European Journal of Operational Research 293 (2021) 681–702

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

The facts on the ground: Evaluating humanitarian fleet management policies using simulation

Liyi Gu[a], Ilya O. Ryzhov[b,*], Mahyar Eftekhar[c]

[a] College of Business, Southern University of Science and Technology, Shenzhen 518055, China
[b] Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA
[c] W. P. Carey School of Business, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:
Received 28 February 2020
Accepted 9 December 2020
Available online 2 January 2021

Keywords:
Humanitarian logistics
Resource allocation
Humanitarian fleet management
Simulation modeling,

ABSTRACT

In humanitarian fleet management, the performance of purchase, assignment, and sales decisions is determined by dynamic interactions between the fleet composition, the time-varying and uncertain demands on the fleet, and the depreciation of the vehicles as they are exploited. We propose to evaluate purchase, assignment, and sales policies in a holistic simulation environment that directly models heterogeneous vehicle attributes and tracks their evolution over time. Using data from a large international humanitarian organization (LIHO), the simulator can identify the rationale behind seemingly ad-hoc decisions by field managers at LIHO. For instance, by selling vehicles later than LIHO recommends, managers are actually reducing their costs; similarly, managers decline to switch vehicles between mission types because the benefits to the operational cost turn out to be marginal at best.

© 2020 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

In humanitarian operations, fleet management is one of the most significant drivers of total operating cost. To give an example, Arsenault et al. (2018) reports that, in 2011, the UN High Commissioner of Refugees (UNHCR) had a fleet of more than 6500 light vehicles, with an estimated operating cost of 130 million USD. Even after a concerted effort to downsize the fleet, in 2017 the organization still owned 5530 light vehicles. Pedraza-Martinez et al. (2011) cites estimates by domain experts that place total worldwide operating costs for humanitarian fleets at 1 billion USD, with high projected growth. These high costs reflect the critical role of vehicles in last mile delivery of humanitarian aid Balcik et al. (2008). Improved fleet management alleviates human suffering both directly (by completing more humanitarian missions) and indirectly (by saving money that can be spent on other humanitarian programs).

In general, fleet management is a well-studied subdomain of operations research (Crainic et al., 2012). However, the humanitarian sector presents very different challenges from the commercial one. Field managers at humanitarian organizations (HOs) are often not trained in, or accustomed to, the use of operations research tools. What is more, the data necessary to calibrate such tools may not be available, e.g., in developing countries with ongoing conflict. Indeed, the UNHCR report by Arsenault et al. (2018) notes that “the exact costs of operating the fleet cannot be accurately measured due to lack of data”. Consequently, implementing the results of various analytical approaches in the field has met with limited success. Eftekhar et al. (2016) presented empirical evidence that field managers do not follow standard policies recommended either by researchers or by international agencies, suggesting that “what seems logical from the headquarters’ perspective may be illogical or inconvenient for the field”.

The authors of the present paper encountered similar statements, expressed by logistics officers and fleet managers at several major international HOs in a series of interviews.[1] To give a specific example, one of these HOs recommends that field managers sell or dispose of vehicles once they are used for either 5 years or 150,000 km, whichever comes first. The consensus among interviewees, however, was that field managers did not follow

* Corresponding author.
E-mail addresses: guly@sustech.edu.cn (L. Gu), iryzhov@rhsmith.umd.edu (I.O. Ryzhov), eftekhar@asu.edu (M. Eftekhar).

[1] We interviewed with two freelance consultants, a logistics officer and a fleet manager at GOAL International, an executive fleet manager and two logistics officers at the International Committee of the Red Cross, a senior advisor of the supply chain management unit at Catholic Relief Services, and two logistics officers at the World Food Program.

https://doi.org/10.1016/j.ejor.2020.12.019
0377-2217/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

this policy in practice, and in fact continued to exploit vehicles for much longer than either of these thresholds, although there appeared to be no single agreed-upon reason for this.

The main contribution of our paper is a holistic simulation environment that models and evaluates the acquisition, assignment, and disposition of multi-attribute vehicles in field operations. We choose to focus on *evaluation*, rather than *optimization*. As mentioned previously, even the higher-quality data collection efforts in this sector may simply not be able to provide reliable inputs for detailed assignment and/or routing models, such as those in Ben-Tal et al. (2011) or Hamedi et al. (2012). Moreover, the assumptions used in formal models often do not reflect the full complexity of the humanitarian context (see, e.g., Gralla et al., 2014 for discussion along these lines), and if they do, the models quickly become analytically (and even numerically) intractable. On the other hand, while field managers often make immediate decisions without rigorous analysis, they also possess a great deal of expertise and intuition (Kahneman et al., 2009) which often leads to successful completion of complex tasks, particularly under time pressure (Hayashi, 2001) and in unstable environments (Khatri & Ng, 2000). With these considerations in mind, we develop a simulation-based approach to counterfactual analysis in humanitarian fleet management that can explain why field managers make certain types of decisions.

1.1. Literature review

The complex settings of humanitarian operations have given rise to a considerable body of analytical work on decision-making in problems such as supply chain integration (Ni et al., 2018; Vanajakumari et al., 2016), facility location (Balcik et al., 2016), prepositioning strategies (Acimovic & Goentzel, 2016; Rawls & Turnquist, 2012; Salmeron & Apte, 2010), inventory pooling mechanisms and field coordination (Ergun et al., 2014; Toyasaki et al., 2017), and aid distribution and last mile delivery (Jahre et al., 2016). Much effort has also been devoted to understanding the challenges that arise when attempting to implement improvements in this sector: these include decentralized decision-making (Pedraza-Martinez et al., 2012), time constraints and suboptimal cost allocations (Dolinskaya et al., 2011), and lack of coordination between actors (Balcik et al., 2010; McClintock, 2009). Some of these challenges can be addressed only at the structural, rather than operational, level; see (Pedraza-Martinez et al., 2020) for an investigation of mechanism design for improved coordination between headquarters and local offices.

In this paper, we specifically consider last mile delivery of humanitarian services from local offices to affected communities (Balcik et al., 2008). Existing analytical and empirical research on this topic has examined fleet sizing (Kunz et al., 2019), vehicle maintenance and replacement (McCoy & Lee, 2014; Pedraza-Martinez et al., 2013), and field vehicle fleet management (Eftekhar et al., 2016; Pedraza-Martinez et al., 2011). However, these various aspects of fleet management are usually studied in isolation. For example, Pedraza-Martinez & Van Wassenhove (2013) studied vehicle disposition (using an analytical model), and arrived at the conclusion that HOs should replace vehicles much earlier than the industry standard. However, this conclusion relied on several simplifying assumptions, one of them being that the monthly mileage[2] of every vehicle is always constant. In practice, the situation is far more complex: the performance of a replacement policy (or other type of policy) is subject to high uncertainty due to unpredictable demand, budget constraints, infrastructure problems and other issues (Eftekhar et al., 2014; McCoy & Lee, 2014). Furthermore, no two vehicles are ever exactly alike: they are acquired at different times, and their individual ages and odometers not only affect performance, but also change dynamically over time. Managers' decisions may also be influenced by these dynamic attributes: for example, they may prefer to place higher loads on newer or older vehicles.

Researchers thus face the following dilemma. If these factors are incorporated into the model, it will no longer admit a tractable solution. Yet, if one ignores them, one runs the risk that the optimal policy for the simplified setting will be suboptimal in reality. One way to resolve this problem is to use simulation to evaluate and compare various policies in a realistic setting. Simulation has been applied in this way to study policies for many applications of interest to the public sector, such as HIV/AIDS prevention (Rauner, 2002) and emergency response (Kaplan, Craft, & Wein, 2002). In the context of humanitarian logistics, this approach will not produce an "optimal" policy, but it allows us to test simple and practicable decision rules, such as might appeal to a field manager. As discussed further down, it can also shed light on how certain choices made by field managers are rational given the circumstances.

Previous applications of simulation to humanitarian operations include agent-based models (Altay & Pal, 2014; Crooks & Wise, 2013) that look for emergent patterns from interactions between AI agents, or discrete-event models that study facility location and configuration in rapid-onset disasters (Sahebjamnia et al., 2017). Most of these papers evaluate static decisions that are made once before the simulation starts; this is also true of many optimization-based approaches such as (Ukkusuri & Yushimito, 2008). When it comes to simulating dynamic decisions that depend on evolving system state variables, there is plenty of research for commercial applications (one landmark study being Simão et al., 2010), but the *humanitarian* literature has mostly been limited to single-attribute inventory management or budget allocation (Beamon & Kotleba, 2006; Chacko et al., 2016; Iakovou et al., 2014). Such settings are not adequate for fleet management, where costs are determined by the management of operating assets (vehicles), in a way that changes over time based on the changing attributes of the fleet. Our paper offers a way to handle these complex dynamics with a great degree of modeling granularity; this is the main distinguishing feature of our work as compared to the existing applied humanitarian logistics literature.

1.2. Overview of approach and findings

As was discussed earlier, our approach focuses on simulation-based policy evaluation. We calibrated and validated the simulator using data provided by a large international humanitarian organization (LIHO). Using detailed data for multi-attribute vehicles in different countries, we designed several modules of the simulator: 1) We treat odometer data as a (censored) stand-in for demand, and develop a stochastic model of non-stationary, attribute-dependent demands over time. The simulated demand trajectories follow the same overall trend as what was observed historically, but incorporate random variation, allowing us to test different "what-if" scenarios. 2) Salvage data are used to calibrate a statistical model for the *depreciation* of vehicles as they are exploited over time; a vehicle is automatically removed from the fleet once it has lost all of its value, but can be sold earlier to redeem a portion of that value. 3) Refueling and maintenance data are used to calibrate statistical models that calculate short-term operating costs (fuel and maintenance). All of these costs are modeled as functions of vehicle attributes, which in turn are impacted by fleet managers' decisions.

[2] We use the word "mileage" informally in this paper to mean "distance travelled", but all of the numbers will actually be in kilometers.

We then deployed the simulator to compare a variety of threshold-based policies for vehicle procurement, assignment, and disposition.[3] We highlight two cases where the results were particularly interesting. First, we compared LIHO's recommended 5 year/150,000 km disposition policy with other combinations of age/odometer thresholds in realistic demand scenarios, and found that LIHO's policy was too quick to dispose of vehicles. The best-performing sales threshold does become lower as the load on the fleet increases, but even for very high loads it is still higher than LIHO's recommendation. We observed this very consistently, even when our model was recalibrated with data from a different developing country on another continent. Thus, we arrive at the opposite conclusion from Pedraza-Martinez & Van Wassenhove (2013), but interestingly, our findings are much closer to what actually happens in the field.

The second case deals with the value of coordination. In practice, LIHO assigns each new vehicle to carry out missions of a particular type, and does not change this assignment for the remainder of the vehicle's lifetime. The literature suggests that it might be preferable to switch types dynamically: (Pedraza-Martinez et al., 2011) argues that improved coordination would reduce the unpredictability of demand on fleets, while (Bhattacharya, Hasija, & Van Wassenhove, 2014) finds that asset transfer between programs in an HO leads to more efficient operations. However, our counterfactual analysis of switching found that the economic benefits are marginal at best, even in artificial "off-sync" scenarios where demand for one mission type increases just as demand for another type ramps down. We are not suggesting that coordination can never be useful, but we believe that our results explain why field managers do not seem to view it as an issue of primary importance: under current operating conditions, the benefits may not be worth the effort.

In both cases where our results disagree with existing literature, they nonetheless agree with current practice. Field managers do not seem to base their decisions on any specific policy, including those considered in our study; nonetheless, the managers' intuitive perception of the situation, based on their experience or other factors, may lead them to reject policies that are clearly suboptimal, as in the case of LIHO's recommended disposition policy. Overall, our results suggest that these intuitive decisions should perhaps be regarded more carefully than has heretofore been the case in the literature. More broadly, the main takeaway should perhaps be that "one-size-fits-all" rules (such as LIHO's sales threshold) are overly simplistic, and attempting to impose them from headquarters may be counterproductive. Instead, humanitarian organizations will benefit from a nuanced approach, where field managers have more authority to decide, e.g., the sales threshold based on the current demand level.

2. Simulation for policy evaluation

Section 2.1 provides a high-level description of the simulation environment. Section 2.2 gives a rigorous discussion of the dynamics of the simulator that enable us to model and track the evolution of fleet attributes.

2.1. Overview of the simulator

Fig. 1 provides a conceptual model of the simulation environment. The main *inputs* of the simulator, to be provided by the user, are as follows:

1. *Fleet composition*. The "state variable" describes the specific attributes of each individual vehicle in the fleet at a given moment. In the LIHO dataset, these are the *age* (in months) and *odometer* (in kilometres) of the vehicle; the vehicle *type* (e.g., make and model, but can also be aggregated by size, engine power or other factors); the *mission type* and *location* to which the vehicle is assigned; the vehicle's *accident history* (e.g., number of accidents); and its *residual value*, which is bounded above by its original purchase price. It is entirely possible for every vehicle to have a unique combination of attributes, with no two vehicles in the fleet being exactly alike.
2. *Policies*. The decision rules to be evaluated include *purchases* of new vehicles, *sales* or disposition of aging vehicles, or *assignment* of vehicles to missions. LIHO's recommended 5 year/150,000 km rule is an example of a sales policy. A simple purchase policy may be to immediately replace any vehicle that is sold; assignment policies may prioritize newer or older vehicles, or attempt to distribute the load on the fleet evenly. Policies should be sufficiently detailed to enable the simulator to automate all purchase, sales and assignment decisions based on the state at any time.

With this starting configuration, the simulator generates *demands* on the fleet from a stochastic simulation model; this *demand model* is a key module of the simulation environment, and should be customized using the available data. In the LIHO dataset, demands have two key attributes, namely *type* (a simple categorical value in our dataset, but potentially could reflect the type of work that is required, the priority of the work, or the degree of danger involved in carrying it out) and *distance* that the vehicle has to travel (in kilometres). The demand model is calibrated using historical data, but individual simulations from the model may deviate from the precise historical values that happened to have been observed.

The simulator then assigns demands to vehicles using the specified assignment policy, updates the attributes of the fleet (for example, assigning a mission to a vehicle will increase its odometer), and repeats the process for the duration of the planning period. In each decision epoch, a *cost* is incurred based on vehicle attributes across the whole fleet. Cost models also require extensive customization based on data. Our approach was to fit a number of statistical models to predict fuel and maintenance costs, as well as on depreciation, as functions of the evolving fleet composition. These models are necessary because our historical records do not provide detailed costs for every possible combination of vehicle attributes, and so costs must be inferred when running simulated scenarios that do not precisely match what was observed. The interaction between different modules of the simulator, during a *single* decision epoch, is illustrated in Fig. 2: the policy takes the state of the fleet and demand as inputs, and produces a cost and a new state as outputs.

At the end of the simulation, the output consists of *total costs* incurred by following the pre-specified policies. These include fuel and maintenance costs, plus purchase costs for any vehicles acquired by following the purchase policy, minus the residual values of any vehicles that were sold before the end of their lifespan as specified by the sales policy. Since the demands are generated using stochastic simulation, one can run the simulator many times with the same starting conditions to estimate the mean performance of a given set of policies. In addition to cost, the simulator also returns the average *completion rates* for all mission types, which is useful for policymakers since minimizing cost is not the only goal in humanitarian contexts.

[3] The LIHO dataset itself is proprietary, but our code for the simulator is available at the following URL: https://bit.ly/3ovV0CU.

