FL3	8	1	95.0	NA	**93.1**	3805	13,172
FL4	80	5	95.0	NA	**93.3**	36,677	214,430

Table 14: Simulation results of SVSL model for Item II, $R_0 = 11$

Depot	Reorder point	Order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	NA	4	97.1	988	5458
WL2	-1	1	NA	4	96.2	1184	5969
WL3	-1	1	NA	4	96.9	1036	2072
WL4	-1	1	NA	4	97.1	967	3014
WL5	-1	1	NA	4	96.9	574	1148
WL6	-1	2	NA	4	96.6	206	3296
WL7	-1	1	NA	4	**92.2**	786	21,172
WL8	-1	2	NA	4	**94.4**	605	8638
WL9	-1	1	NA	4	96.4	194	3104
WL10	-1	1	NA	4	95.0	980	9734
WL11	-1	1	NA	4	96.5	395	4512
WL12	-1	1	NA	4	**94.0**	631	5492
WL13	-1	1	NA	4	97.4	193	1544
WL14	-1	1	NA	15	95.3	1028	8386
WL15	-1	1	NA	15	97.0	932	3332
WL16	-1	1	NA	15	97.6	207	414
WL17	-1	3	NA	15	**90.6**	544	13,056
FL1	3	3	95.0	NA	**94.5**	15,025	33,087
FL2	23	1	95.0	NA	96.6	6758	45,011
FL3	5	1	95.0	NA	98.7	3805	13,172
FL4	20	5	95.0	NA	95.1	36,677	214,430

Table 15: Simulation results of AVSL model for Item II, $R_0 = 18$

Depot	Reorder point	Order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	NA	4	97.9	988	5458
WL2	-1	1	NA	4	97.3	1184	5969
WL3	-1	1	NA	4	97.8	1036	2072
WL4	-1	1	NA	4	97.8	967	3014
WL5	-1	1	NA	4	98.3	574	1148
WL6	-1	2	NA	4	96.6	206	3296
WL7	-1	1	NA	4	**93.1**	786	21,172
WL8	-1	2	NA	4	95.9	605	8638
WL9	-1	1	NA	4	96.9	194	3104
WL10	-1	1	NA	4	96.6	980	9734
WL11	-1	1	NA	4	98.2	395	4512
WL12	-1	1	NA	4	95.7	631	5492
WL13	-1	1	NA	4	99.0	193	1544
WL14	15	1	NA	4	98.2	1028	8386
WL15	7	1	NA	4	99.0	932	3332
WL16	1	1	NA	4	99.5	207	414
WL17	24	3	NA	4	98.5	544	13,056
FL1	3	3	95.0	NA	**94.6**	15,025	33,087
FL2	23	1	95.0	NA	96.6	6758	45,011
FL3	5	1	95.0	NA	98.7	3805	13,172
FL4	19	5	95.0	NA	95.0	36,677	214,430

The simulation results for the DC per model of Item II are displayed in Table 16. The most noticeable result here is the difference in fill rate between the *vNext* model and the VSL models, as it is about 30% lower for the *vNext* model. It is caused by the significant difference between the central reorder points. The difference in central reorder points between the SVSL and AVSL model only causes a difference in fill rate of 1.1%. The DC faces an identical number of orders and total order sizes across all models.

Table 16: Simulation results DC for Item II, $Q_0 = 500$

Model	Reorder point	Achieved fill rate (%)	Total replenishment orders	Total replenishment size
vNext	-135	67.0	54,171	406,041
SVSL	11	96.0	54,171	406,041
AVSL	18	97.1	54,171	406,041

Finally, similar to the previous item, Figures 8, 9 and 10 show the wait time density distributions of Item II per depot for the *vNext*, SVSL and AVSL model, respectively. Note that the DC is disregarded as it provided no further useful insights compared to Item I, lower central reorder points increase the wait time for replenishment orders of the depots. Furthermore, it again holds that the wait time indicates the wait time experienced by the customers at each depot. Moreover, for convenience, the individual observations are disregarded as the density distributions provide sufficient insight.

There are two major differences between the wait time distributions of the *vNext* model and the VSL models. Firstly, for all depots of the *vNext* model, the 75% percentile is at zero days. This is obvious as they all must satisfy a fill rate service measure. The 75% percentile for all depots in the SVSL model are located at the DC-depot lead time L_j. On the other hand for the AVSL, the wait time depots with a positive reorder point have their 75% percentile located at zero, whilst for the remaining depots it is again located at L_j.

Secondly, the SVSL and AVSL model have wait time depots that experience relative frequent wait times of 25 days or more. This is indicated by their longer right tails. These extreme wait times occur due to these depots having a substantial average demand size, additionally some also have a large variance in demand size. This proves difficult to handle for the VSL models. For the AVSL, it was able to allow a positive reorder point at WL17. Therefore, the right tail that is clearly visible for the SVSL model, does not occur for the AVSL model. The *vNext* model clearly has much smaller right tails compared to the VSL models, which is due to the positive reorder points at all the depots needed for the fill rate service measure.

Finally, we do not observe any abnormalities in general. The central lead time was deemed to be unreliable, therefore the contractual lead time of 22 days was used as the constant central lead time, see Table 4, in the inventory allocation and the simulation. Hence, $L_0 = 22$, which caused the right tails around the 25 day mark, as this equals L_0 plus the DC-depot inter-lead time L_j of three days. The right tail for WL17 in the SVSL model is around the 36 days as this again equals $L_0 + L_j$, where $L_j = 14$, see Table 4.

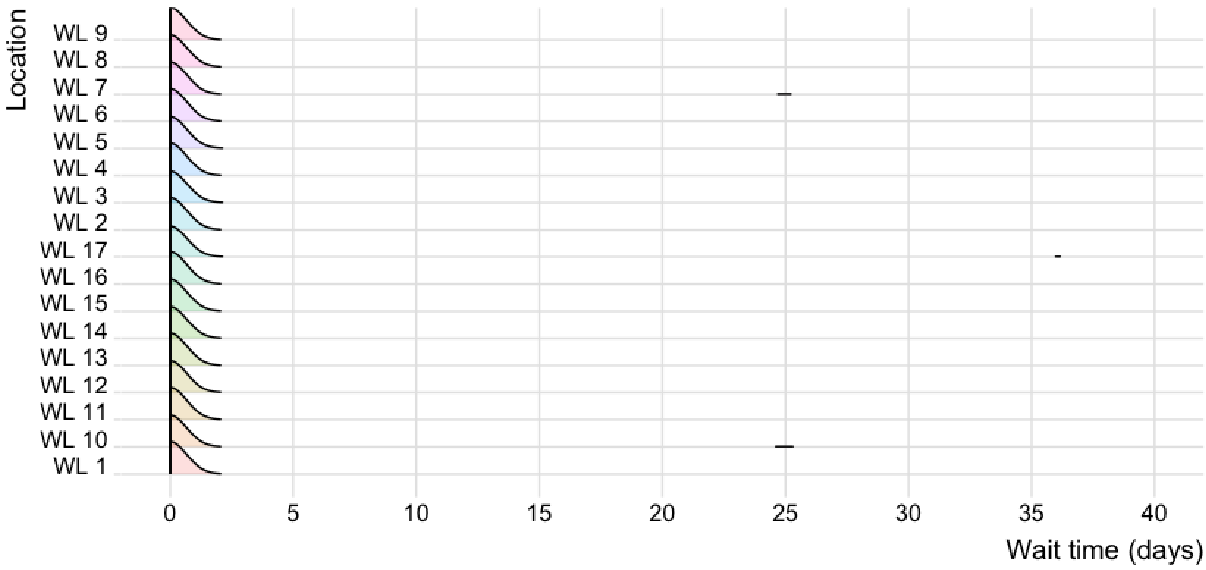

Figure 8: Wait time distribution per location for the *vNext* model of Item II

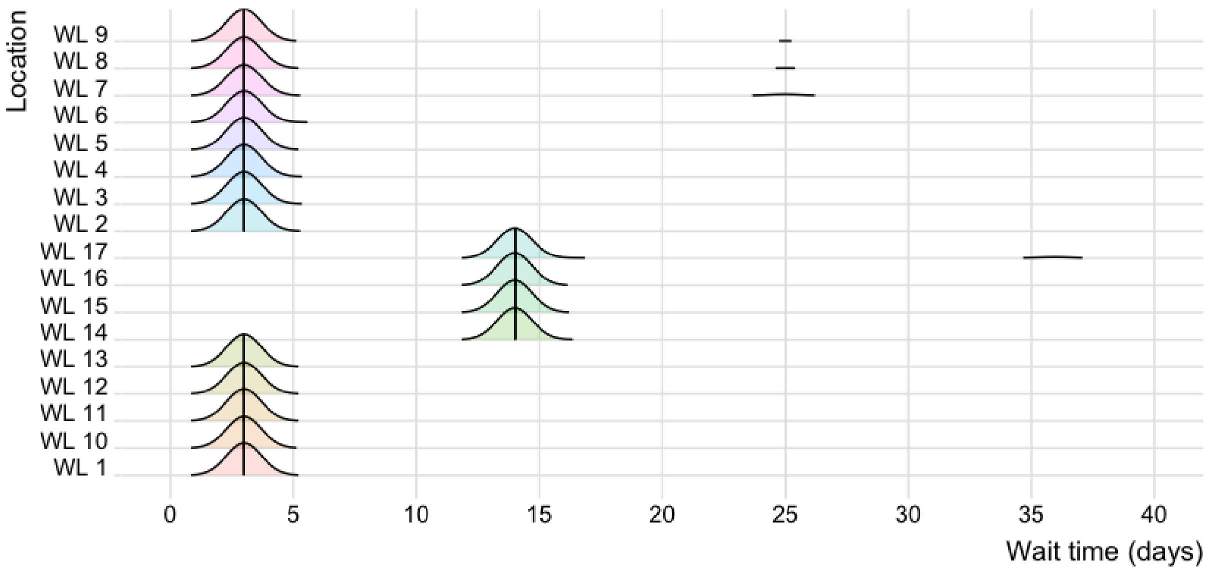

Figure 9: Wait time distribution per location for the SVSL model of Item II

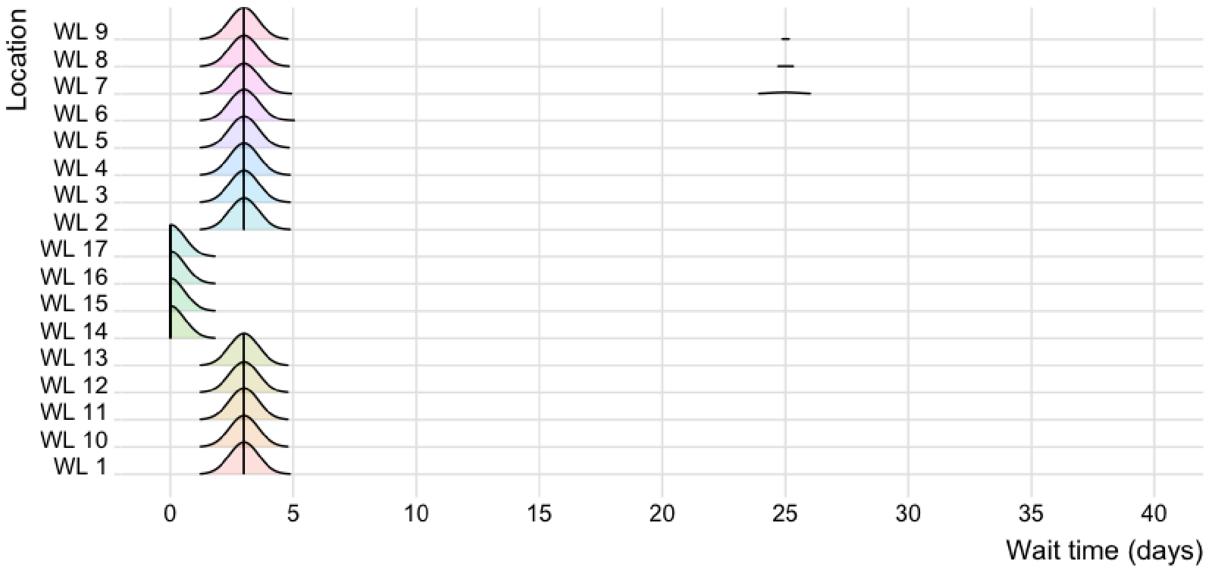

Figure 10: Wait time distribution per location for the AVSL model of Item II

Simulation results for the AVSL model per period

We want to verify the simulation results and provide an insight into the confidence intervals for the achieved service levels by the depots for the AVSL model. Therefore, we separate the total simulation time into five distinct periods of equal length, i.e. 200 years. The length per period provides sufficient time for all depots to experience enough customer arrivals for a fair comparison amongst the periods. Section 4.5.2 described the use of a Wilcoxon Signed-Rank test to determine if there is a significant difference in location between the simulated and historical values. The test requires two dependent samples. Hence, we opt to use this test to determine if the achieved service levels differ significantly between periods.

First, Table 17 displays the achieved service levels for each depot per period for Item I. Moreover, it also presents the 95% confidence intervals and the 2.5% margin of errors used to determine the confidence intervals per depot. To obtain these results, we assumed a normal distribution as it is considered to be the most appropriate. Comparing the periodic results to the simulation results obtained over the entire duration for the AVSL model, see Table 11, we observe similar results. Only WL5 is unable to satisfy its wait time service every single period, this coincides with results over the entire duration. Moreover, the 95% confidence interval of WL5 indicates that it is very unlikely to satisfy its wait time service measure, as the upper bound of the confidence interval is below the 95% probability target of the wait time service measure. The margins of errors across all depots are significantly small, ranging from about 0.0% to 0.6%, indicating that the achieved service levels over the entire simulation duration are reliable.

