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

Liyi Gu Ilya O. Ryzhov Mahyar Eftekhar

August 30, 2019

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 realistic 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.

1 Introduction

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 (Vanajakumari et al., 2016; Ni et al., 2018), facility location (Balcik & Beamon, 2008; Charles et al., 2016), prepositioning strategies (Acimovic & Goentzel, 2016; Salmeron & Apte, 2010; Rawls & Turnquist, 2012), inventory pooling mechanisms and field coordination (Toyasaki et al., 2017; Ergun et al., 2014), and aid distribution and last mile delivery (Balcik et al., 2008; Jahre et al., 2016). However, implementing the results of these analytical approaches in the field has proven challenging, perhaps because the assumptions used in analytical studies often do not reflect the full complexity of the humanitarian context (see, e.g., Gralla et al., 2014 for discussion along these lines). Eftekhar & Van Wassenhove (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. 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 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.

From this, one can see that humanitarian fleet management is subject to very different challenges than commercial fleet management. Field managers at 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. 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 (e.g., those in Ben-Tal et al., 2011 or Hamedi et al., 2012). At the same time, field managers possess a great deal of expertise and intuition (Kahneman & Klein, 2009) which often leads to successful completion of complex tasks, particularly under time pressure (Hayashi, 2001) and in unstable environments (Khatri & Ng, 2000). For these reasons, we do not study optimization algorithms in this paper; instead, we develop a simulation-based approach to counterfactual analysis in humanitarian fleet management that can explain why field managers make certain types of decisions in various settings of interest.

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. Specifically, we consider last mile delivery of humanitarian services from local offices to affected communities (Balcik et al., 2008). The economic performance of a given set of policies (e.g., disposition policies) is subject to high uncertainty due to unpredictable demand, budget constraints, infrastructure problems and other issues (Eftekhar et al., 2014; McCoy & Lee, 2014), and depends on complex interactions between the fleet composition, stochastic and non-stationary demand, and the depreciation of trucks as they are exploited. Existing analytical and empirical research examines various elements of this problem in isolation, such as fleet sizing (Kunz & Van Wassenhove, 2019), vehicle maintenance and replacement (McCoy, 2012; McCoy & Lee, 2014; Pedraza-Martinez & Van Wassenhove, 2013), and field vehicle fleet management (Pedraza-Martinez et al., 2011; 2

---

1We 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.
Eftekhar & Van Wassenhove, 2016); however, the dynamic interactions between these factors are too complex to allow any analytical (or even numerical) treatment of optimal policies. An alternative approach is to use simulation to evaluate and compare various policies in a realistic setting. Simulation has been used in this way to study policies for many applications of interest to the public sector, such as HIV/AIDS prevention (Rauner, 2002), kidney allocation (Zenios et al., 1999), and emergency response (Kaplan et al., 2002).

In humanitarian fleet management, vehicles and demands possess heterogeneous attributes. For example, missions can be categorized by type and location; different types of missions arrive at different rates and impose different loads (kilometers to be traveled) on the fleet. Vehicles are acquired at different times, and their individual ages and odometers not only affect their performance, but also change dynamically over time. Managers’ decisions are influenced by the attributes of the fleet (for example, they may prefer to place higher loads on newer or older vehicles). Our ability to model these attributes for each individual vehicle, track their evolution over time, use them to calculate decisions, and evaluate the impact of these decisions on costs incurred, constitutes the main distinguishing feature of our work as compared to the existing applied humanitarian logistics literature. It also allows us to investigate many dimensions of fleet management within the same simulation environment.

We calibrated and validated the simulator using data provided by a large international humanitarian organization (LIHO). Using detailed data on purchase and maintenance costs, odometers over time, and salvage values, for multi-attribute vehicles in different countries, we design 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. We evaluate various candidate policies for the following problems:

1. Disposition. We compare LIHO’s recommended 5 year/150,000 km disposition policy with
other combinations of age/odometer thresholds in realistic demand scenarios.

2. Mission switching. 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. We counterfactually evaluated the possible benefits of switching types dynamically.

