Journal Title
XX(X):1–16

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Assessing Uncertainty and Risk in an Expeditionary Military Logistics Network

Brandon M. McConnell[1] and Thom J. Hodgson[1] and Michael G. Kay[1] and Russell E. King[1] and Yunan Liu[1] and Greg H. Parlier[1] and Kristin Thoney-Barletta[2] and James R. Wilson[1]

Abstract

Uncertainty is rampant in military expeditionary operations spanning high intensity combat to humanitarian operations. These missions require rapid planning and decision-support tools to address the logistical challenges involved in providing support in often austere environments. The U.S. Army's adoption of an enterprise resource planning (ERP) system provides an opportunity to develop automated decision-support tools and other analytical models designed to take advantage of newly available logistical data. This research presents a tool that runs in near-real time to assess risk while conducting capacity planning and performance analysis designed for inclusion in a suite of applications dubbed the Military Logistics Network Planning System (MLNPS) which previously only evaluated the mean sample path. Logistical data from combat operations during Operation Iraqi Freedom (OIF) drives supply requisition forecasts for a contingency scenario in a similar geographic environment. A nonstationary queueing network model is linked with a heuristic logistics scheduling methodology to provide a stochastic framework to account for uncertainty and assess risk.

Keywords

logistics, capacity planning, queueing, networks, forecasting, risk analysis, nonstationary arrival process, time-dependent arrival rate, dispersion ratio, index of dispersion for counts

1 Introduction

This paper presents a decision-support tool to support military logistics associated with sustaining an expeditionary force in response to the call to action from the 2005 RAND study[1] on logistical issues during the invasion of Iraq. The tool is designed to run in near-real time to develop feasible plans, rapidly assess alternatives, identify logistical capacity requirements to support expeditionary operations, and assess the associated uncertainty and risks relevant to military planners and decision-makers. The model is restricted to a general expeditionary scenario but not limited to an invasion. Modeling efforts focus on repair parts (U.S. class of supply nine, CL IX) but include food and water (CL I) and ammunition (CL V) as they share similar sustainment resources.[2] By design, the tool is focused on sustainment and excludes the time phased force deployment data (TPFDD) problem of getting a deploying units personnel and equipment into the theater of operations. The paper extends Rogers et al.[3] by incorporating uncertainty and the use of a time-dependent variance correction to enable risk analysis.

The rest of this paper is organized as follows: Section 2 motivates the paper, Section 3 introduces related work; Section 4 establishes this research's perspective on risk and presents a stochastic framework for risk analysis; Section 5 outlines the requisition demand forecasting methodology; and Section 6 demonstrates the model on a notional operation. Section 7 provides closing thoughts.

2 Motivation

In March 2003 during Operation Iraqi Freedom (OIF), within a week of the fall of Baghdad, the 3rd Infantry Division experienced equipment readiness for key ground combat systems dropping from 90% to under 70% due to distribution problems for CL IX; the 2005 RAND study examining these problems concluded these distribution problems stemmed from a lack of automated decision-support tools capable of generating analysis fast enough to support the rapid pace of operations.[1]

3 Related Work

Rogers[4] proposes a Military Logistics Network Planning System (MLNPS) that harnesses both the Army's new enterprise resource planning (ERP) system called the Global Combat Support System—Army (GCSS-A) and mission-based forecasting (MBF) to assist decision makers and planners with several aspects of military logistics.[5] As opposed to using historical average usage rates for repair

[1]Industrial & Systems Engineering Department, North Carolina State University, Raleigh, NC. [2]Department of Textile and Apparel, Technology and Management, North Carolina State University, Raleigh, NC.

Corresponding author:
Brandon McConnell, Campus Box 7906, Daniels Hall 400 North Carolina State University, Raleigh, NC 27695-7906.
Email: bmmcconn@ncsu.edu

parts, MBF provides a tailored forecast using stratified sampling that considers unit type, geographical environment, and planned type of operation.[6–11] Some key MLNPS functions include identifying required capacities across the logistics network to support an expeditionary operation, anticipating bottlenecks, conducting what-if analysis, and course of action analysis. MLNPS is a significant contribution as it is the only known decision-support tool designed to use the Army's ERP data.

The MLNPS fundamentally models the logistics network as a large factory with each logistics node represented as a machine in the factory. Supply requisitions (e.g. CL IX) act as jobs that must be processed through this factory. With this approach, MLNPS exploits an engine known as the Virtual Factory (VF) to optimally (or near-optimally) schedule these requisitions across the network in near-real time to minimize the maximum lateness; lateness is the requisition completion time minus the due date. Hodgson et al.[12] introduce the VF in 1998 which has sustained a number of improvements: most notably, Thoney et al.'s[13] addition of batch processing. Rogers[4] finds this heuristic optimization provides a good forecast for real system performance and demonstrates significant insights that would have impacted planners in 2003.

The ability to handle batch processing allows the VF to schedule for multiple network locations while also accounting for transportation between locations since a truck may be modeled as a batch processor itself. Trainor[14] and Melendez[15] use batch processing in conjunction with the VF to address military deployments (also see Hodgson et al.[16]). Rogers et al.[3] extend this idea by focusing on repair parts transiting the military logistics network by using nested batch processors to model requested repair parts being loaded into pallets which are loaded into containers then shipped via surface (ocean) vessels.

Using the VF to process this data along with a MBF for future demand, the MLNPS terminates with a near-optimal sequence to maximize customer (requesting unit) satisfaction by minimizing the maximum lateness. Rather than focusing on the sequencing, Rogers et al.[3] demonstrate the VF provides a forecast of when, where, and how much queuing will occur at various nodes across the logistics network. Validating their model against real performance data from the 2003 invasion of Iraq during Operation Iraqi Freedom (OIF) demonstrates the MLNPS can accurately approximate real logistics network performance. Using this model and mean value parameters, they develop a trial-and-error method to test drive a logistics plan and assess how well it will support a planned expeditionary operation using deterministic outputs from the VF. Their model assesses the logistics network both for the invasion of Iraq and for a notional intervention scenario set in Africa. With two analysts, model setup took 1–2 days from scratch (hours if already created) with run times taking minutes to hours depending on time horizon and the size of the network studied; to our knowledge, this is faster than contemporary models.

MLNPS has tremendous potential as a military logistics model for several key reasons: it supports end-to-end analysis in near-real time (a capability deemed critical by Army leadership) and is designed to incorporate both GCSS-A data and MBF.[17,18] As illustrated by Figure 1, GCSS-A can provide the model with the current location and status of all requisitions currently in the system. The model requires the planned logistics network and necessary constraints as an input from the planner. MLNPS then forecasts how well the network will perform, where significant queuing will exceed thresholds causing sustainment problems, and provides a decision-maker with information to adjust the plan or notify subordinate units of what to expect as the operation unfolds.

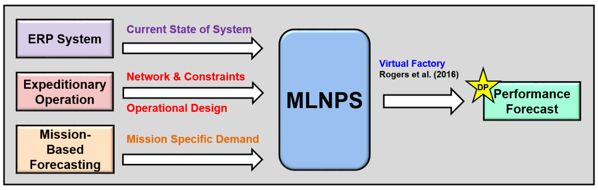


Figure 1. Conceptual overview of the Military Logistics Network Planning System (MLNPS).[3] Note: DP: decision point.

The Army's new ERP system (GCCS-A), with its ability to provide data on demand, provides unprecedented access to data. The Army must harness that data and transform it into useful, actionable risk information for decision-makers. While the uncertainty cannot be altogether eliminated, GCSS-A makes it possible for tools to account for this uncertainty.

4 A Stochastic Framework That Permits Risk Analysis

4.1 Risk in the Military Logistics Context

As a deterministic tool, the MLNPS does not adequately assess uncertainty or permit analysis of risk. The term *risk* is ubiquitous in both military and nonmilitary applications but often lacks precision or quantification.[19] This paper defines risk as a set of possible outcomes with negative consequences. For this definition, assessing risk requires estimation of both the severity and likelihood of those outcomes.[20,21] For a more detailed review of supply chain, military, government, and transportation infrastructure risk, see McConnell[22] (Sec. 3.3).

This view of risk is justified to assess *operational risk* faced by the military logistics network in the form of unanticipated requirements and disruptions from events caused by enemy action, terrain, and weather.[6] Beyond these considerations, a commander faces risks that address timing, performance, or other concerns. These may be measured by any number of qualitative or quantitative metrics that frame how the commander views risk. The framework for risk analysis must be flexible to accommodate different notions of risk and address various sources of risk, even if only through what-if functionality. Further complicating the analysis is the fact that risk has both time and location components that greatly increase the dimensionality – risk can evolve over time as the operational plan unfolds and might be concentrated in different logistic nodes as a function of time.

This research's perspective on risk analysis extends Alderson et al.'s[23] definition of *operational resilience*: the ability of a system to adapt its behavior to maintain

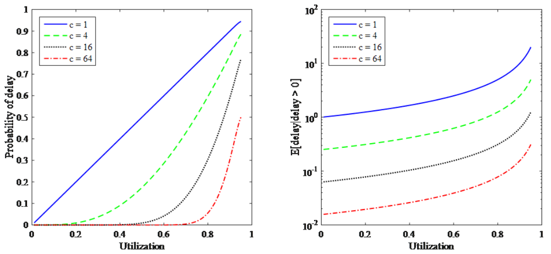


Figure 2. Performance of $M/M/c$ queues (for $c \in \{1, 4, 16, 64\}$) demonstrating it is possible to increase both efficiency and quality through staffing; utilization (ρ) is the arrival rate (λ) divided by the total service rate ($c\mu$), $\rho = \lambda/c\mu$.

continuity of function (or operations) in the presence of disruptions. Alderson et al.[23] link infrastructure resilience to system operation (function) and focus on disruptions. Our model must enable what-if analysis by capturing both severity and likelihood.

4.2 Relevant Advances in Queueing Theory

Queuing theory provides a rich and convenient source of tools to apply to this problem as it can account for uncertainty in both arriving requisitions and processing times through a network. Recent advances also integrate time-varying properties while relaxing the mathematically convenient but not always realistic Markovian assumptions from classical textbooks. Queuing theorists largely concern themselves with two basic questions that align with our stated research objectives: (1) given a stochastic system's properties, estimate its performance (performance analysis) and (2) given a performance target, estimate the system properties required to achieve the desired performance (capacity planning).