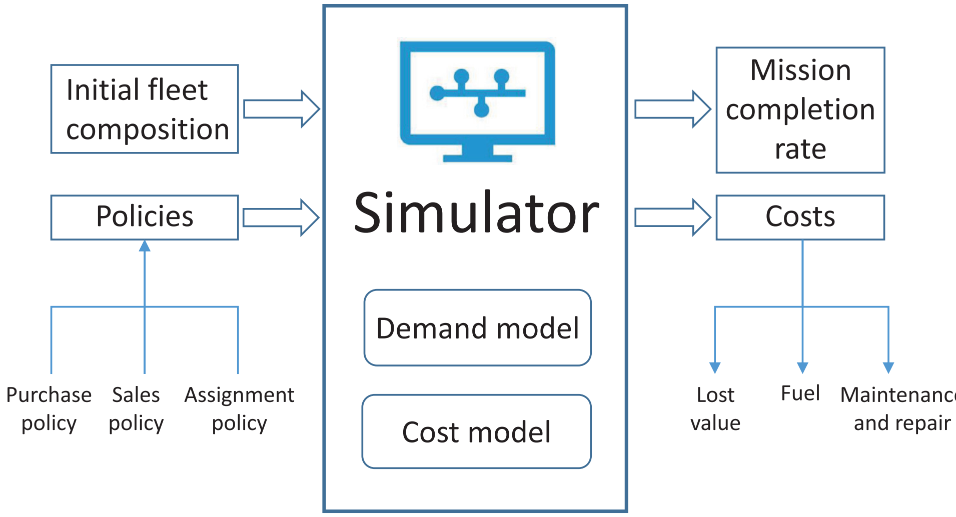


Fig. 1. Conceptual model of the simulation environment.

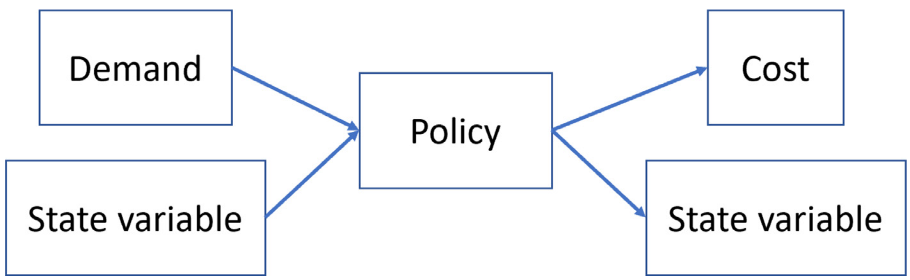


Fig. 2. Illustration of the interactions between different modules of the simulator.

2.2. Dynamics of the fleet composition

In this section, we formalize the dynamics used inside the simulator to model the evolution of fleet attributes over time. We use the framework and notational system of stochastic dynamic resource allocation (see ch. 14 of Powell et al., 2011), in which a "resource" (vehicle) is used to serve "demands" (missions). The state of a single vehicle is defined by an attribute vector a, composed of multiple attributes that may be numerical or categorical:

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} \text{Age} \\ \text{Odometer} \\ \text{Location} \\ \text{Model type} \\ \text{Mission type} \\ \text{Accident history} \\ \text{Residual value} \end{pmatrix} \quad (1)$$

Let $\mathcal{A}$ denote the set of all possible attribute vectors. Let $t = 0, 1, \ldots, T$ be a time index representing the tth "decision epoch," or instant in time when it is necessary to make a purchase, sales, or assignment decision. Our odometer data is aggregated by month (unfortunately, we do not have access to more granular data), so we assume that one month elapses between time $t-1$ and t for each t. However, one could potentially use the same framework to model more frequent decisions.

Let R_{ta} denote the number of vehicles with attribute vector $a \in \mathcal{A}$ at time t, and let $R_t = (R_{ta})_{a\in\mathcal{A}}$ represent the overall vehicle inventory.[4] One may think of R_t as a vector that has very high dimensionality, but is very sparse (as $R_{ta} > 0$ for a very small number of attribute vectors at any given time). Of course, when implementing the simulator, one does not explicitly code R_t as a vector; rather, we use this notation to make our presentation more rigorous. Next, we denote by $\hat{R}_{ta}$ the exogenous (randomly occurring) change in the number of vehicles with attribute vector a between time $t-1$ and time t. For example, such changes could occur due to accidents. We let $\hat{R}_t = (\hat{R}_{ta})_{a\in\mathcal{A}}$ denote all such changes to the fleet.

The attributes of each demand (mission) are given by

$$b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} \text{Location} \\ \text{Mission type} \\ \text{Travel distance} \\ \text{Order} \end{pmatrix},$$

and we similarly denote by $\mathcal{B}$ the set of all attribute vectors b. The first three attributes are self-explanatory. We assume that all missions with the same location and mission type are sorted in some order to be used for assignment, and the fourth attribute describes the (integer-valued) position of a given mission in this ordering. In the simplest case, missions could be sorted in the order in which they become known to the manager; if, however, the manager has advance knowledge of multiple missions, it is possible to order them in other ways, e.g., by travel distance or priority.

The quantity B_{tb} is defined to be the total number of missions with attribute vector b that are known to exist at time t, while $\hat{B}_{tb}$ denotes the number of *new* missions with attribute vector b that first appear at time t. In our study, we assume that demands are not backlogged (any demand that cannot be satisfied immediately is lost), so $B_{tb} = \hat{B}_{tb}$; however, it is straightforward to incorporate

[4] To accommodate purchase decisions within the same modeling framework, one could also add a "dummy" attribute to (1) to represent vehicles that are available to be purchased, or that have been purchased and are scheduled to join the fleet, but that are not yet available for assignment.

backlogs into the model. As before, we let $B_t = (B_{tb})_{b\in\mathcal{B}}$ denote the full vector of all currently-existing demands. Because of the fourth attribute, two missions that exist at the same time cannot have two identical attribute vectors, thus making B_t a binary vector.

The *system state vector* $S_t = (R_t, \hat{B}_t)$ represents all the information that is known to the decision-maker at time t, before the next decision is made. We can now model the decisions themselves. In the time interval of one month, a vehicle can be assigned to multiple missions; for this reason, an assignment decision d that can be applied to a vehicle is represented by an N-vector, where N is the total number of distinct demands (that is, the number of distinct attribute vectors b for which $B_{tb} = 1$) available at time t. These missions are sorted by their b_4 attribute values, and the kth element d_k of the vector d corresponds to a unique existing mission with attribute vector $b^{(k)}$ satisfying $b_4^{(k)} = k$. The statement $d_k = 1$ means that decision d assigns this mission to the given vehicle. If all entries of d are equal to zero, the vehicle remains idle for the entirety of the next time period. Denote $\mathcal{D}^D$ as the set that includes all possible d, and let $\mathcal{D}^M$ be a set of additional decisions not related to specific demands (for example, purchase decisions). Thus, every decision that can possibly be applied to a vehicle is an element of the set $\mathcal{D} = \mathcal{D}^D \cup \mathcal{D}^M$.

Now, let x_{tad} be the number of vehicles with attribute vector a to which we apply the decision d at time t, with $x_t = (x_{tad})_{a\in\mathcal{A},d\in\mathcal{D}}$. This decision variable must satisfy the conditions

$$\sum_{d\in\mathcal{D}} x_{tad} = R_{ta}, \qquad \forall a \in \mathcal{A}, \tag{2}$$

$$\sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}^D} x_{tad} \le B_{tb^{(k)}}, \qquad \forall k = 1,\dots,N, \tag{3}$$

$$x_{tad} \ge 0, \qquad a \in \mathcal{A}, d \in \mathcal{D}, \tag{4}$$

Condition (2) means that, for any a, exactly R_{ta} vehicles are available to act on. Condition (3) means that N distinct missions are available to be assigned. Condition (4) is straightforward. We may also impose problem-specific conditions: for example, we may wish to prohibit assignments where the vehicle and mission are not in the same location, or if the mission is not of the type to which the vehicle has been assigned by LIHO. We may also limit the total number of missions that a single truck can fulfill per month (for example, by setting a maximum daily travel distance).

Let $\mathcal{X}_t$ be the set of all x_t that satisfy (2)-(4). The decisions are determined based on user-specified policies; thus, $x_t = X_t^\pi(S_t)$, where X_t^π is a mapping on the state space into $\mathcal{X}_t$, with the superscript π denoting the "name" of a particular policy. In words, a policy sees the system state S_t at time t and converts this state into a feasible assignment decision x_{tad} for each a and d.

When we act on a vehicle with attribute a using decision d, the attribute vector of the vehicle changes. The new attribute vector $a' = a^M(a,d)$ is calculated using the *transition function* a^M, which has to be explicitly coded. For example, if $d = \{0,0,\dots,0\}$ (that is, no missions are assigned to the vehicle), then $a'_1 = a_1 + 1$, $a'_2 = a_2$, $a'_3 = a_3$, $a'_4 = a_4$, $a'_5 = a_5$, $a'_6 = a_6$, and $a'_7 = a_7 - K$, where K is the decrease in residual value (which must be estimated via a statistical model to be introduced later) of the vehicle resulting from leaving it idle for one month. For notational purposes, we define the indicator function

$$\delta_{a'}(a,d) = \begin{cases} 1, & \text{if } a^M(a,d) = a' \\ 0, & \text{otherwise.} \end{cases}$$

Then, the new fleet composition arising *as a result of our decision* is given by

$$R^x_{ta'} = \sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}} \delta_{a'}(a,d)x_{tad},$$

and the resource transition from t to $t+1$ is given by $R_{t+1} = R^x_t + \hat{R}_{t+1}$ if there are any random changes to the fleet. Again, since we assume that unfulfilled missions are not backlogged, we have $B_{t+1} = \hat{B}_{t+1}$ as mentioned earlier.

Finally, we describe the evaluation of the policy π. Let c_{tad} be the cost of applying decision d to a resource with attribute a at time t. The cost includes maintenance, repair, and purchase costs, which are obtained from statistical models to be discussed later. The cost may also be negative if the vehicle is sold (in that case, the revenue is equal to the vehicle's residual value attribute at the time of sale). The total single-period cost is given by

$$C_t(S_t, x_t) = \sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}} c_{tad}x_{tad},$$

and the performance of the policy is calculated as

$$V^\pi = \mathbb{E}\sum_{t=0}^{T} C_t(S_t, X_t^\pi(S_t)), \tag{5}$$

the expected total cost incurred over the given planning period when policy π is used to calculate decisions. The expectation in (5) is taken over the joint distribution of $\left(\hat{R}_{t+1}, \hat{B}_{t+1}\right)_{t=0}^{T-1}$, and is computationally intractable due to the complex dependence of $(S_t)_{t=1}^T$ on these random quantities. However, it is quite straightforward to estimate (5) through simulation: given an initial state S_0 and a policy π, we can generate M independent trajectories $\left(\hat{R}^m_{t+1}, \hat{B}^m_{t+1}\right)_{t=0}^{T-1}$ for $m = 1,\dots,M$ and report the sample average

$$\bar{V}^\pi = \frac{1}{M}\sum_{m=1}^{M}\sum_{t=0}^{T} C_t(S_t^m, X_t^\pi(S_t^m)), \tag{6}$$

where $\left(S_t^m\right)_{t=0}^T$ is the sequence of states visited in the mth simulation run. The average mission completion rate can be estimated in a similar way (the costs c_{tad} should be redefined accordingly).

3. Development and calibration of cost and demand models

Section 3.1 describes the LIHO data used to calibrate the models that follow. Section 3.2 presents our proposed stochastic model for simulating demands on the fleet. Sections 3.3–3.5 discuss statistical models used to estimate costs due to depreciation, refueling, and maintenance, while Section 3.6 briefly discusses the simulation of accidents. Here, we focus on explaining the models; see the Appendix for a more detailed justification of our modeling choices.

3.1. Description of LIHO data

The LIHO data contains aggregate information about 3846 vehicles across 20 countries during 2004 – 2015. Some descriptive statistics are given in Table 1. The results presented here and in Section 4 primarily focus on a single country, which we refer to as the "focal country" or FC, in which there were 454 vehicles distributed among 16 local offices. We chose FC because it had sufficiently many vehicles assigned to it to allow us to reliably estimate cost and demand models. As a robustness check, we also performed a complete recalibration of our model using data from a second country ("Country 2" or C2); those results can be found in the Appendix. Furthermore, the full 20-country dataset was used to calibrate our depreciation model (accounting for heterogeneities between countries).

In the aggregate cross-sectional dataset, each data point represents one vehicle and its attributes at the end of the observation period, which include 1) unique vehicle ID and the local office where the vehicle was stationed; 2) vehicle model type, mission type given to the vehicle; 3) date registered, date on which odometer is recorded, final odometer; 4) purchase value, sales

Table 1
Descriptive statistics for aggregate cross-sectional dataset.

	Mean	Std. Dev.	Median	Min	Max	No. levels
Country						20
Age	70.91	39.58	74	1	277	
Odometer	99248.86	76782.26	89937.5	5	1102385	
Annual Distance	18177.65	31321.33	14478.51	0	730500	
Purchase Price	37580.63	35329.58	26637.29	1777.14	227500	
Sales Price	14968.04	18756.90	10951.73	0	190896.79	
Vehicle Type						4
% Depreciation	0.58	0.25	0.62	0	1	
Accidents	0.27	0.93	0	0	14	
Total Accident Cost	129.78	2010.42	0	0	89000	
Mission Type						7
Total Repair Cost	353.19	1471.38	0	0	28026.34	
Make						36
Model						154

value (if sold) or booking value (if not sold); 5) total number of accidents during the observation period. We mainly use this dataset to estimate the residual value of vehicles of different model types, mission types, ages and odometers in different countries. Age is obtained by subtracting the starting date of registration from the final date of record.

The second dataset includes monthly traveling distances for individual vehicles (in FC/C2 only, as loads on the fleet are highly individualized by country and even by mission type) over the observation period. Unfortunately, distances for individual missions are not available and we were required to work with aggregate distance traveled by each vehicle in one month. We use this dataset to calibrate the mission arrival process; Section 3.2 presents a probabilistic model for generating individual mission distances from a distribution calibrated using the aggregate data.

The third dataset contains the refueling history of vehicles (again, in FC/C2 only), and provides the refueling cost as well as the age and odometer of the vehicle associated with each refuel. We use this dataset to estimate the fuel cost resulting from completing various missions by vehicles with different attributes. The fourth dataset pertains to the maintenance of vehicles in FC/C2, and is used to estimate the maintenance cost of vehicles with different attributes. For individual vehicles, the cost of each maintenance as well as the age and odometer of the vehicle is recorded. When estimating costs, we eliminated vehicle model types and mission types that do not appear with an adequate number of records in any of the three datasets related to costs. Because many models and mission types appeared very infrequently in the data, our analysis of FC was carried out on one vehicle model type, in one local office, with two possible mission types, involving 160 vehicles in the monthly travel data, 122 vehicles in the refueling dataset and 39 vehicles in the maintenance dataset.

3.2. Demand model

In the LIHO data, the monthly mileage of a vehicle ranges from 20 – 30 to upwards of 4000, and there are also many zeroes that may represent idle vehicles or missing data. This level of variation is consistent for vehicles operating in different sub-delegations or handling different mission types. However, it is difficult to evaluate assignment policies based purely on the historical data. First, because monthly mileage is a consequence of the historical assignment decisions, we cannot directly calculate how a different policy would have performed in the same time frame. Second, the LIHO data only provide aggregate monthly mileage for each vehicle, and does not show how many individual tasks were performed or the size of each task. Third, monthly mileage only provides information about missions that were *completed*, and there is no way to know how many additional tasks there might have been that were visible to the fleet managers, but that could not be completed due to lack of resources or other factors.[5]

The goal of the demand module (see Fig. 1) is to generate, in each time period, a stochastic number of missions whose individual mileage attributes also vary stochastically. The total monthly loads on the fleet should be "realistic," i.e., they should exhibit the same general trends and magnitudes as the historical loads (for example, gradually rising and falling according to historical trends), but they should not be identical to the historical data. This is because 1) we would like to have the flexibility to consider different scenarios, and 2) as discussed above, the historical data may not provide complete information on the total potential load in each month. For example, if we view the historical data as being censored (since we only see mileage for completed missions), we might wish to generate demand trajectories that are consistently *higher* than historical, while following similar trends over time.