3. Centralized procurement. Managers are often able to choose between ordering a new vehicle through headquarters and waiting six months for it to join the fleet, or buying the vehicle locally with a 50% markup. We compared both schemes under realistic demand. In some cases, our results agree with previous work: for instance, we find that procuring vehicles through headquarters ultimately reduces costs, as does Keshvari Fard et al. (2019). In other cases, our results are more surprising. For example, Pedraza-Martinez & Van Wassenhove (2013) studied vehicle disposition (using an analytical model) in a similar context, and recommended replacing vehicles much earlier than the HO’s recommendation; this conclusion relied on various simplifying assumptions, one of them being that the monthly mileage\(^2\) of every vehicle is equal to the same constant. Our simulation results consistently indicated the opposite, namely 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 observe this very consistently in many scenarios, even when our model is completely recalibrated with data from a different developing country on another continent. This observation is much closer to what actually happens in the field.

The mission switching problem presents another surprise. Many researchers have observed that coordination is challenging and often lacking in the humanitarian setting (McClintock, 2009; Balcik et al., 2010). The general consensus is that more coordination would be desirable; thus, Pedraza-Martinez et al. (2011) argues that improved coordination would reduce the unpredictability of demand on fleets, while Bhattacharya et al. (2014) finds that asset transfer between programs in an HO leads to more efficient operations. Pedraza-Martinez et al. (2013) investigates the design of mechanisms to encourage coordination between headquarters and local offices for more efficient procurement. However, we found that the economic benefits of switching 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

\(^{2}\text{In fact, distance is measured in kilometers in the LIHO data; we use the word “mileage” informally in this paper to mean “distance travelled,” but all the numbers are actually in km.}\)
believe that our results explain why field managers do not seem to view it as an issue of primary
importance: under current operating conditions (which may be subject to time constraints and
suboptimal cost allocations; see Dolinskaya et al., 2011), the benefits may not be worth the effort.

In both cases where our results disagree with existing literature, they nonetheless agree with
current practice. In other words, a simulation environment may not produce an “optimal” policy,
but through a greater degree of modeling complexity it helps to show how certain choices made by
field managers are rational given the circumstances. In practice, their decisions do not seem to be
based 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.
Similarly, the lack of improvement from mission switching in our simulation results helps to explain
why, in the field, switching does not seem to be widely practiced or even encouraged. Overall, our
results suggest that these intuitive decisions should perhaps be regarded more carefully than has
heretofore been the case in the literature.

In this way, our work shows the power of simulation as a tool for counterfactual analysis in
humanitarian fleet management. Previous applications of simulation to humanitarian operations
include agent-based models (Crooks & Wise, 2013; Altay & Pal, 2014) 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 evalu-
ate 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 simulat-
ing 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; Iakovou et al., 2014; Chacko et al., 2016). Such settings are not ade-
quate 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,
enabling a detailed counterfactual evaluation of a broad variety of simple, but practical policies in
a realistic humanitarian context.
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

Figure 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. First, we have to describe the vehicles in the fleet, by which we mean, not only the fleet size, but the specific attributes of each individual vehicle. In the LIHO dataset, these are the age of the vehicle (in months) at the start of the time period to be simulated; the odometer 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 to which the vehicle is assigned; the vehicle’s location at the start of the planning period; the vehicle’s accident history (e.g., number of accidents) and the residual value of the vehicle, which is bounded above by its original purchase price. It is entirely possible for every vehicle to have a unique
combination of attributes (that is, no two vehicles in the fleet are exactly alike).

2. Policies. The decision rules to be evaluated may pertain to the purchase of new vehicles, the sales or disposition of aging vehicles, or the assignment of vehicles to missions. LIHO’s recommended 5 year/150,000 km rule is one example of a sales policy; additional policies can be created and compared by considering different values for these thresholds. An example of a simple purchase policy is to immediately replace any vehicle that is sold, thus keeping the fleet size constant. An example of a simple assignment policy is to prioritize new vehicles for incoming missions (as is often done in practice; see Pedraza-Martinez & Van Wassenhove, 2013). Policies should be sufficiently detailed to enable the simulator to automate all purchase, sales and assignment decisions based on the composition of the fleet at any given moment.