Recognizing that the number of servers at a queue is analogous to the capacity of a logistics node, it is clear that queuing theory can assist MLNPS in setting capacities (or estimating required capacities) to achieve better performance. Figure 2 (mimicking Mandelbaum & Momčilović[24]) demonstrates that by increasing the system scale, it is possible to increase quality of service without loss of efficiency – in other words, properly setting the capacity can benefit both quality and efficiency.

We focus on queuing theory that supports staffing time-dependent (nonstationary) queueing systems and non-Markovian (non-exponential) properties, particularly arrival and service processes. Many studies present evidence of time-varying system properties in everything from hospitals, call centers, and even trucks at a seaport.[25–30] It is well known that Markovian approximations to non-Markovian systems can perform poorly, particularly if sufficient variation exists in the arrival process.[31–33] Jennings et al.[34] use an infinite server (IS) and Normal approximation to choose the time-dependent staffing function, $s(t)$, to stabilize the probability of delay for a $G_t/GI_t/s_t$ model with nonstationary non-Markovian arrivals (the first G_t) and nonstationary identical and independent general service times (the GI_t). Liu and Whitt[35] use IS models to develop offered-load (OL) and modified offered-load (MOL) approximations for the time-dependent staffing required to stabilize the expected delay and abandonment probabilities for the $M_t/GI/s_t + GI$ queue. This approach is extended to a feed forward network structure of $M_t/GI/s_t + GI$ queues.[36] He et al.[33] extend established staffing procedures to allow for non-Markovian arrival processes using a heavy-traffic limit that applies to systems where some delay is expected. Their study of the impact and interplay between arrival process variability and service time distributions demonstrates that the variance correction used by Jennings et al.[34] is robust for nonstationary models. Readers wanting a more thorough treatment should see Defraeye & Nieuwenhuyse.[37]

4.3 Modeling the Military Logistics Network

The military logistics network stretching from U.S. depots to the expeditionary theater of operations can be modeled as a queueing network where logistical nodes (or processes) are queues and the requisitions are arriving as orders via GCSS-Army. The simplified model in Figure 3 illustrates how orders are sourced, picked and packed, then shipped to the ordering unit. The model is a feed-forward queuing network as supply requisitions move from the sourcing depot across the network to the ordering unit and do not require rework or depart the network through a lateral exit. In the simplified network shown, requisitions are arrivals in the queuing theory terminology and arrive according to the requisition forecast discussed in Section 5. There are multiple classes of arrivals based on the mode of transportation to be used whether military air, a contracted point-to-point service denoted world-wide express (WWX), or surface shipment (ocean freight). Within each arrival class are many subclasses defined by the specific route that requisition must take through the network, largely defined within a class by the sourcing node and the ordering unit.

We obtain this feed-forward structure through several important assumptions. First, we assume all requisitions are appropriately handled, routed, and moved along the required path from the point of supply to the point of consumption — in other words, we do not model lost, misrouted, or frustrated cargo. This is reasonable due to the sheer size of the network and the immense requisition flow volume. In addition, the military is actively embracing interconnected technologies such as radio-frequency identification, two-dimensional bar codes, satellite-based communications, and others that help prevent these logistical mishaps.[38–42] This paper models repair parts (CL IX) along with food and water (CL I) and ammunition (CL V); other classes of supply are beyond the scope of this paper.

The feed-forward structure implies that by estimating the performance or capacity required at an upstream location, if the departure process is obtainable it also provides the arrival process to the downstream nodes so the analysis can be repeated for downstream locations. Consider the simplified military logistics network depicted as a feed-forward queuing network in Figure 3; supply requisitions arrive at either Defense Depot Susquehanna, Pennsylvania (DDSP), or to an alternative depot acting as a sourcing point (SP1 or SP2). The requisitions from alternative depots still move to DDSP as it is the primary consolidation and containerization point (CCP)[43] for palletizing and containerizing shipments.

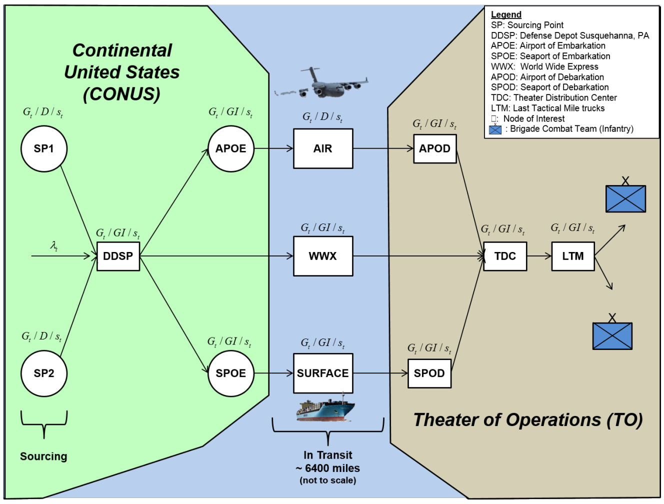


Figure 3. Military logistics network as a queueing network (simplified for illustration).

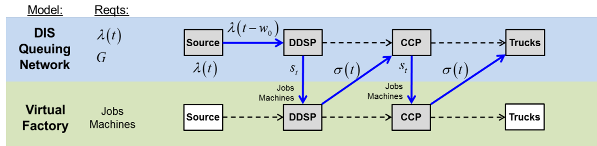


Figure 4. Visualization of the Virtual Factory Delayed Infinite Server (VF-DIS) offered load network model. The models complement each other — our unified approach denoted in bold (blue). The dashed lines represents using the two single model approaches. Note: w_0 represents the average elapsed time from a requisition establishment to the release to the source depot.

As Rogers[4] discusses, most requisitions travel via military aircraft, world-wide express, or surface shipment. Shipments traveling on military aircraft (ocean freight) are palletized (containerized) at DDSP then loaded onto military aircraft (container ships) at the airport (seaport) of embarkation, or APOE (SPOE) for movement into theater where they are offloaded at the airport (seaport) of debarkation, or APOD (SPOD).[43] These shipments are then transported to the theater distribution center (TDC). Requisitions being shipped via WWX travel directly to the TDC. At the TDC, pallets and containers are broken down and moved forward via military transportation to the ordering units. This final process is often referred to as the last tactical mile (LTM) due to the transportation occurring through a designated combat zone.

Using mission-based forecasting (MBF) and the data from GCSS-Army, it is possible to estimate the arrival processes to the network shown in Figure 3. Since MBF is not available for all unit types, Section 5 presents a data-driven approach to forecasting requisitions for a given scenario. This forecast provides an estimated nominal time-varying arrival rate to the upstream sourcing nodes.

4.4 Virtual Factory Delayed-Infinite Server Feed Forward Model

4.4.1 Network Perspective To assess risk and answer the research objectives, this paper uses a tandem Virtual Factory Delayed Infinite Server (VF-DIS) offered load approximation for each location; the model alternates between using the VF and the DIS approximations for each location moving across the network from upstream to downstream. The appeal lies in leveraging each submodel's strengths. The Virtual Factory is efficient and can handle nested batch processing (e.g., a multipack inside a pallet on a truck) as well as a number of realistic constraints, and with small refinements, the DIS model[35] provides a computationally efficient method to evaluate performance and predict capacity requirements while communicating a sense of the uncertainty in the predictions. This tandem approach uses each model to the maximum potential while avoiding each model's weaknesses.

The approach starts with the data-driven requisition forecast and the performance target of average delay desired for each of the network locations. These performance targets are obtained from stakeholders, proposed for the sake of continued analysis, or derived from senior leader interactions. The forecast conforms to requirements described in Section 4 but in general is both nonstationary and non-Markovian. We assume an empirical or fitted theoretical distribution is available for the processing time at each location in the network to account for stochastic variation; these may be general (non-exponential).

Figure 4 offers a visualization of the VF-DIS process with a small portion of the network. It depicts the network as two complementary models, the DIS and the VF approaches, with the bold (blue) line representing the VF-DIS hybrid solution approach for an analyst conducting risk analysis and military logistics planning. The dashed line represents the model workflow if only using a single model in isolation from the other. Taking the logistics network, processing logic, and requisition forecast, the $\lambda(t)$, as inputs, VF-DIS first uses the VF at the upstream nodes to assess the arrival process and the resulting departure process given the default time-dependent capacity plan, the s_t, which can be initialized via a simple constant capacity. With the average delay targets for the upstream nodes, the DIS model calculates the time dependent capacity required to meet the targets by providing both a time dependent average (the s_t) and the stochastic variability around that average. These capacity plans (functions over time) are inputs to the VF which accounts for location-specific policies, logic, and schedules; the VF then returns the time dependent departure process represented as $\sigma(t)$. Since the network is feed-forward, the departure process is the downstream arrival process so the model advances downstream once the upstream planning is complete and repeats the process until it reaches the most downstream nodes.

The VF can model location-specific policies via its processor logic. These are important because some locations do not work weekends, and resources vary according to specific schedules driven by real-world considerations. Another example stems from a working policy used by a container packing location (one of several batch processes): a container is considered full when it reaches its effective capacity, reaches the minimum capacity to send and there are no orders left to pack, or when it has been sitting open (and partially filled) for at least three days. The VF's processor logic conveniently incorporates these nuances.[3]

Planners may obtain the performance targets for each node from senior leader guidance, experts, or simply a staff estimate as part of the planning process. If not derived from successful historical performance or leader guidance, staff planners may choose to use a range of targets they believe to be feasible and present them with their resulting impacts to the staff and/or leadership for analysis and decision. The calculations this choice permits is the science, but choosing an appropriate performance target is clearly part of the art of the planning method described here.

The VF-DIS performs the same steps at every location. If that location is not of interest or cannot be impacted by available decisions, there is no need to linger performing analysis after obtaining the departure process. However, when the VF-DIS model reaches a node of interest such as the TDC or the LTM trucks, it may be desirable to identify a logistics plan for that location to meet senior leader performance guidance and assess the associated risk. Section 4.4.2 provides the technical details for this process.

4.4.2 Node Perspective Consider an arbitrary node in the logistics network. Let w be the average delay taken as the performance target for this node and denote the average arrival rate on day t as $\lambda(t)$; the VF provides this nonstationary arrival rate using the data-driven requisition forecast discussed in Section 5. The logistics node is represented as a $G_t/GI/s_t$ queueing model where the processing time, S, at this location is independent and identically distributed (i.i.d.) according to the general distribution F_S. Figure 5 labels this Model 1. Define $\sigma(t)$ as the average departure rate on day t for the associated departure process. Denote $Q(t)$ as the queue length at time t. Let $B(t)$ be the number of busy servers at time t with mean $m(t) = \mathrm{E}[B(t)]$ and variance $v(t)$; $B(t)$ is approximately

$$B(t) \sim \text{Normal}(m(t) + 1/2, z(t)\, m(t)), \tag{1}$$

with

$$\mathrm{E}[B(t)] = \int_0^{(t-w)^+} \lambda(t - w - x)\bar{F}_S(x)dx, \tag{2}$$

where the notation $(x)^+ = \max\{x, 0\}$, and the service time complementary cumulative distribution function, $\bar{F}_S(x) = 1 - F_S(x)$. Equation 2 calculates the average capacity over time required to maintain the performance target w.