We found that the monthly mileages for vehicles did not exhibit any significant correlation between locations and mission types. For this reason, we assume that demands are independent across all location/mission type combinations, and so we estimate an independent demand model for each such combination separately from the others. In the following discussion, we take one mission type in the capital city of FC to illustrate how the model works.

To model the time-varying behaviour of the demand, we construct a stochastic process $(L_t)_{t=0}^{T}$ that takes on positive integer values. The value of L_t for given t can be viewed in terms of the number of different humanitarian "projects" (distinct development efforts) that are currently active and can generate tasks for the fleet to perform. Although this is typical in humanitarian logistics, the LIHO dataset provides no information about any such projects, so the process (L_t) is a modeling construct rather than an empirically observed quantity. One could also view (L_t) as a kind of "latent fleet size", that is, the number of vehicles that we *should* have on hand in order to complete all the tasks. The actual fleet size is modeled with the resource variable R_t and may be completely different from L_t; in particular, the actual fleet size may be lagging behind the "latent" one, if the fleet managers make delayed reactions to sudden growth in demand.

The process (L_t) is modeled as a $G/G/\infty$ queue with batch arrivals. Each batch represents an active "project", with the batch size representing the number of vehicles needed in order to complete all the tasks. The "service time" in this system represents the lifetime of the project; once this time runs out, the project disappears and stops generating tasks, leading to a reduction in

[5] We expect there to be a strong correlation between the observed load on the fleet and the total size of all visible missions, so we are using the former as a stand-in for the latter. This is consistent with other work in this area, and in any case, the LIHO dataset does not provide any other more precise information about demand.

demand. The number of “servers” is infinite because there is no limit on how many projects may be active at the same time. Thus, as the number of busy servers fluctuates over time, we will see some periods of very high demand.

Because our goal is to generate realistic demands that resemble the overall trajectory of the historical loads, the interarrival times and batch sizes are bootstrapped from the data. Since, in the monthly data, we can see the exact time periods when new vehicles enter the fleet, we simply use the times elapsed between two such “arrivals”. Usually, multiple new vehicles are added to the fleet at the same time, providing us with the batch size. Thus, we are treating historical purchases and fleet sizes as observations of L_t, though it is possible that these observations are actually censored. Unfortunately, if censoring is present, we have no way to know *which* observations are censored since there may well have been months when all of the visible demand was met by the historical fleet. However, bootstrapping from the data provides us with a rough trajectory for L_t over time, and the stochasticity of the simulation can be used to generate scenarios that deviate from the data in other ways (e.g., with higher demand).

The service times for the queueing system could also be obtained by bootstrapping from the observed vehicle lifetimes; however, in the data, these lifetimes are right-censored since many vehicles are still in the fleet at the end of the observation period. For this reason, we used a Weibull distribution for the service time, and computed the parameters using maximum likelihood estimation with censoring.

The total mileage of all new missions that become visible at time t (for the given, fixed mission type/location pair) is then calculated as

$$D_t = \sum_{\ell=1}^{L_t} Y_\ell, \qquad Y_\ell = \sum_{n=1}^{N_\ell} X_n, \tag{7}$$

where N_ℓ is a positive integer representing the number of distinct tasks generated by project ℓ in time period t, the values X_n for $n = 1, \dots, N_\ell$ represent the mileages required to complete each individual task, and Y_ℓ then represents the total mileage generated by project ℓ. We require the flexibility to model demand at the level of individual tasks in order to compare different assignment policies (for example, a policy that assigns tasks in order of increasing mileage, with the quickest tasks being completed first).

Since we use the historical fleet size as a stand-in for L_t, we can treat the monthly mileage of each vehicle as an observation of Y_ℓ. Although the data do not tell us anything explicit about N_ℓ or X_n, we can estimate distributions for these random variables by interpreting Y_ℓ as a compound Poisson random variable, that is, $N_\ell \sim \mathit{Poisson}(\lambda)$ and $X_n \sim \mathit{Gamma}(\alpha, \gamma)$. Under a particular transformation of $(\lambda, \alpha, \gamma)$, this becomes the well-known Tweedie distribution (Zhang et al., 2013), which is widely used in applications (e.g., in insurance claims modeling, see Smyth & Jørgensen, 2002) where it is necessary to “reconstruct” the individual components of an observed sum. The Tweedie distribution also allows $(\lambda, \alpha, \gamma)$ to depend on covariates such as the fleet size.[6] The Appendix gives the estimated values of these parameters for various cases that we considered.

Fig. 3 provides an illustration of how the output of the demand model is “realistic” without being identical to the real data. First, Fig. 3(a) shows two realizations of the demand-generating process (L_t) compared to the actual historical fleet size (again, for this particular location/mission type combination). All three trajectories grow over time, and so the simulation output has the same general trend as historical. However, each individual simulated scenario deviates quite a bit from the historical trajectory; in particular, the number of “active projects” generated in the “high” scenario is consistently much greater than the historical fleet size. This trajectory is useful if we believe that the historical data are significantly under-reporting the visible demands. Second, Fig. 3(b) fixes the process L_t to have the same values as the historical data, and uses (7) to randomly generate individual tasks. In other words, we are comparing the simulated values of Y_ℓ against the historical values. Again, we see that, while the simulation output is not identical to the historical data, all of the trajectories follow the same general trend. Of course, when L_t is also simulated, we will expect to see greater deviation from historical.

We close this discussion by reiterating that demand simulation must take two diametrically opposed concerns into consideration. On one hand, we want the simulation output to resemble the historical data, as otherwise the results of policy evaluation may not be relevant to LIHO. On the other hand, we also want to encourage a certain amount of deviation from historical, both because we do not want to “overfit” our results to the one set of numbers that happened to have been observed, and also because the dataset itself is not completely reliable with respect to the demand.

3.3. Cost models: Depreciation

From (1), recall that our resource model tracks the residual value of every vehicle over time. This value serves as a constraint on the vehicle’s lifespan: once it reaches zero, the vehicle is automatically removed from the fleet. Residual value also plays two roles in cost modeling. First, it determines our revenue in the event that we sell the vehicle. Second, at the final time T, the residual value of every vehicle remaining in the fleet is subtracted from the total cost, to avoid a bias in favour of policies that deliberately make fewer purchases late in the planning period.

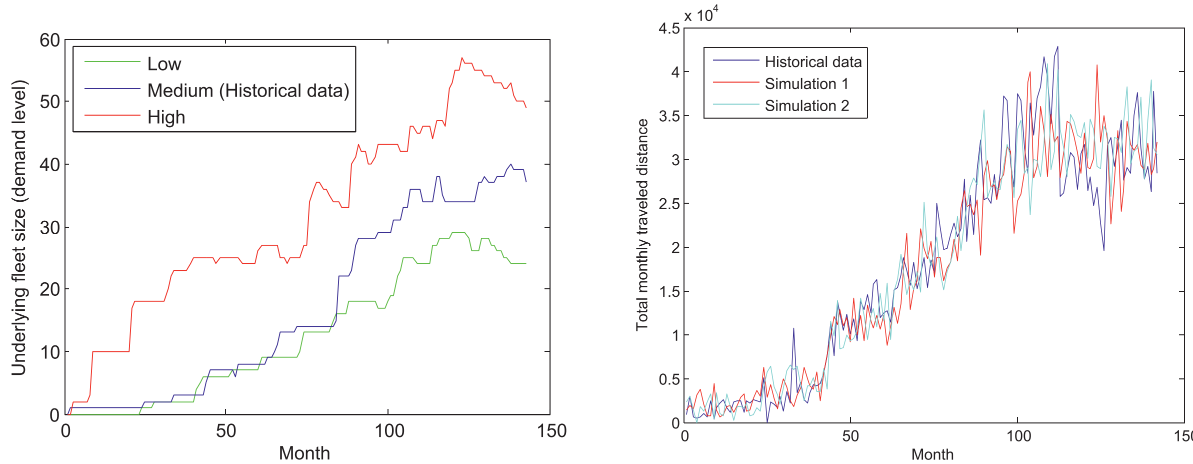
We formulate and estimate a statistical model that can be used to calculate changes in residual value as a function of vehicle attributes and assignment decisions. Essentially, this model serves as part of the transition function a^M. Specifically, we use the zero-intercept Tobit model

$$\mathrm{Dep\%} = \beta_1 \mathrm{Age} + \sum_j \beta_{2j} \log(\mathrm{Odometer}) \times \mathrm{MissionType}_j + \beta_3 \mathrm{NumAcc}, \tag{8}$$

where Dep% is the depreciation expressed as a percentage of the original purchase price, $\mathrm{MissionType}_j$ is a dummy variable that is equal to 1 if the vehicle is assigned to the jth distinct mission type, and NumAcc is the total number of accidents in the vehicle’s history at the time of disposition. The Tobit model is used because the LIHO dataset contains an inflated number of zeroes (and no negative numbers) reported as residual values; we assume that any negative value for the right-hand side of (8) indicates that the vehicle has been rendered unusable and can only be salvaged.

We use the percentage loss rather than the actual residual value because the initial purchase price of a vehicle depends on factors such as accessories, payload, special design etc., that are not explicitly captured in (8). By modeling the loss relative to the original purchase price, we implicitly keep these factors in the estimated residual value. The age and odometer variables are self-explanatory: both should affect the residual value negatively, or the percentage loss positively. However, the effect of odometer should vary with the mission type on which the vehicle operates, as some mission types inflict more wear on the vehicle, resulting in lower residual value. We apply a log transformation to the odometer because we expect residual values to drop more sharply in the early

[6] Eftekhar et al. (2014) points out that such a dependence may indicate that the demand is censored; unfortunately, as stated earlier, we are not able to see which particular observations are censored. It is also possible that, if there is a large number of projects, some individual tasks can be combined into a single “trip”, leading to a slightly smaller number of tasks.



(a) Sample paths of (L_t). (b) Sample paths of (D_t).

Fig. 3. Simulated demands compared with historical data.

Table 2
Estimation results for the depreciation model (8).

	Independent variable:
	Depreciation%
Age	2.215×10^{-3} ***
log(Odometer)×Mission2	0.04845***
log(Odometer)×Mission4	0.05399***
NumAcc	0.04512***
Observations	3846 (454 in FC)
Log Likelihood	−56,724.86
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3
Estimation results for the fuel cost/km model (9).

	Dependent variable:
	Cost per Km
Odometer	1.632×10^{-7} ***
Age	2.506×10^{-4} ***
Mission4	0.009759**
Constant	0.1292***
Observations	11,040
R^2	0.301
Adjusted R^2	0.293
Residual Std. Error	0.029 (df = 10917)
F Statistic	38.486*** (df = 122; 10917)
Note:	*p<0.1; **p<0.05; ***p<0.01

stages of a vehicle's life than in the later stages (an additional 100 km traveled should have less impact if the vehicle has already accrued 100,000 km). Due to the small number of accidents in the data, we assume that the number of accidents has the same linear impact on depreciation across all mission types.

In (8), the intercept is set to zero, because a vehicle's residual value is initialized to its purchase price (a new vehicle has not yet incurred any depreciation). Note that (8) includes interaction terms between log (Odometer) and the mission types, but does not include either main effect. The main effects for mission types are not included because (8) is a zero-intercept model (essentially the reasons for omitting the intercept term also apply to these effects). The main effect of log (Odometer) is not included because this would create linear dependence among independent variables; as an alternative, one could omit one value of j from (8) and include the main effect of log (Odometer) instead. Since the primary purpose of this model is to generate costs inside the simulator, this is not a major issue.

For maximum accuracy, we used the full 20-country cross-sectional data to calibrate our residual value model. Thus, we estimated a version of (8) that contained four-way interactions between vehicle odometer, mission type, vehicle type, and country. However, since our simulations focused on FC, most of these terms were not actually used inside the simulator. Table 2 shows the estimation results only for those terms relevant to FC, with two mission types (numbered 2 and 4 in the data), a single vehicle type, and a single location. The coefficients for all the regressors are positive and significant; note the heterogeneity between mission types.

3.4. Cost models: Fuel

Each data point in the LIHO refueling dataset represents one refueling for one vehicle. We make the assumption (in the absence of any information whatsoever in this regard) that the tank is filled on each refueling, and thus the amount of fuel purchased corresponds to the amount that has been consumed during the distance traveled since the previous refueling of that vehicle (which we can obtain from the monthly odometer data). Then, for each refueling record, we can calculate the fuel cost per kilometer between two refuels of the same vehicle and relate this quantity to the vehicle's mission type, age and odometer at that time. We propose the linear model

$$\text{FuelCostPerKm} = \beta_0 + \beta_1 \text{Odometer} + \beta_2 \text{Age} + \sum_j \beta_{3j} \text{MissionType}_j, \tag{9}$$

with additional fixed effects for each vehicle ID. To improve estimation quality, we removed the outliers and only considered data with traveled distance above 300 km.

Table 3 shows the OLS regression result of the model. We can see that, since all coefficients are significant, there exists a baseline cost of 0.1292 per km traveled, each extra kilometer on the odometer adds 1.632×10^{-7} to this cost, each extra year on the age adds 2.506×10^{-4}, and if the vehicle operates on mission type 4, an additional 0.009759 is added on the per km cost. Note that, even if the cost/km follows a linear model, the resulting fuel costs are *not* linear in the distance traveled: if the vehicle is new or has a low odometer, its fuel efficiency is at a higher level, reflected by

Table 4
Estimation results for the cumulative maintenance/repair cost model (10).

	Dependent variable:
	Cumulative costs
Age×Mission2	25.553***
Age×Mission4	15.011***
Odometer×Mission2	0.014***
Odometer×Mission4	0.0083*
$\text{Odometer}^2\times$Mission2	3.21×10^{-7} ***
$\text{Odometer}^2\times$Mission4	7.06×10^{-7} ***
Observations	735
Log Likelihood	− 5,386.827
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$

the lower fuel cost per kilometer. The cost incurred for a fixed distance increases with the age and utilization of the vehicle.

3.5. Cost models: Repair/maintenance

In the LIHO data, vehicles do not appear to follow a strict schedule of maintenance and repairs. Rather, they appear to receive several "levels" of maintenance approximately every 5000, 15, 000 and 50, 000 km, but also other services at seemingly unsystematic times and odometer readings. From this, we conjecture that field managers do schedule maintenance and repairs based on age and odometer, but that any such schedule is only followed roughly. We formulate the linear model

$$\begin{aligned}\text{CC} = \sum_j \beta_{1j}\text{Age}\times\text{MissionType}_j + \beta_{2j}\text{Odometer}\times\text{MissionType}_j \\ + \beta_{3j}\text{Odometer}^2\times\text{MissionType}_j, \qquad (10)\end{aligned}$$

where CC is the *cumulative* maintenance/repair costs of one vehicle that were incurred since it entered the fleet. We use a zero-intercept model since the cumulative cost for a brand-new vehicle should equal zero; for the same reason, we do not include main effects for the various mission types. We also do not include main effects for age and odometer, but (10) is equivalent to the model where one mission type is removed and the main effects are included. We include the nonlinear term Odometer^2 since we expect that the vehicle will receive more repairs as it is utilized more. Interaction terms between age/odometer and mission types are included because vehicles operating on different mission types may have different maintenance schedules, and more strenuous missions may require more repairs. Although accidents are not explicitly included in (10), the model can be viewed as indirectly incorporating costs due to accidents in an average sense since repair costs are present in the data.