With this starting configuration, the simulator then begins to generate 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 km). Using the historical data, we can estimate the probability distributions of demands with different attributes, to give ourselves the flexibility of generating demands that resemble the historical data in a general sense, but may deviate from the precise historical values that happened to have been observed.

After generating the demands, the simulator automatically 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 comprise the other key module of the simulator, and also require extensive customization based on the 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.

Once the required number of decision epochs has been simulated, the environment returns the
total costs that were incurred by following the pre-specified set of policies over the course of the simulation run. Costs are determined by adding the total fuel and maintenance costs that were computed by the cost module, plus the costs of any new vehicles that were 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, running the simulator multiple times with the same starting conditions will produce different cost values; one can then perform many runs and apply standard output analysis techniques to estimate the mean performance of the given set of policies. In addition to cost, the simulator also returns the average completion rates for all mission types, which may be useful as a second objective for policymakers since minimizing cost is not the only goal in humanitarian contexts.

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, 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:

\[
a = \begin{pmatrix}
    a_1 \\
    a_2 \\
    a_3 \\
    a_4 \\
    a_5 \\
    a_6 \\
    a_7
\end{pmatrix} = \begin{pmatrix}
    \text{Age} \\
    \text{Odometer} \\
    \text{Location} \\
    \text{Model type} \\
    \text{Mission type} \\
    \text{Accident history} \\
    \text{Residual value}
\end{pmatrix}
\]  

Let \( \mathcal{A} \) denote the set of all possible attribute vectors. Let \( t = 0, 1, ..., T \) be a time index representing the \( t \)th “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.\(^3\) One may think of \( R_t \) as a vector that has very

\(^3\)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.
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
$\hat{R}_t = (\hat{R}_{ta})_{a \in A}$ 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 $B$ 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 backlogs into the model. As before, we let $B_t = (B_{tb})_{b \in B}$ 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 $k$th 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 \( D^D \) as the set that includes all possible \( d \), and let \( D^M \) 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 \( D = D^D \cup D^M \).

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

\[
\sum_{d \in D} x_{tad} = R_{ta}, \quad \forall a \in A, \quad (2)
\]
\[
\sum_{a \in A} \sum_{d \in D^D} x_{tad} \leq B_{q(k)}, \quad \forall k = 1, \ldots, N, \quad (3)
\]
\[
x_{tad} \geq 0, \quad a \in A, d \in D, \quad (4)
\]

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 \( X_t \) 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 \( X_t^\pi \) is a mapping on the state space into \( X_t \), with the superscript \( \pi \) 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 \( d = \{0, 0, \ldots, 0\} \) (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

\[
R_{ta}^{\pi} = \sum_{a \in A} \sum_{d \in D} \delta_{ad}^\pi (a,d)x_{tad},
\]

and the resource transition from \( t \) to \( t + 1 \) is given by \( R_{t+1} = R_t + \hat{R}_{t+1} \) 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 \( \pi \). 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

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

and the performance of the policy is calculated as

\[
V_\pi = \mathbb{E}\left[ \sum_{t=0}^{T} C_t(S_t, X_t^\pi(S_t)) \right],
\]

(5)

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

\[
\bar{V}_\pi = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=0}^{T} C_t(S_t^m, X_t^\pi(S_t^m)),
\]

(6)

where \( (S_t^m)^T_{t=0} \) is the sequence of states visited in the \( m \)th 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 modules

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. Our presentation here focuses on data from a single developing country, but the analysis is repeated on a second country in the Appendix.

3.1 Description of LIHO data

The LIHO data contains aggregate information about 3846 vehicles in 21 different countries between 2004 and 2015. 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. However, 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 21-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. The attributes used in our study include:

- Unique vehicle ID and local office where the vehicle was stationed;
- Vehicle model type, mission type given to the vehicle;
- Date registered, date on which odometer is recorded, final odometer;
- Purchase value, sales value (if sold) or booking value (if not sold);
- 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.