Define $\{A(t),\, t \geq 0\}$ as the counting process that tracks the number of arrivals (events) by time t, then the arrival process *index of dispersion* (for counts), $I(t)$, is the variance-to-mean ratio of the cumulative number of arrivals (events) as given by Equation (3). If the node sees M_t arrivals according to a nonhomogeneous Poisson process (NHPP), $I(t) = 1,\, t > 0$.

$$I(t) = \frac{\text{Var}(A(t))}{\mathrm{E}[A(t)]},\, t > 0 \tag{3}$$

The arrival process is *overdispersed* if $I(t) > 1$ and *underdispersed* if $I(t) < 1$. If the arrival process to Model 1 is significantly overdispersed then a naïve implementation of the DIS offered load approximation will underestimate risk to the decision-maker because while the predicted average will be true, the model will underestimate the variance. Using the Sudan scenario from Rogers et al.[3], Figure 6 provides an estimated dispersion for the LTM trucks revealing the LTM arrivals are over five times more variable than a NHPP. In other applications, healthcare clinics often see underdispersion (≈0.4–0.6) in appointment-based systems but overdispersion (≈1.5–2.5) in emergency departments or at call centers.[44–47]

This research assumes that with GCSS-A data an analyst can estimate the dispersion for a given location (Section 5 enables estimating the dispersion using multiple sample paths for the requisition forecast).

If arrivals occur according to a NHPP, then $I(t) = 1,\, \forall t$, and for a fixed t, $B(t)$ is a Poisson random variable with mean $\mathrm{E}[B(t)]$ (2).[35,48] This implies $m(t) = v(t)$. If the arrival process is not a NHPP, then $m(t) \neq v(t)$ and the heuristic risk correction factor (RCF), $\tilde{z}(t)$, enables the approximation

$$v(t) \approx \tilde{z}(t)m(t), \tag{4}$$

with the correction factor

$$\tilde{z}(t) = \max\{z(t), 1\}, \tag{5}$$

$$z(t) = 1 + \frac{(c_a^2(t) - 1)}{E[S]} \int_0^\infty [1 - F_S(x)]^2 dx, \tag{6}$$

where

$$c_a^2(t) \approx \frac{\text{Var}(A(t-w) - A(t-w-\eta))}{\int_{t-\eta}^{t} \lambda(u-w)du}, \tag{7}$$

for a chosen $\eta > 0$.

The capacity recommendation follows the square root staffing (SRS) rule

$$s_\gamma(t) = \left\lceil \mathrm{E}[B(t)] + \delta_\gamma \sqrt{\text{Var}(B(t))} \right\rceil, \tag{8}$$

establishing a buffer to hedge against stochastic variability by establishing a probability γ and associated quality of service parameter δ_γ such that $P(\mathrm{N}(0,1) > \delta_\gamma) = \gamma$, which exploits the Normal approximation to the Poisson.

4.4.3 Discussion Figure 5 graphically depicts the node specific approximation. Model 2 ($M_t/GI/s_t$ queue) can approximate Model 1 (see Figure 5) as it captures the time-dependent fluctuations in the arrival process with the same deterministic mean function while retaining Model 1's non-exponential service time distribution and time-varying capacity. The only difference is in the arrival process variability. Convenience motivates the transition to Model 2 and requires a correction to account for the arrival variability. There is no reliable method to obtain analytical approximations for the capacity required over time to meet performance targets for Model 1 and use of simulation will not permit the analysis to be performed in near-real time. From the network perspective, Model 2 enables Poisson superposition at a downstream node that receives requisitions from multiple upstream nodes; the aggregate arrival process is then a nonhomogeneous Poisson process (NHPP).

To identify the capacity to achieve the average delay target w for Model 2, Liu and Whitt[35] show that the Delayed

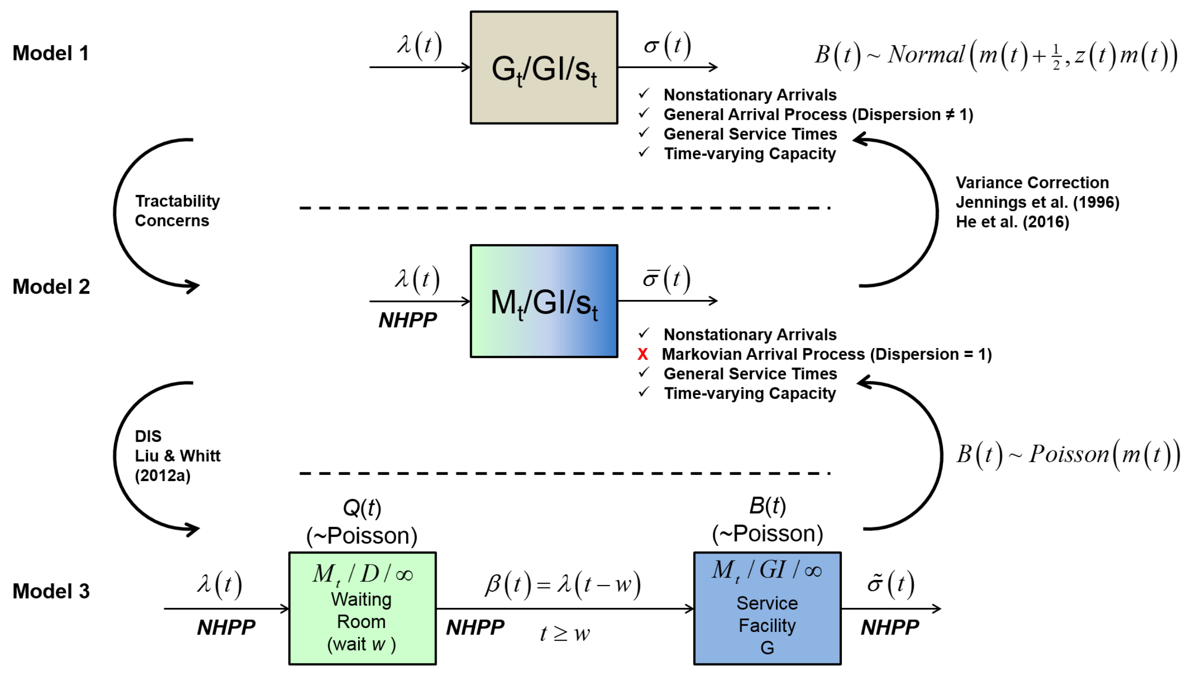


Figure 5. Queueing models for a location. Model 2 approximates Model 1 by focusing on the time-dependent behavior. Model 3 is the Delayed Infinite Server (DIS) offered load approximation for the $M_t/GI/s_t$; Liu & Whitt[35] show Model 3 approximates Model 1. Within Model 3, the contents of the first two queues, $Q(t)$ and $B(t)$ respectively, are independent Poisson random variables for a fixed t. See Table 1 for departure rate, $\tilde{\sigma}(t)$.

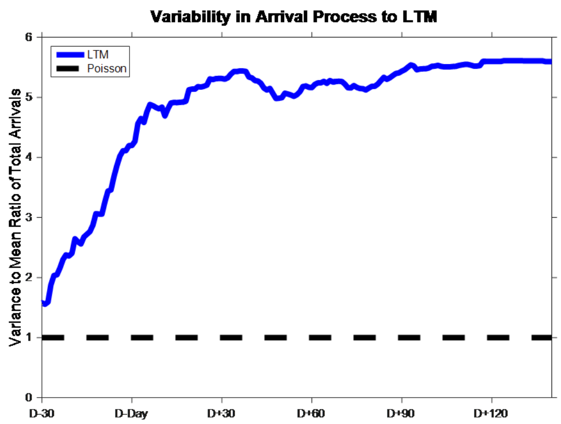


Figure 6. LTM arrival dispersion for the Sudan scenario is over 5 times more variable as a NHPP (dashed line). Graph shows arrival dispersion in 463L pallet equivalent units (PEU) from Equation 9.

Infinite Server (DIS) offered load approximation works well for systems such as the military logistics system where requisitions may be expected to wait for some (even small) amount of time before processing. The DIS model, depicted as Model 3 in Figure 5, approximates Model 2 by using two infinite capacity queues in series. This presentation of the DIS model omits the abandonment process implying the assumption that there is no lost, misrouted, or frustrated cargo. The first queue represents the waiting space in Model 2; the second queue represents the service facility of Model 2.

The infinite capacity implies the departure process from each queue in Model 3 is also a NHPP which is computationally and mathematically critical in the feed-forward network. The idea behind Model 3 is simple. Requisitions arrive to the first queue (waiting area) and wait a deterministic amount of time equal to the target average delay, w, before continuing to the second queue. This implies an average arrival rate of $\beta(t) = \lambda(t - w)$, $t \geq w$, at the second queue which is just a deterministic time shift.

At the second queue, requisitions immediately begin processing according to the general service time distribution $G = F_S$. The objective is to determine the number in the second queue at time t, $B(t)$, which serves as the first order approximation of the number of busy servers in Model 2 while maintaining an average delay of w. In simpler terms, Model 3 approximates Model 2 by having all requisitions wait the desired average delay then simply observes how many busy servers would be in the second queue over time. Based on the mathematics of the infinite server queue, this is obtained via direct calculation using Equations 1–2 applying known results for the $M_t/GI/\infty$ queue (see Theorem 1, Eick et al.[48]).[35]

While Equation 2 calculates the average capacity over time required to maintain the performance target, the mathematics of Model 3 fully specify the approximate distribution of this predicted capacity requirement. Section 6 exploits this idea and uses Monte Carlo methods to estimate what is not available analytically. Table 1 summarizes the remaining analytical formulas for the DIS offered load approximation.

The DIS approximation prediction for Model 2 with M_t arrivals implies $B(t) \sim \text{Poisson}(\text{E}[B(t)])$. Correcting for the variability of the G_t arrival process requires application of a result from the stationary $G/G/\infty$ queueing model as a heuristic which is consistent with Jennings et al.[34](see

Table 1. Model 3 DIS approximations for Model 2 (assumes M_t arrivals).