The LIHO dataset only provides maintenance data for 39 vehicles in FC, and testing with a mixed linear model suggests that the level of between-group variability is sufficient to warrant the inclusion of a random effect representing vehicle ID. The estimation results are shown in Table 4. We can see that the coefficients of the quadratic terms are positive and significant, indicating that maintenance/repair costs do indeed grow more quickly as the odometer increases.

3.6. Generation of accidents

In Section 3.3, we estimated the impact of accidents on the residual value of a vehicle. In order to include accidents in our simulations, we also require a probabilistic model of how likely they are to occur in a single decision epoch. Unfortunately, our data are not sufficient to estimate such a model; for the one location and two mission types considered in Section 3.3, we have records of only 12 accidents. For this reason, we constructed an artificial model in which the accident occurrence probability is linearly related to a vehicle's age and odometer. The coefficients of this linear model can be viewed as tunable parameters, and we chose them so that the total numbers of accidents in our simulations resembled those in the data (we refer to this as the "base case"). In addition, we tested other parameter values with 2x and 4x the base accident probability, and in the Appendix we also perform additional sensitivity analysis on accident severity. We also ran a version of this model in which the accident probability was flat (did not increase with age and odometer), but the results were not substantially different.

4. Analysis, results and insights

We present a case study calibrated to the LIHO data. Recall from Section 3 that, due to the specifics of the LIHO data, we focus on a single location with mission types 2 and 4. The various policies that were compared are described in Section 4.1. The first case considered in Section 4.2 focuses on sales and assignment policies under stable (stationary) demand. Section 4.3 considers realistic demand (in the sense discussed in Section 3.2) and introduces purchase policies. Section 4.4 investigates the impact of allowing vehicles to change their assigned mission type (a practice not currently implemented by LIHO). An additional case, focusing on centralized procurement, is deferred to the Appendix.

4.1. Description of policies

Recall from Section 2.2 that a policy provides a way to calculate a decision when given any system state. Thus, to run the simulator, the user must choose policies for purchasing, assigning, and selling vehicles. We consider a number of simple and intuitive choices for each of these categories.

Purchase policies. The simplest purchase policy is *pure replacement*, where we purchase a new vehicle only when an existing vehicle is removed from the fleet (either sold, or disposed after reaching zero residual value). Under this policy, the fleet size is constant. We mainly consider this policy in Section 4.2, where demand is assumed to be stationary.

We also consider simple "reactive" policies that purchase new vehicles when the recent mission completion rate appears to be "low" (i.e., falls below some tunable threshold) or when the utilization of the existing fleet appears to be "high". In most cases, we assume that new vehicles join the fleet instantaneously upon request; however, in the Appendix, we investigate the issue of lead time.

Assignment policies. We assume that managers cannot anticipate the arrival of new missions and must assign them to vehicles in the order in which they appear (are generated by the simulator). The following simple assignment rules are considered:

- *Balance*. This rule assigns an incoming mission to the vehicle that currently (based on previously assigned missions) has the least traveling distance assigned to it for the month. Essentially, this rule attempts to balance the monthly load on the fleet.
- *Least/Most Odometer*. An incoming mission is assigned to the vehicle with the least/most mileage on the odometer.
- *Oldest/Newest*. An incoming mission is assigned to the vehicle with the largest/smallest age attribute.
- *Myopic*. Assigns an incoming mission to a vehicle to minimize the immediate cost of the assignment (i.e., assigning the mission to this vehicle incurs less cost than assigning it to any other vehicle), calculated by adding fuel, maintenance, and depreciation costs.

Among these, Balance and Least Odometer can be viewed as workload-balancing rules, which have been widely studied in commercial transportation (Matl et al., 2017). The Balance policy attempts to evenly divide the monthly load on the fleet, whereas Least Odometer attempts to balance total odometers. The Newest policy is inspired by tendencies observed in practice (Pedraza-Martinez & Van Wassenhove, 2013).

We impose a limit on the number of missions that can be assigned to a single vehicle. First, we calculate the maximum monthly traveling distance of any vehicle in the LIHO data, and divide this quantity by 30 to obtain a cap on daily traveling distance (approximately 120 km). The “travel distance” attribute of each generated mission (the value X_n in (7)) can be divided by this daily cap to obtain the number of days required by the mission. A vehicle cannot take more than 30 days’ worth of missions; if it is unable to receive any more assignments, the next best matching vehicle (depending on the assignment policy) is used. If no vehicles are eligible to receive the next incoming mission, it is simply dropped (not completed).

Sales policies. LIHO recommends selling a vehicle once it has reached *60 months* in age or *150,000 km* of odometer, whichever comes first. We consider other combinations of these thresholds, such as 40 mos./100,000 km (40/100), 100 mos./225,000 km (100/225), 140 mos./300,000 km (140/300), as well as the option to run the vehicle into the ground (RIG), which means that we continue to exploit it until it reaches zero value and is automatically removed from the fleet.

It can be seen that most of these policies are based on thresholds of various state attributes (such as age and odometer), or quantities derived from these (such as utilization). We adopt this approach because it is simple and practicable, and would easily appeal to a field manager. Indeed, in the case of sales policies, our definitions are motivated by current practice.

4.2. Case 1: Stable demand, sales/assignment policies

In this case, we assume that demand is stationary: the process (L_t) in (7) is set to a constant, but the Tweedie distribution (calibrated using real data; see the Appendix for the exact parameter values) is still used to randomly generate individual tasks. A pure replacement policy is used to procure new vehicles, thus keeping the fleet size constant.

We consider a time horizon of 200 months. The initial fleet consists of four vehicles, of which two are new while the other two record 20 mos./30,000 km and 40 mos./60,000 km on age and odometer, respectively. All vehicles are assumed to have been initially purchased for \$30,000, which is also used as the fixed purchase price of all new vehicles. Five levels of stationary demand were considered by setting $L_t \equiv i$ for $i \in \{1, 2, 3, 4, 5\}$. For levels 1 and 2, the demand is lower than the fleet capacity, meaning that the vehicles will tend to be under-utilized. Levels 3 and 4 are roughly equal to the fleet capacity (level 3 is slightly lower, but can occasionally generate high loads), and level 5 is above the fleet capacity, meaning that the fleet will not be able to complete all the missions.

We run every sales policy together with every possible assignment policy. The estimated performance (6) of a sales policy is reported using the best-performing assignment policy. We found that, for any fixed demand level, there was very little variation in completion rates between policies. Completion rates were close to 1 for demand levels 1–4, and around 77% for demand level 5. For this reason, we use total cost as the primary performance metric in this case.

Fig. 4 shows the results of the comparison with accidents not included (i.e., accident probability was set to zero). We conducted sufficiently many simulation runs to obtain statistically significant differences between policies. When demand is very low, the best performance is achieved by the RIG policy, which never sells vehicles; however, as the demand increases, the 140/300 policy starts to perform better (levels 2–3), and at levels 4–5, 100/225 becomes the best. Thus, 1) the optimal sales threshold is always later/higher than the one recommended by LIHO, and 2) the optimal threshold moves earlier/lower as the demand increases.

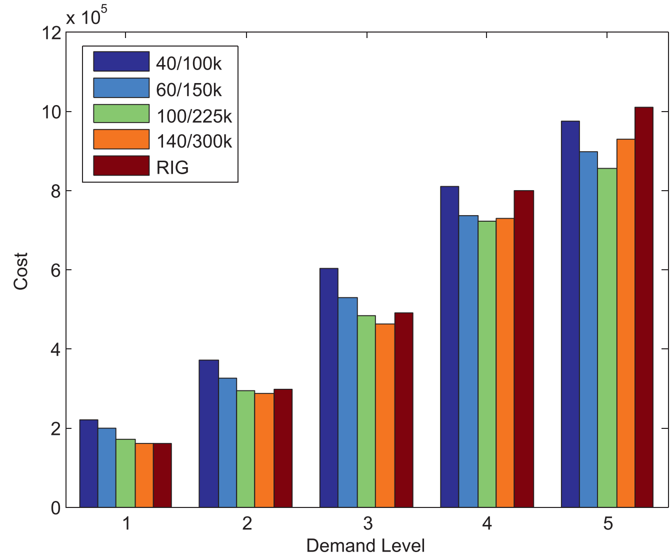


Fig. 4. Cost comparison of sales policies under demand levels 1–5.

The first observation dovetails with the historical practice of LIHO field managers, who generally continued to exploit vehicles after the 60/150 threshold. We saw this in our data, where over 70% of dispositions occurred after at least one of these thresholds (with ages reaching 250 months and odometers exceeding 180,000 km), and of the remaining 30% many were listed as having zero age, suggesting that they were not sold but transferred elsewhere in the HO. Our interviewees agreed that the 60/150 rule was rarely followed, and similar observations were made by Pedraza-Martinez & Van Wassenhove (2013) in an empirical study.

This behaviour can be explained in terms of the tradeoff between utilization and residual value, first highlighted by Eftekhar & Van Wassenhove (2016) in the humanitarian context. Costs related to vehicle value (purchase and depreciation) are mostly incurred early on in the vehicle’s lifetime; on the other hand, operational costs (fuel, maintenance and repair) become steeper late in the vehicle’s lifetime. When the demand is low, value-related costs account for a greater share of the total cost and thus we prefer to continue exploiting the fleet rather than making new purchases. When the demand is high, the fleet is exploited more heavily and it becomes better to avoid keeping vehicles with high age/odometer in the fleet. Even then, however, the optimal tradeoff is made after LIHO’s recommended threshold.[7]

Next, we examine both sales and assignment policies and bring accidents into the picture. Table 5 reports optimal sales/assignment policy combinations for all demand levels and accident frequencies. With regard to sales, the pattern is virtually identical to what we saw in Fig. 4: RIG and/or 140/300 are optimal for low demand

[7] We also considered thresholds based on only age (e.g., 40 mos., 60 mos. etc) or only odometer. The results for odometer-based policies were very similar to what is shown here; the best-performing threshold was 300k and, for higher demand, 225k. Performance was slightly more sensitive to age, with the best choice starting at 140 mos. at demand level 1 and reducing to 60 at level 5. We did not observe any scenario in which it was optimal to sell before LIHO’s threshold.

Table 5
Cost comparison of sales/assignment policies with varying levels of accidents.

Accident Level	Demand Level	Sales	Assignment	Total cost	Avg. no. of accidents
0	1	RIG	Myopic	151,519	0
	2	140/300	Myopic	286,123	0
	3	140/300	Least Odometer	460,906	0
	4	100/225	Newest	721,593	0
	5	100/225	Balance	854,286	0
Baseline	1	RIG	Myopic	162,034	1.433
	2	140/300	Myopic	289,548	2.672
	3	140/300	Least Odometer	466,112	3.979
	4	100/225	Newest	728,315	4.967
	5	100/225	Balance	862,691	6.202
2x Baseline	1	140/300	Myopic	163,841	2.182
	2	140/300	Myopic	292,427	5.437
	3	140/300	Least Odometer	470,577	7.908
	4	100/225	Newest	734,835	9.804
	5	100/225	Balance	871,176	12.410
4x Baseline	1	RIG	Myopic	166,533	4.813
	2	140/300	Myopic	298,577	10.699
	3	140/300	Least Odometer	479,387	15.088
	4	100/225	Newest	747,526	19.514
	5	100/225	Balance	889,165	24.729

levels, followed by 100/225 as the demand increases. This holds for all accident frequencies.

Assignment policies exhibit the following pattern across all accident levels: for low demand levels Myopic is always preferable, whereas Least Odometer/Newest are the best choices for medium demand levels, and Balance is the best for very high demand. When the demand is low, Myopic essentially concentrates the demand on a portion of the fleet, exploiting those vehicles very heavily while leaving the others idle. This is optimal in low-demand settings because residual value is the most important cost driver. As the demand increases, it becomes necessary to use all the vehicles to complete the missions, and so workload-balancing rules[8] start to perform better as they essentially control the growth of maintenance costs across the fleet, while also making vehicles reach the sales threshold later. For very high demand, all assignment policies become very similar as all the vehicles are fully utilized and have very similar monthly mileages; however, the Balance policy has a slight edge since, when the fleet is running at close to full capacity, only small missions can be "squeezed into" the loads, and the Balance policy will tend to assign more of these missions.

If cost is the main performance criterion, it appears that the presence of accidents in the simulator does not substantially change our conclusions regarding sales and assignment policies. For this reason, accidents are mostly omitted from the remainder of our study. If the field manager is concerned with issues other than cost (for example, personnel safety), reducing the sales threshold will mitigate the risk of accident somewhat (the number of accidents grows with demand, but switching from 140/300 to 100/225 slows the growth), but does not eliminate it entirely.

4.3. Case 2: Realistic demand, assignment/purchase policies

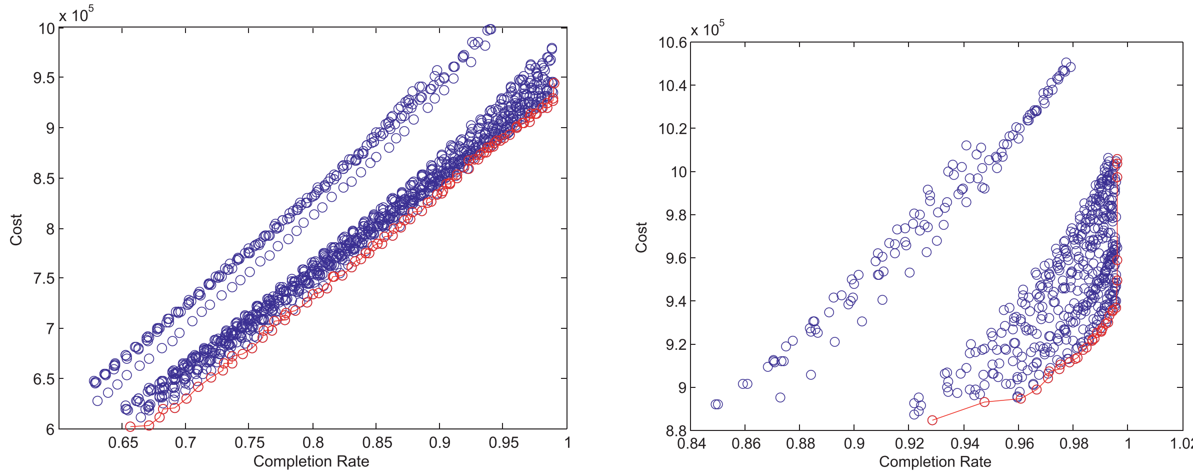
We now consider nonstationary demand, based on the model in Section 3.2 calibrated to 143 months of historical data for mission type 2 at a single local office. We focus on the "high" demand trajectory shown in Fig. 3, which demonstrates the same rising trend as in the historical data, but represents a hypothetical situation where the data under-reported the actual demands.

Our dataset tells us when new vehicles were purchased historically; however, because the historical fleet size was also used to calibrate the demand process in Section 3.2, the historical purchase schedule will tend to appear as if it is ordering too many vehicles relative to the simulated demand. For a more informative comparison, we construct two reactive purchase policies. The first policy purchases a new vehicle when the average mission completion rate over the past three months dips below a tunable threshold α (we label this policy "COMP" for "completion"), while the second policy purchases a new vehicle when the average vehicle utilization over the past three months, calculated as the number of days needed to complete all missions divided by the total vehicle-day units in the fleet, is above a tunable threshold β (we label this policy "UTIL" for "utilization"). Thus, high α and low β lead to more purchases, while low α and high β lead to fewer.

