DOI: https://doi.org/10.46354/i3m.2019.mas.014

SIMULATED OPERATING CONCEPTS FOR WHOLESALE INVENTORY OPTIMIZATION AT NAVAL SUPPLY SYSTEMS COMMAND

Sean Teter[a], Emily Craparo[b], Javier Salmerón[c]

[(a),(b),(c)]Operations Research Department, Naval Postgraduate School, Monterey, CA, U.S.A.

[(a)]sean.m.teter.mil@mail.mil, [(b)]emcrapar@nps.edu, [(c)]jsalmero@nps.edu

ABSTRACT
Naval Supply Systems Command Weapon Systems Support (NAVSUP WSS) serves as the Navy's inventory control point, managing approximately 375,000 line items. Constrained by funding, NAVSUP WSS uses the Wholesale Inventory Optimization Model (WIOM) to maximize customer service. Since demand distributions for different parts change over time, NAVSUP WSS reruns WIOM quarterly. However, large changes to the solution create an administrative burden. To deal with this problem, referred to as churn, WIOM has a persistence parameter that can discourage change from one run to the next, but it is inherently at odds with customer service performance. This research develops the Comparative Optimized Results Simulation to explore the system's performance under different persistence settings and periodicities of running WIOM. The research finds that periodicities greater than quarterly significantly degrade customer service, and increasing the persistence parameter dramatically improves churn while only marginally degrading customer service.

Keywords: inventory management; discrete event simulation; wholesale inventory optimization model

1. INTRODUCTION

Naval Supply Systems Command, Weapon Systems Support (NAVSUP WSS) serves as the main inventory control point for the Navy. The command manages over 375,000 unique line items (NAVSUP 2018) used in the repair of ships, submarines, Navy and Marine Corps aircraft, and associated weapons systems. The effective management of this supply chain is essential in maintaining readiness of the fleet to operate and conduct combat operations around the world.

Like any organization, NAVSUP WSS has a limited set of resources with which to conduct its operations. The biggest constraint is financial. Given limited budgetary means, NAVSUP WSS strives to maximize support to the warfighter. The predominant metric used to measure customer support is fill rate. When NAVSUP WSS receives a requisition, one of two things can happen. Either the requisition is filled immediately with stock on hand, or the requisition is backordered. The fill rate metric shows the relationship between the number of requisitions filled immediately on receipt and the number of requisitions that are backordered. Fill rate is defined mathematically as follows:

Fill rate = Requisitions Filled / Requisitions Received.

For example, if 50 requisitions were received in a given period, and 43 of them were filled immediately and 7 were backordered, then a fill rate of 86% was achieved for this period. The above calculation can be applied to a specific item or to a group of items. When it is applied to a group of items, it can be done in one of two ways. First, the fill rate can be calculated as an average of all the individual item fill rates. Or, the fill rate can be calculated with the above equation without regard to what the particular item is. This is also called demand weighting, because it is equivalent to a weighted average of item fill rates, weighted according to the demands of the individual items. In this study, we use demand weighted fill rate unless specifically noted otherwise.

In the past NAVSUP WSS used commercially-developed optimization software to maximize their achieved fill rate given their budget constraints. Developed by MCA Solutions, the Service Planning and Optimization (SPO) was effective but had shortcomings. First, it was a "black box" to the users at NAVSUP WSS, who did not have access to the models and algorithms SPO used to develop its solutions. SPO did not have the ability to accept budget as a constraint. Therefore, NAVSUP WSS had to run SPO iteratively, adjusting a fill rate constraint until a satisfactory budget figure was reached. Additionally, SPO was expensive, costing around $800,000 per year in licensing fees. In order to replace SPO with a better-functioning optimization tool at reduced cost, Naval Postgraduate School faculty developed the Wholesale Inventory Optimization Model (WIOM) (Salmeron and Craparo 2017). WIOM is a mixed-integer linear program designed to maximize a function closely related to fill rate, for the wholesale inventory managed by NAVSUP WSS. Roth (2016) used simulation modeling to conclude that WIOM 3.51 was in fact superior to SPO in maximizing fill rates. NAVSUP WSS sunset SPO and began using WIOM in April of 2017. While WIOM performs well compared to SPO, NAVSUP WSS identified further features they would like to be incorporated into WIOM. First, WIOM 3.51 did not use demand weighting. Instead, it had two settings that could be used. First, WIOM could treat each National Item Identification Number (NIIN) equally. This is not desirable because it ignores the relative importance of NIINs with high demand. Alternatively, WIOM could give preferential treatment to NIINs that were assigned to specific groups called level setting strategy indicators (LSSIs). By assigning high-demand NIINs to a certain LSSI and then assigning that LSSI a high weight, NAVSUP WSS could mitigate the demand weighting issue. Additionally, NAVSUP WSS could use a series of business rules to create low-demand cutoff points, choosing to leave very low demand NIINs out of the optimization altogether. In order to address this concern, WIOM was revised to use demand weighting, and incorporated this change into the WIOM 4.1 release.