In order to estimate the costs coherently for vehicles of different attributes, 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 from the cross-sectional dataset, 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 3000, 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 provides 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.4

The goal of the demand module (see Figure 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

---

4We 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.
trends), but they should not be identical to the historical data 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 first construct a stochastic process $\left(L_t\right)_{t=0}^T$ 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. This is typical in humanitarian logistics; however, the LIHO dataset provides no information about any such projects, so the process $\left(L_t\right)$ is a modeling construct rather than an empirically observed quantity. One could also think about $\left(L_t\right)$ 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 $\left(L_t\right)$ is modeled as a $G/G/\infty$ queue with batch arrivals. Each new batch represents a new active “project,” for which we should have a certain number of vehicles (represented by the batch size) 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 demand. The number of “servers” is infinite because we do not impose any 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

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

where $N_{\ell}$ is a positive integer representing the number of distinct tasks generated by project $\ell$ in time period $t$, the values $X_n$ for $n = 1, ..., N_{\ell}$ represent the mileages required to complete each individual task, and $Y_{\ell}$ then represents the total mileage generated by project $\ell$. We would like the flexibility to model demand at the level of individual tasks because this may affect the performance of various assignment policies (for example, consider a simple policy that always assigns tasks in order of increasing mileage, with the quickest tasks being completed first).

Since we are using the historical fleet size as a stand-in for $L_t$, we can treat the monthly mileage of each vehicle as an observation of $Y_{\ell}$. Although the data do not tell us anything explicit about $N_{\ell}$ or $X_n$, we can estimate distributions for these random variables by interpreting $Y_{\ell}$ as a compound Poisson random variable, that is, $N_{\ell} \sim \text{Poisson} (\lambda)$ and $X_n \sim \text{Gamma} (\alpha, \gamma)$. Under a particular transformation of $(\lambda, \alpha, \gamma)$, this becomes the well-known Tweedie distribution (Zhang, 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 historical fleet size.\footnote{Eftekhar et al. (2014) points out that such a dependence may constitute evidence that the demand is censored; unfortunately, as stated earlier, we are not able to see which particular observations are censored. It may also be the case that, if there is a large number of projects, the managers may be able to combine some individual tasks together into a single “trip,” leading to a slightly smaller number of tasks.}

Figure 2 provides an illustration of how the output of the demand model is “realistic” without being identical to the real data. First, Figure 2(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, Figure 2(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 $Y_t$ 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

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

where Dep\% is the depreciation expressed as a percentage of the original purchase price, MissionType$_j$ is a dummy variable that is equal to 1 if the vehicle is assigned to the $j$th 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 stages of a vehicle’s life than in the later stages (an additional 100 km traveled should have
Independent variable:

<table>
<thead>
<tr>
<th>Depreciation%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>log(Odometer)$\times$Mission1</td>
</tr>
<tr>
<td>log(Odometer)$\times$Mission2</td>
</tr>
<tr>
<td>NumAcc</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Table 1: Estimation results for the depreciation model (8).

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 21-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 1 shows the estimation results only for those terms relevant to FC (with two mission types, a single vehicle type, and a single location). The coefficients for all the regressors are positive and significant; note the heterogeneity between mission types.
Dependent variable: Cost per Km

<table>
<thead>
<tr>
<th></th>
<th>Cost per Km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odometer</td>
<td>$1.632 \times 10^{-7}$ ***</td>
</tr>
<tr>
<td>Age</td>
<td>$2.506 \times 10^{-4}$ ***</td>
</tr>
<tr>
<td>Mission2</td>
<td>0.009759**</td>
</tr>
<tr>
<td>Constant</td>
<td>0.1292***</td>
</tr>
</tbody>
</table>

Observations 11,040

R² 0.301

Adjusted R² 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

Table 2: Estimation results for the fuel cost/km model (9).

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

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

(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 2 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 $1.632 \times 10^{-7}$ to this cost, each extra year on the age adds $2.506 \times 10^{-4}$, and if the vehicle operates on mission type 2, 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 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, 15000 and 50000 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

\[ CC = \sum_j \beta_{1j} \text{Age} \times \text{MissionType}_j + \beta_{2j} \text{Odometer} \times \text{MissionType}_j + \beta_{3j} \text{Odometer}^2 \times \text{MissionType}_j, \]

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 3. 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, for the purpose of sensitivity analysis and also because there may have been accidents that were not recorded in the data. 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 Results and insights

We present a case study in which our simulator compared a number of purchase, assignment, and sales policies using the models in Section 3 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 two mission types.