To compare the achieved service levels across the periods for Item I, Table 18 displays the p-values of the Wilcoxon Signed-Rank test between each pair of periods. The test pairs the achieved service levels of the depots between the different periods, thus we can evaluate whether the difference in location between the different periods over all depots is significant. Note that the results are symmetric as comparing one period with another period achieves the same result as its inverse. Moreover, the test cannot compare a period with itself, thus these results are denoted as ‘NA’. These apply to all similar tables in the remainder of Section 6.2. The p-values implicate that we should reject the alternative hypothesis of the Wilcoxon Signed-Rank test for all periods. Therefore, the achieved service levels at the depots do not differ significantly across the distinct periods. Hence, the obtained results for Item I are considered to be reliable.

Table 17: Periodic simulation results and confidence intervals of Item I for the AVSL model

Depot	Achieved service level (%) in the					95% confidence	2.5% margin
	1^{st} period	2^{nd} period	3^{rd} period	4^{th} period	5^{th} period	interval (%)	of error (%)
WL1	98.9	98.8	100.0	98.7	99.0	[98.9 - 99.3]	0.207
WL2	100.0	100.0	100.0	100.0	97.9	[99.2 - 99.9]	0.365
WL3	100.0	100.0	100.0	100.0	100.0	[100.0 - 100.0]	0.000
WL4	97.0	100.0	100.0	97.2	100.0	[98.2 - 99.5]	0.625
WL5	**93.9**	**92.1**	**93.7**	**93.3**	**93.3**	[**93.0 - 93.6**]	0.274

Table 18: P-values of the Wilcoxon Signed-Rank test comparing the periodic service levels of Item I for the AVSL model

	1^{st} period	2^{nd} period	3^{rd} period	4^{th} period	5^{th} period
1^{st} period	NA	1.000	0.423	0.789	1.000
2^{nd} period	1.000	NA	0.371	0.789	1.000
3^{rd} period	0.423	0.371	NA	0.181	0.181
4^{th} period	0.789	0.789	0.181	NA	0.855
5^{th} period	1.000	1.000	0.181	0.855	NA

Note: The test can not be applied between the n^{th} period with itself. Hence, these are denoted as NA.

For Item II, Table 19 presents the achieved service levels for each depot per period together with their 95% confidence intervals and the 2.5% margin of errors, again we assume a normal distribution to obtain these results. These results mostly coincide with the simulation results over the entire duration displayed Table 15. The results over the entire duration in combination with the computed confidence intervals indicate that it is unlikely for WL7 and FL1 to satisfy their service measures. However, Table 15 indicated that FL4 does suffice its fill rate service measure over the entire simulation duration, yet the confidence interval of FL4 indicates that the lower bound is just below the fill rate target of 95%. The margins or errors diverge more compared to those of Item I and range from about 0.1% to 1.2%. We consider these still to be relatively small and therefore the achieved service levels over the entire simulation duration are deemed to be reliable.

Finally, Table 20 displays the the p-values of the Wilcoxon Signed-Rank test between each pair of periods to compare the achieved service levels across the periods for Item II. Given these p-values we reject the alternative hypothesis of the Wilcoxon Signed-Rank test for all periods, except between first and second period. For these periods, the test indicates that there is a significant difference between the achieved service levels across the depots. However, provided that this only holds for these periods, we accept the simulation results and also conclude that in general the achieved service levels at the depots do not differ significantly across the periods for Item II. Therefore, the obtained results are reliable.

Table 19: Periodic simulation results and confidence intervals of Item II for the AVSL model

Depot	Achieved service level (%) in the					95% confidence	2.5% margin
	1^{st} period	2^{nd} period	3^{rd} period	4^{th} period	5^{th} period	interval (%)	of error (%)
WL1	97.8	98.6	99.4	97.8	96.0	[97.4 - 98.4]	0.488
WL2	97.0	96.9	96.5	98.3	97.9	[97.0 - 97.6]	0.292
WL3	98.7	97.7	98.5	97.4	96.4	[97.4 - 98.1]	0.350
WL4	96.9	99.0	97.9	98.4	96.9	[97.4 - 98.2]	0.375
WL5	98.3	98.1	97.4	98.3	99.2	[98.0 - 98.5]	0.248
WL6	97.4	95.1	97.2	100.0	**93.2**	[95.6 - 97.6]	1.009
WL7	**91.9**	**93.0**	**93.2**	97.0	**90.2**	[**92.1 - 94.1**]	0.978
WL8	96.4	96.9	**94.6**	97.3	**94.4**	[95.4 - 96.4]	0.519
WL9	**93.5**	**94.7**	100.0	97.9	100.0	[96.0 - 98.4]	1.178
WL10	**93.5**	96.6	96.4	97.8	99.0	[95.9 - 97.5]	0.799
WL11	97.5	98.6	100.0	96.4	98.8	[97.7 - 98.8]	0.534
WL12	95.9	98.1	**94.4**	**93.6**	97.0	[95.1 - 96.5]	0.728
WL13	97.1	100.0	100.0	97.7	100.0	[98.4 - 99.5]	0.572
WL14	98.5	97.6	99.6	97.2	98.4	[97.9 - 98.6]	0.355
WL15	98.9	98.9	99.5	99.5	98.4	[98.9 - 99.2]	0.175
WL16	100.0	100.0	100.0	100.0	97.4	[99.0 - 99.9]	0.461
WL17	**94.7**	98.4	100.0	100.0	99.1	[97.6 - 99.3]	0.864
FL1	95.0	**94.9**	**94.3**	**94.7**	**94.2**	[**94.5 - 94.8**]	0.134
FL2	96.7	97.3	97.1	96.1	96.0	[96.4 - 96.9]	0.221
FL3	98.1	98.7	99.0	99.1	98.7	[98.6 - 98.9]	0.155
FL4	**94.7**	95.2	**94.8**	95.5	**94.9**	[**94.9** - 95.2]	0.125

Table 20: P-values of the Wilcoxon Signed-Rank test comparing the periodic service levels of Item II for the AVSL model

	1^{st} period	2^{nd} period	3^{rd} period	4^{th} period	5^{th} period
1^{st} period	NA	**0.031**	0.054	0.097	0.780
2^{nd} period	**0.031**	NA	0.533	0.614	0.173
3^{rd} period	0.054	0.533	NA	0.984	0.121
4^{th} period	0.097	0.614	0.984	NA	0.412
5^{th} period	0.780	0.173	0.121	0.412	NA

Note: The test can not be applied between the n^{th} period with itself. Hence, these are denoted as NA.

Using the global minimum for the AVSL model

As described in Section 6.1.1, the AVSL model cannot determined an optimal solution due to the bisection-based optimisation method. Therefore, we will compare the simulation results of the determined solution, as given in Tables 11 and 15, to the solution at the global minimum, which is the optimal solution according to the AVSL model. Hence, we want to determine if the inventory allocation at the global minimum also provides the best performance through simulation.

Table 21 displays the simulation results for the DC of both items at the global minimum, depicted in Section 6.1.1. For both items, the number of total orders and total order size remains unchanged. However, for Item I, the fill rate increases by about 20% which naturally follows from the significant increase of the central reorder point. The contrary is true for Item II, the central reorder point decreased slightly and therefore the fill rate decreased by 1.1%. These results are compared to those in Tables 12 and 16, respectively.

Table 21: Simulation results DC for Item I and II of the AVSL model at global minimum

Model	Reorder point	Order quantity	Achieved fill rate (%)	Total replenishment orders	Total replenishment size
Item I	5	5	99.8	4024	5721
Item II	11	500	96.0	54,171	406,041

For Item I, Table 22 displays the simulation results for the solution at the global minimum. First, comparing the reorder points between the two different solutions, see Table 11, we notice that all depots now have a reorder point of -1, except for WL4. Due to the DC-depot inter-lead time of 14 days, WL4 still has a reorder point of one as it is otherwise not able to satisfy its wait time target of four days. This also explains the difference between the global minimum of the SVSL and AVSL models in Figure 2. Finally, for this solution, the AVSL model is able satisfy all the service measures. Therefore, increasing the central reorder point and decreasing the reorder points for which it holds that $L_j < WT_j^T$, proves to be the best solution.

Table 22: Simulation results of AVSL model at global minimum for Item I, $R_0 = 5$

Depot	Reorder point	Order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	4	99.8	447	894
WL2	-1	1	4	100.0	224	448
WL3	-1	1	4	100.0	187	187
WL4	1	1	4	99.4	177	354
WL5	-1	1	4	99.9	3001	3838

Table 23 displays the simulation results for the solution at the global minimum of Item II. Following Figure 4, the only difference between the reorder points is the decrease of the central reorder point. All local reorder points remain unchanged. Therefore, comparing these results to Table 15, all service levels have either decreased or remained unchanged. Furthermore, due to these decreasing, WL8 and WL12 now fail to meet their service measure, as was also the case for the SVSL model in Table 14. Therefore, it seems that the non-optimal solution performs better than the global minimum due to the higher central reorder point. Hence, these results indicate that the optimal central reorder point, as determined by the AVSL model, is too low for all depots to attain their service measures. It suggests that during optimisation the DC does not anticipate the correct replenishment order sizes over its lead time. Since the average demand sizes are significantly larger than the order quantities at most depots, see Table 4. We will further analyse and elaborate on this issue in Section 6.3.

Table 23: Simulation results of AVSL model at global minimum for Item II, $R_0 = 11$

Depot	Reorder point	Order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	NA	4	97.1	988	5458
WL2	-1	1	NA	4	96.2	1184	5969
WL3	-1	1	NA	4	96.9	1036	2072
WL4	-1	1	NA	4	97.1	967	3014
WL5	-1	1	NA	4	96.9	574	1148
WL6	-1	2	NA	4	96.6	206	3296
WL7	-1	1	NA	4	**92.2**	786	21,172
WL8	-1	2	NA	4	**94.4**	605	8638
WL9	-1	1	NA	4	96.4	194	3104
WL10	-1	1	NA	4	95.0	980	9734
WL11	-1	1	NA	4	96.5	395	4512
WL12	-1	1	NA	4	**94.0**	631	5492
WL13	-1	1	NA	4	97.4	193	1544
WL14	15	1	NA	15	98.2	1028	8386
WL15	7	1	NA	15	99.0	932	3332
WL16	1	1	NA	15	99.5	207	414
WL17	24	2	NA	15	98.5	544	13,056
FL1	3	2	95.0	NA	**94.5**	15,025	33,087
FL2	23	1	95.0	NA	96.6	6758	45,011
FL3	5	1	95.0	NA	98.7	3805	13,172
FL4	20	5	95.0	NA	95.1	36,677	214,430

6.2.2 All items

The simulation results over all items for the *vNext*, SVSL and AVSL models are shown in Tables 24, 25 and 26, respectively. For all tables, the minimum (Min.), average (Avg.) and maximum (Max.) are determined per item, whilst the final column (Total) is the summation over all items. Again, a fill rate or wait time service measure is achieved when a depot has a service level of 95%.

The tables indicate that the SVSL model experiences difficulties in achieving the wait time service measures, whilst the AVSL model experiences difficulties with the fill rate service measures. The vNext performs best on the depot level, 80.4% of all depots satisfy their service measure. The SVSL model performs worst, with only 76.3% of all depots achieving their service measure. For the AVSL model, 77.1% of all depots achieve their target. The AVSL model also outperforms the SVSL model when only considering the WL depots, as 82.0% of all wait time service measures are achieved. The SVSL model only outperforms the AVSL model regarding solely the FL depots by 28.8%. These results partially contradict the simulation results of the individual items, as there the AVSL model performed best. The DC fill rates are entirely in line with the individual item results. The fill rates for the VSL models are, on average, about 15% higher compared to *vNext* model.

The average service levels for the *vNext*, SVSL and AVSL models are 97.0%, 95.1% and 95.4%, respectively. This indicates that all models do perform adequately. Moreover, the depots that fail their service measure, generally miss their target only slightly. The FL and WL depots separately achieve an average service level of 95.9% and 94.5% for the SVSL model. For the AVSL model, the FL depots attain an average service level of 94.4% and the WL depots of 95.6%.

The *vNext* model performed better in general compared to the VSL models. However, it does allocate significantly more inventory, see Section 6.1.2, to achieve its 3.3% better performance across all depots. Moreover, as indicated by the results for the individual items in Section 6.1.1, the AVSL model in general achieves a non-optimal solution. That said, the results in Section 6.2.1 indicated that the optimal solution can perform worse than the non-optimal solution due to it providing a lower central reorder point. Therefore, it is crucial to determine the adequate central reorder point as these significantly affect the service measures attained at the depots. Finally, considering these results, the VSL models, particularly the AVSL model, do not underperform compared to the baseline, the *vNext* model.