NAVSUP WSS has an additional concern with WIOM (and SPO before it): churn. Churn is the change between solutions from one model run to the next. NAVSUP WSS runs the optimization model once every quarter. In the three months between model runs, the number of requisitions received changes the demand parameters that feed into WIOM. Subsequently, the optimization problems are quite different and considerably differing solutions are possible. Indeed, if multiple optimal (or near-optimal) solutions exist, churn may occur even in the absence of changes to the input data. Churn creates an administrative burden in contracting and can reduce senior leadership's confidence in optimization efforts. To deal with the problem, Salmeron and Craparo (2017) included a term in WIOM's objective function that calculates a churn penalty. This term contains two penalty parameters. One is indexed by NIIN, allowing the user to adjust the relative importance of each NIIN within the churn term. The other is a global persistence parameter that reflects the overall importance of the churn term. This study focuses on the global persistence parameter; for simplicity we use the term "persistence parameter" hereafter. The persistence parameter rewards a solution for maintaining legacy values from one model run to the next. The parameter is not an on/off switch; rather, it is a continuous parameter that can be set from zero to an arbitrarily large number. At zero, the persistence parameter is "off." As the parameter increases, the model more strongly prefers to retain incumbent solutions. Additionally, there is an inherent tradeoff between churn reduction and achieved fill rate. The higher the persistence parameter, the less important fill rate becomes in the objective function. This paper explores this tradeoff via discrete event simulation following work by Teter (2018).

The remainder of the paper is organized as follows: Section 2 presents a literature review of related inventory models, to include optimization and simulation methods. Section 3 describes the methodology goals, data and an introduction to the simulation metamodel. Section 4 explores the effects of periodicity and persistence settings on fill rate. Finally, our conclusions are presented in Section 5.

2. LITERATURE REVIEW

2.1. Inventory Management

Wholesale inventory management is concerned with finding strategies to meet demand requirements from customers at an acceptable service level and an acceptable cost level. Many different models have been proposed, but the two we will discuss are the order-point, order-quantity (s,Q) model and the classic inventory model.

Order-point, order-quantity models are discussed in Silver et al. (1998). In an (s,Q) system, two parameters are used to make decisions on stock replenishment. The first is the reorder point, s. As an item's stock level decreases, a reorder is triggered once the item's inventory position decreases to the level of the reorder point. Inventory position is defined as the quantity on hand plus the quantity on order minus the quantity in a backordered status (i.e., owed to customers). The second parameter is the order quantity Q. This is the quantity of material ordered every time there is a reorder. When a reorder is placed, the time it takes for this order to arrive is known as the lead time. A key feature of an (s,Q) system is that each reorder is triggered by a low inventory position, not low inventory on hand. This prevents the system from placing extra orders when there is already an order due-in that will replenish stock sufficiently. Silver et al. provide an analogy: "A good example of ordering on the basis of inventory position is the way a person takes aspirin to relieve a headache. After taking two aspirin, it is not necessary to take two more every five minutes until the headache goes away. Rather, it is understood that the relief is 'on order'— aspirin operates with a delay" (Silver et al. 1998).

WIOM uses the (s,Q) system to model NAVSUP WSS's wholesale inventory. However, NAVSUP WSS only determines reorder points. The quantity of the reorders is decided by Navy Enterprise Resource Planning (ERP), and is treated as input by NAVSUP WSS, who then strives to maximize effectiveness by deciding on appropriate reorder points.