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 compares centralized and decentralized procurement. Section 4.5 investigates the impact of allowing vehicles to change their assigned mission type (a practice not currently implemented by LIHO).
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 Section 4.4, 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.** These rules assign an incoming mission to the vehicle with the least/most mileage on the odometer.

- **Oldest/Newest.** These rules assign an incoming mission 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 across all vehicles. 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.

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) 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 composition 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 fix each successive sales policy and run it together with every possible assignment policy. The smallest total cost across all assignment policies, as estimated by (6), is then reported as the performance of that particular sales policy. We found that, for any fixed demand level, there was very little variation in completion rates between policies; thus, demand levels 1-4 resulted in completion rates close to 1, and demand level 5 resulted in a completion rate of around 77%. For this reason, we use total cost as the primary performance metric in this case.

Figure 3 shows the results of the comparison with accidents not included (i.e., the accident generation probability was set to zero). The performance differentials between any two sales policies are all statistically significant as a result of running sufficiently many simulation runs. 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. Two observations can be made from these results:

1. The optimal sales threshold is always later/higher than the one recommended by LIHO;

2. However, the optimal threshold moves earlier/lower as the demand increases.

The first observation appears to dovetail 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
<table>
<thead>
<tr>
<th>Accident Level</th>
<th>Demand Level</th>
<th>Sales</th>
<th>Assignment</th>
<th>Total cost</th>
<th>Avg. no. of accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>RIG</td>
<td>Myopic</td>
<td>151519</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>140/300</td>
<td>Myopic</td>
<td>286123</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>140/300</td>
<td>Least Odometer</td>
<td>460906</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100/225</td>
<td>Newest</td>
<td>721593</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>100/225</td>
<td>Balance</td>
<td>854286</td>
<td>0</td>
</tr>
<tr>
<td>Baseline</td>
<td>1</td>
<td>RIG</td>
<td>Myopic</td>
<td>162034</td>
<td>1.433</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>140/300</td>
<td>Myopic</td>
<td>289548</td>
<td>2.672</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>140/300</td>
<td>Least Odometer</td>
<td>466112</td>
<td>3.979</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100/225</td>
<td>Newest</td>
<td>728315</td>
<td>4.967</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>100/225</td>
<td>Balance</td>
<td>862691</td>
<td>6.202</td>
</tr>
<tr>
<td>2x Baseline</td>
<td>1</td>
<td>140/300</td>
<td>Myopic</td>
<td>163841</td>
<td>2.182</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>140/300</td>
<td>Myopic</td>
<td>292427</td>
<td>5.437</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>140/300</td>
<td>Least Odometer</td>
<td>470577</td>
<td>7.908</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100/225</td>
<td>Newest</td>
<td>734835</td>
<td>9.804</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>100/225</td>
<td>Balance</td>
<td>871176</td>
<td>12.410</td>
</tr>
<tr>
<td>4x Baseline</td>
<td>1</td>
<td>RIG</td>
<td>Myopic</td>
<td>166533</td>
<td>4.813</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>140/300</td>
<td>Myopic</td>
<td>298577</td>
<td>10.699</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>140/300</td>
<td>Least Odometer</td>
<td>479387</td>
<td>15.088</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100/225</td>
<td>Newest</td>
<td>747526</td>
<td>19.514</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>100/225</td>
<td>Balance</td>
<td>889165</td>
<td>24.729</td>
</tr>
</tbody>
</table>

Table 4: Cost comparison of sales/assignment policies with varying levels of accidents.

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.\(^6\)

\(^6\)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 performance was obtained from high thresholds such as 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.
Next, we examine both sales and assignment policies and bring accidents into the picture. Table 4 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 Figure 3: RIG and/or 140/300 are optimal for low demand 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 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 a single local office. We focus on the “high” demand trajectory shown in Figure 2, which demonstrates the same rising trend as in the historical data, but represents a hypothetical situation where the data under-reported the actual demands.