Table 24: Simulation results of all items for the *vNext* model

	Min.	Avg.	Max.	Total
Customers	167	10,618.987	8,517,494	105,595,202
Demand	167	35,116.947	17,998,964	349,202,923
FL depots	1	3.786	26	37,652
FL depots achieved service measure	0	3.044	21	30,270
WL depots	0	0.000	0	0
WL depots achieved service measure	0	0.000	0	0
Fill rate DC (%)	0.0	66.5	100.0	NA

Table 25: Simulation results of all items for the SVSL model

	Min.	Avg.	Max.	Total
Customers	167	10,618.987	8,517,494	105,595,202
Demand	167	35,116.947	17,998,964	349,202,923
FL depots	1	0.521	26	5205
FL depots achieved service measure	0	0.339	20	3386
WL depots	1	3.246	24	32,447
WL depots achieved service measure	0	2.536	24	25,350
Fill rate DC (%)	0.0	90.9	100.0	NA

Table 26: Simulation results of all items for the AVSL model

	Min.	Avg.	Max.	Total
Customers	167	10,618.987	8,517,494	105,595,202
Demand	167	35,116.947	17,998,964	349,202,923
FL depots	1	0.521	26	5205
FL depots achieved service measure	0	0.241	14	2411
WL depots	1	3.246	24	32,447
WL depots achieved service measure	0	2.662	24	26,602
Fill rate DC (%)	0.0	79.4	100.0	NA

Verifying the simulated demand statistics

To verify the simulation parameters we perform Wilcoxon Signed-Rank tests, as described in Section 4.5.2. Table 27 displays the demand statistics for the simulation, which we compare to the actual demand statistics in Table 2. Moreover, the table also displays the results of the performed tests, describing the test statistic, p-value and confidence interval per variable. The test is performed on the actual values versus the simulated values, which is important as the test uses the difference between these paired values.

The test results indicate that the null hypothesis can be rejected for all variables with 99% certainty, as all p-values are smaller than 0.01. Hence, the distribution differences between the actual and simulation values are not symmetric around zero and, therefore, differ significantly in location, implicating that the simulation variables do correctly portray the actual values. However, upon inspecting the 95% confidence intervals, the results indicate that the absolute differences are only slight. Comparing the simulated averages to the actual averages in Table 2, only a difference of small margin is observed considering the vast number of depots in the distribution system. The confidence interval indicates that the average inter-arrival times only differs about one day. Moreover, as the tests compared the actual versus the simulated values, and as shown by the simulated demand statistics, the simulation actually provided a slightly increases demand frequency.

The opposite holds for the average demand size and variance of demand size, the actual values exceed the simulated ones. As mentioned in Section 4.5.2, this is due to the use of the empirical distribution for the demand sizes and therefore these results follow our expectation as continuous sampling from deterministic values is expected to decrease the variance.

The results per type of depot are disregarded, as these do not differ in any regard from the overall results, i.e. all null hypotheses are rejected. Thus, there is a significant difference in location between the simulated and actual values even per type of depot. However, similar to the overall results, the differences are minimal. Therefore, even though the tests indicate a significant difference, we accept the simulation as valid as the differences are minimal and indicate that the distribution systems experienced increased amounts of demand during the simulation period.

Table 27: Simulated demand statistics and Wilcoxon Signed-Rank test statistics

	Min.	Avg.	Max.	Test statistic	p-value	95% confidence interval
Demand frequency (yearly)	0.167	2.804	4087.852	306,229,049	**0.000**	[-0.003, -0.003]
Avg. inter-arrival time (days)	0.089	936.301	2186.228	380,287,975	**0.000**	[0.500, 0.840]
Avg. demand size	1.000	2.485	1481.000	38,389,520	**0.000**	[-0.004, -0.003]
Var. demand size	0.000	19.080	63,921.100	85,181,577	**0.000**	[0.233, 0.254]

Simulation results for the AVSL model per period

Finally for the simulation results, we want to verify the use our starting set-up for our simulation, as discussed in Section 4.5.2 and presented in Section 6.2. Moreover, we want to provide an overall insight into the confidence intervals for the achieved service levels per type of depot for the AVSL model. Therefore, similarly to individual items in Section 6.2.1, we split the total simulation duration into five distinct periods of equal length, i.e. 200 years. Moreover, we again make use of a Wilcoxon Signed-Rank test to determine if there is a significant difference between the achieved service levels of the distinct periods.

Let us first analyse the achieved service levels per type of depot per period, the 95% confidence intervals and the 2.5% margin of errors per type of depot, all presented in Table 28. Note that we again assume a normal distribution to obtain the confidence intervals as, similarly to Section 6.2.1, this is considered to be most appropriate. We observe that on average all depots achieve their service measure per period. Moreover, the average achieved service level remains constant over the different periods. However, the results do show that on average the FL depots fail to achieve their fill rate service measure, contrary to the WL depots with their wait time service measure.

Furthermore, the lower bound of the 95% confidence intervals is above the service measure target of 95.0%, indicating that about 95% of the depots should satisfy their service measure as they should be contained within the confidence interval. This seems to be a discrepancy with the result in Table 26, as it indicated that only 77.1% of depots satisfy their service measure. However, upon investigating the quantiles of the achieved service levels, the 24.0% quantile is at exactly 95.0% and thus confirming our earlier findings. The margin of errors are on average significantly small. However, we do note a significant difference between the FL and WL depots, as the average margin of errors for the WL depots is almost twice that of the FL depots. In general, the results indicate that the achieved service levels over the entire simulation duration are reliable and therefore displayed service levels throughout Section 6 are accurate.

Finally, we compare the achieved service levels across the periods through Table 18, displaying the p-values of the Wilcoxon Signed-Rank test for each pair of distinct periods. Note that, similarly to Section 6.2.1, the results are symmetric and the results denoted as 'NA' are due to the test not being able to compare a period with itself. Given the p-values, we should reject the alternative hypothesis of the Wilcoxon Signed-Rank test for all periods, with the exception of period 4 with periods 1 through 3. The test indicates that period 4 seems to have significant differences in achieved service levels with all remaining periods except period 5. However, given that the first period only differs significantly from the fourth period, the provided starting set-up of the simulation provided a fair and unbiased start to the simulation and did not significantly affect the results. Moreover, given that the depots do not differ significantly across most periods, we deem the obtained service levels across the entire distribution system to be reliable and do not depend on randomness.

Table 28: Statistics of the periodic simulation results for the AVSL model

	Depots	Min.	Avg.	Max.
Achieved service level (%) 1^{st} period	All	**0.0**	95.4	100.0
	FL	**0.0**	**94.4**	100.0
	WL	**0.0**	95.6	100.0
Achieved service level (%) 2^{nd} period	All	**0.0**	95.4	100.0
	FL	**0.0**	**94.4**	100.0
	WL	**0.0**	95.5	100.0
Achieved service level (%) 3^{rd} period	All	**0.0**	95.4	100.0
	FL	**0.0**	**94.5**	100.0
	WL	**0.0**	95.5	100.0
Achieved service level (%) 4^{th} period	All	**0.0**	95.4	100.0
	FL	**0.0**	**94.4**	100.0
	WL	**0.0**	95.6	100.0
Achieved service level (%) 5^{th} period	All	**0.0**	95.4	100.0
	FL	**0.0**	**94.5**	100.0
	WL	**0.0**	95.6	100.0
95% confidence interval (%)	All	[**0.0 - 0.0**]	[95.0 - 95.8]	[100.0 - 100.0]
	FL	[**0.0 - 0.0**]	[**94.2 - 94.7**]	[100.0 - 100.0]
	WL	[**0.0 - 0.0**]	[95.1 - 96.0]	[100.0 - 100.0]
2.5% margin of error (%)	All	0.000	0.437	6.351
	FL	0.000	0.258	1.519
	WL	0.000	0.465	6.351

Table 29: P-values of the Wilcoxon Signed-Rank test comparing the periodic service levels for the AVSL model

	1^{st} period	2^{nd} period	3^{rd} period	4^{th} period	5^{th} period
1^{st} period	NA	0.650	0.304	**0.003**	0.089
2^{nd} period	0.650	NA	0.542	**0.000**	0.268
3^{rd} period	0.304	0.542	NA	**0.040**	0.346
4^{th} period	**0.003**	**0.000**	**0.040**	NA	0.497
5^{th} period	0.089	0.268	0.346	0.497	NA

Note: The test can not be applied on the n^{th} periods themselves. Hence these are denoted as NA.

6.3 Tests on robustness

In this section we analyse robustness tests performed on the AVSL model and identify the important parameters that influence its performance. Furthermore, we want to analyse why the performances of the AVSL model are sometimes subpar. This section follows the order of Section 4.5.3, in which all tests and expectations are discussed. Any adjustments made on the data or model parameters compared to the baseline results in Section 6.1 and 6.2, are indicated per part of the analysis.

6.3.1 Different wait time distributions

As indicated in Section 4.5.3, we want to test a heavier-tailed wait time distribution and determine its effect on the inventory allocation. The *Gamma* distribution is assumed to be the standard distribution of the wait time as indicated in Section 4.4.1. Therefore, we compare its performance to that of the *LogNormal* distribution using the AVSL model for Items I and II.

Figures 11 and 12 display the optimisation range and inventory allocation of Item I, respectively, for the different wait time distributions. The optimisation ranges indicate only a slight deviation at the lower bound of the optimisation domain. Here, the *LogNormal* distribution provides a greater total stock, i.e. the sum over all reorder points. However, the AVSL model achieves an identical total stock for both wait time distributions at different non-optimal solutions, i.e. the highlighted points. It is noteworthy that both distributions produce an almost identical relation between the total stock and central reorder point, which is likely due to the integrality of the reorder points.

The second figure depicts the difference between the determined solutions. For the *Gamma* distribution, a $R_0 = 0$ is proposed, while an $R_j = 1$ is proposed for WL5. The opposite holds for the *LogNormal* distribution. Hence, the wait time distribution did not significantly influence the inventory allocation results of Item I.

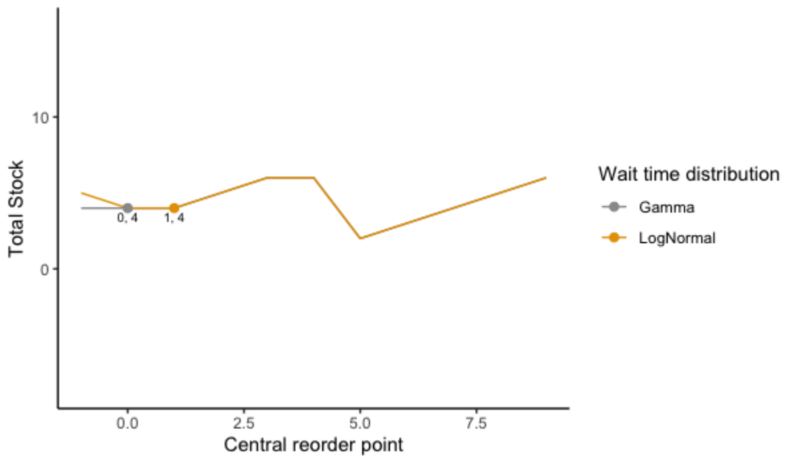

Figure 11: Total stock versus central reorder point for Item I of the AVSL model per wait time distribution

Note: The highlighted points indicate the determined solution per wait time distribution

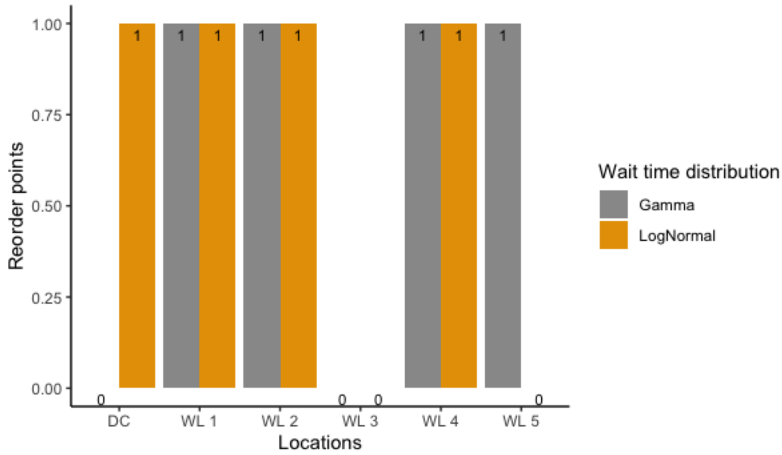

Figure 12: Reorder points at all warehouses for Item I of the AVSL model per wait time distribution

For Item II, Figures 13 and 14 depict the optimisation range and inventory allocation of Item II for the different wait time distributions, respectively. Contrary to Item I, here the first figure does indicate a significant change in the relation between the total stock and central reorder point for the *LogNormal* distribution compared to the *Gamma* distribution. The deviation only occurs on a sub-space of the domain as the choice of distribution only affects the optimisation on this sub-space. On the sub-space, the *LogNormal* distribution behaves as expected. It allocates more stock locally and only after the central reorder point has significantly increased compared to the *Gamma* distribution, is it able to reduce the local reorder points. However, both wait time distributions produce the identical non-optimal solution, as the *LogNormal* distribution synchronises with the *Gamma* distribution around the global minimum.

Due to both wait time distributions producing the identical solution, the second figure displays the obvious identical inventory allocation across all warehouses. As previously stated in Section 6.1, the bisection-based optimisation approach causes the proposed solution to be non-optimal as the relation between the total stock and central reorder point is not unimodal.