A special case of the (s,Q) system is the classical inventory model discussed in Tersine (1994). The classical inventory model uses an (s,Q) system but with a very rigid set of assumptions. Among other things that are not relevant to our purposes, the classical inventory model assumes the following: deterministic and constant demand; constant deterministic lead time; reorders arrive as a whole lot of size Q; and backorders are not allowed, since constant demand and lead time allow backorders to be avoided with certainty. The resulting system creates a characteristic saw-tooth

pattern. This inventory model is used primarily as a means to estimate an order quantity that minimizes cost, known as the economic order quantity (Silver et al. 149-197). Since NAVSUP WSS treats the order quantity as a given input from ERP, we are not concerned with that aspect of the model. However, the model has some unique qualities that we will use when establishing initial conditions for our simulation. Specifically, a result of the model is that the average amount of inventory on hand is equal to Q/2. Furthermore, the inventory on hand at any given time is distributed uniformly from zero to Q.
Bachman et al. (2016) combine discrete event simulation with optimization for inventory models to manage items with demand that is either infrequent or highly variable. Simulation-based optimization is also proposed by Köchel and Nieländer (2005) to define optimal inventory policies in multi-echelon systems.

2.2. Discrete Event Simulation

Discrete event simulation is addressed in detail in Law (2015). Discrete event simulations are those that advance time from one discrete event to the next. These events may change the state of the system being represented, and the system cannot change during the time between events. Law presents several concepts to understand such a simulation: System state; Simulation Clock; Event List; and, Initialization Routine.
We develop a simulation using this next-event time advance principle. Events in the system are arranged in time in an event list. The simulated time moves forward from one event to the next according to the events' arrangement in time. The current event is evaluated, state changes to the system are made as necessary, and the simulation moves to the next event in time while the simulation clock is updated.

2.3. Previous WIOM Simulation Study

Roth (2016) conducted a comparative simulation study between three different optimization methods: simple calculation (a heuristic), SPO, and WIOM. Using a discrete event simulation and testing across five types of material, Roth concluded that WIOM was the best performing of these three alternatives. However, Roth's simulation relies on several strong assumptions:

- NIIN demand probability distributions are known and unchanging through time;
- NIIN demands arrive in quantities of one only;
- Demands are uncorrelated between NIINs.

In addition to these assumptions, the simulation models a lengthy warm-up period of 400,000 days to reach steady state. Due to these assumptions and warm-up period, Roth's simulation would be ineffective to try to model short-term performance of the system with frequent WIOM runs and changes in estimated demand distributions every quarter.

3. METHODOLOGY

3.1. Goals

The work creates a discrete event simulation that uses historical requisitions as input and requires no warm-up period. We call this simulation the Comparative Optimized Results Simulation (CORS). By using historical data and not requiring a warm-up period, CORS allows for multiple runs of WIOM during the test period. We conduct a series of experiments using the simulation and analyzes the output in order to gain insight into: (a) the relative tradeoff between churn and fill rate using differing settings for the persistence parameter, and (b) the effect of WIOM periodicity on fill rate.
In practice, NAVSUP WSS has historically used a set of business rules to help it overcome limitations in SPO. These business rules include mandating minimum and maximum reorder points for some NIINs, which restrict the range of solutions that SPO can use.
Additionally, WIOM accounts for churn by use of the persistence parameter and accounts for low demand items by using demand weighting. Therefore, no additional business rules will be used in this study. While exploring differing concepts of operations for NAVSUP WSS, we do not explore all possible periodicities. Running WIOM and implementing its solution is administratively burdensome, and organizationally NAVSUP WSS wants to maintain a normal battle rhythm (Ellis et al. 2017). For this reason, we assume that WIOM can only be run quarterly, semiannually, or annually.
Our work is limited to non-nuclear consumable material. Modeling repairable material is more complex and not addressed in this study.
CORS does not attempt to model all aspects of inventory management. Therefore, while the model delivers insight into performance, it only does so relatively. We are only comparing between simulations and claiming which operating condition performed better.
Using deterministic demand gives great flexibility to explore the effects of different concepts of operations that a long term steady state simulation does not. However, by using deterministic demand we are essentially restricted to one data point and a trace simulation. Thus, our conclusions are inherently limited. We can say that one concept of operations performed better than another in the simulation, but only for the given set of demands.