7Under demand level 4, the Newest rule can be viewed as balancing the residual value of the fleet since vehicles with higher residual value tend to receive higher utilization. The Least Odometer rule also performs very similarly to Newest at this demand level.
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 \( \alpha \) (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 \( \beta \) (we label this policy “UTIL” for “utilization”). Thus, high \( \alpha \) and low \( \beta \) lead to more purchases, while low \( \alpha \) and high \( \beta \) lead to fewer.

Figure 4 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 \( \alpha \), suggesting that the manager could easily use \( \alpha \) as the “knob” for achieving a desired tradeoff between cost and completion rate. Including accidents in the model did not substantially change the shape or composition of the Pareto front in Figure 4(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 \( \beta \) 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, the COMP policy 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” when the load on the fleet appears to be growing, instead of reacting to a perceived shortage of vehicles; 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: centralized purchases with lead times

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. Purchasing locally can be 50% more expensive than purchasing through the HQ (Besiou et al., 2014), creating a dilemma for fleet managers.

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 a setting where the manager can instantly obtain
Table 5: Comparison of costs between purchasing locally and ordering predictively for stable demand.

<table>
<thead>
<tr>
<th>Demand level</th>
<th>Purchase locally (no lead time, 50% markup)</th>
<th>Order predictively (6-month lead time)</th>
<th>% Cost saved from predictive ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>189519</td>
<td>158339</td>
<td>16.5%</td>
</tr>
<tr>
<td>2</td>
<td>336855</td>
<td>285701</td>
<td>15.2%</td>
</tr>
<tr>
<td>3</td>
<td>520906</td>
<td>460403</td>
<td>11.6%</td>
</tr>
<tr>
<td>4</td>
<td>787328</td>
<td>720886</td>
<td>8.4%</td>
</tr>
<tr>
<td>5</td>
<td>978276</td>
<td>848291</td>
<td>13.3%</td>
</tr>
</tbody>
</table>

new vehicles with a 50% markup in cost.

Table 5 reports the results of the comparison; since completion rates were similar under both 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, with typical savings of around 10% relative to the decentralized scheme.

We also conducted a similar comparison for the realistic demands in Section 4.3. Here, in order for the two schemes to produce comparable completion rates, the decentralized setting uses the COMP policy ($\alpha = 0.8, 0.81, ..., 0.99$). 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

Figure 5: Comparison of centralized and decentralized purchases.
aims to preventively increase the fleet size, we use it as the purchase policy (\(\beta = 0.6, 0.61, ..., 0.99\)). The resulting Pareto fronts for both schemes are shown in Figure 5, 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%).

4.5 Case 4: 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.

For illustrative purposes, we first consider an artificial scenario that one might expect to be biased in favour of switching. 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 Figure 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 1 is 100%, the average completion rate of type 2 is below 100%, and the utilization for vehicles that currently operate on type 1 is under 70% (ensuring that we can afford to switch some of them), the oldest vehicle that operates on type 1 is switched to type 2.\(^8\) Similar criteria are used to switch from type 2 to type 1. If no switching occurs either way, a

\(^8\)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.
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.

Figures 6(b)-6(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 Figure 6. We can observe that, for the most part, switching seems to offer very little benefit. The most improvement that we see is in Figure 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 40000 (from 30000), 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 Figure 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). Figure 7(a) shows the trajectory of \((L_t)\) for each of the two mission types; we see that type 1 generates heavy demand in the first half of the planning period, then gradually ramps down, while type 2 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 1 to type 2. However, the Pareto fronts in Figure 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 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. Figure 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

![Graph showing realistic demand levels for two mission types.](image)

![Graph showing cost and completion rate comparison for switching vs. not switching.](image)

Figure 7: Comparison of switching vs. non-switching under historical demand.
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, 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 Figure 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. Figure 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 in humanitarian fleet management. The simulator uses several modules, which are calibrated using real-world data, to evaluate policies for the purchase, assignment, and disposition of vehicles, in terms of both cost and mission completion rate. This environment models vehicles with heterogeneous attributes, and simulates how these attributes change over time in response to decisions that are made. 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 cost benefits at best: it is indeed possible to use them more efficiently and thereby reduce the fleet size, but the cost savings will be largely nullified by the accompanying increase in operational costs.