Fig. 5 shows the costs and completion rates, in the high-demand scenario, for different assignment/sales combinations together with each of the two reactive policies. The Pareto fronts in each subplot are highlighted in red. For the COMP policy, out of 86 points on the Pareto front, 64 use the Balance assignment policy and 32 use the 100/225 sales policy (all other Pareto-optimal sales policies are 140/300 and RIG). The Pareto-optimal points are nearly equidistant and monotonic in α, suggesting that the manager could easily use α as the "knob" for achieving a desired trade-off between cost and completion rate. Including accidents in the model did not substantially change the shape or composition of the Pareto front in Fig. 5(a), but only moved it upward; for this reason, accidents are omitted from this discussion. For the UTIL policy, more points are concentrated on the right as most of the low β values produce very high completion rates. There are 31 Pareto points, of which 23 use the Least Odometer assignment policy, while 12 and 16, respectively, use the 140/300 and RIG sales policies.

Overall, COMP has a higher preference for Balance and 100/225, while UTIL leans toward Least Odometer and 140/300 or RIG. These results are consistent with the relationship between the demand level and the optimal assignment/sales combinations that we observed in Section 4.2. Under COMP, purchase decisions lag behind the demand to some extent, and so this scenario is closer to the high-demand scenario in Section 4.2, where Balance and 100/225 were indeed optimal. On the other hand, UTIL acts "preventively"

[8] Under demand level 4, the Newest rule essentially balances the residual value of the fleet since vehicles with higher residual value tend to have higher utilization. The Least Odometer rule also performs very similarly to Newest at this demand level.



(a) COMP policy: $\alpha = \{0.6, 0.61, ..., 0.99\}$.

(b) UTIL policy: $\beta = \{0.8, 0.81, ..., 0.99\}$.

Fig. 5. Cost and completion rate comparison of reactive purchase policies under high demand.

when the load on the fleet appears to be growing. As a result, the fleet size is more closely matched to the demand, producing results that resemble medium-demand levels in Section 4.2, where Least Odometer and 140/300 were preferable. Despite this distinction, the Pareto points for COMP and UTIL achieve similar costs under the same completion rate, though COMP provides more flexibility for trading off the two objectives.

4.4. Case 3: Switching of mission types

Current practice at LIHO locks in a single mission type for every new vehicle, and no vehicle is ever observed to complete missions from two different types. However, it is possible for a particular *vehicle type* to be feasible for multiple mission types: in the data, we frequently see different vehicles of the same type serving different mission types. It is natural to ask whether any improvement in efficiency can be achieved by allowing "switching" of mission types, or reassignment of vehicles from one type to another. Intuitively, allowing switching can help to use existing vehicles more efficiently without having to purchase new ones, which would reduce the cost of purchasing new vehicles and then incurring massive depreciation in the early stages of their lifespan. Since the data from FC include one vehicle type that is allowed to operate on two mission types, we can use our simulator to evaluate the effectiveness of switching.

We first consider an artificial, but illustrative scenario where switching might be expected to do well. The initial fleet composition is the same as in Section 4.2, but the underlying demand levels (trajectories of L_t) are now two sine functions rounded to the nearest integer (see Fig. 6(a)). Since switching should be most useful when the demand is high for one mission type and low for the other, one of the sine functions is lagged by half of its period, causing the two demand levels to be completely off-sync. There is still stochasticity in the demands; the distributions of (X_n) are calibrated using data for the two mission types.

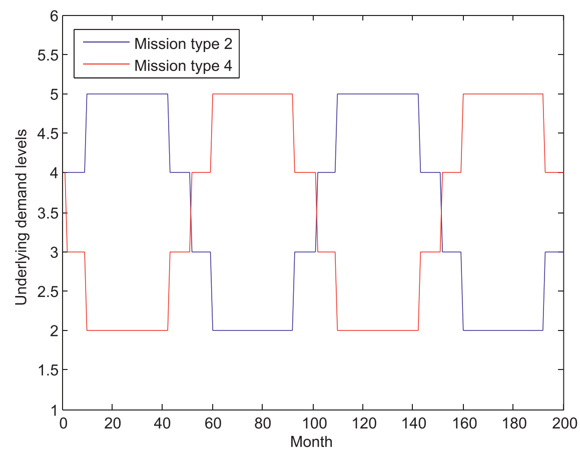
The switching policy works as follows. Every month, a switching decision is made based on the 3-month average mission completion rates and vehicle utilizations for both mission types: if the average completion rate for type 2 is 100%, the average completion rate of type 4 is *below* 100%, and the utilization for vehicles currently operating on type 2 is under 70% (i.e., we have vehicles to spare), the oldest vehicle that operates on type 2 is switched to type 4.[9] Similar criteria are used to switch from type 4 to type 2. If no switching occurs either way, a purchase decision is then made using the COMP logic. To obtain depreciation due to switching, we compute (8) for a hypothetical vehicle that has been running on the new mission type, and has also been given the same distance to travel; we then take the difference in the estimated residual values to get the monthly reduction after switching, thus bypassing all previous mission type history.

Figs. 6(b)–(d) show the performance of different assignment/sales combinations under the switching policy described above, compared to applying the COMP policy independently to both mission types without switching. To avoid clutter, only the Pareto-optimal combinations are reported in Fig. 6. We can observe that, for the most part, switching seems to offer very little benefit. The most improvement that we see is in Fig. 6(c), where switching can achieve about a 3% reduction in cost with the same completion rate. We repeated the same experiment with‘ the purchase price increased to 40,000 (from 30,000), the idea being that higher purchase prices should lead to greater savings in value-related cost. Surprisingly, however, we found that this scenario did not yield any advantage for the switching policy; on the contrary, the advantage observed in Fig. 6(c) disappeared. Essentially, this happened because, when the purchase cost is increased, the RIG policy becomes optimal in the nonswitching case. In other words, instead of switching vehicles, it is more efficient to just run them longer under their assigned mission type.

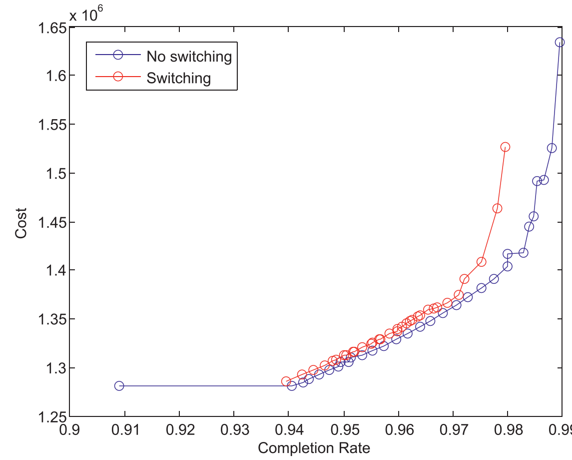
Next, we considered a scenario where the demand levels for both mission types were calibrated using LIHO data (with the model of Section 3.2). Fig. 7(a) shows the trajectory of (L_t) for *each* of the two mission types; we see that type 2 generates heavy demand in the first half of the planning period, then gradually ramps down, while type 4 steadily ramps up over the course of 143 months. Potentially, one might expect switching to offer some benefit when the two demand levels cross, since we could then switch some vehicles from type 2 to type 4. However, the Pareto fronts in Fig. 7(b) show that this is not at all the case: any savings obtained from switching are vanishingly small.

To obtain further insight into this surprising lack of improvement, we closely analyzed two Pareto-optimal policies with very similar completion rates: one policy used switching, the Balance

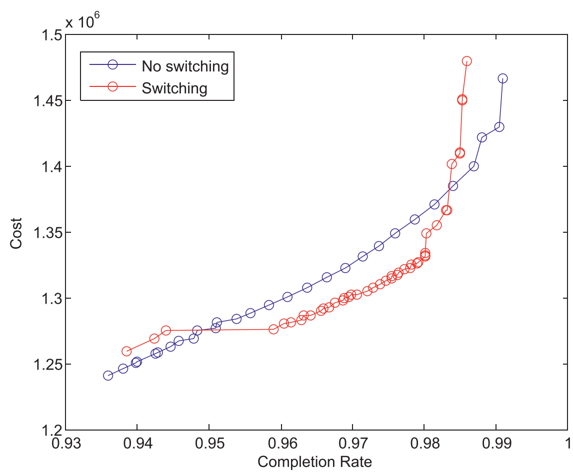
[9] We also considered other ways of choosing the vehicle to be switched (for example, based on least or most odometer), but this did not substantially change the conclusions.



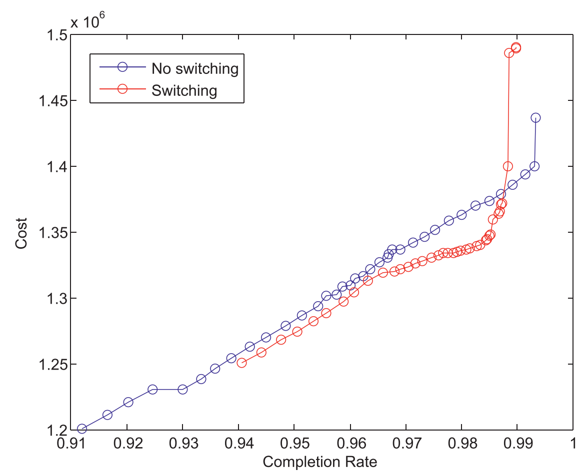
(a) Demand, period = 100.



(b) Performance, period = 20.

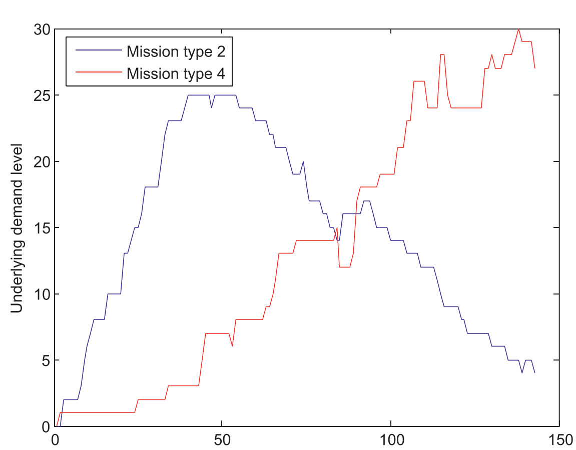


(c) Performance, period = 60.

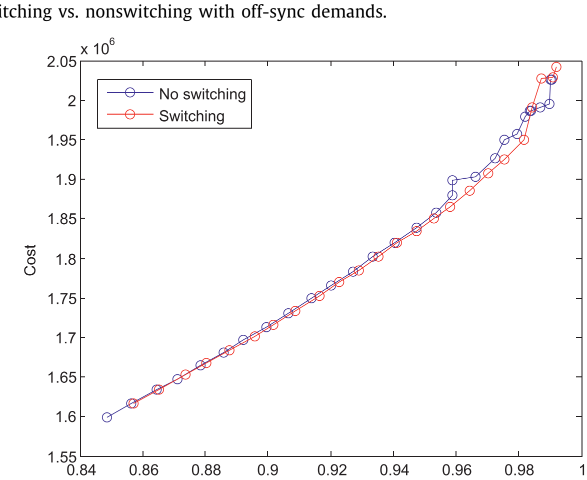


(d) Performance, period = 100.

Fig. 6. Demand levels and performance in switching vs. nonswitching with off-sync demands.



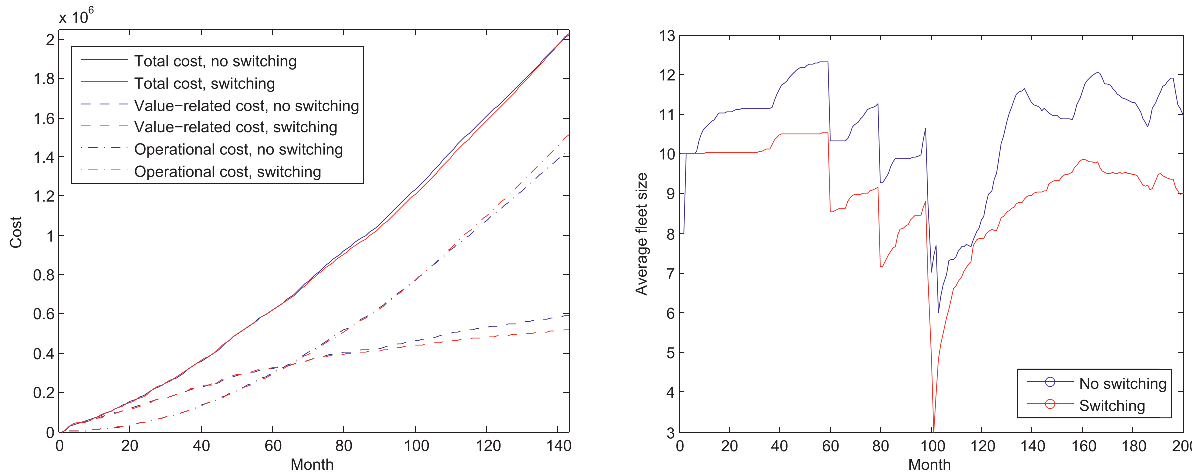
(a) Realistic demand levels for two mission types.



(b) Costs and completion rates, switching vs. not switching.

Fig. 7. Comparison of switching vs. non-switching under historical demand.

assignment policy, 140/300 sales, and $\alpha = 0.99$, while the other policy used the Least Odometer assignment policy, 140/300 sales, $\alpha = 0.99$ and no switching. Fig. 8(a) shows how total costs for both policies grow over the course of the planning period. Each cost trajectory is decomposed into *value-related cost* (purchase, depreciation) and *operational cost* (fuel, repair). Early on, both types of costs grow quite similarly for both policies (note that, as expected, value-related costs account for a greater share of the total earlier on). The savings from switching are seen from month 80 onwards in the value-related cost, reflecting the fact that some purchases have been avoided entirely by using switching. At the same time, switching has incurred a correspondingly greater operational cost,



(a) Costs over time, switching vs. not switching.

(b) Fleet size over time, switching vs. not switching.

Fig. 8. Comparison of switching vs. non-switching under historical demand.

reflecting the fact that existing vehicles are being exploited more heavily; as a result, the savings have been essentially negated. In some very specific cases, it may be possible to achieve (small) total savings by switching (as we saw in Fig. 6(c)), but in general there is no significant improvement. We believe that this offers a compelling explanation for what is typically practiced at LIHO.

It is worth noting that switching may have an extra benefit in terms of fleet size. Fig. 8(b) shows the fleet size over time for the two cases analyzed in the initial setting (the dip around month 100 occurs due to the 100/225 sales policy). It can be observed that the switching policy consistently keeps the total fleet size (both types combined) well below that of the non-switching policy. While this does not translate to big savings in the cost of operating the fleet, it may have other benefits such as requiring less storage space, smaller maintenance crews, and so on.

5. Conclusion

We have presented a holistic simulation environment for counterfactual analysis of purchase, assignment, and disposition of heterogeneous vehicles in humanitarian fleet management. In the context of the LIHO data, we used the simulator to obtain the following practical insights:

1. If cost is the primary objective, it is better to exploit vehicles longer than recommended by LIHO headquarters. The sales threshold should be reduced when the demand is higher, but there is no scenario in which the 60/150 threshold is cost-optimal.
2. When demand is low, it is better to assign it to a portion of the vehicles while leaving the others idle. For high demand, the best assignment policies are those that attempt to balance the load in some way.
3. Switching vehicles between mission types has marginal benefits at best: although they are used more efficiently, the cost savings are largely nullified by the accompanying increase in operational costs.