3.2. Data

The set of data provided includes historical requisitions. The data are a record of all demands that NAVSUP WSS received during 2013 through 2017. The next set of data provided consists of historical candidates files. These files contain information for all the NIINs that were input into SPO for each quarter. Also provided are historical wholesale data files (Ellis 2017). These files have the majority of the data elements needed to run in WIOM, but they do not include the budget category, which is necessary to classify a NIIN as a particular type of material. After excluding repairable and nuclear items, and items with inconsistent requisition data, our dataset contains approximately 3,800 maritime NIINs and over 100,000 requisitions.

3.3. Metamodel

As input, CORS requires requisition data and WIOM outputs for each quarter of the time period being tested. To obtain the necessary WIOM outputs, we start by running WIOM for the first quarter in the time period. This run uses the candidates file for the first time period developed above, the budget figure, and the persistence parameter we are exploring. The second WIOM run for the next sequential quarter requires all the same input data plus the first WIOM solution, as it uses this information to enforce persistence. The third WIOM run requires the second WIOM solution, etc. After repeating the process for all available quarters we have a library of WIOM output. This WIOM output contains both the optimal reorder points (ROPs) and the NIIN characteristics CORS requires; namely, each NIIN's lead time (LT) and order quantity (Q). This library of 18 WIOM outputs is fed into CORS, along with the requisition data. CORS then performs its simulation and outputs system performance in terms of fill rate. Figure 1 illustrates the process.

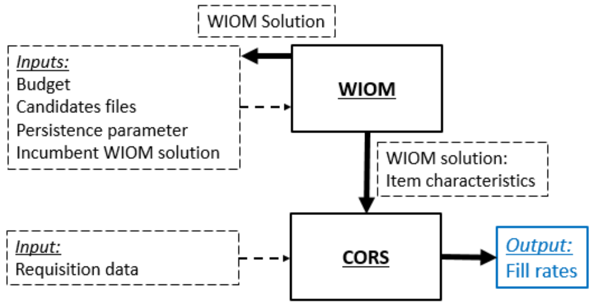

Figure 1: Metamodel Relationships

Simulating one NIIN at a time, CORS maintains an event queue with events aligned in time to trigger demand arrivals, order arrivals, and parameter changes due to new WIOM input. Each event triggers a particular logic sequence that examines the current state of the system and makes appropriate changes to the system and event queue. The simulation tracks the following system characteristics (state variables):

- Order Quantity (Q)
- Reorder Point (ROP)
- Lead Time (LT)
- Quantity On-Hand (Q_O/H)
- Inventory Position (IP)
- Time (t)

CORS output information that can be used to calculate fill rate in a variety of ways. First, the model outputs the overall fill rate for each NIIN for the entire simulation. Next the model outputs aggregate data for all NIINs that can be used to calculate the fill rates for a number of time frames. For each month, the total number of requisitions filled (across all NIINs) and the total number of requisitions received are both recorded. With these data, aggregate demand weighted fill rates can be calculated for any periodicity that is a multiple of months (i.e., quarterly, annually, etc.). Finally, the model outputs the average length all backordered requisitions stayed in a backorder status.

We implement the simulation logic in the R programming language (R Core Team 2016) to run CORS. A formal representation of the model, as well as additional logic to record statistics of system performance, appear in Teter (2018).

4. ANALYSIS

4.1. Operating Concepts Explored

We have two items we wish to explore: run periodicity and the persistence parameter. We only consider periodicities of quarterly, semi-annually, and annually. For the persistence parameter, we choose to use parameters that roughly correlate to none (0.0), low (0.1), medium (1.0), and high (5.0).

Additionally, we explore the possibility of a hybrid approach, where WIOM is run every quarter, but with different persistence parameters. In this hybrid idea, persistence is turned off in one model run per year in order for the solution to "reset" and adapt to any drift that has occurred in the demand distributions. The other three quarters the persistence parameter is set at the low, medium, or high level. Thus, the total number of concepts tested in our experiments is 15.

4.2. Effect of Periodicity in Fill Rate

Table 1 shows the list of settings for the 15 designs and the resulting "overall" (average over all NIINs) fill rates achieved under each design, as simulated in CORS. Note that WIOM does not directly maximize fill rate; rather, it minimizes a nonlinear penalty associated under-achieving fill rate goals. Nonetheless, overall fill rate provides a simple aggregate figure of merit by which to judge system performance. The results indicate clear differences in fill rate achievement between periodicities, with the greatest difference between Annual and Quarterly designs (close to 10%). Differences between persistence levels (within the same periodicity) are almost negligible, noting that higher levels have the desired effect of significantly reducing churn (to be shown later).