4. If reasonably accurate forecasts of future demands are available, one can order new vehicles through headquarters and entirely compensate for the resulting long lead times. However, this imposes a greater burden of data collection and analysis on the field manager.

While these findings are valuable for LIHO and similar organizations, it is our view that the simulation-based approach presented in this paper is more broadly useful beyond the LIHO case. Using a single environment, we have investigated a broad variety of research questions in humanitarian fleet management, which previously had been studied independently using various mutually incompatible models. Our approach provides a level playing field for evaluating results obtained from such models together with practices observed in the field; in certain cases, we can identify the rationale behind “the facts on the ground,” which we believe can lead to a more productive dialogue between researchers and practitioners.
References


Table 6: Residual value estimation results for C2.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Depreciation Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>$2.215 \times 10^{-3} , ***$</td>
</tr>
<tr>
<td>NumAcc</td>
<td>$0.04512 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission1 × VehicleType4</td>
<td>$0.0355 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission2 × VehicleType1</td>
<td>$0.0021 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission2 × VehicleType2</td>
<td>$0.0405 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission2 × VehicleType3</td>
<td>$0.0232 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission3 × VehicleType4</td>
<td>$0.0333 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission4 × VehicleType1</td>
<td>$0.0303 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission4 × VehicleType2</td>
<td>$0.0520 , ***$</td>
</tr>
<tr>
<td>log (Odometer) × Mission4 × VehicleType3</td>
<td>$0.0489 , ***$</td>
</tr>
<tr>
<td>Observations</td>
<td>3846 (235 in C2)</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

6 Appendix: 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 4) in this discussion. However, Cases 1-3 can all be considered.

6.1 Estimation results for cost models

Since the depreciation model from Section (3.3) was estimated using data from 21 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

$$\text{Dep}\% = \beta_1 \text{Age} + \sum_{i,j} \beta_{2ij} \log (\text{Odometer}) \times \text{MissionType}_i \times \text{VehicleType}_j + \beta_3 \text{NumAcc},$$
and the results are summarized in Table 6. 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 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 7 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:

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

The results are shown in Table 8. 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.

### 6.2 Simulation results for C2, vehicle type 2

We repeated Cases 1-3 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 Figure 9. The overall pattern is similar to what we observed in the FC study, 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 9 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.
Figure 9: Cost comparison of sales policies under demand levels 1-5.

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 4.3 were implemented. The Pareto fronts for both policies are shown in Figure 10. For COMP,

<table>
<thead>
<tr>
<th>Accident Level</th>
<th>Demand Level</th>
<th>Policies</th>
<th>Cost</th>
<th>NumOfAccidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>Myopic, RIG</td>
<td>118217</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Myopic, RIG</td>
<td>204220</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, RIG</td>
<td>307332</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>449596</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>550307</td>
<td>0</td>
</tr>
<tr>
<td>Baseline</td>
<td>1</td>
<td>Most Odometer, RIG</td>
<td>126206</td>
<td>1.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Myopic, RIG</td>
<td>211178</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, RIG</td>
<td>324512</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>459015</td>
<td>6.97</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>558606</td>
<td>6.13</td>
</tr>
<tr>
<td>2x Baseline</td>
<td>1</td>
<td>Oldest, RIG</td>
<td>131178</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Myopic, RIG</td>
<td>219344</td>
<td>7.62</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>332864</td>
<td>7.95</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>468697</td>
<td>14.11</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>567274</td>
<td>12.53</td>
</tr>
<tr>
<td>4x Baseline</td>
<td>1</td>
<td>Most Odometer, RIG</td>
<td>140256</td>
<td>6.98</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Myopic, 140/300</td>
<td>226891</td>
<td>11.32</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>342903</td>
<td>15.63</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Least Odometer, 140/300</td>
<td>488761</td>
<td>27.19</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>583912</td>
<td>24.77</td>
</tr>
</tbody>
</table>

Table 9: Comparison between different levels of accidents.
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.