Performance Feature	DIS Approximation (for a fixed t)
Queue Length, $Q(t)$	$\sim$Poisson with mean E$[Q(t)] = \int_0^{t\wedge w} \lambda(t-x)dx$
Number of Busy Servers, $B(t)$	$\sim$Poisson with mean E$[B(t)] = \int_0^{(t-w)^+} \lambda(t-w-x)\bar{F}_s(x)dx$
Departure Rate, $\tilde{\sigma}(t)$	$\sim$NHPP with time-varying rate $\tilde{\sigma}(t) = \int_0^{(t-w)^+} \lambda(t-w-x)dF_S(x)$
Total Number in System*, $X(t)$	$X(t) = Q(t) + B(t)$

Notes: **System* refers to a specific node; $t \wedge w = \min\{t, w\}$; $(t-w)^+ = \max\{t-w, 0\}$.

Sec. 6) and He et al.[33](see Sec. 3). This paper is the first to use a time-shifted variance correction to integrate the DIS approach.

The function $\tilde{z}(t)$ (5) does not allow a reduction in variance (an optional modeling assumption due to lack of dispersion data — this avoids a false sense of certainty) and can increase the variance for $B(t)$ using the time-dependent generalization of the heavy-traffic peakedness (6) taken from Whitt's[49] treatment of the stationary $G/G/\infty$ model which characterizes the variance-to-mean ratio of the steady-state number of busy servers. Since Model 1 is nonstationary, (6) is a heuristic. Equation (6) assumes a stationary service time distribution but this can be relaxed. Equation (7) is a time-dependent generalization of the asymptotic variability parameter and is similar in form to (3); this paper time-shifts the asymptotic variability parameter by w to account for the DIS approximation. The parameter η determines the dispersion estimate in the local interval $[t-\eta, t]$; this paper uses a timestep of one day and numerical evaluations confirmed $\eta = 1$ is a good choice for this application. The intuition is that both the arrival variability (7) and the service time variability (6) affect the variance of the capacity prediction (number of busy servers). Specifically, the tail of the service time distribution drives the impact on the variance with the term $\int_0^\infty [1 - F_{S_t}(x)]^2 dx$ in (6).

Since we estimate the arrival variability parameter (7) empirically, it can be undefined when $\int_{t-\eta}^{t} \beta(u)du = 0$ which occurs during lulls in arrivals on η consecutive days. When this occurs, we define $c_a^2(t) = 0$ for this special case such that the RCF $z(t) = 1$ which aligns with engineering intuition.

This RCF (5) corrects the variance in the approximation for Model 2 so the final results may serve as an approximation for Model 1. The Normal approximation to the Poisson implies that instead of $B(t) \sim \text{Poisson}(m(t))$ which underestimates risk, the required capacity is actually approximated by $B(t) \sim \text{Normal}(m(t) + 1/2, z(t)m(t))$. Adding a half to the mean function corrects for the conversion between a discrete distribution to a continuous one but may be omitted in practice if desired. This continuity correction may lead to a positive bias in the capacity forecast when the nominal requirements are relatively low and the continuity correction has a larger relative impact on the mean. In Section 6 which demonstrates this technique, $B(t)$ is a Normal distribution truncated on the interval $[0, +\infty)$ to prevent the sampled $z(t)$ from pushing probability below zero; this is equivalent to $(B(t)|B(t) > 0)$. Figure 5 displays an overview of the entire process to predict capacity requirements for a single location. The VF is responsible for both the arrival rate to each node, $\lambda(t)$, and determining the departure rate $\sigma(t)$ as it is designed to handle location specific packing policies, work schedules, and other realism constraints.

Much of the queueing literature cited is motivated by staffing requirements for call centers and often employs (8) over a discretized time horizon. In the case of expeditionary military logistics such continuous control is not likely to be possible, even over long subintervals. Section 6 addresses this concern and also demonstrates a technique to use the VF-DIS model structure to generate possible capacity plans for a location given the time-varying risk information.

Combined with the tandem VF-DIS approach for the network, this approximation provides a framework to describe the average requirements that fluctuate over time as well as the stochastic variation around that deterministic prediction. In short, this includes both severity and likelihood in the predictions.

While the feed-forward structure imparts computational efficiency and suggests a simple sequential approach, the time-varying (nonstationary) property and presence of non-Markovian (non-exponential) arrival and service processes greatly complicate the analysis. The reader will notice the fundamental modeling unit traversing the network is changing as well and there is no common unit for capacity (number of servers); Rogers[4] uses number of requisitions per day for DDSP, 463L pallets[50] per day for the APOE, and forty foot container equivalents (FEU) per day for surface freight, and even the number of medium truck companies as a capacity unit. Queueing theory has no easy way of accounting for these unit changes.

To simplify the analysis, the model employs a 463L pallet equivalent unit (PEU) as the base unit of capacity to be used throughout the network. For reference, Figure 7 portrays airmen loading a 463L pallet holding multipack containers into a military aircraft. This unit is reasonable as it readily converts to twenty foot container (TEU) and forty foot container (FEU) equivalent units widely used in the logistics practitioner community. The CCP packs requisitions and multipack boxes into containers. The TDC requires an equivalency unit as it breaks down both pallets and containers and organizes them with individual requisitions for onward transport. Given a particular Army truck company equipment list, known as the modified table of organization and equipment (MTOE)[43], it is possible to convert between PEU and a collection of logistics distribution resources.

Establishing the PEU unit also provides an opportunity to account for the tare weight. One can calculate the number of PEUs arriving (or departing) location m on day t by taking the total weight and volume for that day and location and dividing by the effective maximum capacity of the 463L pallet given by Equation 9 below. The larger of the two

Figure 7. Depiction of a 463L pallet with multipack containers in Al Taqqadum, Iraq (Jan. 9, 2009). Seabees assigned to Naval Mobile Construction Battalion (NMCB) 7 and airmen of the Air Force Expeditionary Logistics Readiness Squadron Detachment 4 maneuver 463L pallets into an Air Force C-17 for transportation.[50]

determines the 463L pallet equivalent capacity required; note for resources the smaller of the two is the offered resource capacity. Table 2 lists the 463L maximum and effective capacity. The use of cubic inches preserves integrality in the model for computational efficiency and adequately models the many requisitions with less than a unit cubic foot volume.

$$P_{mt} = \max\left\{\frac{TotalWT_{mt}}{95\%\text{ PEU WT}}, \frac{TotalCU_{mt}}{85\%\text{ PEU CU}}\right\} \quad (9)$$

Using 95% of the maximum weight accounts for tare weight. Using 85% of the maximum volume accounts for the inevitable voids between contents that prevent using all of the physical container volume. Whether a tractor trailer or a shipping container, spaces are considered full at the 85% volume utilization point.

Table 2. Maximum and effective capacity of the 463L pallet.

	Weight (lbs)	Volume (cu in)
100% 463L Capacity	10000	838080
Effective PEU	9500 (95%)	712368 (85%)

5 Generating a Data-Driven Demand Forecast

The MLNPS provides a convenient set of tools for analyzing performance of logistical courses of action as well as a means of identifying the capacities needed to hit performance targets. Both of these capabilities require forecasted demand from certain classes of supply such as food and water, ammunition, and repair parts over the studied time horizon. Since MBF, driven by consumption data, is not available for platforms outside of Army aviation, we use available modern combat data as a surrogate to generate the demand forecast. Acknowledging this is supply-side data and therefore a faulty signal for true demand, this data-driven process mimics some MBF techniques such as stratifying on unit type and mission intensity in order to generate the best demand forecast possible without true MBF.

5.1 Data-Driven Approach

It is critical to use modern combat data to get a representative picture of modern combat repair part demand. To forecast demand for different missions in varying operational environments, the process presented here would be replicated using data generated under those conditions. While order data may not be a perfect demand signal, when properly characterized this data provides an initial approximation of the demand required in absence of the consumption-driven MBF. The model focuses on food and water, ammunition, and repair parts primarily because the data describe repair parts and they all share common resources across the distribution network.

An author from the 2005 RAND study[1] provided OIF repair part requisitions and U.S. Transportation Command (TRANSCOM) provided data on all requisitions DDPS processed in 2003. Together these datasets provide the weight, volume, sourcing depot, requisition date, requesting unit and location, and other factors for all repair part orders processed by DDSP and those destined for units in Kuwait and Iraq. For specific operational characteristics, the process relies on data from the first 87 days of OIF which consists of 647,189 individual requisitions (11.6 million parts) spanning the two weeks prior to crossing the line of departure through the end of May 2003. The DDSP-specific data contains the 7.7 million requisitions from 2003. Both datasets have a time precision of days which aligns with the chosen time step for this work. Due to similar terrain and environmental factors the OIF dataset is sufficient for the analysis which considers a notional operation in Sudan.

5.2 Process Overview — A Sample Path Approach

Generating a demand forecast requires three key inputs– the task organization, concept of the operation, and the timeline–which are estimates produced during the military planning process. The task organization lists what units are conducting the operation and may change over time. The specific tasks (missions) for these units are found in the concept of the operation. These products provide the analyst with the specific details (who, what, when, where, and why) for the operation.

The timeline is essential to breaking down the time horizon of interest into distinct missions for the units. The forecast must account for the fact that repair part demand is different by both unit type (think infantry versus aviation units) and type of operation (preparing for combat versus combat). This is accomplished by defining three operational intensity levels (OIL) then using the OIF invasion data to characterize each type of units demand under those operational descriptions to describe pre-combat (OIL 1), steady-state operations out of an established base (OIL 2), and direct (major) combat operations or high intensity conflict (OIL 3). Since these levels are qualitative in nature, Table 3 provides recommended guidelines for identifying these levels using historical data.

Table 3. Guidelines to Identify Operational Intensity Levels.