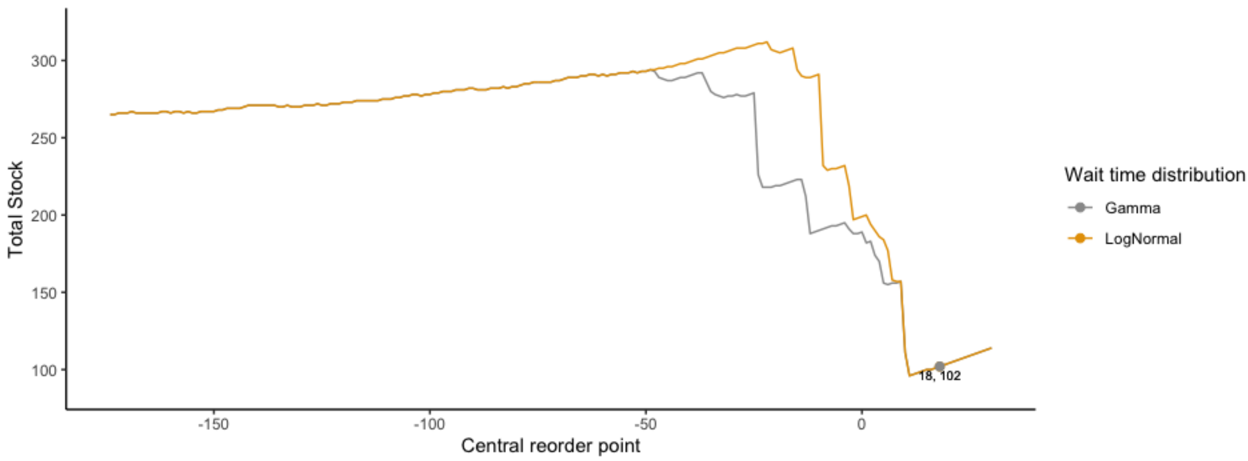

Figure 13: Total stock versus central reorder point for Item II of the AVSL model per wait time distribution

Note: The highlighted points indicate the determined solution per wait time distribution

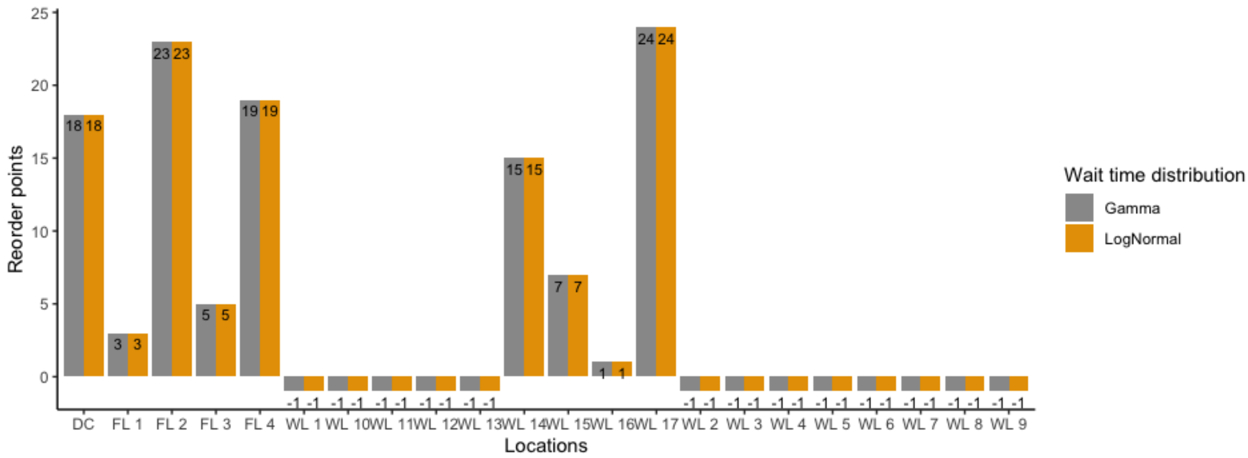

Figure 14: Reorder points at all warehouses for Item II of the AVSL model per wait time distribution

6.3.2 Different wait time targets

The AVSL model is developed to be flexible in its selection of wait time targets. Therefore, we want to test its performance using different wait time targets on Items I and II. The baseline presented in Sections 6.1 and 6.2 uses a wait time target of four days, i.e. $WT_j^T = 4$ for all $j \in W_i$. We want to analyse the case in which almost all locations have a DC-depot inter-lead time greater than its set wait time target, i.e. $L_j > WT_j^T$, and the case in which the wait time target is significantly larger than most inter-lead times. Therefore, we will analyse the cases in which it holds that $WT_j^T = 2$ and $WT_j^T = 7$ for all $j \in W_i$. With these cases we want to evaluate the AVSL model's flexibility in its wait time target selection.

Figure 15 and 16 display the optimisation range and inventory allocation results for Item I, respectively. They indicate that assigning a larger wait time target decreases the total stock, which is as expected. However, decreasing the wait time target does not lead to an increase in total stock compared to the original solution of the AVSL model. The optimisation range does show that with $WT_j^T < L_j$, the AVSL model prefers to stock locally.

The inventory allocation further highlights these findings. For a wait time target of seven days, the AVSL model only stocks the DC and WL4, the depot for which it holds that $L_j = 14$. These results are unsurprising and thus follow our expectation.

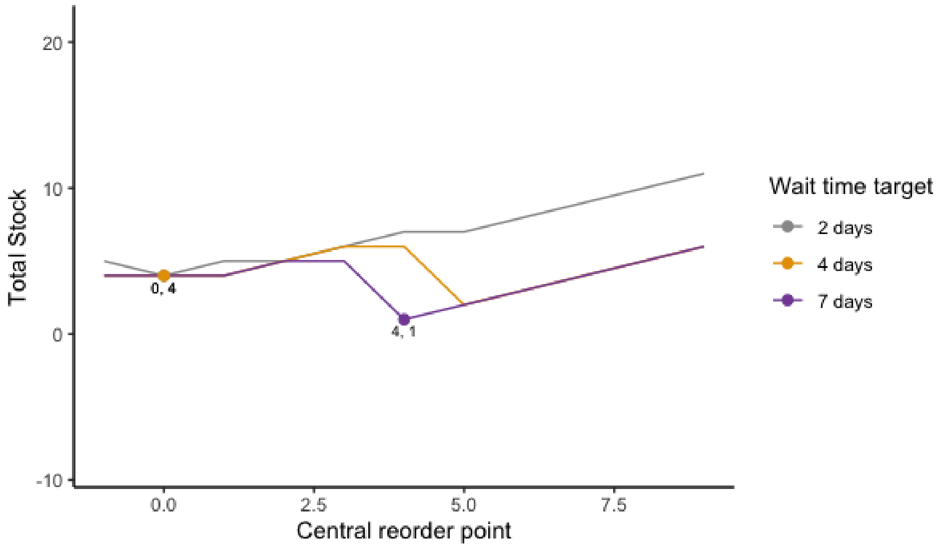

Figure 15: Total stock versus central reorder point for Item I of the AVSL model per wait time target

Note: The highlighted points indicate the determined solution per wait time target

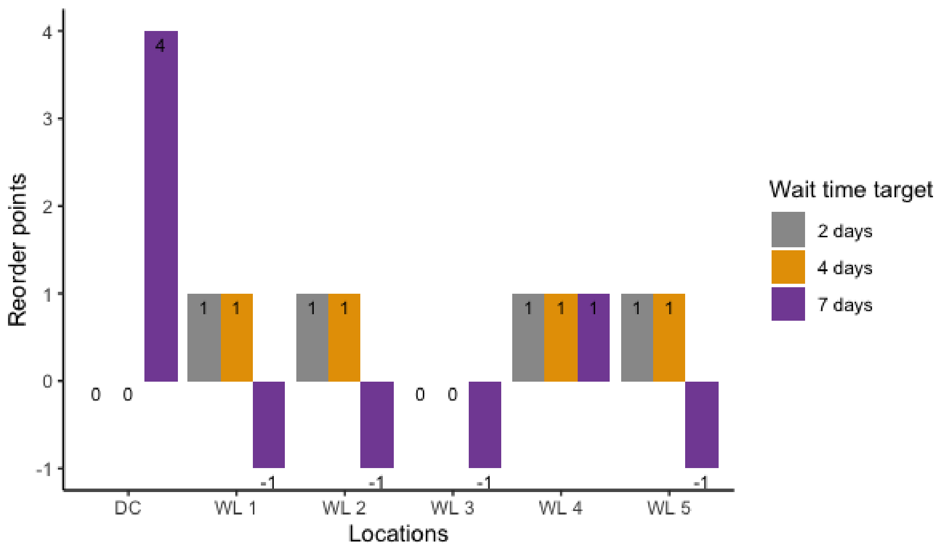

Figure 16: Reorder points at all warehouses for Item I of the AVSL model per different wait time target

The simulation results for Item I are displayed in Tables 30 and 31 of the AVSL model with $WT_j^T = 2$ and $WT_j^T = 7$, respectively. As previously indicated, these results are compared to Table 11, which displays the simulation results for Item I results with a four-day wait time target. The results between a wait time target of two and four days are nearly identical. Both fail to meet the target at WL5. Whilst, unsurprisingly, using a wait time target of seven days allows all depots to satisfy their wait time service measure.

Table 30: Simulation results for Item I of AVSL model with $WT_j^T = 2$, $R_0 = 0$

Depot	Reorder point	Order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	1	1	2	99.1	447	894
WL2	1	1	2	99.6	224	448
WL3	0	1	2	100.0	187	187
WL4	1	1	2	98.9	177	354
WL5	1	1	2	**93.1**	3001	3838

Table 31: Simulation results for Item I of AVSL model with $WT_j^T = 7$, $R_0 = 4$

Depot	Reorder point	Order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	7	98.4	447	894
WL2	-1	1	7	99.1	224	448
WL3	-1	1	7	95.5	187	187
WL4	1	1	7	100.0	177	354
WL5	-1	1	7	99.5	3001	3838

For Item II, Figures 17 and 18 display the optimisation range and inventory allocation results of the AVSL model per wait time target. The optimisation ranges implicate similar results compared to Item I. However, an important observation here is that for a wait time target of two days the AVSL model does not consider stocking centrally across entire optimisation range. This is observed as the relation between the total stock and central reorder point is almost monotonically increasing for the entire domain. We expected that using $WT_j^T < L_j$ would lead to an increase in local stock. However, with a large enough central reorder point it was expected for the AVSL to still determine a solution in which not every depot has to be stocked. Instead, the proposed solution is at the lower bound for the AVSL model with $WT_j^T = 2$. For the seven-day wait time target, the results are identical to those of Item I.

The inventory allocation results show that for a two-day wait time target, all depots are stocked. For $WT_j^T = 7$, the allocation is almost identical to the initial results with a four-day wait time target, with the only difference being the central reorder point and the reorder point at FL4. Finally, if we compare the inventory allocation results of the AVSL model with a wait time target of two days to the *vNext* model for Item II, see Figure 5, we do see a decrease in total stock. The central reorder point is significantly decreased. Moreover, due to the wait time service measure at the wait time depots, the local reorder points of the fill rate depots do not increase that drastically and leads to a decrease in total stock.

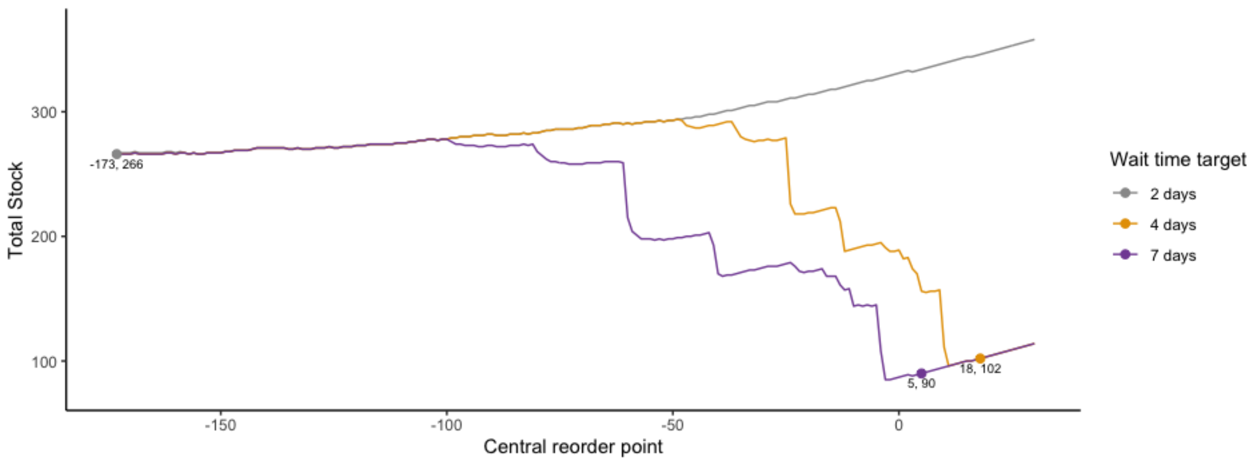

Figure 17: Total stock versus central reorder point for Item I of the AVSL model per wait time target
Note: The highlighted points indicate the determined solution per wait time target

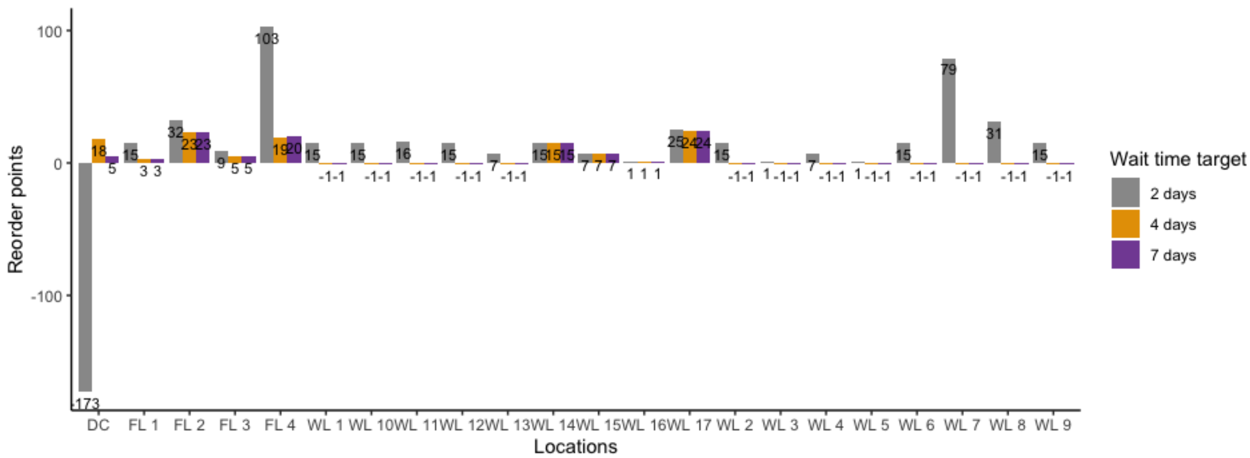

Figure 18: Reorder points at all warehouses for Item II of the AVSL model per different wait time target

Tables 32 and 33 display the simulation results of Item II for the AVSL model with $WT_j^T = 2$ and $WT_j^T = 7$, respectively. Again, we compare these to the initial results for the AVSL model with a four-day wait time target, see Table 15. The results indicate that setting a four-day wait time provides the best performance for Item II. It allows most depots to achieve their service measure, with the exception of WL7 and FL1, see Section 6.1.1.