An interesting question is when the overall simulated fill rates start to diverge under different periodicities. For example, in the case of designs 4 and 9, the overall difference of about 10% does not occur until month 20 (see Figure 2). In our design comparisons, divergence takes even longer to take place. Based on what we see here, we expect at least two quarters before any impact of a WIOM implementation is experienced. This makes intuitive sense as the average lead time across the NIINs tested is a little more than a year.

Table 1: Overall Fill Rates by Design

Design	Periodicity	Persistence	Overall fill rate
1	Annual	0.0	51.77%
2		0.1	51.88%
3		1.0	51.85%
4		5.0	51.74%
5	Semi-annual	0.0	58.34%
6		0.1	58.08%
7		1.0	58.45%
8		5.0	58.01%
9	Quarterly	0.0	61.57%
10		0.1	61.16%
11		1.0	61.43%
12		5.0	60.90%
13	Annual, then Quarterly	0.0, then 0.1	61.53%
14		0.0, then 1.0	61.46%
15		0.0, then 5.0	61.32%

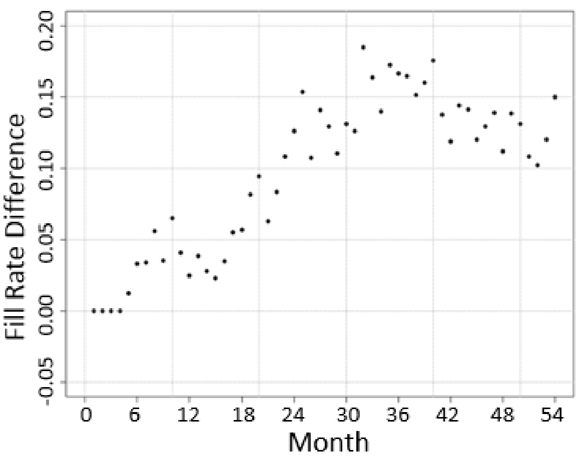

Figure 2: Monthly Differences in Simulated Overall Fill Rates between Designs 4 and 9

4.3. Effect of Persistence Level

The persistence parameter does not directly set a certain level of churn. Rather, it is a change in the weighting of the objective function for the WIOM optimization model. So, we must first analyze the effect of the persistence parameter on churn, and then analyze the effect on fill rate performance.

It is important to note here that we are comparing churn, which is calculated in WIOM, against simulated fill rate performance, which is not calculated exactly: The optimization uses a closed-form approximation amenable to calculations. The simulation calculates it more accurately, as shown by Roth (2016). The purpose here is not to compare the relative values of the two terms in WIOM's objective function. Rather, our goal is to compare churn against simulated system performance. Having shown that annual and semi-annual concepts perform poorly, we restrict the persistence analysis to quarterly periodicities only.

4.3.1. Persistence Level on Churn

The persistence parameter in WIOM enforces persistence by applying a penalty when the safety stock of a NIIN differs from the previous safety stock level. The safety stock is the expected quantity on hand when an order arrives. The penalty (*Total Churn*) can be defined by the following expression:

$$\textit{Total Churn} := \sum_{i \in I} \frac{\left|\hat{S}_i^0 - S_i\right|}{\hat{S}_i^0 + 1}$$

where I is the set of all NIINs, $\hat{S}_i^0$ is the safety stock of NIIN i in the incumbent solution, and is S_i is the new (optimized) safety stock.

We note the churn penalty for a single item i is proportional to the relative magnitude of the change. For example, a change of solution from 9 to 10 incurs a penalty of 0.1, while a change from 9 to 19 incurs a penalty of 1.0. To compare the churn across our quarterly designs, we compute *Total Churn* for every quarter, and take the mean value across the simulation time period for each concept of operation design. The results of these calculations are shown in Table 2.

Table 2: Total Churn by Quarterly Design

Design	Persistence	*Total Churn*
9	0.0	5,673
10	0.1	600
11	1.0	154
12	5.0	50

The designs using a constant persistence parameter every quarter show a clear reduction in churn with increasing persistence parameter. The highest persistence parameter tested has, on average, less than 1% the churn present with the parameter set to 0.0.