Operational Intensity	Distinguishing Features
Level 1	☐ Prior to crossing line of departure ☐ Not conducting combat operations ☐ Preparation for combat operations
Level 2	☐ Operations conducted from established and secured base or fixed location ☐ Operations take on a steady-state, routine nature during this period
Level 3	☐ Conducting invasion ☐ Major combat operations / high intensity conflict

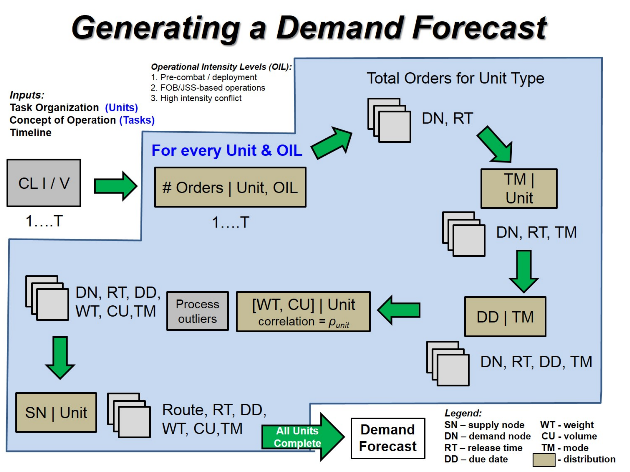


Figure 8. Process overview to generate a demand forecast. More details available in McConnell.[22] The irregularly shaped (blue) box indicates the steps required for every unit type and OIL.

Figure 8 graphically shows the forecasting process which begins by generating requisitions for the food, water, and ammunition supplies that must sustain the units in the model. Then a specific workflow addresses repair part demand for every unit for every OIL. This workflow determines the number and timing of requisitions, how those requisitions are routed to the ordering unit, the required delivery date, order weight and volume, and the sourcing depot. Once this has occurred for every unit for every operational intensity level, the concatenation of these requisitions constitutes the forecast. Due to how the forecast employs probability distributions, the resulting forecast is a projected sample path for that time horizon.

5.3 Modeling the Demand Process

5.3.1 Food, Water (CL I), & Ammunition (CL V) Unit size plays an obvious role in estimating CL I requirements. Since bottled water was the primary source of potable water during OIF and this research focuses on initial expeditionary operations, we generate pallets of bottled water to supply the units originating at the TDC.[1,4] This assumption may vary across the time horizon as are the starting locations of those pallets depending on the logistics plan. Army Tactics, Techniques, and Procedures (ATTP) 4-41, Army Field Feeding and Class I Operations recommends ration cycles and feeding plan guidelines (see Sec. 5.3 of Rogers[4]) to identify the capacity required to haul CL I to the units so the remaining capacity may be allocated to other classes of supply.

Assuming an average Brigade Combat Team (BCT) strength of 4,500 Soldiers with attached enablers that increase the troop count by 10% yields approximately 4,950 personnel per BCT. Consistent with doctrinal water planning guides, assume 7.27 gallons of water per person per day is required in an arid environment to account for hydration, hygiene, and feeding purposes. The standard planning factor estimates a pallet may hold up to 228 gallons of bottled water.[51] Using a 10% breakage planning factor, this yields a daily water requirement of 173 pallets of water per day for a BCT.[4] A pallet holds 576 Meals-Ready-to-Eat (MREs) and assuming initial rations are three MREs per day, the daily food requirement is approximately 26 pallets per day. These rations are nonperishable. Based on planning factors provided by the Army's Training & Doctrine Command Analysis Center–Fort Lee (TRAC-LEE, a logistical analysis center), ammunition requirements would be approximately 60% of the water weight and 40% of the water volume.[4]

5.3.2 Generating Repair Part (CL IX) Requisitions This section describes the workflow depicted in Figure 8 by the irregular shaded (blue) box. This process occurs for each unit with generated requisitions being collected into a large list along with the CL I and CL V requisitions. Depending on multiple factors such as time horizon, task organization, and mission, this process generates a significant number of requisitions. For reference, the Sudan scenario in Section 6 generates requisitions for 61 days prior to LD and 90 days of operations after. In addition to the almost 2.3 million requisitions that competed for DDSP resources during the OIF invasion used as a surrogate for global, non-Sudan demand, the random requisitions generated for just the Sudan operation is on the order of 294,149 requisitions on average (20 samples with sample standard deviation 11,538).

Analysis of the OIF invasion data clearly identified that the number of daily requisitions varies both with the type of unit and based on that unit's OIL. For a given unit, the model assumes each day as independent and identically distributed within an OIL as the data showed weak autocorrelations. This assumption is not limiting and can be relaxed. The OIF invasion data permits modeling the 8 unit types listed in Table 4. For each unit, the sub-timeline of each OIL determines the number of requisitions per day. If an IBCT has an OIL schedule as depicted in Figure 9, the number of requisitions released on days 1 through 14 requires the appropriate estimated probability distribution for OIL 1. In Figure 9, $Nrel^{(j)}$, $j = 1, 2, 3$, is the daily number of requisitions ordered (released by GCSS-A) for OIL level j. Table 4 lists unit types the model can support using the OIF data taken from 6 March–31 May 2003; interested readers may refer to McConnell[22] (p. 92–94) for distribution and sample size details.

After generating a unit's requisition volume across the timeline of interest, the process randomly selects each requisition's mode of transportation according to a specified distribution which identifies the route. The Sudan operation

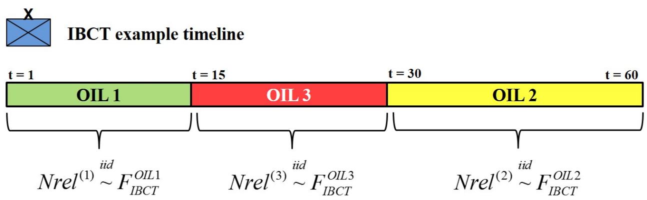


Figure 9. Example for generating number of requisitions by day for IBCT. Note: $F_{unit_type}^{OIL_level}$ is the estimated probability distribution for the number of requisitions required by day for a particular unit type at a specific OIL.

Table 4. Unit types supported by model. Relies on historical data taken from OIF invasion 6 March – 31 May 2003.

Unit	Operational Intensity Level (OIL)
IBCT	1, 2, 3
HBCT†	1, 2, 3
AVN BN	1, 2, 3
EN BN	single level only
DIV HQ	single level only
Sustainment BDE	single level only
Misc Enabler BN	single level only
3x Truck CO	single level only

Notes: † now called ABCT, also used as surrogate for SBCT.

Table 5. Transportation mode distribution taken from OIF data.

Transportation Mode	Probability
Military Air	0.7340
World Wide Express (WWX)	0.1292
Surface (Ocean)	0.1368

in Section 6 employs the distribution from Table 5 taken from the OIF invasion data.

The allotted time to deliver the requisition in days, known as the standard delivery time (SDT), is generated based on a requisition's transportation mode ($TransM$). This time is added to the release time (RT) to calculate the due date (DD) according to Equation (10) where i indexes each requisition. In Equation (11), each standard delivery time (conditional on transportation mode) is modeled with a discrete uniform distribution (denoted DU). Since standard delivery times vary by requisition priority and service-specific processes, assuming the actual standard delivery times found in Table 6 from the Department of the Army Pamphlet (DA PAM) 710-2-1 (Using Unit Supply System) are the upper bounds (b_{TransM}) and allowing up to an approximate 70% reduction (a_{TransM}) to account for varying priority designations helps to account for individual prioritization.[52] For more details on requisition priorities see Rogers[4] (p. 31) and DA PAM 710-2-1[52] (Ch. 2 and Table 2.2).

$$DD_i = (SDT_i|TransM_i) + RT_i \quad (10)$$

$$(SDT_i|TransM_i) \overset{\text{iid}}{\sim} \text{DU}(a_{TransM}, b_{TransM}) \quad (11)$$

The VF requires orders to have both a weight and volume to properly capture the details of processes such as movement by truck, packing a 463L pallet, and breaking

Table 6. Bounds for Standard Delivery Times (SDT) in days modeled by the discrete uniform (DU) distribution in Equation 11. Transportation mode distribution taken from OIF invasion data.

Transportation Mode	(a_{TransM}, b_{TransM})	Probability
Military Air	(12, 18)	0.734
World Wide Express (WWX)	(10, 14)	0.129
Surface (Ocean)	(52, 75)	0.137

down and sorting packages in a forty foot container. The OIF invasion data provides an estimate of these marginal distributions. Reality requires them to be correlated. Using a Gaussian copula (see Ross[53]) allows using a different target correlation based on the unit type from Table 4 while respecting the correct marginal distributions for weight and volume (see p 94–95 of McConnell[22]). Sampled weights or volumes that exceed that requisition's transportation mode get reassigned to the maximum capacity for the offending weight or volume; this is reasonable as this only occurs less than one-fifth of one percent of the time in the data.

Each requisition is assigned to a sourcing depot conditional on the requesting unit according to the empirical probability mass function found for each unit types source depot location. These empirical distributions can vary based on order content; since the model does not specify individual types of repair parts, this approach permits capturing the unit type variation of supply depots while keeping the level of detail at overall requisition attributes. As explained by Rogers[4], certain Defense Logistics Agency (DLA) distribution centers have specialized stock—such as communications equipment at Tobyhanna, Pennsylvania—and these specializations affect these sourcing distributions.

After specifying the unit (demand node), transportation mode, and sourcing depot (supply node), it is a simple lookup from the appropriate route matrix that stores the route for a requisition going from that depot to that unit with that transportation mode. Collectively these route matrices comprise the route library which is an output of the logistical network modeling process. After identifying each requisitions route, the entire requisition forecast is completed by aggregating the CL I / CL V requisitions, each units forecast, and the forecast for global requisitions that will compete for DDSP resources. A comparable set of requisitions that routed through DDSP in OIF provides a start point for the non-expeditionary requisitions that share continental U.S. (CONUS) resources upstream.[4]

This procedure (Figure 8) provides a data-driven approach to generating requisition forecasts since MBF is not yet available for all Army units and platforms. The result is a sample path approach which requires replication to assess the uncertainty in the forecast itself which is the arrival process to the queueing network described in Section 4.

6 Risk-based Expeditionary Logistics Planning for a Notional Operation

Armed with the methodology from Section 4, this section applies the VF-DIS framework on a notional operation. The section uses the same fictional scenario used by Rogers[4] where an Infantry Brigade Combat Team (IBCT) and a

Table 7. Task Organization for Sudan mission.[4]

Task Organization
1 x Infantry Brigade Combat Team (IBCT)
1 x Stryker Brigade Combat Team (SBCT)
1 x Engineer Battalion (Construction Effects)
1 x Aviation Battalion (Attack & Lift)
1 x Sustainment Brigade
1 x Division Headquarters
3 x Battalions of Miscellaneous Enablers

Stryker Brigade Combat Team (SBCT) are conducting operations from South Sudan into Sudan against the self-styled Islamic State in Iraq and Syria (ISIS). The SBCT will operate in the outlying Darfur region while the IBCT operates in the capital of Khartoum.[4] A Division Headquarters conducts command and control for the operation and the units are provided with enablers that include engineer and aviation units; Table 7 presents the Task Organization used to generate the forecast for supply requisitions using the procedure outlined in Section 5. In his analysis, Rogers[4] evaluates different courses of action (COAs) based on potential locations for the TDC. While the techniques from this section could assist with evaluating the COAs presented by Rogers, this section focuses on Rogers selected option, Sudan COA 1, locating the TDC in Juba, the capital of South Sudan.