The two-day wait time target seems to be unattainable as 9 out of 21 depots are unable to achieve their either their fill rate or wait time service measure. WL3 and WL5 are the depots with the lowest expected and varying demand sizes, see Table 4, whilst WL17 is one of the depots with the highest expected demand size. Moreover, all the FL depots fail to meet their service measure. The results imply that due to the lower wait time target, the AVSL model stocks locally. However, the central reorder point is reduced too drastically to replenish the depots on time. Therefore, this mainly affects the FL depots and the wait time depots with smaller demand sizes as they have a lower likelihood of triggering a replenishment order at the DC.

For the seven-day wait time target, the results are unexpected. One expects that increasing the number of days customers are allowed to wait increases the likelihood of depots attaining their service measures. However, the contrary yields as, compared to the initial simulation results, more depots fail to achieve their service measure. The WL depots that perform subpar have significant variance in demand size, compared to the remaining ones. For the FL depots, FL1 and FL4 fail to satisfy their service measure and have the greatest demand frequency.

Table 32: Simulation results for Item II of AVSL model with $WT_j^T = 2$, $R_0 = -173$

Depot	Reorder point	Order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	15	1	NA	2	96.9	988	5458
WL2	15	1	NA	2	96.9	1184	5969
WL3	1	1	NA	2	**91.0**	1036	2072
WL4	7	1	NA	2	98.0	967	3014
WL5	1	1	NA	2	**93.6**	574	1148
WL6	15	2	NA	2	97.6	206	3296
WL7	79	1	NA	2	96.9	786	21,172
WL8	31	2	NA	2	96.9	605	8638
WL9	15	1	NA	2	96.4	194	3104
WL10	15	1	NA	2	**92.7**	980	9734
WL11	16	1	NA	2	98.2	395	4512
WL12	15	1	NA	2	95.6	631	5492
WL13	7	1	NA	2	98.4	193	1544
WL14	15	1	NA	2	**94.4**	1028	8386
WL15	7	1	NA	2	95.6	932	3332
WL16	1	1	NA	2	97.1	207	414
WL17	25	3	NA	2	**91.9**	544	13,056
FL1	15	3	95.0	NA	**92.6**	15,025	33,087
FL2	32	1	95.0	NA	**93.1**	6758	45,011
FL3	9	1	95.0	NA	**92.9**	3805	13,172
FL4	103	5	95.0	NA	**93.7**	36,677	214,430

Table 33: Simulation results for Item II of AVSL model with $WT_j^T = 7$, $R_0 = 5$

Depot	Reorder point	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand	
WL1	-1	1	NA	7	96.5	988	5458
WL2	-1	1	NA	7	95.9	1184	5969
WL3	-1	1	NA	7	96.5	1036	2072
WL4	-1	1	NA	7	97.5	967	3014
WL5	-1	1	NA	7	96.3	574	1148
WL6	-1	2	NA	7	96.6	206	3296
WL7	-1	1	NA	7	**91.9**	786	21,172
WL8	-1	2	NA	7	**94.2**	605	8638
WL9	-1	1	NA	7	95.4	194	3104
WL10	-1	1	NA	7	**94.8**	980	9734
WL11	-1	1	NA	7	**93.2**	395	4512
WL12	-1	1	NA	7	**93.2**	631	5492
WL13	-1	1	NA	7	97.4	193	1544
WL14	15	1	NA	7	98.4	1028	8386
WL15	7	1	NA	7	99.1	932	3332
WL16	1	1	NA	7	100.0	207	414
WL17	24	3	NA	7	99.3	544	13,056
FL1	3	3	95.0	NA	**94.4**	15,025	33,087
FL2	23	1	95.0	NA	96.5	6758	45,011
FL3	5	1	95.0	NA	98.7	3805	13,172
FL4	20	5	95.0	NA	**94.8**	36,677	214,430

Therefore, these findings implicate that the performance of the AVSL model is significantly affected by the central reorder point. However, it is evident that the AVSL model tends to excessively reduce the central reorder point due to insufficient consideration of incoming replenishment order sizes, thus negatively impacting the service levels at the depots.

6.3.3 Alteration of demand frequency and size

The influence of the demand frequency and size on results on the VSL models has been mentioned throughout the the current section. To further investigate their effect on the inventory allocation and performance of the AVSL model, we adjust the demand frequency and size of WL2 for Item I, see Table 34, whilst conserving its initial order quantity. The new demand frequency is 1.000, i.e. the depot now experiences one customer arrival per year. The new average demand size is set to 20. Note that we distinguish between two demand cases. Namely, case 1 where all other parameters remain unchanged and thus WL2 has no demand size variance. However, for case 2, we also add significant demand size variance at WL2. These separate cases aid us in highlighting the difficulties that arise in the AVSL model.

Table 34: Demand statistics for WL2 of Item I with adjusted demand frequency and sizes

	Contractual lead time (days)	Order quantity	Avg. time between demand (days)	Demand frequency (per year)	Avg. demand size	Var. demand size
Initial WL2	3	1	1825.000	0.200	2.000	-
Case 1 WL2	3	1	365.000	1.000	20.000	-
Case 2 WL2	3	1	365.000	1.000	20.000	312.500

Figure 19 displays the inventory allocation results for the AVSL model with the initial demand and two demand cases at WL2. Naturally, the central reorder point experiences a significant increase. However, for case 1, with no demand size variance, the increase seems to be insufficient. The inventory position at the DC will never equal 20, the average demand size at WL2. Therefore, it will be impossible to replenish WL2 within the set wait time. For case 2, the increase is sufficient to satisfy replenishment orders from WL2 as the demand size variance causes replenishment to be within or below the domain of inventory position. Due to the increase of the central reorder point for both cases, all local reorder points drop to -1. Except for WL4 as it has a DC-depot inter-lead time of 14 days, causing the reorder point to remain unchanged compared to the solution of the AVSL model with initial demand parameters at WL2.

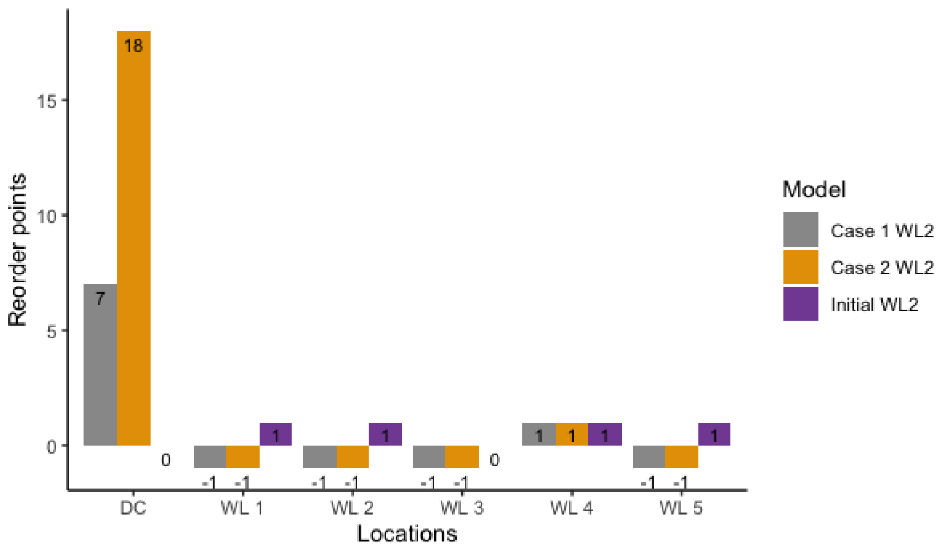

Figure 19: Reorder points at all warehouses for Item I of the AVSL model with different demand cases for WL2

The simulation results of cases 1 and 2 are displayed in Tables 35 and 36, respectively. These indicate that a significant increase in demand frequency and size for WL2, causes it to be unable to satisfy its service measure. As already mentioned, the central reorder point for case 1 is too low to replenish WL2 from stock on hand. The simulation results confirm this observation and show a wait time service level of

0% at WL2. Due to the disproportionate central reorder point, the wait time service levels at all depots, except WL4, are negatively affected. For case 2, only the service level at WL2 is significantly affected. Hence, the addition of demand size variance caused the AVSL model to anticipate greater replenishment orders, leading to a greater central reorder point. Therefore, the remaining depots service levels are unaffected by the increase in demand frequency and size at WL2.

Table 35: Simulation results of case 1 for WL2 of Item I for the AVSL model, $R_0 = 7$

Depot	Reorder point	Order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	4	**88.1**	447	894
WL2	-1	1	4	**0.0**	1026	20,520
WL3	-1	1	4	**88.2**	187	187
WL4	1	1	4	99.4	177	354
WL5	-1	1	4	**87.9**	3001	3838

Table 36: Simulation results of case 2 for WL2 of Item I for the AVSL model, $R_0 = 18$

Depot	Reorder point	Order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
WL1	-1	1	4	97.1	447	894
WL2	-1	1	4	**65.7**	1026	21,010
WL3	-1	1	4	96.3	187	187
WL4	1	1	4	99.4	177	354
WL5	-1	1	4	95.7	3001	3838

These results indicate that a high demand frequency and a large demand size, whilst maintaining the initial order quantity, is troublesome for the AVSL model. This issue is particularly highlighted through case 1. The central reorder point is unable to increase sufficiently, as it does not anticipate for the demand size at WL2. The mean of the demand over the central lead time, $\mathbb{E}[D_j(L_0)]$, equals about 2.57, whilst its variance, $Var[D_j(L_0)]$, is zero. In combination with the low order quantity, the DC does not expect constant replenishment orders of such magnitude. This is confirmed by the fact that $\delta_j(k) = 1$ for $k = 3$ at WL2 in (2), i.e. the DC expects WL2 to order at most three times its order quantity. This is significant as for all $k > 3$, $\delta_j(k) = 1$. Therefore, as the probability of WL2 ordering a replenishment order of exactly three times its order quantity equals one, $s_j^{ord}(k) = 1$ for $k = 3$, and $s_j^{ord}(k) = 0$ for all other k in (3).

These probabilities affect the expected value and variance of $\zeta_{0,j}$ in (20) and (21), which in turn affect the delay Δ_j in (23) - (28). Since the cumulative probability of the inventory level at the depot, $Pr[IL_j \leq y]$, in (31) always indicates that a customer must wait for $R_j = -1$, the Δ_j and $Pr[IL_j \leq y]$

affect the wait time condition in (35) to such a degree that the DC is only able to cover the expected demand over the central lead time instead of the actual demand sizes. Ultimately resulting in the wait time condition implicating that the currently determined central and local reorder points will suffice the wait time target. However, during simulation we observe that the determined solution fails to serve the customers at WL2 within the set wait time target, negatively impacting the service levels at most depots and especially the service level at WL2.

Adding variance to the demand size in case 2, increased $Var[D_j(L_0)]$ significantly. Therefore, the DC has to frequently expect larger replenishment orders from WL2, which is represented by a wider spread of probabilities for $s_j^{ord}(k)$. Even though, as stated by Berling and Farvid (2014), the likelihood of a depot ordering multiple batches is diminishing in k; the DC accounts for these varying demand sizes and is better able to, yet it is still insufficient. Hence, as expected increasing the demand frequency and size significantly affects the results of the AVSL model. Moreover, the computed probabilities throughout the optimisation are heavily dependent on the provided order quantities.

6.3.4 Adjusting the order quantities

As indicated throughout the preceding sections, the order quantities greatly affect the performance of the wait time service measure. Moreover, Section 6.3.3 displayed the challenges that arise if the order quantity at a depot is significantly lower than most of the customer demand sizes. Consequently, the DC is unable to anticipate the significant replenishment orders demanded by the depot, resulting in an inventory allocation by the AVSL model within the distribution system that is insufficient. Even though, the preceding section adjusted the demand frequency and sizes, as mentioned in Section 5, the statistics in Table 2 also indicate that the average demand size at the WL depots is significantly greater than their average order quantity. Therefore, we will analyse the effect on the performance of the AVSL model by modifying the order quantities at all depots, such that the modified order quantity at a depot covers at least 80% of its customer demand sizes.

First, we will analyse the effect of modifying the order quantities on the performance for Item II. For Item I, the demand sizes have little to no variance and do not differ drastically from the order quantities, therefore the modification is unlikely to have a significant impact. Thus, we will investigate the effects of the order quantity modification on cases 1 and 2 for WL2 of Item I with altered demand frequency and sizes from Section 6.3.3. Finally, we evaluate the implications on the performance of the entire distribution system for the AVSL model.