Next, we look at hybrid concept designs, 13, 14, and 15, separating average churn rates when the persistence parameter is equal to zero and when it is not (Table 3).

Table 3: Total Churn for Hybrid Designs

Design	Persistence	Average with zero persistence	Average with positive persistence	Overall
13	0.0, then 0.1	8,816	572	2,512
14	0.0, then 1.0	9,391	155	2,328
15	0.0, then 5.0	8,925	49	2,137

We make two observations. First, in the quarters when persistence above 0.0 is used, average churn for designs 13, 14, and 15 is very similar to average churn for designs 10, 11, and 12, respectively (see Table 2). The next observation is that the large overall average churn for the hybrid designs comes from the annual runs with persistence set to 0.0. In these designs churn is very high during the annual "reset" of the WIOM solution but effectively reduced during other quarters.

The above analysis shows that the persistence parameter reduces churn. However, this definition of churn is abstract and mathematical, and there is no immediate understanding of what its values mean to the system. An alternate way to express churn that is more intuitive is to define it as the proportion of NIINs that had any change in safety stock. While WIOM does not use this definition (nor does it pursue such a goal in the objective function), we expect this measurement to decrease in concert with WIOM's definition of churn, and we wish to know if it does not. Using this alternate definition of churn as a proportion, we calculate the average across the simulation period for the different designs in Table 4. As expected, increasing the persistence parameter reduces the proportion of NIINs that have a change in safety stock. However, the reduction is less dramatic than that reflected in the churn formula. The churn formula calculated churn at persistence parameter level 5.0 as less than 1% of the churn at persistence parameter 0.0. Using this alternate definition, the fraction of items with any churn is reduced by about 75% when the persistence parameter is increased from 0.0 to 5.0.

Table 4: Churn Redefined as % of Items with Change

Design	Persistence	Average items with churn
9	0.0	39.99%
10	0.1	30.72%
11	1.0	16.12%
12	5.0	10.35%
13	0.0, then 0.1	36.73%
14	0.0, then 1.0	27.34%
15	0.0, then 5.0	23.10%

Table 5: Churn as Change in Absolute Dollar Value

Design	Persistence	Average churn ($ million)
9	0.0	6.29
10	0.1	4.68
11	1.0	2.83
12	5.0	1.95
13	0.0, then 0.1	5.49
14	0.0, then 1.0	4.55
15	0.0, then 5.0	4.08

A third way to define churn is by dollar value. For any given NIIN, we can define a change in the stock cost as the absolute value of the change in the solution times the unit cost of that NIIN. This dollar value is sometimes referred to as "execution cost." As before, this is a value of interest to the analyst but not one that WIOM pursues by design. Tables 5 shows the results, and we observe similar behavior as in the two previous definitions for churn.

4.3.2. Churn versus Fill Rate

One of the fundamental questions of this study is the trade-off between churn and fill rate performance. For this analysis we use WIOM's original calculation of churn. We start by looking at the relationship between churn value and fill rate for our seven quarterly designs. A graph of these points is presented in Figure 3. However, it is important to note that we are graphing the simulated fill rates achieved over the time period. We are not attempting to find the Pareto curve of efficient solutions, which would be applicable to the two components of the objective value calculated by WIOM. Rather, we are trying to get an idea of the trade-off of between fill rate performance and churn achieved in a production-type environment.

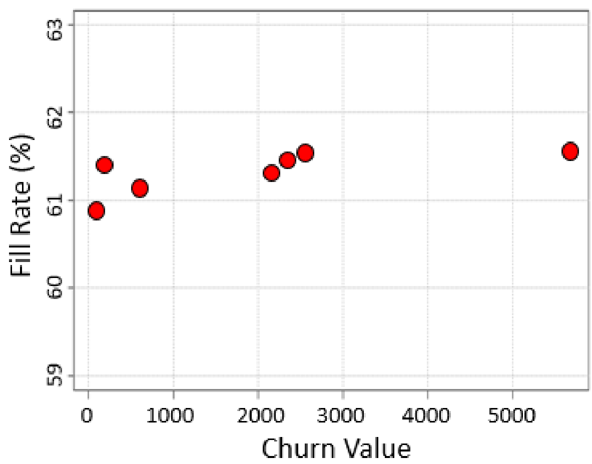

Figure 3: Fill Rate by Churn Value

It appears that there is a very slight increase (improvement) in fill rate associated with an increase (degradation) in churn, which is what we expect. But, we have few data points and the increase is very slight. Reductions (improvement) in churn are very "cheap" in terms of fill rate for these levels of persistence parameter for this set of historical demand.