While these findings are valuable for LIHO and similar organizations, the simulation-based approach presented in this paper is more broadly useful, since it enables us to jointly study research questions that previously had only been studied independently using various mutually incompatible models. In certain cases, we can identify the rationale behind “the facts on the ground”, leading to a more productive dialogue between researchers and practitioners.

The main limitation of our work is the scarcity of data in the humanitarian sector. Although our model can accommodate a wide variety of fleet attributes, our implementation is by necessity limited to those covered by the LIHO data. Thus, we did not have data about individual missions or deliveries, nor did we have access to vehicle maintenance histories or inventories of spare parts. If such data become available, they may be used to elaborate further on the results presented here.

At a higher level, some of the findings in this paper challenge our current understanding of optimal decision-making in the field. We find that, in many complex situations, field managers are making effective decisions without in-depth analysis, perhaps even without a clear understanding of how the result was obtained (Tazelaar & Snijders, 2013), as may occur in “naturalistic decision making” A promising avenue for future work is to adopt methods of behavioral operations management to control for the element of intuition in complex humanitarian operations settings.

Appendix A. Additional results for FC

Section A.1 presents one additional case using the FC data. The focus of this case is on centralized vs. decentralized vehicle procurement. Section A.2 presents additional sensitivity analysis on the severity of accidents (i.e., the magnitude of their effect on residual value).

A1. Centralized purchases with lead times (FC)

Previously, we assumed that new vehicles join the fleet as soon as they are ordered. In practice, this corresponds to a situation where field managers purchase new vehicles from the local market. However, it often happens that managers have the additional option to purchase vehicles through headquarters, in which case there may be a lead time, perhaps as long as six months, before the vehicle will become available. This is known as a “centralized procurement model”, discussed in detail by Eftekhar et al. (2014) and Kunz & Van Wassenhove (2019). Under such a model, the HO has a long-term contract with the manufacturer, providing a discount that typically does not depend on the purchase quantity. As a consequence, purchasing locally can be 50% more expensive than purchasing through the HQ (Besiou, Pedraza-Martinez, & Van Wassenhove, 2014), creating a dilemma for fleet managers, who generally do not have sufficient bargaining power to obtain discounts on the local market.

Table 6
Comparison of costs between purchasing locally and ordering predictively for stable demand.

Demand level	Purchase locally			Order predictively
	10%	30%	50%	
1	165,519	177,519	189,519	158,339
2	298,123	320,331	336,855	285,701
3	472,906	496,906	520,906	460,403
4	739,328	763,328	787,328	720,886
5	879,084	928,680	978,276	848,291

Let us return to the setting of stable demand from Section 4.2. Consider a hypothetical situation where the manager has some partial knowledge of demand in the near future: specifically, the manager knows the underlying value of L_t, but not the random number of tasks generated from L_t or their random magnitudes. (The manager does know the distributional parameters of these random variables, as these can be estimated from data.) Using this information, the manager can make a crude forecast of the next time that a vehicle in the current fleet will reach zero residual value; if this is predicted to occur within the next six months, the manager orders a new vehicle for \$30,000. We then simulate all possible sales/assignment policy combinations together with this method of purchasing, and report the lowest cost achieved for five demand levels. Likewise, we simulate all possible sales/assignment policies for three settings where the manager can instantly obtain new vehicles with a cost markup of 10%, 30% and 50%, respectively.

Table 6 reports the results of the comparison; since completion rates were similar under all of the purchasing schemes, we focus on cost. Since the manager does not have perfect knowledge of the future, the savings obtained from centralized purchases are partially offset by increased utilization costs (since there may now be times when the remaining vehicles in the fleet are forced to take on more missions while waiting for a new vehicle to arrive). Nonetheless, there is a clear net gain from centralized purchases even when the markup is small. In the more realistic situation where the markup is 50%, we typically observe savings of around 10% relative to the decentralized scheme.

We also conducted a similar comparison for the realistic demands in Section 4.3 in the case of 50% markup. Here, in order for the two schemes to produce comparable completion rates, the decentralized setting uses the COMP policy ($\alpha = 0.8, 0.81, \ldots, 0.99$). In the centralized setting, the purchase policy anticipates future increases in demand (again, we assume that the fleet manager knows the trajectory of L_t) by keeping the fleet size at a higher level than what may be needed at the moment. Since UTIL also aims to preventively increase the fleet size, we use it as the purchase policy ($\beta = 0.6, 0.61, \ldots, 0.99$). The resulting Pareto fronts for both schemes are shown in Fig. 9, and a fair comparison can be made by looking at the cost figures under a fixed completion rate. Again, centralized purchases result in substantial savings (about 15%).

A2. Sensitivity analysis of accident severity

We revisit Case 1 in Section 4.2 and provide some additional results to complement Table 5, in which the effect of frequency of accidents on total cost was examined. Here, we use the baseline accident frequency, but vary the severity of accidents. In this context, "severity" refers to the coefficient of the variable NumAcc in Table 2. We consider two scenarios in which this number is cut in half and doubled, respectively, relative to its original estimated value. One can think of these scenarios as being optimistic and pessimistic estimates for "light" and "severe" accidents.

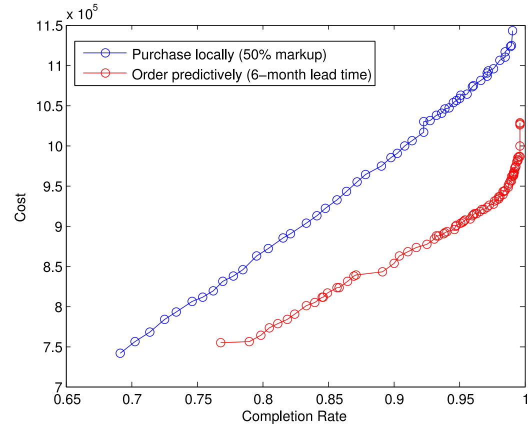


Fig. 9. Comparison of centralized and decentralized purchases.

The results are shown in Table 7. They are largely the same as those in Table 5, with the same best-performing sales/assignment combination for each demand level. It is interesting to note that, in some cases, increasing the severity may actually result in a slightly smaller average number of accidents. This occurs because more severe accidents cause vehicles to be replaced sooner, and new vehicles are slightly less likely to suffer accidents.

Appendix B. Simulation results for Country 2

In this section, we investigate the generalizability of our results by recalibrating our models from Sections 3.2–3.5 on data from a second developing country ("Country 2" or C2) where sufficiently good fuel, repair, and maintenance data were available. The fleet in C2 consists of 235 vehicles total and contains multiple vehicle types, with the most data available for types 2 and 3; Type 3 is the same vehicle type used in our analysis of FC (strong transporter, e.g., Land Cruiser or Pajero), while Type 2 represents light vans/minibuses. Thus, we are able to perform a double robustness check: first, we can run new simulations for C2, and second, we can also test whether our conclusions continue to hold on Type 2 vehicles. Unfortunately, C2 does not have enough data to allow detailed demand estimation for more than one mission type (namely Type 2), so we do not study mission type switching (Case 3) in this discussion. However, Cases 1 and 2, as well as the additional case in Section A.1, can all be considered.

B1. Estimation results for cost models

Since the depreciation model from Section 3.3 was estimated using data from 20 countries, we simply report those coefficients that are relevant to C2. The cross-sectional data for C2 contains 4 mission types and 4 vehicle types, so the model is

$$\text{Dep\%} = \beta_1 \text{Age} + \sum_{i,j} \beta_{2ij} \log\left(\text{Odometer}\right) \times \text{MissionType}_i \times \text{VehicleType}_j + \beta_3 \text{NumAcc}, \quad \text{(B.1)}$$

and the results are summarized in Table 8. Note that the coefficients for age and number of accidents are the same as in Section 3.3, since only a single estimation was performed on the full dataset. However, there is significant heterogeneity in the impact of odometer depending on the vehicle and mission type involved.

Next, we estimate the fuel cost model using C2 data. We use the same model as in Section 3.4, but add dummy variables representing vehicle types since our previous analysis focused on a

Table 7
Cost comparison of sales/assignment policies with varying accident severity.

Accident Severity	Demand Level	Sales	Assignment	Total cost	Avg. no. of accidents
0.5x Baseline	1	RIG	Myopic	160,413	1.437
	2	140/300	Myopic	287,947	2.700
	3	140/300	Least Odometer	463,557	3.922
	4	100/225	Newest	724,938	4.942
	5	100/225	Balance	858,525	6.262
2x Baseline	1	RIG	Myopic	163,315	1.334
	2	140/300	Myopic	292,630	2.760
	3	140/300	Least Odometer	469,942	3.862
	4	100/225	Newest	734,846	5.008
	5	100/225	Balance	871,490	6.172

Table 8
Residual value estimation results for C2.

	Dependent variable:
	Depreciation Percentage
Age	2.215×10^{-3} ***
NumAcc	0.04512***
log(Odometer) × Mission1 × VehicleType4	0.0355***
log(Odometer) × Mission2 × VehicleType1	0.0021***
log(Odometer) × Mission2 × VehicleType2	0.0405***
log(Odometer) × Mission2 × VehicleType3	0.0232***
log(Odometer) × Mission3 × VehicleType4	0.0333***
log(Odometer) × Mission4 × VehicleType1	0.0303***
log(Odometer) × Mission4 × VehicleType2	0.0520***
log(Odometer) × Mission4 × VehicleType3	0.0489***
Observations	3846 (235 in C2)
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$

Table 9
Fuel cost estimation results for C2.

	Dependent variable:
	Cost per Km
Odometer	3.194×10^{-8} *
Age	5.930×10^{-5} *
Mission2	-0.1284***
Mission4	-0.1268***
VehicleType2	0.0084*
VehicleType3	0.0529***
Constant	0.1955***
Observations	4747
R^2	0.5507
Adjusted R^2	0.5428
Residual Std. Error	0.03132 (df = 4663)
F Statistic	68.87*** (df = 83; 4663)
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$

single type. The model thus becomes

$$\text{FuelCostPerKm} = \beta_0 + \beta_1\text{Odometer} + \beta_2\text{Age} + \sum_j \beta_{3j}\text{MissionType}_j + \sum_k \beta_{4k}\text{VehicleType}_k.$$

Table 9 shows the results. As before, age and odometer are positively correlated with fuel cost.

Next, we estimate the maintenance cost model. We found that vehicle types do not appear to have a significant correlation with maintenance/repair costs. Therefore, we assume that vehicles working on the same mission type follow similar maintenance/repair schedules, and the model remains unchanged from Section 3.5:

$$\text{CC} = \sum_j \beta_{1j}\text{Age} \times \text{MissionType}_j + \beta_{2j}\text{Odometer} \times \text{MissionType}_j + \beta_{3j}\text{Odometer}^2 \times \text{MissionType}_j.$$

Table 10
Maintenance/repair cost estimation results for C2.

	Dependent variable:
	CumuCost
Age	34.638***
Odometer	0.021***
Odometer^2	1.165×10^{-7} ***
Observations	520
Log Likelihood	−4,074.901
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$

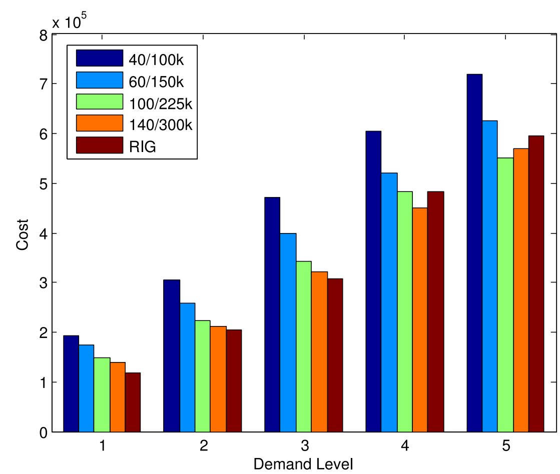


Fig. 10. Cost comparison of sales policies under demand levels 1–5.

The results are shown in Table 10. Since C2 only has enough data on mission type 2, the MissionType component of the interaction terms is omitted; we see, however, that the coefficients are overall similar to what we saw in Section 3.5, with the main difference being that the nonlinear effect of odometer is less pronounced. Thus, if enough data can be gathered to perform estimation reliably, one can recalibrate the cost module of the simulator. If, additionally, monthly mileage data are available, one can calibrate the demand module and repeat the analysis of Section 4.

B2. Simulation results for C2, vehicle type 2

We repeated three cases from our study of FC on the C2 data with mission type 2 and vehicle type 2. First, let us consider Case 1 (comparison of sales policies), in which the demand is stable, and five demand levels are considered with an initial fleet size of four vehicles. In other words, the experimental setup is the same, but costs are now calculated according to the C2 models. First, we summarize the results of the base case (no accidents) in Fig. 10. The overall pattern is similar to what we observed in the FC study,

Table 11
Comparison between different levels of accidents.

Accident Level	Demand Level	Policies	Cost	NumOfAccidents
0	1	Myopic, RIG	118,217	0
	2	Myopic, RIG	204,220	0
	3	Least Odometer, RIG	307,332	0
	4	Balance, 140/300	449,596	0
	5	Balance, 100/225	550,307	0
Baseline	1	Most Odometer, RIG	126,206	1.81
	2	Myopic, RIG	211,178	4.13
	3	Least Odometer, RIG	324,512	5.76
	4	Balance, 140/300	459,015	6.97
	5	Balance, 100/225	558,606	6.13
2x Baseline	1	Oldest, RIG	131,178	3.68
	2	Myopic, RIG	219,344	7.62
	3	Least Odometer, 140/300	332,864	7.95
	4	Balance, 140/300	468,697	14.11
	5	Balance, 100/225	567,274	12.53
4x Baseline	1	Most Odometer, RIG	140,256	6.98
	2	Myopic, 140/300	226,891	11.32
	3	Least Odometer, 140/300	342,903	15.63
	4	Least Odometer, 140/300	488,761	27.19
	5	Balance, 100/225	583,912	24.77

Table 12
Cost comparison with varying accident severity.

Accident Severity	Demand Level	Policies	Total cost	Avg. no. of accidents
0.5x Baseline	1	Most Odometer, RIG	123,922	1.800
	2	Myopic, RIG	208,553	4.253
	3	Least Odometer, RIG	321,128	6.354
	4	Balance, 140/300	454,285	6.928
	5	Balance, 100/225	554,513	6.214
2x Baseline	1	Most Odometer, RIG	131,274	1.772
	2	Myopic, RIG	217,887	3.847
	3	Least Odometer, RIG	331,648	5.003
	4	Balance, 140/300	468,985	6.963
	5	Balance, 100/225	566,888	6.110

except that we now exploit vehicles for slightly longer: RIG is preferred for levels 1–3, followed by 140/300 at level 4 and 100/225 at level 5.

Next, Table 11 compares optimal sales and assignment policies for varying accident frequencies, as was done in our study of FC. The results are very similar to FC and consistent with the base case: Myopic and RIG are generally optimal for lower demand, while 140/300 and 100/225 (together with load balancing) are preferred when the demand is high. The optimal sales thresholds move slightly earlier with higher likelihood of accident, as this makes vehicles wear out faster and has an effect similar to increasing the load on the fleet. At low demand levels, the Oldest and Most Odometer policies, like Myopic, have the effect of concentrating utilization on a portion of the available vehicles. We can also see the impact of sales policies on safety: by using 100/225 sales instead of 140/300, accidents are actually reduced between demand levels 4 and 5.