Before analysing the effects of utilising the modified order quantities, Table 37 displays the statistics of the initial and modified order quantities. For the FL depots, we only observe a slight increase in the average order quantity. However, as expected, the average order quantity for the WL depots increased significantly, as it almost tripled in value. Comparing these modified order quantities to the demand statistics in Table 2, we observe that the average order quantity of the WL depots is now greater than their average demand size. This was already true with the initial order quantities for the FL depots.

Table 37: Order quantity statistics of the initial and modified data specified per depot type

	Depots	Min.	Avg.	Max.
Initial order quantity	All	1.000	1.407	325.000
	FL	1.000	3.502	325.000
	WL	1.000	1.071	69.000
Modified order quantity	All	1.000	3.017	1481.000
	FL	1.000	4.128	410.000
	WL	1.000	2.839	1481.000

Table 38 displays the simulation results for Item II with the modified order quantities. Note that the original order quantities, as given in Table 4, are also provided. Moreover, the simulation results with the initial order quantities are provided in Table 15. The modified order quantities experienced a significant increase, especially at the depots with a large demand size variance. The results indicate that, the current inventory allocation by the AVSL model allows all depots to satisfy their service measure. Therefore, the modification of the order quantities significantly impacted the performance of the distribution system, as previously WL7 and FL1 were unable to attain their service measures.

The most noticeable differences between the initial inventory allocation and the current one with the modified order quantities are the reorder points at the DC and WL7. The central reorder point increased significantly, indicating that it anticipates greater replenishment order sizes. Furthermore, the local reorder point of 23 at WL7 suggests that the models wants to suffice most the customers at this depot through stock on hand. Yet, for the infrequent customers with larger demand sizes, it permits them to wait. Therefore, the AVSL model determined an intermediate approach at this depot, allowing WL7 to achieve its service measure through a combination of using stock on hand and back orders.

Table 38: Simulation results of Item II for the AVSL model with modified order quantities

Warehouse	Initial order quantity	Reorder point	Adjusted order quantity	Fill rate target (%)	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
DC	500	34	500	NA	NA	98.7	48,954	406,006
WL1	1	-1	4	NA	4	99.3	988	5458
WL2	1	-1	8	NA	4	98.8	1184	5969
WL3	1	-1	2	NA	4	99.3	1036	2072
WL4	1	-1	2	NA	4	99.7	967	3014
WL5	1	-1	2	NA	4	98.8	574	1148
WL6	2	-1	16	NA	4	98.1	206	3296
WL7	1	23	80	NA	4	98.3	786	21172
WL8	2	-1	32	NA	4	99.0	605	8638
WL9	1	-1	16	NA	4	100.0	194	3104
WL10	1	-1	12	NA	4	98.9	980	9734
WL11	1	-1	16	NA	4	99.2	395	4512
WL12	1	-1	16	NA	4	98.6	631	5492
WL13	1	-1	8	NA	4	99.5	193	1544
WL14	1	14	8	NA	4	99.0	1028	8386
WL15	1	7	4	NA	4	99.7	932	3332
WL16	1	1	2	NA	4	99.5	207	414
WL17	3	23	24	NA	4	98.7	544	13056
FL1	3	4	2	95.0	NA	95.2	15,025	33,087
FL2	1	18	12	95.0	NA	95.8	6758	45,011
FL3	1	4	4	95.0	NA	95.2	3805	13,172
FL4	5	19	6	95.0	NA	95.4	36,677	214,430

Note: For the DC the 'Achieved service level', 'Total customers' and 'Total demand' are the 'Achieved fill rate', 'Total replenishment orders' and 'Total replenishment size', respectively. These apply to the replenishment orders from the depots.

For demand case 1 and 2 for WL2 of Item I, Tables 39 and 40 display their simulation results with modified order quantities, respectively. Note that the respective simulation results with the initial order quantities are provided in Tables 35 and 36. For case 1 of WL2, we observe a significant improvement in the achieved wait time service levels across all depots. For the initial order quantities, only WL4 was able to satisfy its service measure, whilst now with the modified order quantities only WL5 is unable to achieve its wait time target. Therefore, the modification of the order quantities provided a significant improvement in the achieved service levels for case 1.

However, for case 2 of WL2, although the central reorder point is significantly increased due to the order quantity of WL2 increasing, the service level at WL2 did not improve as significantly. It is still unable to satisfy its wait time service measure. Due to the introduced variance in case 2, about 20% of its customers arrive with a demand size of 50. The DC still does not account for replenishment order of this size, therefore these customers cannot be served within the wait time target as the modified local order quantity only covers 80% of the demand sizes at WL2. Hence, accounting for the additional customers that WL2 is unable to serve within the set wait time target, the achieved service level is only 73.0%. Implicating that even while using the modified order quantities, if the average customer demand sizes differ significantly between depots, the AVSL model still faces difficulties allocating sufficient stock to the distribution system.

Similarly to Section 6.3.3, the expected demand over the central lead time, $\mathbb{E}[D_j(L_0)]$, remains about 2.57, even if its variance, $Var[D_j(L_0)]$, is now positive in case 2 for WL2. Due to increasing the order quantity, the DC now accounts for a single replenishment order of size 25 for this depot. However, due to the low demand over the effective lead time, the DC does not anticipate for the 20% of customers that demand twice the order quantity of WL2. This is confirmed as $\delta_j(k) = 1$ for $k = 1$ at WL2 in (2), therefore the DC expects WL2 to order its order quantity at most once during the central lead time. Hence, the probability of WL2 ordering exactly its order quantity equals one, $s_j^{ord}(k) = 1$ for $k = 1$, and $s_j^{ord}(k) = 0$ for all other k in (3). These probabilities again affect the delay Δ_j and $Pr[IL_j \leq y]$ such that the wait time condition in (35) implicates that the currently determined central and local reorder points will suffice the wait time target. Resulting in a solution that fails to serve part of its customers within the set wait time target at WL2.

Table 39: Simulation results of case 1 for WL2, Item I for the AVSL model with modified order quantities

Warehouse	Initial order quantity	Reorder point	Adjusted order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
DC	5	0	5	NA	53.4	4824	25,793
WL1	1	1	2	4	98.4	447	894
WL2	1	32	20	4	99.1	1026	20,520
WL3	1	0	1	4	100.0	187	187
WL4	1	1	2	4	98.9	177	354
WL5	1	1	1	4	**91.9**	3001	3838

Note: For the DC the ‘Achieved service level’, ‘Total customers’ and ‘Total demand’ are the ‘Achieved fill rate’, ‘Total replenishment orders’ and ‘Total replenishment size’, respectively. These apply to the replenishment orders from the depots.

Table 40: Simulation results of case 2 for WL2, Item I for the AVSL model with modified order quantities

Warehouse	Initial order quantity	Reorder point	Adjusted order quantity	Wait time target (days)	Achieved service level (%)	Total customers	Total demand
DC	5	36	5	NA	91.3	4824	25,793
WL1	1	-1	2	4	98.0	447	894
WL2	1	-1	25	4	**73.0**	1026	21,010
WL3	1	-1	1	4	96.8	187	187
WL4	1	1	2	4	99.4	177	354
WL5	1	-1	1	4	96.8	3001	3838

Note: For the DC the 'Achieved service level', 'Total customers' and 'Total demand' are the 'Achieved fill rate', 'Total replenishment orders' and 'Total replenishment size', respectively. These apply to the replenishment orders from the depots.

The simulation results with the altered order quantities for all the items are displayed in Table 38. Comparing these results to those obtained with the initial order quantities, see Table 8, we observe that now 86.1% of all depots are able to satisfy their service measure, which is an increase of about 9.0%. Moreover, with the modified order quantities 92.0% of WL depots are able to satisfy their wait time service measure, compared to only 82.0% for the WL depots with the initial order quantities. However, the performance across the FL depots dropped by about 1.9%, indicating that these depots perform worse with the modified order quantities.

Table 41: Simulation results of all items for the AVSL model with modified order quantities

	Min.	Avg.	Max.	Total
Customers	167	10,618.987	8,517,494	105,595,202
Demand	167	35,116.947	17,998,964	349,202,923
FL depots	1	0.521	26	5205
FL depots service measure achieved	0	0.231	14	2313
WL depots	1	3.246	24	32,447
WL depots service measure achieved	0	3.011	24	30,094
Fill rate DC (%)	0.0	86.2	100.0	NA

In general, we can state that the AVSL model is highly sensitive to the selection of order quantities. The model prefers larger order quantities for the WL depots, these permit the DC to anticipate larger replenishment orders instead of only anticipating the demand of a depot across the central lead time. With the modified order quantities, the probabilities $\delta_j(k)$ and consequently $s_j^{ord}(k)$ are better able to reflect the probability of a wait time depot ordering the actual demand sizes. Therefore, the wait time condition in (35) is better able to adequately represent the wait time scenario for a customer at that depot, resulting in a better estimation of the necessary central and local inventory. However, for the FL depots, due to their significantly greater demand over the central lead time compared to the WL depots, the model prefers lower order quantities as this permits a higher reorder point at the depots and thus being able to serve sufficient customers through stock on hand. Hence, properly determining the order quantity at every depot is crucial to the performance of the AVSL model.

6.4 The EVSL model

In Section 4.3.3 we proposed an experiment in which we allow local reorder points R_j of the wait time depots, for $j \in W_i$, below -1. This experiment is introduced to further highlight why it is disadvantageous to allow for local reorder points below minus one and let customers wait upon other customer arrivals. To facilitate these reorder points, we introduced *Assumption* 22 and altered the data accordingly, see Section 5.2. The results are therefore obtained using the modified data. The service measure targets are identical to the AVSL model, $\beta_j^T = 0.95$, $WT_j^T = 4$ and $P_j^T = 0.95$.

Let us now analyse the results of the experiment. Table 42 displays the inventory allocation results of the EVSL model. Comparing these results to Table 8 of the AVSL model, we observe a significant decrease in central reorder points and total stock, defined as the sum over all reorder points. The total stock decreased by 30.8%. This significant decrease in total stock causes a 6.5% decrease in total cost. On the other hand, the computation times are nearly identical.

Table 42: Inventory allocation results of all items for the AVSL model

	Min.	Avg.	Max.	Total
Central reorder point	-1988	1.155	1646	11,544
Total stock	-1187	6.562	2831	65,590
Inventory cost (ZAR)	0.00	28,891.00	4,510,061.00	288,765,649.00
Computation time (s)	0.000	0.108	175.937	1075.196

The decrease in total stock can mainly be explained by the decrease in central reorder points. These were able to be significantly decreased due to the constant demand size equal tp one for all customers at the wait time depots. To visualise the difference in allocation, Figure 20 displays the inventory allocation of Item II for the AVSL model versus the EVSL model. It shows that the EVSL model prefers to stock locally and use a very low central reorder point. Consequently, the local stock at the FL depots has to increase significantly, yet the total stock equals -2 compared to 102 for the AVSL model. The EVSL model determined that by placing stock locally it can suffice the wait time service measures despite the increased demand frequency.

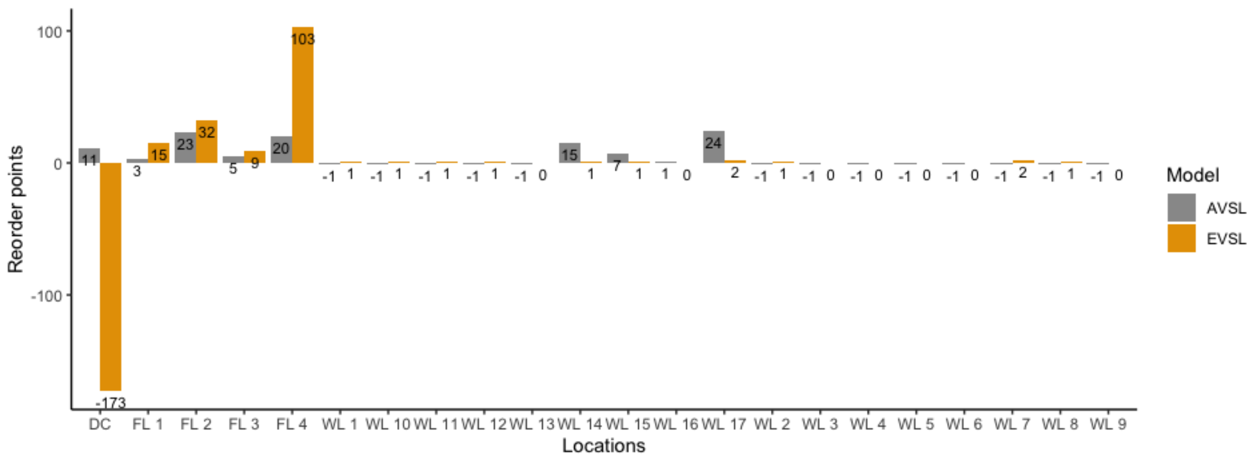

Figure 20: Reorder points at all warehouses of Item II for the AVSL and EVSL models

To confirm the results above, Table 43 presents the statistics regarding the local reorder points per type of depot for the AVSL and EVSL models. However, these indicate that the averages of local reorder point across all depots and for the WL depots are lower for the EVSL model compared to the AVSL model. Upon further inspection of the results, the EVSL model only allocated an $R_j > 0$ to 1611 WL depots compared to 5661 for the AVSL model. Moreover, the EVSL model only has 5285 WL depots for which $R_j = -1$ holds compared to the 12795 WL depots for the AVSL model. Hence, due to the modification of the demand frequencies and sizes, the EVSL model prefers to stock locally as for the gross of the WL depots it holds that $R_j = 0$, as this holds for 25,465 WL depots. Therefore, the inventory position will always be positive as the order quantities are positively defined.