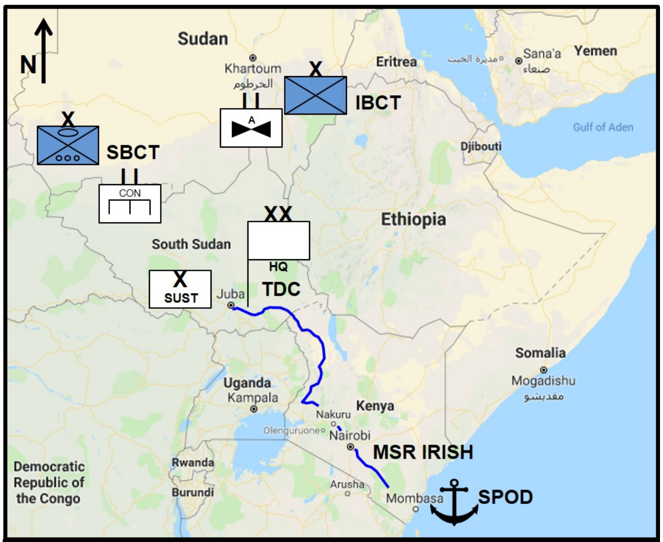
6.1 Focusing on the Last Tactical Mile (LTM)

Applying the techniques from Section 4 to the final plan recommended by Rogers[4] illustrates the contribution of this methodology. The LTM trucks—the ground units that transport supplies from the TDC to the BCT Supply Support Activities (SSAs)—are the ideal candidate for this demonstration for several reasons. The LTM trucks are the logistical link to the units which implies this may be a location where a theater commander has the most control as they are not necessarily constrained by ties to airports, seaports, or other infrastructure or process restrictions, including CONUS effects. Practically, their geographical proximity to the units also incurs more risk. Given the feed-forward structure of the network, the LTM trucks are the final node to analyze and identify the required capacity for operations; the process to analyze other nodes is almost identical except that no other node is the furthest downstream resource. This property creates a few technical challenges that are readily overcome but not present at any other place in the feed-forward network. In simpler terms, if analysis of LTM trucks is possible, it is possible to do this for any node upstream.

6.2 Overview of the Notional Operation

The operational details and timeline are identical to Rogers[4] scenario; Figure 10 provides a visual summary. The SPOD is located in the port of Mombasa, Kenya. The TDC is located in Juba, South Sudan, with the APOD nearby. Though not depicted in Figure 10, the CONUS network is also identical to LTC Rogers scenario.

The logistics network in Figure 10 is identical to the one used by LTC Rogers to enable direct comparison with key details summarized here. For more details on the



Timeline by Phase:	Location of Critical Nodes:
PH I: D-60 – D-1	SPOD: Mombasa, Kenya
PH II: D Day – D+27	APOD: Juba, South Sudan
PH III: D+28 – D+89	TDC: Juba, South Sudan
PH IV: D+90 – D+140	

Figure 10. Sudan COA 1 Overview with Main Supply Route (MSR) IRISH annotated. Note: PH: phase.

logistics network, the reader is encouraged to see Ch. 3, of Rogers (2016). To model the logistics network, we make the following assumptions:

1. CONUS dedicated truck routes depart six days a week (Monday through Saturday). Trucks travel seven days a week but may not deliver on a Saturday or Sunday.
2. CONUS dedicated trucks have unlimited capacity as DLA can quickly acquire additional trucks. Similarly, there are always enough trucks to move requisitions from source depots to CCPs or from a CCP to the APODs/SPODs.
3. Restocking the sourcing depots does not require the same resources used by the distribution system.
4. A CONUS depot sources all orders then ships them to their destination.
5. Once a multipack, pallet, or container has started loading, it remains at the location until full or the maximum waiting period has been met (whichever is soonest). Pallets wait up to three days with containers waiting up to fifteen. Multipacks, pallets and containers are considered full when they reach 95% of maximum weight capacity or 85% maximum volume capacity. If at least one of these criteria are met, prior to sending, any requisitions that could fit in the remaining space are pulled from the queue and packed (on a revised slack basis) to ship forward to the next location.
6. All pallets and containers contain repair parts for different units unless the route to a unit does not support break bulk operations. This ensures the model does not ship near empty containers.
7. If a requisition can fill an entire multipack, pallet, or container, the logistics node builds it unit-pure (no other unit orders included). Unit pure multipacks,

pallets, and containers do not require breakdown and sorting operations at the TDC and advance directly to the LTM trucks.

8. All air shipments use 463L pallets and all ocean freight employ 40 foot containers. These pallets and containers are comprised of multipack boxes and individual parts that are too large to fit in the multipacks.
9. All surface (ocean) freight travels on commercial ships. This is not a limiting assumption but fits the Sudan scenario.
10. Orders for the Sudan scenario follow the same distribution of transportation modes as occurred in OIF. This is not a limiting assumption as this is an input to the forecasting procedure.

Unlike Rogers[4], times to process requisitions are stochastic and are akin to service time distributions. The mean roundtrip times from Rogers[4] are kept fixed (10 days for IBCT, 8 for SBCT) but this research assumes LTM convoys may sometimes return up to one day early if there are good conditions but may be delayed significantly if adversely impacted by weather or mechanical failures. Without data to estimate these deviations, we assume the LTM roundtrip time is distributed via a generalized Beta distribution having the form $a + (b - a)\text{Beta}(\alpha_1, \alpha_2)$ with first shape parameter $\alpha_1 = 1.2$, and second shape parameters $\alpha_2 = 12$.[54] The generalized lower and upper bounds result from the assumption that roundtrips to support the maneuver units take a minimum of 90% of the mean roundtrip time from Rogers[4] and no more than twice the mean time.

6.3 Forecasting Required Capability to Achieve Performance Target at Location (LTM)

With the network capacities determined by Rogers et al.[3] for this scenario, the VF readily provides the sample average arrival rate in PEU to the LTM trucks using 55 sample paths. The reader is reminded multiple sample paths are necessary due to our method of forecasting repair part demand; with a different and perhaps more direct forecasting method–such as consumption data-driven MBF–the average scenario demand over time might be more accessible which would require only one run of the VF instead of the multiple runs required in this work. The analysis presented used 55 sample forecasts.

The DIS model requires a performance target for this node. Based on Rogers'[4] findings, this section uses a performance target that requires the capacity needed to achieve a requisition average (not time average) delay of seven days. The sample dispersion at the LTM is both time-varying and greater than 1 (≈ 5) which requires the RCF to adjust the forecasted variance. Without this correction, the model would underestimate risk.

The forecasted requirements presented in Figure 11 include the risk correction for LTM required capacity given in PEU. The solid bold (blue) line marks the average with shaded regions denoting the probability the required capacity is in that range on any given day. The graph is not smooth for good reasons. The LTM node is located furthest downstream and is subject to the accumulated effects of every nodes schedule nuances (e.g. some logistics nodes in CONUS do not operate on Saturdays and Sundays). With the timeline fixed, ship schedules, dedicated trucking routes, and long convoy round trip times create a very jagged forecast.

This forecast provides the framework which allows analyzing potential outcomes in a stochastic (probabilistic) sense because it provides a fully specified distribution for required capacity for every day. With this in place, it is possible to rapidly compute probabilities, calculate expectations, or even generate realizations via Monte Carlo methods. This implies that if something can be calculated or generated then it is possible to get stochastic descriptions for any metrics of interest.

6.4 Generating Courses of Action

The queueing theory that permits construction of the forecast depicted in Figure 11 implies a continuous or near-continuous control of that logistics node but that is hardly possible in the military logistics context, especially under expeditionary conditions. It may be possible to plan for significant capacity changes once per phase. Perhaps a commander can adjust resources once in the planning horizon or maybe not at all. The forecast generated with the VF-DIS model is useful for generating some default capacity options for a specific location.

Developing multiple options is attractive as the computational efficiency of these models permits rapid detailed analysis of each option and permits comparing them over time. One approach is to simply look at the requirements forecasted by Figure 11 and visually set the capacities with intuition or external knowledge about the plan; this technique might develop a plan to ensure the LTM trucks have 70 PEU for Phase I and II then only 20 PEU for Phase III and IV (presumably freeing up some capacity for other missions including a reserve). A more detached technique would be to simply use Figure 11's forecast to calculate the daily value-at-risk ($\text{VaR}_{0.95}$), also known as the 95th quantile, and plan each phase to receive the capacity set to the phases time-averaged $\text{VaR}_{0.95}$. Alternatively, a constant capacity throughout the planning horizon may be appropriate. The analysis proceeds with these three plans, though one can evaluate any given plan.

Figure 12 provides a visualization of these competing options over time with the constant option set to 78 PEU consistent with Rogers'[4] final recommended plan for the LTM trucks. The chart overlays the three LTM options against the backdrop of average required capacity, the 75th quantile for required capacity, and the 95th quantile for required capacity to convey a sense of the stochastic variation that exists about the average. These options serve to demonstrate the flexibility of this approach and the capability to evaluate any given logistical capacity plan.