4.4. Additional Persistence Exploration

Based on the results of the quarterly concepts from our original experimental design, we observe only a small trade-off relationship between churn value and simulated fill rate. However, we know that at some level a larger trade-off exists. The annual and semi-annual designs effectively have churn-free solutions in the quarters that WIOM is not run. These designs have clear degradation in fill rate compared to the quarterly designs. Therefore, there must be some threshold of churn improvement that causes greater levels of simulated fill rate degradation. However, the persistence parameters we explored did not create churn reduction that crossed that threshold. We therefore conduct a new experiment with higher settings of the persistence parameter to find this threshold and find a steeper trade-off between churn and fill rate.

We add three new concepts of operation to our experiment. We use quarterly runs with the persistence parameter set at 10, 100, and 1000. For this analysis we exclude the hybrid designs. Our results show that increasing the persistence parameter above 5.0 only marginally decreases churn, and increasing it over 10.0 has a negligible effect. This is shown by a clear “knee” in the curve shown in Figure 4.

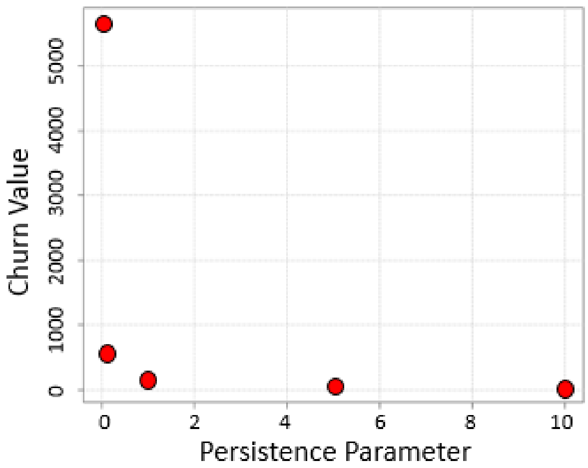

Figure 4: Extended Testing of Churn

As the increase in persistence parameter has little effect on churn, it also has little effect on fill rate performance, which remains fairly constant around 60%.

5. CONCLUSIONS

This work develops a simulation model, CORS, in order to explore the effects of different concepts of operation for wholesale inventory optimization. These concepts of operation vary in terms of the periodicity that WIOM is executed and the persistence parameter used. We explore a variety of different concepts of operations using CORS and we measure system performance for each design in terms of simulated fill rate and churn.

Through the course of this research we have gained several key insights into NAVSUP WSS’s wholesale inventory system. The first insight is that it takes time for different implementation concepts to differentiate in terms of fill rate. Even very clearly different solutions take at least six months to produce different fill rates. It takes even longer for the magnitude of the difference to become clear. This insight is important because it reminds us to be cautious in judging the performance of the system in the short term.

Our next key insight into the system is that WIOM solutions have a short shelf life. The system changes sufficiently over time that there are clear degradation to fill rate performance for semi-annual designs and dramatic degradation for annual designs. While different solutions take time to diverge, it is important for the optimization model to be able to adjust to changes in the underlying demand structure quickly. We see no reason to recommend a change to the quarterly periodicity that NAVSUP WSS currently uses.

Perhaps our most important finding is that, for the historical demand considered, churn can be drastically reduced without sacrificing system performance in terms of fill rate. By implementing the use of the persistence parameter, NAVSUP WSS can gain significant improvement in churn, which reduces administrative burden in contracting and improves the ability to explain WIOM results to senior leadership. All this improvement can be gained without sacrificing fill rate performance and support to the fleet.

Our final finding is somewhat unexpected: It appears that WIOM has a limit to how far it can enforce persistence. Beyond a certain point, increasing the persistence parameter has no practical effect on churn. Even increasing the persistence parameter several orders of magnitude has virtually no effect on churn. This may be due to side constraints on WIOM reorder points preventing them to match legacy values. It could also be due to WIOM’s budget constraint: If the incumbent solution is too costly for the current budget, the lowest feasible value of churn will be strictly positive.