Table 43: The local reorder points per type of depot for the AVSL and EVSL models

Model	AVSL			EVSL		
Depots	All	FL	WL	All	FL	WL
Min.	-1.000	-1.000	-1.000	-3.000	-2.000	-3.000
Avg.	1.591	10.06	0.233	1.435	10.96	-0.092
Max.	958.000	958.000	700.000	958.000	958.000	86.000

These results indicate that the EVSL model generally stocks locally contrary to the AVSL model. Furthermore, the number of depots with a reorder point below -1 is only 87. Hence, the EVSL model behaves as expected and does not prefer wait time depots with reorder points below -1, even when the demand frequency is significantly increased due to the data modification, see Section 5.2. It demonstrated that having customers wait upon other customers to arrive for slow-moving items is not optimal, which follows our expectation.

7 Conclusion

The main research question posed in this paper is: How can the inventory, especially of slow-moving items, be reduced within a multi-echelon distribution system whilst providing a proper and flexible solution? To address this question, we proposed two distinct models. Both models implement fill rate and wait time as their service measures, assigning a fill rate target to depots with fast-moving items and a wait time target for depots with slow-moving items.

The Simple-Various Service Level (SVSL) model restricts the local reorder points to always equal minus one, leading to the majority of customers being served through back orders and assigning minimal local stock. It requires the wait time targets always to be greater than the DC-depot inter-lead times. On the other hand, the Advanced-Various Service Level (AVSL) model allows local reorder points to be greater than minus one, thus also permitting them to be non-negative. This allows for more adaptable wait time targets can be set, permitting customers to be served from stock on hand. These models extend the pre-existing *vNext* model, which only utilises a fill rate service measure.

To better address the main research question, we composed four sub-research questions. The first sub-research question explored the types of wait time approximation methods. In the scope of the current research problem, Berling and Farvid (2014) presented the most suitable wait time approximation method, as it is able to determine the specific wait time per depot and can account for negative central reorder points.

The second sub-research question focuses on the additional necessary assumptions for the introduced VSL models. These assumptions mainly concerned the limitations of the chosen wait time approximation methods and the model-specific restrictions. An important additional assumption, as stated by Berling and Farvid (2014), is that of constant central lead times. Furthermore, we elaborated on the choice of distributions, where assume that all demand over time follows a *Gamma* distribution.

Thirdly, we conducted performance tests on the VSL models, comparing them to the performance of the *vNext* model, which we regarded as the baseline. The VSL models demonstrated a significant reduction in stock within the distribution system. Notably, the SVSL model achieved the greatest inventory reduction, while the AVSL model saved most on the inventory costs. However, general performances during the simulation were subpar across all models. All of them encountered difficulties throughout the simulation. As a result, no model significantly out performed one other. It is worth noting that the VSL models, in general, did not underperform compared to the *vNext* model. Both VSL models effectively implemented wait time as a service measure and successfully reduced the inventory within the distribution system.

Robustness tests were conducted specifically on the AVSL model to address the fourth and final sub-research question. The simulation results hinted at the significance of the size of order quantities, as these directly affect the reorder points both locally and centrally. The robustness tests clearly revealed the importance of determining an appropriate order quantities, and how these affect the central and local reorder points and therefore the performances. The tests highlighted the difficulties the AVSL model experiences in scenarios where the variance of the demand sizes is significant and when the order quantity is significantly lower than the average demand size. The tests provided valuable insights of the AVSL model's behaviour across various scenarios and presented the areas of improvement.

In conclusion, incorporating wait time as a service measure within multi-echelon optimisation is determined to be viable. The proposed models offer satisfactory solutions for a diverse range of items. Importantly, they did not underperform compared to the baseline model. Hence, utilising a wait time service measure provides a proper solution, leading to significant reduction in stock. The VSL models excel when the demand size distribution coincides with the demand over the central lead time distribution, as this allows the DC to best approximate the incoming replenishment order sizes and allocate its stock accordingly. The models' performances are heavily influenced by the order quantities, therefore these need to be chosen carefully. Moreover, an alternative optimisation method has to be used in order to guarantee optimal solutions at the global minima. The wait time service measure demonstrates to be flexible in its target selection, as the AVSL model was the best-performing model allowing wait time targets below the inter-lead time of the depots. Further research on wait time service measures is recommended to gain a deeper understanding of critical features impacting performance and to explore strategies for addressing these issues within multi-echelon distribution systems.

8 Discussion

With the current research we aimed to reduce the inventory within multi-echelon distribution systems whilst providing a proper and flexible solution. Though the exploration of two distinct models, the SVSL and AVSL models, we pursued the use of both fill rate and wait time service measures to achieve the desired reduction in inventory.

A significant limitation encountered during our research, pertained the provided order quantities used in the computations of the models. These values were determined by Gordian Logistic Experts in consideration with Company A. Therefore, as previously mentioned, the order quantities included preferences set by Company A instead of being the optimally selected values. We provided various insights into the effects of the order quantities on the inventory allocation by the models. Consequently, these non-optimal order quantities significantly impacted the performance of the models, particularly in scenarios where the actual demand sizes were significantly greater than the assigned order quantities.

Furthermore, the introduced models applied a bisection-based method to determine the optimal inventory allocation. We determined that the unimodality assumption of this method was violated by the AVSL model, resulting in non-optimal inventory allocations. Additionally, the in some cases the used methods were unable to adequately represent the actual replenishment order sizes of the depots at the DC. This prevented the DC to anticipate larger replenishment orders as it could only account for their demand over the central lead time, limiting the central reorder point to suffice the wait time depots replenishment orders. Thus, affecting the overall performance of the models.

Despite these limitations, our research provided useful insights into incorporating wait time as a service measure within multi-echelon optimisation. The AVSL model demonstrated significant reductions in stock within the distribution system and saving on inventory cost. We positively demonstrated the viability of combining fill rate and wait time service measures to reduce the inventory within a multi-echelon distribution system. Moving forward, further research should focus on adequately representing the demand sizes of customers at wait time depots within the proposed formulas. Furthermore, an alternative to the bisection-based approach should be considered. The new approach should be able to consistently achieve the global minimum, whilst refraining from iterating over all possible solutions. Finally, the order quantity at a depot should be selected with consideration to its demand sizes and selected service measure.

References

Axsäter, S. (1993). Chapter 4 continuous review policies for multi-level inventory systems with stochastic demand. In *Logistics of Production and Inventory*, volume 4 of *Handbooks in Operations Research and Management Science*, pages 175–197. Elsevier.

Axsäter, S. (2003a). Approximate optimization of a two-level distribution inventory system. *International Journal of Production Economics*, 81:545–553.

Axsäter, S. (2003b). Supply chain operations: Serial and distribution inventory systems. *Handbooks in operations research and management science*, 11:525–559.

Axsäter, S. (2015). *Inventory control*, volume 225. Springer.

Axsäter, S., Marklund, J., and Silver, E. A. (2002). Heuristic methods for centralized control of one-warehouse, n-retailer inventory systems. *Manufacturing & Service Operations Management*, 4(1):75–97.

Axsäter, S., Olsson, F., and Tydesjö, P. (2007). Heuristics for handling direct upstream demand in two-echelon distribution inventory systems. *International Journal of Production Economics*, 108(1-2):266–270.

Berling, P. and Farvid, M. (2014). Lead-time investigation and estimation in divergent supply chains. *International Journal of Production Economics*, 157:177–189.

Chew, E. P. and Tang, L. C. (1995). Warehouse-retailer system with stochastic demands—non-identical retailer case. *European Journal of Operational Research*, 82(1):98–110.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 23(3):289–303.

de Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J., and Schade, K. (2018). A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3):955–983.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., and Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how*, volume 488. Springer.

Diks, E. B. and De Kok, A. (1998). Optimal control of a divergent multi-echelon inventory system. *European journal of operational research*, 111(1):75–97.

Doğru, M. K., De Kok, A., and van Houtum, G.-J. (2009). A numerical study on the effect of the balance assumption in one-warehouse multi-retailer inventory systems. *Flexible services and manufacturing journal*, 21(3):114–147.

Forsberg, R. (1997). Exact evaluation of (r, q)-policies for two-level inventory systems with poisson demand. *European journal of operational research*, 96(1):130–138.

Gallego, G., Özer, Ö., and Zipkin, P. (2007). Bounds, heuristics, and approximations for distribution systems. *Operations Research*, 55(3):503–517.

Geelen, M., Flapper, S., Basten, R., Driessen, M., and Gordian Logistic Experts (2019). A single-item, two-echelon inventory model with order fill rate constraints, batching and compound poisson demand. *Master Thesis*.

Gordian Logistic Experts (2021). Multi-echelon implementation: the vNext model. Unpublished.

Grob, C. and Bley, A. (2018). Comparison of wait time approximations in distribution networks using (r, q)-order policies. *arXiv preprint arXiv:1801.09617*.

Harris, F. W. (1913). How many parts to make at once.

Kiesmüller, G. P., De Kok, A., and Dabia, S. (2011). Single item inventory control under periodic review and a minimum order quantity. *International Journal of Production Economics*, 133(1):280–285.

Oracle Corporation (1995). *Java SE*, version 1.8. https://www.oracle.com/java/.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rey, D. and Neuhäuser, M. (2011). Wilcoxon-signed-rank test. In *International encyclopedia of statistical science*, pages 1658–1659. Springer.

Sherbrooke, C. C. (1968). Metric: A multi-echelon technique for recoverable item control. *Operations research*, 16(1):122–141.

Whitt, W. (1982). Approximating a point process by a renewal process, i: Two basic methods. *Operations Research*, 30(1):125–147.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Wolff, R. W. (1982). Poisson arrivals see time averages. *Operations Research*, 30(2):223–231.

Appendix A - Central wait time for effective lead time

In this appendix, we elaborate on the computation of the central wait time used to determine the effective lead time per depot, L_j^{eff}, in (6) and (7). For L_j^{eff}, it is assumed that all depots experience the identical stochastic delay Δ. We determine the mean and variance of the delay through the methods defined by Axsäter (2003a). One of their main assumptions states that the demand over the central lead time is normally distributed, thus the computations are based on the normal distribution. The goal is to estimate the mean and the standard deviation of the central delay, μ_Δ and σ_Δ, respectively.

First, the average wait time can be computed as:

$$\mathbb{E}[W] = \frac{\mathbb{E}[IL_0]^-}{\sum_{j=1}^{M} \mu_j}, \tag{49}$$

where $\mathbb{E}[IL_0]^-$ represents the expected number of back orders at the DC during the central lead time L_0.

Let q be the greatest common divisor of the local order quantities, $q = gcd(Q_1, \ldots, Q_{M_i})$. Then, Axsäter (2003a) provide the following equations for the expected back orders:

$$\mathbb{E}[IL_0]^- = \frac{(\sigma(D_0(L_0))^2}{Q_0 - q}\left[H\left(\frac{R_0 + q - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))}\right) - H\left(\frac{R_0 + Q_0 - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))}\right)\right], \tag{50}$$

with $H(x) = \frac{1}{2}[(x^2+1)(1-\Phi(x)) - x\varphi(x)]$. Moreover, if $q = Q_0$ holds, then

$$\mathbb{E}[IL_0]^- = \sigma(D_0(L_0)G\left(\frac{R_0 + Q_0 - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))}\right), \tag{51}$$

with $G(x) = \varphi(x) - x(1-\Phi(x))$, being the loss function of the normal distribution.

Note that in both equations $\sigma(D_0(L_0)) = \sqrt{Var[D_0(L_0)]}$. However, as mentioned before, Axsäter (2003a) assumed normally distributed demand. Thus, they compute the $Var[D_0(L_0)]$ as follows:

$$Var[D_0(L_0)] = \sum_{j=1}^{M} \sum_{k=-\infty}^{\infty} (kQ_j - \mu_j \mathbb{E}[L_0])^2 p_{j,k}(L_0), \tag{52}$$

here $p_{j,k}(L_0)$ is the probability of depot j placing k orders of size Q_j. This is in fact the only term that differs from the terms used in (5), since it is assumed here that the demand over the central follows a normal distribution. We can compute this probability as follows:

$$\begin{aligned} p_{j,k}(L_0) = \frac{\sigma_j \mathbb{E}[L_0]}{Q_j}&\left[G\left(\frac{(k-1)Q_j - \mu_j \mathbb{E}[L_0]}{\sigma_j \mathbb{E}[L_0]}\right) + G\left(\frac{(k+1)Q_j - \mu_j \mathbb{E}[L_0]}{\sigma_j \mathbb{E}[L_0]}\right)\right. \\ &\left. -2G\left(\frac{kQ_j - \mu_j \mathbb{E}[L_0]}{\sigma_j \mathbb{E}[L_0]}\right)\right]. \end{aligned} \tag{53}$$

For the variance of the wait time, they proposed the following equation:

$$Var[W] = (\sigma_\Delta)^2(1 - Pr(W=0)) + \mu_\Delta \mathbb{E}[W] - \mathbb{E}[W]^2, \tag{54}$$

where μ_Δ and σ_Δ are the mean and standard deviation of the time between the arrival time and demanded time of an item, i.e. the delay Δ. For which it is noted by Axsäter (2003a) that $W = (\Delta)^+$. Also, $Pr(W=0)$ is the probability of the wait time being zero.