6.5 Evaluating Multiple Courses of Action

Regardless of the complexity or the number of the capacity plans, the visualization provided by Figure 12 is not enough. Commanders want to understand the impacts of these plans in meaningful terms that include both expected performance and an understanding of the uncertainty involved. With the VF and the DIS models, an analyst can evaluate backlogs,

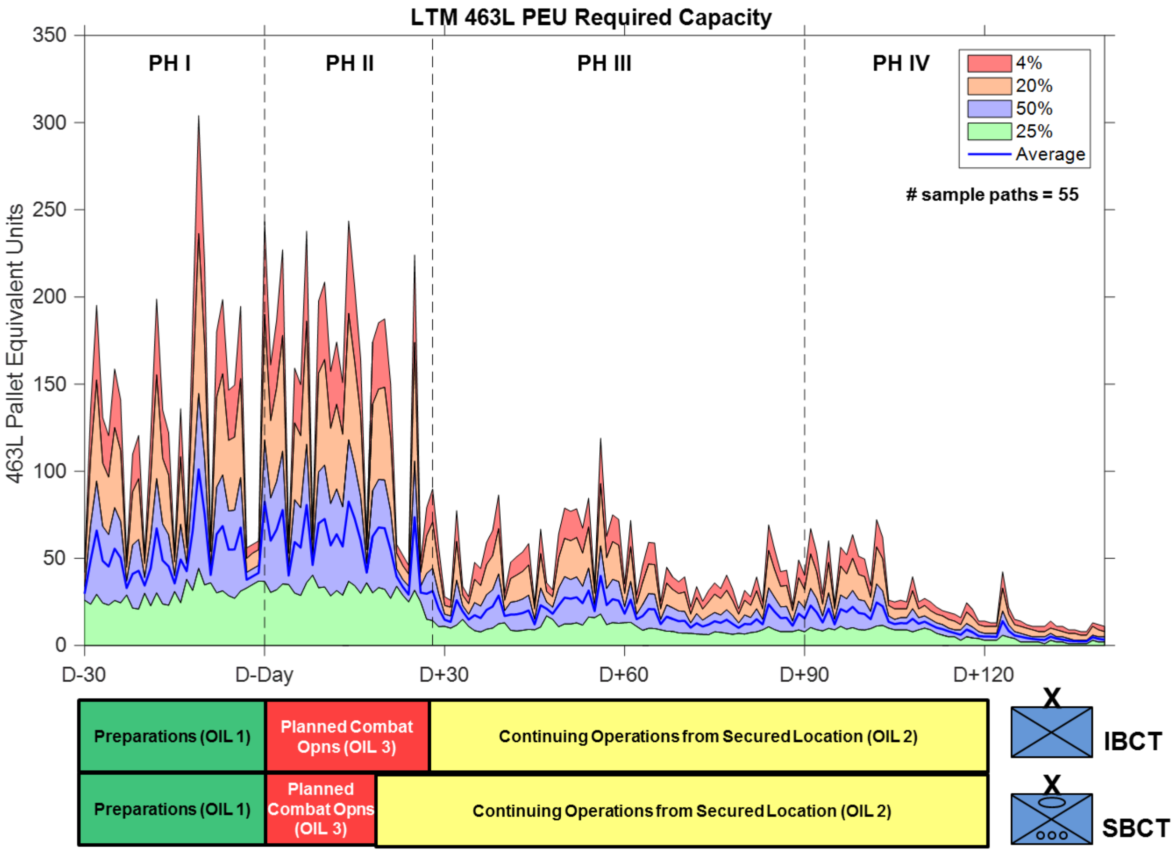


Figure 11. Forecasted Capacity Required at LTM in 463L PEU to Achieve Target Performance of 7 day Average Delay (55 sample paths). Dashed vertical lines denote phases of the operation. BCT operational intensity level timelines displayed below graph for reference.

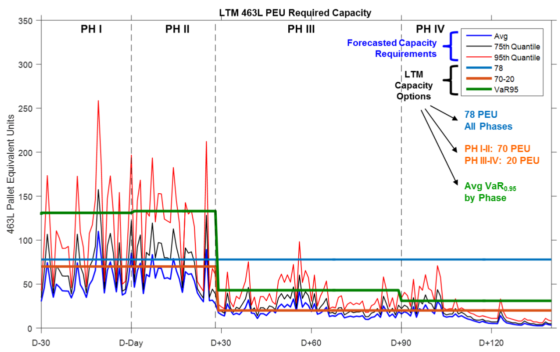


Figure 12. Generating Options for the LTM trucks in the Sudan scenario.

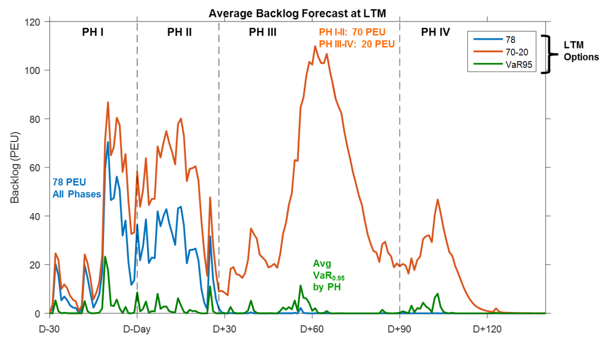


Figure 13. Average Backlog in PEU at LTM trucks for the LTM capacity plans from Figure 12.

delay, lateness, utilization, and other measures by location, by day, by requesting unit, transportation mode, or any combination of these. The model equips the analyst to dig for insights. We present some examples of initial insights that may be constructed by default to inform the decision-maker. Figure 13 shows average backlog over time for the three options as an example.

This approach extends Rogers et al.[3] by not only evaluating average delay across the time horizon but also by making stochastic information available either directly calculated analytically using the distributions (by day) or via Monte Carlo methods which are both quick and accessible with modern computers. Though Figure 13 currently shows average backlog only, it is just as easy to present confidence intervals, quantile bands, or another stochastic visualizations for a chosen metric of interest. Presenting risk-based information that depicts the uncertainty coupled with delay predictions provides a more complete understanding of the tradeoffs between multiple courses of action. Senior leaders seek to understand the risks faced and how to mitigate them

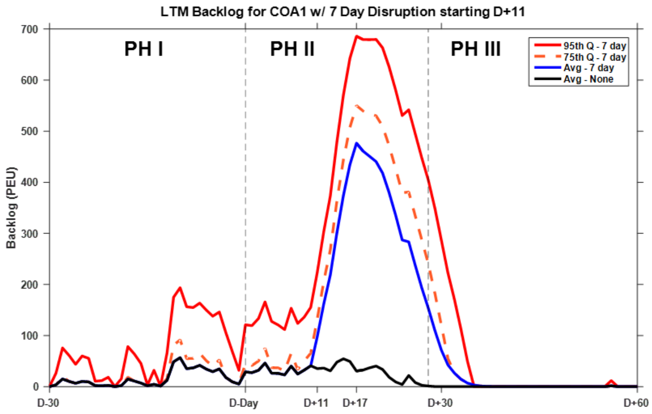


Figure 14. Daily LTM backlog with & without a 7 day disruption (starting D+11) for Sudan COA 1 scenario in Rogers[4] for 78 PEU LTM plan.

— to that end it is critical to estimate how bad things can actually get both by location and over time.

6.6 Further Evaluation of a Specific Plan

After evaluating a set of plans over time, we arbitrarily select the constant capacity plan taken from Rogers et al.[3] for further analysis. With the outputs of the VF and DIS models, an analyst can evaluate backlog, delay, lateness, and other constructed metrics over time for a specific plan and under multiple what-if scenarios. Motivated by the massive sandstorm that resulted in a seven day disruption to CL IX part resupply early in the invasion of Iraq in 2003, Rogers[4] evaluates the impact of a complete disruption of the resupply vehicles for this Sudan scenario. It is possible to perform a risk-based analysis of the constant 78 PEU capacity plan for LTM trucks from Rogers[4] under a complete disruption of the resupply vehicles.

Rogers[4] shows the value of performing what-if analysis on a given plan to assess how it performs under different potential outcomes. This VF-DIS model permits the same analysis but also shows information beyond the average by connecting potential outcomes with their likelihoods. Figure 14 illustrates this by showing the average backlog at the LTM trucks with and without a seven day disruption starting at D+11; the figure also depicts how bad the backlog can get by showing the 75th and 95th quantiles. The peak backlog with a 7 day disruption will be almost 480 PEU (7.4 times the no disruption peak). By taking into account the variation around the average, an analyst can forecast there is a 75% chance the peak backlog with this 7 day disruption would be less than 550 PEU (8.7 times the no disruption peak) if the LTM trucks have 78 PEU capacity. Similarly, there is a 95% chance the peak backlog would be no more than 685 PEU (11 times the peak when there is no disruption peak). The recovery times are also available from Figure 14.

Because of the underlying distribution forecast by day, Monte Carlo methods are readily available to assess not only average behavior over time for a specified logistics plan but the full distribution of worst-case behavior. Traditional average worst case metrics used in finance such as conditional value at risk can be readily computed but the computational speeds enable looking at more than the conditional expectation of the worst case scenarios (e.g. worst 5%). Further, the worse case distribution of a particular metric is available.[22]

The utility of risk-based measures for what-if analysis cannot be overstated. These tools enhance the MLNPS and extend the possible depth of analysis. Using multiple demand forecasts (sample paths) required multiple runs of the VF to estimate the sample arrival rate as well as the sample dispersion. If the U.S. Army continues to develop MBF beyond aviation units, this analysis would require only one run with the VF if GCSS-Army data provided the sample dispersion for the variance correction. Multiple sample paths are currently required to obtain the required capacity forecast (Figure 11) but with an approach such as MBF that does not rely on sample paths, this is obtainable with a single run of the VF.

By exploiting the strengths of both the DIS model and the VF as well as the feed-forward network structure, this approach maximizes its computational advantages. After obtaining the forecasted requirements to meet a logistics target (Figure 11), the subsequent analysis is computationally efficient using either analytical results, VF output, or simple Monte Carlo methods which are extraordinarily fast in our experience just working with MATLAB.[55]

7 Conclusion

This research contributes to the military expeditionary logistics planning problem in several ways. First, since MBF is not operational for the majority of Army platforms, formations, missions, or operational environments, we use supply-side modern combat data taken from OIF and apply MBF-style techniques, namely stratified sampling, to generate the best forecast possible for the demand signal. This forecast accounts for unit type and operational mission. Based on an operational scenario, we improve the MLNPS capabilities by adding techniques to account for uncertainty and assess risk; we achieve this improvement using a sample path-based forecasting approach, incorporating recent advances in time-varying queueing networks such as the DIS approximation, and fusing these capabilities with the strengths of the VF using the tandem approach. The VF excels at realistic tasks such as properly packing multipacks, pallets, and containers, timing the shipments, and accommodating real-world schedules; the queueing model integrates both time-varying properties and overdispersion. This is the first paper to use a time-shifted variance correction to account for overdispersion in a DIS setting.

These enhancements to the MLNPS center on two fundamental tasks: (1) given a plan, estimate the plans performance, and (2) given a target performance, find the required plan. This research provides a data-to-decision support process that yields a framework for assessing risk as both the severity and likelihood of possible outcomes become available via analytical calculation or Monte Carlo methods.

Acknowledgements

The authors would like to thank Marc Robbins from RAND, Joseph Faris from DLA Distribution, and LTC John Hiltz & Team from TRANSCOM; this paper would not have been possible without their assistance. We are indebted to the skilled professionals at ERDC, TRAC-LEE, and TRAC-MRY for taking time to showcase their analytical tools as well as provide invaluable feedback for both this paper and our military logistics research. A special thanks to Walt DeGrange for sharing his military logistics expertise.