The results found by this study became one input that led to the decision of maintaining the current periodicity by which NAVSUP WSS sets quarterly inventory levels.

It is also important to discuss what we did not find. Our first important caveat concerns the lack of reduction in fill rate with increases in the persistence parameter. In this particular case, we observed that the limit that persistence could be enforced was above the critical threshold where it would impact simulated fill rates. In this way, we could increase the persistence parameter to an arbitrarily large number and not affect fill rates. However, we do not have evidence that this is true generally. It may well be that this is simply a coincidence of this particular type of material, for these demands, and at this budget level.

The next important caveat is that our conclusions are based on only 4.5 years of data. We showed that simulated fill rates did not degrade with increases in the persistence parameter for this time period only. We also showed that the difference between a good and bad concept of operations takes time to develop. It is possible that some level of persistence does impact long-term fill rates when viewed from a longer term horizon.

ACKNOWLEDGMENTS
The authors thank Naval Supply Systems Command for the support of this research.

REFERENCES
Bachman T, Williams P, Cheman K, Curtis J, Carrol R (2016) PNG: Effective Inventory Control for Items with Highly Variable Demand, *Interfaces*, Vol. 46, No. 1, January–February 2016, pp. 18–32.

Ellis D, Cardillo J, Motter, M (2017) Information regarding NAVSUP WSS operations and WIOM implementation provided to author in person via personal communication during site visit to NAVSUP WSS, June 13–15.

Köchel P and Nieländer U (2005) Simulation-based optimisation of multi-echelon inventory systems, *International Journal of Production Economics*, Vol. 93–94, pp. 505-513.

Law A (2015) *Simulation Modeling and Analysis*, 5th ed. (McGraw-Hill Education, New York, NY)

NAVSUP (2019) NAVSUP Enterprise. Accessed April 15, 2019, https://www.navsup.navy.mil/public/navsup/enterprise/.

Roth G (2016) A simulation of alternative for wholesale inventory replenishment. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/handle/10945/48587.

Salmeron J, Craparo E (2017) Wholesale Inventory Optimization Model. Naval Postgraduate School, Monterey, CA.

Silver E, Pyke D, Peterson R (1998) *Inventory Management and Production Planning and Scheduling*, 3rd ed. (John Wiley and Sons, Hoboken, NJ).

Tersine R (1994) *Principles of Inventory and Materials Management*, 4th ed. (PTR Prentice Hall, Upper Saddle River, NJ).

Teter S (2018) Simulated operating concepts for WIOM implementation. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/58269

AUTHORS BIOGRAPHY

Lieutenant Commander Sean Teter (United States Navy) is currently serving as an Operations Research Logistics Analyst and Distribution Planner for U.S. Transportation Command. His work focuses on the practical application of quantitative techniques to military logistics and supply problems. Operationally, he has served at sea onboard the ballistic missile submarine USS WEST VIRGINIA and the amphibious assault ship USS BATAAN. LCDR Teter received his B.A. from Florida State University and his M.S. from the Naval Postgraduate School.

Dr. Emily Craparo is an Associate Professor of Operations Research (OR) at the Naval Postgraduate School (NPS). Her research and teaching interests focus on mathematical optimization and the design of efficient algorithms. Dr. Craparo has provided analytical and modeling support to the US Army, Navy, and Marine Corps on a variety of topics related to decision making under uncertainty and resource constraints. Her academic publications include topics such as sensor placement, optimal operation of energy systems, and capital planning under uncertainty. Prior to joining the NPS OR faculty, Dr. Craparo completed National Research Council Postdoctoral Fellowship at NPS. Dr. Craparo received her S.B., S.M., and Ph.D. from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology (MIT).

Dr. Javier Salmerón is a Professor of Operations Research at the Naval Postgraduate School (NPS) where he teaches graduate courses in optimization and operations management. His research focuses in the area of applied optimization, where he has developed multiple tools to guide decisions in both civilian and military problems. He has published in multiple peer-reviewed journals and is the recipient of several research and teaching awards. Before his time with NPS, Dr. Salmerón worked for Spanish electric utility Iberdrola and was adjunct professor at the Statistics School of Complutense University of Madrid.