To determine the mean and standard deviation of the delay, we have to perform the following in between step:

$$\propto = -\frac{\mu_\Delta}{\sigma_\Delta} = \Phi^{-1}(Pr(W=0)). \tag{55}$$

for which we can derive the probability of the wait time being zero as

$$Pr(W = 0) = \frac{\sigma(D_0(L_0)}{Q_0 - q} \left[G\left(\frac{R_0 + Q_0 - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))} \right) - G\left(\frac{R_0 + q - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))} \right) \right], \tag{56}$$

and if again $q = Q_0$, then it changes to

$$Pr(W = 0) = \Phi \left(\frac{R_0 + Q_0 - \mathbb{E}[D_0(L_0)]}{\sigma(D_0(L_0))} \right). \tag{57}$$

Finally we can compute the μ_Δ and σ_Δ,

$$\sigma_\Delta = \frac{\mathbb{E}[W]}{G(\propto)} \tag{58}$$

$$\mu_\Delta = - \propto * \sigma_\Delta. \tag{59}$$

As mentioned at the start, these are the expected delay and standard deviation of the delay at the DC, respectively. The mean corresponds to the $\mathbb{E}[\Delta]$ in (6), i.e. $\mathbb{E}[\Delta] = \mu_\Delta$. Furthermore, the standard deviation is used for the $Var[\Delta]$ in (7), i.e. $Var[\Delta] = \sigma_\Delta^2$.

Appendix B - Differences in notation between Grob and Bley (2018) and Berling and Farvid (2014)

This appendix explains the differences in notation in detail between the papers of Grob and Bley (2018) and Berling and Farvid (2014). The current paper mainly complies with the notation introduced by Grob and Bley (2018). Let us first revise the introduced notation in Section 4.2.1:

$\mu_{\zeta_{0,j},p}$ The expected demand for depot p at the DC over time period L_0, i.e. the individual contribution of depot p to $\mu_{\zeta_{0,j}}$.

$\sigma^2_{\zeta_{0,j},p}$ The variance of the demand for depot p at the DC over time period L_0, i.e. the individual contribution of depot p to $\sigma^2_{\zeta_{0,j}}$.

Moreover, Berling and Farvid (2014) introduced the following variables regarding (20) and (21):

$\mu_{\zeta_{0,j}}$ The expected value of $\zeta_{0,j}$.

$\sigma^2_{\zeta_{0,j}}$ The variance of $\zeta_{0,j}$.

Note that these variables correspond to $\mathbb{E}[\zeta_{0,j}]$ and $Var[\zeta_{0,j}]$ in (20) and (21), respectively.

Then Berling and Farvid (2014) introduce the following equations for $\mu_{\zeta_{0,j}}$ and $\sigma^2_{\zeta_{0,j}}$:

$$\mu_{\zeta_{0,j}} = \mu_{\zeta_{0,j},1} + \ldots + \mu_{\zeta_{0,j},M} = \sum_{j=1}^{M} \mu_{\zeta_{0,j},p}, \tag{60}$$

with

$$\mu_{\zeta_{0,j},p} = \mu_j L_0 \qquad \text{for } p \neq j, \tag{61}$$

$$\mu_{\zeta_{0,j},j} = \sum_{k=0}^{\infty} kQ_j g_j(kQ_j) \qquad . \tag{62}$$

$$\sigma^2_{\zeta_{0,j}} = \sigma^2_{\zeta_{0,j},1} + \ldots + \sigma^2_{\zeta_{0,j},M} = \sum_{j=1}^{M} \sigma^2_{\zeta_{0,j},p}, \tag{63}$$

with

$$\sigma^2_{\zeta_{0,j},p} = \sum_{k=0}^{\infty} (\mu_{\zeta_{0,j},p} L_0 - kQ_p)^2 g_p(kQ_p). \tag{64}$$

Again, note that (60) and (63) correspond to (15) and (16), respectively. Here, $g_j(u)$ is defined by Berling and Farvid (2014) as the probability that exactly u units are ordered by depot j. Therefore, it can be written out as:

$$g_j(u) = \begin{cases} \delta_j(0), & \text{if } u = 0 \\ \delta_j(k) - \delta_j(k-1), & \text{if } u = kQ_j \text{ for } k > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{65}$$

$\delta_j(n)$ is identically defined in both papers when assumed that L_0 is constant, as done by Berling and Farvid (2014). Therefore, the only difference in notation that remains, arises the definition of the function $g_j(u)$ and $s_j^{ord}(k)$, see (3). The latter is defined as the probability that depot j orders exactly k batches of Q_j. It can be seen in (62) and (64) that for $g_j(u)$, the only values evaluated for u are non-negative multiples of Q_j as $k \in [0, \infty)$. Hence, the functions $g_j(u)$ and $s_j^{ord}(k)$ provide identical results. Therefore, the complete written out equations in (20) and (21) utilise the equivalent equations of (62) and (64), respectively.

Appendix C - Correction of the error in Berling and Farvid (2014)

In this appendix, we elaborate on the discrepancy between our formulas and the introduced ones by Berling and Farvid (2014). More specifically, it regards the second indicated case in Section 4.2.1, involving (25) and (26).

The formulation of Berling and Farvid (2014) for the second case is given as:

$$\text{Case 2:} \quad 0 \le R_0 < Q_j$$

$$\mathbb{E}[\Delta_j] = \frac{L_0}{Q_0}\left(Q_j - R_0 - 2 + \sum_{x=Q_j}^{R_0+Q_0}\left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty}\left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right) f_j(\zeta_{0,j})\right)\right), \tag{66}$$

$$\mathbb{E}[\Delta_j^2] = \frac{L_0^2}{Q_0}\left(Q_j - R_0 - 2 + \sum_{x=Q_j}^{R_0+Q_0}\left(\sum_{\zeta_{0,j}=x-Q_j+1}^{\infty}\left(1 - \frac{x-Q_j}{\zeta_{0,j}}\right)^2 f_j(\zeta_{0,j})\right)\right). \tag{67}$$

$$\tag{68}$$

Note that the difference between their formulation above and our formulation in (25) and (26) are the terms $\frac{Q_j-R_0-2}{Q_0}$ and $\frac{Q_j-R_0-1}{Q_0}$. These terms are the result of the following probability:

$$Pr[R_0 + 1 \le IP_0 \le Q_j - 1]. \tag{69}$$

The difference applies to the provided explanation by Berling and Farvid (2014). They indicate that

$$IP_0 \sim \text{Unif}(R_0 + 1, R_0 + Q_0). \tag{70}$$

Hence, the inventory position follows a discrete uniform distribution.

A discrete uniform distribution, Unif(a, b), has the following PDF and CDF:

$$f_{IP}(x) = \begin{cases} \frac{1}{b-a+1} = \frac{1}{Q_0}, & \text{if } a \le x \le b, \text{ i.e. } R_0 + 1 \le x \le R_0 + Q_0 \\ 0, & \text{otherwise,} \end{cases} \tag{71}$$

$$F_{IP}(x) = \begin{cases} 0, & \text{if } x < a, \text{ i.e. } x < R_0 + 1 \\ \frac{\lfloor x \rfloor - a + 1}{b-a+1} = \frac{\lfloor x \rfloor - R_0}{Q_0}, & \text{if } x \in [a, b], \text{ i.e. } x \in [R_0 + 1, R_0 + Q_0] \\ 1, & \text{if } x > b, \text{ i.e. } x > R_0 + Q_0. \end{cases} \tag{72}$$

Note the cases for which it holds that $a = R_0 + 1$ and $b = R_0 + Q_0$ are also presented, as these represent the domain of the discrete uniform distribution at hand.

Using the CDF we determine the probability in (69) as

$$\begin{aligned} Pr[R_0 + 1 \le IP \le Q_j - 1] &= Pr[IP \le Q_j - 1] - Pr[IP \le R_0] \\ &= \frac{Q_j - 1 - R_0}{Q_0} - 0 \\ &= \frac{Q_j - R_0 - 1}{Q_0}. \end{aligned} \tag{73}$$

Moreover, an alternative method to compute this probability is by using the PDF and the discreteness of the distribution:

$$
\begin{aligned}
Pr[R_0 + 1 \leq IP \leq Q_j - 1] &= \sum_{x=R_0+1}^{Q_j-1} Pr[IP = x] \\
&= \sum_{x=R_0+1}^{Q_j-1} \frac{1}{Q_0} \\
&= \sum_{x=1}^{Q_j-1-R_0} \frac{1}{Q_0} \\
&= \frac{Q_j - R_0 - 1}{Q_0}. \qquad (74)
\end{aligned}
$$

Both of these computations prompt the identical result. Therefore, we determine that the probability in (69) is equal to $\frac{Q_j - R_0 - 1}{Q_0}$, as shown by (73) and (74). Hence, we utilise this probability in (25) and (26), therefore correcting the original probability presented by Berling and Farvid (2014).

Appendix D - *Gamma* distribution discretisations

This appendix provides an in-depth example our *Gamma* distribution discretisation elaborates on why it would be problematic if peaks occur in between integer values in the PDF of the *Gamma* distribution. The cases when this poses issues, is usually if the variance is significantly small. If we take the probabilities at the integer values of a *Gamma* distribution for this case, the PDF provides probabilities that are too low to ever attain a large enough fill rate in (11). Moreover, in such cases, summing over the probabilities of the integer values does not lead to a cumulative probability of one. Therefore, we introduced (46) to discretise the *Gamma* distribution accordingly with $\alpha = 0.6$.

We first provide a *Gamma* distribution with a very low variance, let $k = 125$ and $\theta = 0.02$; Figure 21 illustrates the PDF of this distribution. The figure indicates that the peak of the distribution is between two and three. Table 44 indicates that he probabilities at these integer values are $Pr[X = 2] = 0.123$ and $Pr[X = 3] = 0.163$. Given the remainder of the probabilities provided by the PDF, we encounter the issue that these do not add up to a cumulative probability of one; not allowing the possibility to attain the fill rate service measure. Therefore, the probabilities have to be adjusted. Thus, we assign the probability values as: $Pr[X = 2] = 0.336$ and $Pr[X = 3] = 0.664$. The adjusted probabilities use the entire distribution and assign the entire probability mass to each integer value based on the selected α-value. This provides an adjustable method to discretise the *Gamma* distribution as desired.

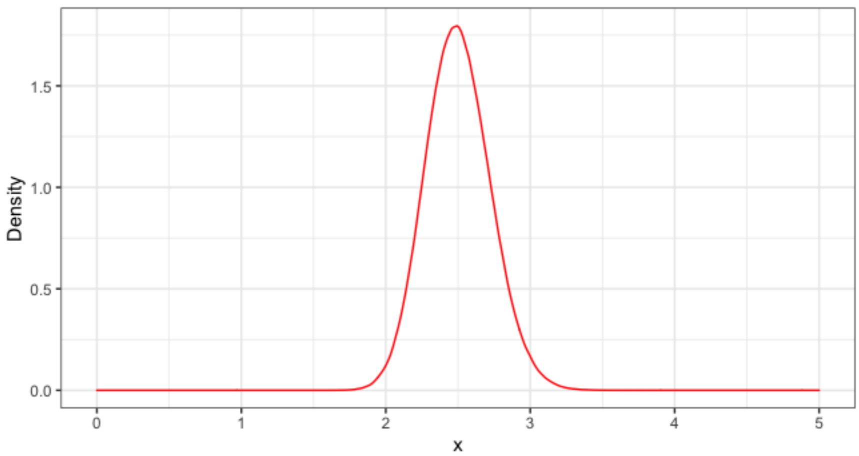

Figure 21: Density graph of a *Gamma*(125, 0.02) distribution with its peak at 2.5

Table 44: Probabilities of a *Gamma*(125, 0.02) distribution

x	Probability through PDF	Probability through discretisation
0	0.000	0.000
1	0.000	0.000
2	0.123	0.336
3	0.163	0.664
4	0.000	0.000
5	0.000	0.000

The modification in (46) does alter the probabilities for the general *Gamma* distributions as well. Figure 22 presents the PDF of a *Gamma* distribution with $k = 12.5$ and $\theta = 0.2$. Note that compared to the previous distribution, the variance increased ten fold. Table 45 provides the probabilities at the relevant integer values for this *Gamma* distribution. We observe that the PDF assigns a greater probability at 2 compared to the discretised probability. This is the affect of the selected α value. As discussed in Section 4.4.2, we opted for a more conservative value. Hence, it assigns greater probabilities to the larger integer values.

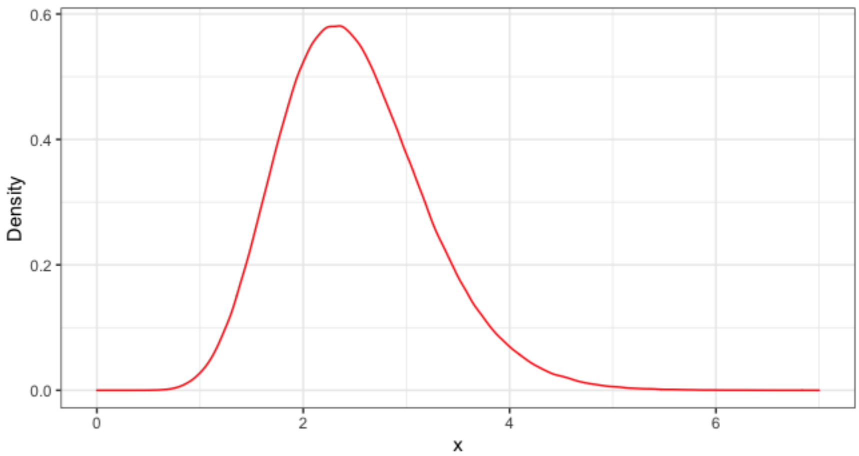

Figure 22: Density graph of a *Gamma*(12.5, 0.2) distribution with its peak at 2.5

Table 45: Probabilities of a Gamma(12.5, 0.2) distribution

x	Probability through PDF	Probability through discretisation
0	0.000	0.000
1	0.027	0.038
2	0.525	0.442
3	0.374	0.411
4	0.069	0.097
5	0.006	0.010
6	0.000	0.001
7	0.000	0.000