Disclaimer

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Army, the Department of Defense, or the United States Government.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

Funding

This research was funded, in part, by a grant from the Army Research Office grant # W911NF1910055.

References

1. Peltz E, Halliday J, Robbins M et al. Sustainment of Army Forces in Operation Iraqi Freedom: Battlefield Logistics & Effects on Operations, 2005. RAND Corporation.
2. Department of Defense. Joint Publication (JP) 4-09, Department of Defense Distribution Operations, 2013. http://dtic.mil/doctrine/new_pubs/jp4_09.pdf, C-12.
3. Rogers M, McConnell B, Hodgson T et al. A Military Logistics Network Planning System (MLNPS). *Military Operations Research* 2018; 23(4): 5–24.
4. Rogers M. A logistic planning system for contingency missions to identify a feasible and efficient logistical footprint, 2016. PhD Dissertation, Department of Industrial & Systems Engineering, North Carolina State University, http://www.lib.ncsu.edu/resolver/1840.16/10968.
5. GCSS-Army website. http://gcss.army.mil/. Accessed: 2016-07-11.
6. Parlier G. *Transforming U.S. Army Supply Chains*. New York: Business Expert Press, LLC, 2011. 39–50, 145–149.
7. Office of the Secretary of Defense (OSD), Logistics and Materiel Readiness (LM&R). Maintenance Value Chain: POC Design Updates presentation, 2014. 6, 9, 11, 19–27.
8. Deputy Assistant Secretary of Defense for Maintenance Policy and Programs (DASD(MPP)). Maintenance Value Chain: Capstone Report, 2014.
9. Parlier G. Mission-based forecasting: Demand forecasting for military operations. *Foresight* 2016; 43: 32–37.
10. Schwartz B, McConnell B and Parlier G. How data analytics will improve logistics planning, 2018. White Paper, go.ncsu.edu/LogPlanGaps.
11. Evans G, Biles W and Bae K. *Analytics, Operations, and Strategic Decision Making in the Public Sector*. Hershey, PA: IGI Global, 2019. Ch 13, 286–289.
12. Hodgson T, Cormier D, Weintraug A et al. Note. satisfying due dates in large job shops. *Management Science* 1998; 44(10): 1442–1446.
13. Thoney K, Hodgson T, King R et al. Satisfying due-dates in large multi-factory supply chains. *IIE Transactions* 2002; 34(9): 803–811.
14. Trainor T. On scheduling military deployments, 2001. PhD Dissertation, Department of Industrial & Systems Engineering, North Carolina State University, https://proxying.lib.ncsu.edu/index.php/login?url=https://search.proquest.com/docview/230809668?accountid=12725.
15. Melendez B. On scheduling delivery in a military deployment scenario, 2001. PhD Dissertation, Operations Research Graduate Program, North Carolina State University, http://www.lib.ncsu.edu/resolver/1840.16/5010.
16. Hodgson T, Melendez B, Thoney K et al. The Deployment Scheduling Analysis Tool (DSAT). *Mathematical and Computer Modeling* 2004; 39(6): 905–924.
17. Shelton H. Joint Vision 2020. Office of the Chairman of the Joint Chiefs of Staff, The Pentagon, Washington, D.C., 2000.
18. Smith B. The mandate to revolutionize military logistics. *Air & Space Power Journal* 2007; 21(2): 90–125.
19. Heckman I, Comes T and Nickel S. A critical review on supply chain risk – definition, measure, and modeling. *Omega* 2015; 52: 119–132.
20. Kaplan S and Garrick B. On the quantitative definition of risk. *Risk Analysis* 1981; 1(1): 11–27.
21. Department of the Army. Army Techniques Publication (ATP) 5-19 Risk Management, 2017. Washington, D.C., Accessed 30 Nov 2017, http://www.apd.army.mil/epubs/DR_pubs/DR_a/pdf/web/atp5_19.pdf.
22. McConnell B. Assessing uncertainty & risk in an expeditionary military logistics network, 2018. PhD Dissertation, Operations Research Graduate Program, North Carolina State University, http://www.lib.ncsu.edu/resolver/1840.20/35098.
23. Alderson D, Brown G and Carlyle W. Operations Models of Infrastructure Resilience. *Risk Analysis* 2015; 35(4): 562–586.
24. Mandelbaum A and Momčilović P. Queues with many servers: The virtual waiting-time process in the qed regime. *Mathematics of Operations Research* 2008; 33(3): 561–586.
25. Buffa E, Cosgrove M and Luce B. An integrated work shift scheduling system. *Decision Sciences* 1976; 7(4): 620–630.
26. Gans N, Koole G and Mandelbaum A. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* 2003; 5(2): 79–141.
27. Green L, Kolesar P and Whitt W. Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management* 2007; 16(1): 13–29.
28. Chen X, Zhou X and List G. Using time-varying tolls to optimize truck arrivals at ports. *Transportation Review Part E: Logistics and Transportation Review* 2011; 90: 1–6.
29. Izady N and Worthington D. Setting staffing requirements for time dependent queuing networks: The case of accident and emergency departments. *European Journal of Operational Research* 2012; 219(3): 531–540.
30. Armony M, Israelit S, Mandelbaum A et al. Patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems* 2015; 5(1): 146–194.
31. Liu Y and Whitt W. Algorithms for time-varying networks of many-server fluid queues. *INFORMS Journal on Computing* 2014; 26(1): 59–73.

32. Aras A. Heavy-traffic gaussian limits for queues with nonstationary arrivals, and nonexponential service and patience times, 2016. PhD Dissertation, Operations Research Graduate Program, North Carolina State University, http://www.lib.ncsu.edu/resolver/1840.16/11022.
33. He B, Liu Y and Whitt W. Staffing a service system with non-poisson nonstationary arrivals. *Probability in the Engineering & Informational Sciences* 2016; 30(4): 593–621.
34. Jennings O, Mandelbaum A, Massey W et al. Server staffing to meet time-varying demand. *Management Science* 1996; 42(10): 1383–1394.
35. Liu Y and Whitt W. Stabilizing customer abandonment in many-server queues with time-varying arrivals. *Operations Research* 2012; 60(5): 1551–1564.
36. Liu Y and Whitt W. Stabilizing performance in networks of queues with time-varying arrival rates. *Probability in the Engineering and Information Sciences* 2014; 28(4): 419–449.
37. Defraeye M and Nieuwenhuyse V. Staffing and scheduling under nonstationary demand for service: A literature review. *Omega* 2016; 58: 4–25.
38. Hodge J. Logistics Automation Support of Joint Vision 2020 and Army Transformation, 2001. Strategy Research Project, US Army War College, Carlisle Barracks, PA.
39. Kross W. Iraqi Freedom: Triumph of Precision-Guided Logistics. *Army Logistician* 2003; 35(5): 22–23.
40. Juskowiak T and Wharton J. Joint and Expeditionary Logistics for a Campaign-Quality Army. *Army Logistician* 2004; 36(5): 2–8.
41. Cholek B and Anderson M. Distribution-Based Logistics in Operation Iraqi Freedom. *Army Logistician* 2007; 39(2): 2–8.
42. National Research Council. *Force Multiplying Technologies for Logistics Support to Military Operations*, 2014. National Academies Press, http://www.nap.edu/catalog/18832, 6, 30–31, 130–146.
43. Department of Defense. Joint Publication (JP) 1-02, Department of Defense Dictionary of Military and Associated Terms, 2016. https://fas.org/irp/doddir/dod/jp1_02.pdf, A-12, A-27, A-124, A-170.
44. Kim S and Whitt W. Are call center and hospital arrivals well modeled by nonhomogeneous poisson processes? *Manufacturing & Service Operations Management* 2014; 16(3): 464–480.
45. Kim S, Vel P, Whitt W et al. Poisson and non-poisson properties in appointment-generated arrival processes: The case of an endocrinology clinic. *Operations Research Letters* 2015; 43(3): 247–253.
46. Whitt W and Zhang X. A data-driven model of an emergency department, 2015. Technical Report.
47. Liu R, Kuhl M, Liu Y et al. Modeling and simulation of nonstationary non-poisson processes. *INFORMS Journal on Computing* 2017; Forthcoming, Accessed 26 Jul 2018, https://yunanliu.wordpress.ncsu.edu/files/2018/03/CIATAmain.pdf.
48. Eick S, Massey W and Whitt W. The physics of the $M_t/G/\infty$ queue. *Operations Research* 1993; 41(4): 731–742.
49. Whitt W. Understanding the efficiency of multi-server service systems. *Management Science* 1992; 38(5): 708–723.
50. Lavender M. US Navy Seabees and Air Force Airmen maneuver 463L pallets into an Air Force C-17 aircraft. U.S. Navy photo, Al Taqqadum, Iraq, (9 Jan 2009), accessed 7 Nov 2017, https://commons.wikimedia.org/wiki/File:US_Navy_090109-N-1120L-075_Seabees_and_Air_Force_airmen_maneuver_463L_pallets_into_an_Air_Force_C-17_aircraft.jpg, 2009.
51. US Army Combined Arms Support Command (CASCOM). Water Planning Guide, 2008. Fort Lee, VA, Accessed 29 Aug 2017, http://www.quartermaster.army.mil/pwd/Publications/Water/Water_Planning_Guide_rev_103008_dtd_Nov_08_(5-09).pdf.
52. Department of the Army. Department of the Army Pamphlet (DA PAM) 710-2-1, 2016. Washington, D.C., Accessed 25 Feb 2019, https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/p710-2-1_Web_FINAL.pdf.
53. Ross S. *Simulation*. 5 ed. San Diego, CA: Academic Press, 2013. 103–105.
54. Kuhl M, Ivy J, Lada E et al. Univariate input models for stochastic simulation. *Journal of Simulation* 2010; 4(2): 81–97.
55. MATLAB website. https://www.mathworks.com/products/matlab.html. Accessed: 2019-02-20.

ORCID iD

Brandon M. McConnell https://orcid.org/0000-0003-0091-215X
Thom J. Hodgson https://orcid.org/0000-0002-8077-4780
Michael G. Kay https://orcid.org/0000-0002-1359-8270
Russell E. King https://orcid.org/0000-0003-4576-6600
Yunan Liu https://orcid.org/0000-0001-9961-2610
Greg H. Parlier https://orcid.org/0000-0002-6837-9120
Kristin Thoney-Barletta https://orcid.org/0000-0002-8374-8633
James R. Wilson https://orcid.org/0000-0002-6255-4485