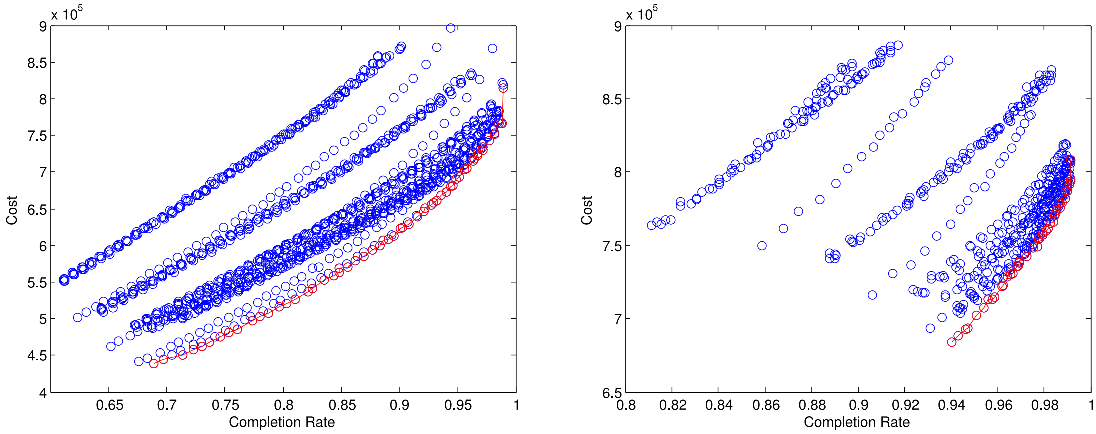
Table 12 repeats the analysis of Table 7, which varied the severity of accidents. Again, we use the baseline accident frequency from Table 11, but run two scenarios where the coefficient of NumAcc in (B.1) is half and double, respectively, of its estimated value in Table 8. The results are virtually unchanged from the baseline scenario in Table 11.

Next, we consider Case 2, in which realistic demand data (in this case, for mission type 2) is used to evaluate purchase, assignment and sales policies jointly. The same reactive purchase policies from Section 4.3 were implemented. The Pareto fronts for both policies are shown in Fig. 11. For COMP, 49 out of 58 points on the Pareto front use the Balance assignment policy, and 41 out of 58 use 140/300 sales; for UTIL, 24 out of 46 Pareto-optimal points use the Least Odometer assignment policy and 26 out of 46 use 140/300 sales (with another 20 using RIG). This is fairly consistent with the results observed for FC, where COMP behaves similarly to a high-demand scenario (hence the preference for Balance) while UTIL behaves similarly to a medium/high-demand level.

Table 13
Comparison of costs between purchasing locally and ordering predictively for stable demand.

Demand level	Purchase locally			Order predictively
	10%	30%	50%	
1	121,217	127,217	133,217	117,517
2	208,717	213,554	213,554	203,184
3	312,741	322,843	322,843	307,425
4	461,596	485,596	509,596	448,206
5	575,105	613,046	625,046	546,630

Finally, we consider the case from Section A.1, which compares centralized procurement with a 6-month lead time vs. decentralized procurement with no lead time but higher purchase cost. Table 13 reports the results of the comparison for stable demand, assuming that the fleet manager has rough knowledge of the demand over the next six months. Just as for FC, we see that centralized procurement is better for all five demand levels; however, the savings are smaller overall. This is due to the fact that the nonlinear behaviour of the maintenance cost (which causes steep increases late in vehicles' lifetimes) is less pronounced in C2, so it is generally less prohibitive to keep vehicles longer.



(a) COMP policy: $\alpha = \{0.6, 0.61, ..., 0.99\}$. (b) UTIL policy: $\beta = \{0.8, 0.81, ..., 0.99\}$.

Fig. 11. Cost and completion rate comparison of reactive purchase policies in C2.

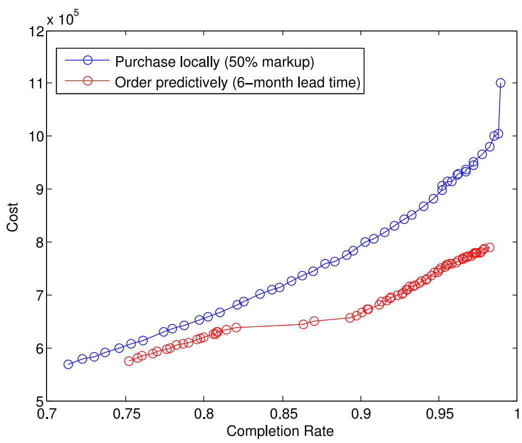


Fig. 12. Comparison of centralized and decentralized purchases.

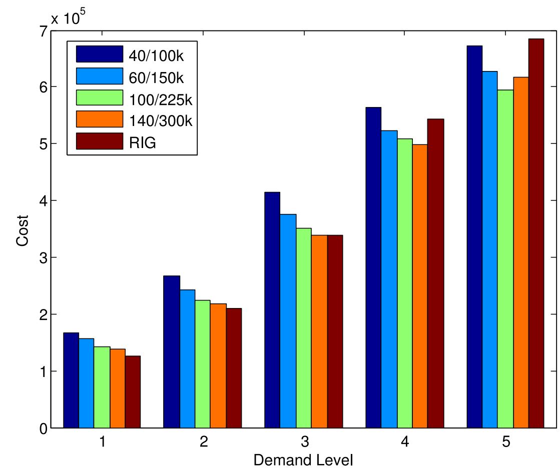


Fig. 13. Cost comparison of sales policies under demand levels 1–5.

For this reason, we make fewer purchases under decentralized procurement and so the difference between purchase prices is of less importance. Note that, in certain cases (low demand levels), different models may incur identical costs, because the demand is so low that no new purchases are made. Overall, however, the general observation that centralized procurement is preferable, given the ability to order anticipatively, is still valid.

Fig. 12 compares the performance of centralized vs. decentralized procurement (with 50% markup for local purchases) under realistic demand for C2. The same conclusions apply.

B3. Simulation results for C2, vehicle type 3

Finally, we repeat the previous simulations for mission type 2 and vehicle type 3, i.e., we remain in C2 but consider a different vehicle type. The cost models are unchanged from Section B.1 as they already considered both vehicle types. Fig. 13 reports the results of Case 1; the conclusions are very similar to those in Section B.2.

Next, Table 14 compares optimal sales and assignment policies for varying accident frequencies, with results that are consistent with both FC and the other vehicle type from C2. Myopic and RIG are optimal for lower demand, while 140/300 and 100/225 (together with load balancing) are preferred when the demand is high. Table 15 varies the accident severity and obtains virtually the same results as the baseline in Table 14.

In Case 2, Fig. 14 shows that, for COMP, 87 out of 94 Pareto-optimal points use the Balance assignment policy, while 42 out of 94 use 140/300 sales (and 37 out of 94 use RIG). For UTIL, 25 out of 51 Pareto-optimal points use Balance assignment (and 24 out of 51 use Least Odometer), while 25 out of 51 use 140/300 sales (and 16 use RIG). These results are quite consistent with Section B.2 as well as with the main study of FC.

Table 16 reports the results of procurement analysis for stable demand, while Fig. 15 does the same for realistic demand. These results are similar to those observed for FC since the same vehicle type is considered in both cases.

B4. Discussion

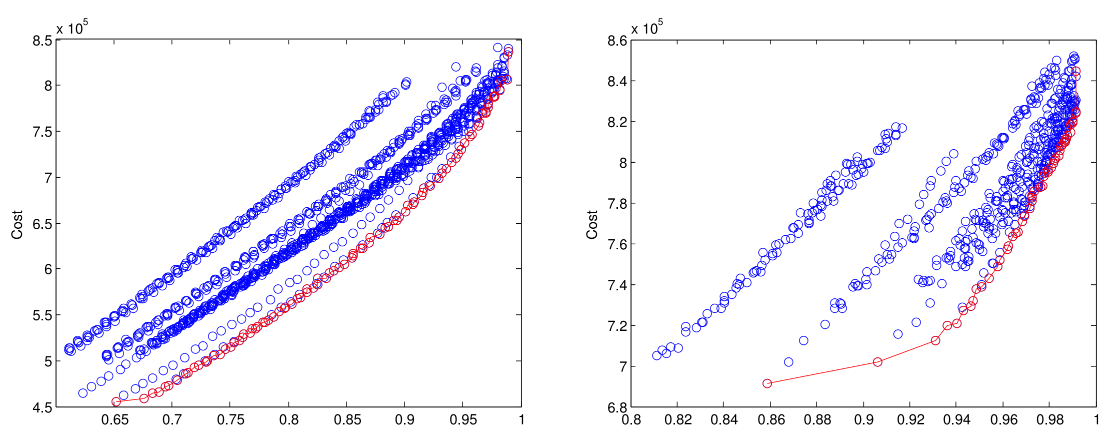
Overall, the results in Sections B.2–B.3 are similar both to each other and to our earlier results for FC. When vehicle type 2 is used (Section B.2), the results behave as if the demand were “lower”, i.e., the patterns that we typically see for demand level 1 persist for demand level 2, and the transition from Myopic to Least Odometer (or from RIG to 140/300) takes place later than for FC. In general, C2 imposes less severe penalties on operational costs for vehicles late in their lifetimes, allowing us to run them longer than was the case in FC. The results for type-3 vehicles tend to be in between the results for type 2 and the original results for FC (which also considered type-3 vehicles).

Table 14
Comparison between different levels of accidents.

Accident Level	Demand Level	Policies	Cost	NumOfAccidents
0	1	Myopic, RIG	126,743	0
	2	Least Odometer, RIG	210,614	0
	3	Least Odometer, 140/300	337,772	0
	4	Balance, 140/300	497,793	0
	5	Balance, 100/225	594,451	0
Baseline	1	Myopic, RIG	129,126	1.76
	2	Least Odometer, RIG	216,130	4.06
	3	Least Odometer, 140/300	343,170	3.99
	4	Balance, 140/300	507,293	7.02
	5	Balance, 100/225	602,801	6.17
2x Baseline	1	Myopic, RIG	131,466	3.47
	2	Least Odometer, RIG	221,666	7.88
	3	Least Odometer, 140/300	348,397	7.85
	4	Balance, 140/300	516,776	14.03
	5	Balance, 100/225	611,406	12.53
4x Baseline	1	Myopic, RIG	136,836	6.92
	2	Least Odometer, RIG	233,815	14.78
	3	Least Odometer, 140/300	359,035	15.71
	4	Newest, 100/225	535,641	19.67
	5	Balance, 100/225	627,917	24.72

Table 15
Cost comparison with varying accident severity.

Accident Severity	Demand Level	Policies	Total cost	Avg. no. of accidents
0.5x Baseline	1	Myopic, RIG	127,968	1.809
	2	Least Odometer, RIG	213,362	4.060
	3	Least Odometer, 140/300	340,471	3.988
	4	Balance, 140/300	502,519	6.982
	5	Balance, 100/225	598,661	6.221
2x Baseline	1	Myopic, RIG	131,670	1.764
	2	Least Odometer, RIG	221,582	3.790
	3	Least Odometer, 140/300	348,536	3.978
	4	Balance, 140/300	516,329	6.846
	5	Balance, 100/225	611,043	6.129



(a) COMP policy: $\alpha = \{0.6, 0.61, \ldots, 0.99\}$. (b) UTIL policy: $\beta = \{0.8, 0.81, \ldots, 0.99\}$.

Fig. 14. Cost and completion rate comparison of reactive purchase policies in C2.

These differences illustrate the ability of our simulator to fine-tune its recommendations to the realities of each country and local office. However, the big-picture tendencies observed in this study do not differ from those obtained for FC in any major way.

Appendix C. Justification of cost and demand models

In this section, we provide additional arguments in support of our chosen cost and demand models relative to certain alternatives. Section C.1 provides additional validation of the cost models in Sections 3.3–3.5. Section C.2 provides additional discussion of the demand model.

Table 16
Comparison of costs between purchasing locally and ordering predictively for stable demand.

Demand level	Purchase locally			Order predictively
	10%	30%	50%	
1	126,743	126,743	154,220	126,743
2	210,614	210,614	246,337	210,614
3	338,186	338,186	375,032	337,335
4	509,793	533,793	552,154	496,243
5	619,249	668,845	665,147	590,504

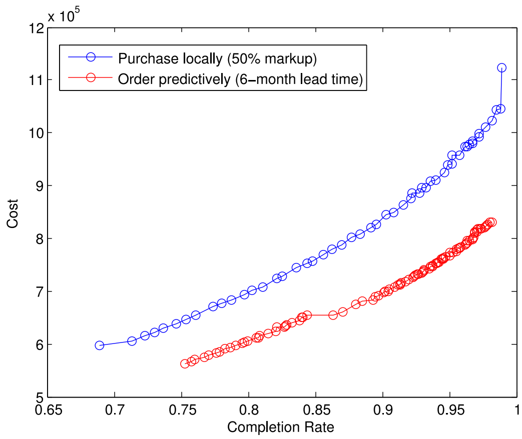


Fig. 15. Comparison of centralized and decentralized purchases.

C1. Cost models

The statistical models in Sections 3.3–3.5 can be validated by comparing against alternative specifications or model types. We would like to choose the model with the best explanatory power, which for a linear regression model can be measured in terms of R^2. At the same time, the model must reflect certain important aspects of the real-world problem: for example, a zero-intercept model should be used for residual value, because the initial value of the vehicle (with age and odometer set to zero) should be equal to the purchase price. A model with an intercept variable will typically appear to produce a better fit to the data, but such a model also runs the risk of creating arbitrage opportunities where the value of a completely new vehicle is estimated to be higher than the purchase price. Even if such a model happens to have lower squared error on the observed values, it is not useful for our purposes.

First, as an alternative to the residual value model (8), we consider an ordinary linear regression model without the Tobit property. That is, residual values of 0 in the data are treated as the actual observed value of the vehicles, rather than negative values censored at zero as in Section 3.3. Table 17 presents the estimation results for this model, including relevant terms for both FC and C2. These results are quite similar to those in Tables 2 and 8; furthermore, the R^2 value for this model is quite high (unfortunately an analogous quantity is not readily available for the Tobit model). Since there are only 108 vehicles out of 3728 in the data that had zero value when sold, the inclusion or omission of the Tobit property does not significantly change the results. Potentially, one could use the ordinary least squares model, if one felt that the high R^2 was a strong argument in favour.

To demonstrate that this alternative model does not affect the conclusions, we tested it on Case 1 from Section 4.2 (stable demand with pure replacement). Table 18 reports the simulated costs for various sales policies under the fuel and maintenance models used in Sections 3.4–3.5. While the numbers themselves are slightly different from those reported in Section 4.2, the conclusions are unchanged: the best policy is Myopic under low demand, transitioning to load-balancing policies as the demand increases.

Next, we considered a number of variations on the fuel model of Section 3.4. The model (9) is relatively simple, since it is estimated separately for each country and only includes three features in FC (six in C2 due to the greater variety of mission and vehicle types). To provide further justification for this model, we compared it to several alternatives with some of these features removed or transformed; for example, we considered replacing Odometer by $\log(\text{Odometer})$, which increases more slowly for high odometer values. The R^2 values for these models are given in Table 19. Overall, the variations do not substantially change the performance of the model. The very last variant, in which an interaction term between odometer and mission type is considered, slightly improves performance in FC (only in the third decimal point) but not in C2; furthermore, in C2, the estimated coefficients of this interaction are not statistically significant. Thus, there are no strong arguments in favour of replacing the base model.

Similarly, we considered several variations of the maintenance cost model of Section 3.5. Recall that this is a zero-intercept model with a random effect representing vehicle ID. Two types of R^2 values are available for mixed-effect models: marginal R^2, which represents the proportion of variance explained by the fixed effects only, and conditional R^2, which represents the proportion of variance explained by the full model. Both types of values are given in Table 20 for the base model as well as the variants; the first variant does not apply to C2 since there is only one mission type in that country (thus, the variant will be identical to the base model). Overall, we see that the base model has a high conditional R^2 (above 0.96 in both countries), and none of the variants improves on this value. Furthermore, some of the variants produce estimates that are unreasonable for simulation. For example, replacing odometer and squared odometer by $\log(\text{Odometer})$ results in negative coefficients for the latter, which would imply the obviously incorrect conclusion that cumulative cost decreases with usage. Adding higher-degree polynomial terms is also undesirable since it has the effect of producing extreme cost values for very small or very large odometers. In other words, there are no strong arguments in favour of replacing the base model.

C2. Demand model

The demand module of our simulator serves a somewhat different purpose than the cost module. When it comes to cost, the main concern is realism, because we use cost as the main performance measure, and our conclusions about various policies depend on the cost calculations used to evaluate them. The demand module, however, is designed with more flexibility in mind: one can try to mimic historical demand, as in Fig. 3(a), but since the historical demand itself is censored, one may also be interested in settings where demand is higher than historical, as well as other scenarios of interest such as the “off-sync” configuration in Fig. 6(a). Furthermore, one will not be able to evaluate assignment policies without a means of decomposing the observed monthly demand into individual missions, something that our model in (7) allows.

Having said this, we discuss here a simple alternative model that does not have the full capability of the model in Section 3.2, but is useful as an illustration. As in Section 3.2, the monthly demand is generated from the trajectory of an underlying process (L_t) which can be viewed as a “latent fleet size” roughly measuring the workload on the HO; that is, L_t is the number of vehicles that we should have on hand at time t in order to complete all or most of the missions. The alternative model fixes L_t to the historical fleet size, and generates monthly demand according

Table 17
Estimation results for OLS residual value model.

	Dependent variable: Percentage.Depreciation
Age	2.2146×10^{-3} ***
Accident	0.04511^{***}
log(Odometer)×FC × Mission 2 × VehicleType3	0.04845^{***}
log(Odometer)×FC × Mission 4 × VehicleType3	0.05399^{***}
log(Odometer)×C2 × Mission 1 × VehicleType4	0.03546^{***}
log(Odometer)×C2 × Mission 2 × VehicleType1	0.0021
log(Odometer)×C2 × Mission 2 × VehicleType2	0.0405^{***}
log(Odometer)×C2 × Mission 2 × VehicleType3	0.02324^{***}
log(Odometer)×C2 × Mission 3 × VehicleType4	0.03330^{***}
log(Odometer)×C2 × Mission 4 × VehicleType1	0.03028^{***}
log(Odometer)×C2 × Mission 4 × VehicleType2	0.05199^{***}
log(Odometer)×C2 × Mission 4 × VehicleType3	0.04892^{***}
Observations	3846
R^2	0.928
Adjusted R^2	0.924
Residual Std. Error	0.175 (df = 3660)
F Statistic	252.127^{***} (df = 186; 3660)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 18
Cost comparison for Case 1 under the OLS residual value model.

Demand	40/100	60/150	100/225	140/300	RIG	Best
1	198,266	183,952	164,540	156,510	145,568	Myopic
2	333,543	299,002	278,733	275,178	275ins,322	Myopic
3	534,124	480,741	451,460	440,014	489,049	Least Odometer
4	732,882	692,772	685,123	705,234	885ins,866	Newest
5	879,922	830,668	816,790	892,508	1,161,100	Balance

Table 19
Performance comparison for several variants of the fuel cost model.

Model Description	R^2 value (FC)	R^2 value (C2)
Base model (Section 3.4)	0.301	0.551
Odometer replaced by log (Odometer)	0.300	0.550
Age removed	0.298	0.548
Adding Odometer × Mission Type	0.306	0.551

to $D_t = \beta_0 + \beta_1 L_t + \varepsilon_t$, where $\beta_0 = 3760$ and $\beta_1 = 817.62$ are estimated using ordinary least squares regression ($R^2 = 0.87$), and ε_t is generated from a normal distribution with zero mean and variance proportional to L_t.

Fig. 16 presents an illustration of simulated monthly demands under the alternative model compared to historical ones, as in Fig. 3(b). While the simulations are generally close to the historical values, both sample paths consistently overestimate demand at the beginning of the time horizon. The main reason is the high estimated value of β_0 in the alternative model; since this model does not have the flexibility to generate varying numbers of individual tasks, its output fluctuates more closely around the intercept in the early stages. For the same reason, the alternative model also cannot be used inside the simulator, which requires individual missions in order to compare assignment policies. This is the primary motivation for the use of a compound random variable model.

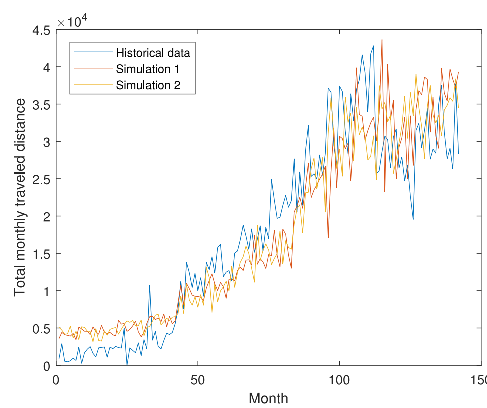


Fig. 16. Trajectory of monthly demand under alternative model.

We close this discussion with an explanation of how the Tweedie distribution in our demand model was calibrated. As discussed in Section 3.2, the parameters $(\lambda, \alpha, \gamma)$ of the distribution can be made to depend on the value of L_t, reflecting the fact that, in the LIHO data, fleet size appears to be correlated with distance travelled. As explained in Zhang et al. (2013), the precise dependence can be obtained as follows. Given four parameters $(\tau_0, \tau_1, p, \phi)$ estimated numerically from data, one calculates $\mu =$

Table 20
Comparison of R^2 values for several variants of the maintenance model.

Model Description	Marginal (FC)	Conditional (FC)	Marginal (C2)	Conditional (C2)
Base model (Section 3.5)	0.874	0.967	0.894	0.965
Age × Mission Type replaced by Age	0.877	0.967	–	–
Odometer replaced by log (Odometer)	0.684	0.880	0.868	0.951
Higher power of Odometer	0.896	0.967	0.906	0.965

Table 21
Estimated parameters of Tweedie model for various cases.

Case	τ_0	τ_1	p	ϕ
Section 4.2	6.8803	0.2465	1.1608	171.0340
Section 4.3	7.4961	-0.0199	1.2377	74.7419
Appendix B	7.6258	-0.0116	1.3983	47.9957

$e^{\tau_0+\tau_1 L_t}$ and uses this quantity to compute $\lambda = \frac{\mu^{2-p}}{\phi(2-p)}$, $\alpha = \frac{2-p}{p-1}$, and $\gamma = \phi(p-1)\mu^{p-1}$. Table 21 provides the estimated parameter values for various cases used in our study. Section 4.4 uses the same Tweedie parameters as Section 4.3 for both mission types. Likewise, Section Appendix B uses the same Tweedie parameters for both vehicle types.

References

Acimovic, J., & Goentzel, J. (2016). Models and metrics to assess humanitarian response capacity. *Journal of Operations Management, 45*(1), 11–29.

Altay, N., & Pal, R. (2014). Information diffusion among agents: implications for humanitarian operations. *Production and Operations Management, 23*(6), 1015–1027.

Arsenault, M., Rouleau, E., Fernandez Gomez, J. M., Somodi, C., & Yanni, P.-Y. (2018). Evaluation of UNHCR's global fleet management. UNHCR Evaluation Service (ES/2018/13).

Balcik, B., & Beamon, B. M. (2008). Facility location in humanitarian relief. *International Journal of Logistics: Research and Applications, 11*(2), 101–121.

Balcik, B., Beamon, B. M., Krejci, C. C., Muramatsu, K. M., & Ramirez, M. (2010). Coordination in humanitarian relief chains: Practices, challenges and opportunities. *International Journal of Production Economics, 126*(1), 22–34.

Balcik, B., Beamon, B. M., & Smilowitz, K. (2008). Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems, 12*(2), 51–63.

Beamon, B. M., & Kotleba, S. A. (2006). Inventory management support systems for emergency humanitarian relief operations in South Sudan. *The International Journal of Logistics Management, 17*(2), 187–212.

Ben-Tal, A., Chung, B. D., Mandala, S. R., & Yao, T. (2011). Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation Research Part B: Methodological, 45*(8), 1177–1189.

Besiou, M., Pedraza-Martinez, A. J., & Van Wassenhove, L. N. (2014). Vehicle supply chains in humanitarian operations: Decentralization, operational mix, and earmarked funding. *Production and Operations Management, 23*(11), 1950–1965.

Bhattacharya, S., Hasija, S., & Van Wassenhove, L. N. (2014). Designing efficient infrastructural investment and asset transfer mechanisms in humanitarian supply chains. *Production and Operations Management, 23*(9), 1511–1521.

Chacko, J., Rees, L. P., Zobel, C. W., Rakes, T. R., Russell, R. S., & Ragsdale, C. T. (2016). Decision support for long-range, community-based planning to mitigate against and recover from potential multiple disasters. *Decision Support Systems, 87*, 13–25.

Charles, A., Lauras, M., Van Wassenhove, L. N., & Dupont, L. (2016). Designing an efficient humanitarian supply network. *Journal of Operations Management, 47–48*, 58–70.

Crainic, T. G., & Laporte, G. (2012). *Fleet management and logistics*. Springer Science & Business Media.

Crooks, A. T., & Wise, S. (2013). GIS And agent-based models for humanitarian assistance. *Computers, environment and urban systems, 41*, 100–111.

Dolinskaya, I. S., Shi, Z. E., Smilowitz, K. R., & Ross, M. (2011). Decentralized approaches to logistics coordination in humanitarian relief. In T. Doolen, & E. V. Aken (Eds.), *Proceedings of the 2011 Industrial Engineering Research Conference*.

Eftekhar, M., Masini, A., Robotis, A., & Van Wassenhove, L. N. (2014). Vehicle procurement policy for humanitarian development programs. *Production and Operations Management, 23*(6), 951–964.

Eftekhar, M., & Van Wassenhove, L. N. (2016). Fleet management policies for humanitarian organizations: Beyond the utilization-residual value trade-off. *Journal of Operations Management, 44*(1), 1–12.

Ergun, O., Gui, L., Heier Stamm, J. L., Keskinocak, P., & Swann, J. (2014). Improving humanitarian operations through technology-enabled collaboration. *Production and Operations Management, 23*(6), 1002–1014.

Gralla, E., Goentzel, J., & Fine, C. (2014). Assessing trade-offs among multiple objectives for humanitarian aid delivery using expert preferences. *Production and Operations Management, 23*(6), 978–989.

Hamedi, M., Haghani, A., & Yang, S. (2012). Reliable transportation of humanitarian supplies in disaster response: Model and heuristic. *Procedia Social and Behavioral Sciences, 54*, 1205–1219.

Hayashi, A. M. (2001). When to trust your gut. *Harvard Business Review, 79*(2), 59–65.

Iakovou, E., Vlachos, D., Keramydas, C., & Partsch, D. (2014). Dual sourcing for mitigating humanitarian supply chain disruptions. *Journal of Humanitarian Logistics and Supply Chain Management, 4*(2), 245–264.

Jahre, M., Kembro, J., Rezvanian, T., Ergun, O., & Hapnes, S. J. (2016). Integrating supply chains for emergencies and ongoing operations in UNHCR. *Journal of Operations Management, 45*, 57–72.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychological Association, 64*(6), 515–526.

Kaplan, E. H., Craft, D. L., & Wein, L. M. (2002). Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy of Sciences, 99*(16), 10935–10940.

Khatri, N., & Ng, H. A. (2000). The role of intuition in strategic decision making. *Human Relations, 53*, 57–86.

Kunz, N., & Van Wassenhove, L. N. (2019). Fleet sizing for UNHCR country offices. *Journal of Operations Management, 65*(3), 282–307.

Matl, P., Hartl, R. F., & Vidal, T. (2017). Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science, 52*(2), 239–260.

McClintock, A. (2009). The logistics of humanitarian emergencies: Notes from the field. *Journal of Contingencies and Crisis Management, 17*(4), 295–302.

McCoy, J. H., & Lee, H. L. (2014). Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management, 23*(6), 965–977.

Ni, W., Shu, J., & Song, M. (2018). Location and emergency inventory pre-positioning for disaster response operations: Min-max robust model and a case study of Yushu earthquake. *Production and Operations Management, 27*(1), 160–183.

Pedraza-Martinez, A. J., Hasija, S., & Van Wassenhove, L. N. (2020). Fleet coordination in decentralized humanitarian operations funded by earmarked donations. *Operations Research (to appear)*.

Pedraza-Martinez, A. J., Stapleton, O., & Van Wassenhove, L. N. (2011). Field vehicle fleet management in humanitarian operations: a case-based approach. *Journal of Operations Management, 29*(5), 404–421.

Pedraza-Martinez, A. J., & Van Wassenhove, L. N. (2012). Transportation and vehicle fleet management in humanitarian logistics: challenges for future research. *EURO Journal on Transportation and Logistics, 1*(1–2), 185–196.

Pedraza-Martinez, A. J., & Van Wassenhove, L. N. (2013). Vehicle replacement in the International Committee of the Red Cross. *Production and Operations Management, 22*(2), 365–376.

Powell, W. B. (2011). *Approximate dynamic programming: Solving the curses of dimensionality (2nd ed.)*. New York: John Wiley and Sons.

Rauner, M. S. (2002). Using simulation for AIDS policy modeling: benefits for HIV/AIDS prevention policy makers in Vienna, Austria. *Health Care Management Science, 5*(2), 121–134.

Rawls, C. G., & Turnquist, M. A. (2012). Pre-positioning and dynamic delivery planning for short-term response following a natural disaster. *Socio-Economic Planning Sciences, 42*(1), 46–54.

Sahebjamnia, N., Torabi, S. A., & Mansouri, S. A. (2017). A hybrid decision support system for managing humanitarian relief chains. *Decision Support Systems, 95*, 12–26.

Salmeron, J., & Apte, A. (2010). Stochastic optimization for natural disaster asset prepositioning. *Production and Operations Management, 19*(5), 561–574.

Simão, H. P., George, A., Powell, W. B., Gifford, T., Nienow, J., & Day, J. (2010). Approximate dynamic programming captures fleet operations for Schneider National. *Interfaces, 40*(5), 342–352.

Smyth, G. K., & Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *ASTIN Bulletin, 32*(1), 143–157.

Tazelaar, F., & Snijders, C. (2013). Operational risk assessments by supply chain professionals: Process and performance. *Journal of Operations Management, 31*(1–2), 37–51.

Toyasaki, F., Arikan, E., Falagara Sigala, I., & Silbermayr, L. (2017). Disaster relief inventory management: Horizontal cooperation between humanitarian organizations. *Production and Operations Management, 6*(26), 1221–1237.

Ukkusuri, S. V., & Yushimito, W. F. (2008). Location routing approach for the humanitarian prepositioning problem. *Transportation Research Record, 2089*(1), 18–25.

Vanajakumari, M., Kumar, S., & Gupta, S. (2016). An integrated logistics model for predictable disasters. *Production and Operations Management, 25*(5), 791–811.

Zhang, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing, 23*(6), 743–757.