Finally, we consider Case 3, which compares centralized procurement with a 6-month lead time vs. decentralized procurement with no lead time but a 50% markup in purchase cost. Table 10 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. 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. For this

<table>
<thead>
<tr>
<th>Demand level</th>
<th>Purchase locally (no lead time, 50% markup)</th>
<th>Order predictively (6-month lead time)</th>
<th>% Cost saved from predictive ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>133217</td>
<td>117517</td>
<td>11.8%</td>
</tr>
<tr>
<td>2</td>
<td>213554</td>
<td>203184</td>
<td>4.9%</td>
</tr>
<tr>
<td>3</td>
<td>322843</td>
<td>307425</td>
<td>4.8%</td>
</tr>
<tr>
<td>4</td>
<td>509596</td>
<td>448206</td>
<td>12.0%</td>
</tr>
<tr>
<td>5</td>
<td>625046</td>
<td>546630</td>
<td>12.5%</td>
</tr>
</tbody>
</table>

Table 10: Comparison of costs between purchasing locally and ordering predictively for stable demand.
reason, we make fewer purchases under decentralized procurement and so the difference between purchase prices is of less importance. The general observation that centralized procurement is preferable, given the ability to order anticipatively, is still valid.

Figure 11 compares the performance of centralized vs. decentralized procurement under realistic demand for C2. The same conclusions apply.

6.3 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 6.1 as they already considered both vehicle types. Figure 12 reports the results of Case 1; the conclusions are very similar to those in Section 6.2.

Next, Table 11 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.

In Case 2, 13 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
<table>
<thead>
<tr>
<th>Accident Level</th>
<th>Demand Level</th>
<th>Policies</th>
<th>Cost</th>
<th>NumOfAccidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>Myopic, RIG</td>
<td>126743</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Least Odometer, RIG</td>
<td>210614</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>337772</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>497793</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>594451</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand Level</th>
<th>Policies</th>
<th>Cost</th>
<th>NumOfAccidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1</td>
<td>Myopic, RIG</td>
<td>129126</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Least Odometer, RIG</td>
<td>216130</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>343170</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>507293</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>602801</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand Level</th>
<th>Policies</th>
<th>Cost</th>
<th>NumOfAccidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x Baseline</td>
<td>1</td>
<td>Myopic, RIG</td>
<td>131466</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Least Odometer, RIG</td>
<td>221666</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>348397</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Balance, 140/300</td>
<td>516776</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>611406</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand Level</th>
<th>Policies</th>
<th>Cost</th>
<th>NumOfAccidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x Baseline</td>
<td>1</td>
<td>Myopic, RIG</td>
<td>136836</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Least Odometer, RIG</td>
<td>233815</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Least Odometer, 140/300</td>
<td>359035</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Newest, 100/225</td>
<td>535641</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Balance, 100/225</td>
<td>627917</td>
</tr>
</tbody>
</table>

Table 11: Comparison between different levels of accidents.

Section 6.2 as well as with the main study of FC.

Table 12 reports the results of Case 3 for stable demand, while Figure 14 does the same for realistic demand. These results are quite close to those observed for FC since the same vehicle type
(a) COMP policy: \( \alpha = \{0.6, 0.61, ..., 0.99\} \).

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

Figure 13: Cost and completion rate comparison of reactive purchase policies in C2.

Figure 14: Comparison of centralized and decentralized purchases.

is considered in both cases.

<table>
<thead>
<tr>
<th>Demand level</th>
<th>Purchase locally (no lead time, 50% markup)</th>
<th>Order predictively (6-month lead time)</th>
<th>% Cost saved from predictive ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>154220</td>
<td>126743</td>
<td>17.8%</td>
</tr>
<tr>
<td>2</td>
<td>246337</td>
<td>210614</td>
<td>14.5%</td>
</tr>
<tr>
<td>3</td>
<td>375032</td>
<td>337335</td>
<td>10.1%</td>
</tr>
<tr>
<td>4</td>
<td>552154</td>
<td>496243</td>
<td>10.1%</td>
</tr>
<tr>
<td>5</td>
<td>665147</td>
<td>590504</td>
<td>11.2%</td>
</tr>
</tbody>
</table>

Table 12: Comparison of costs between purchasing locally and ordering predictively for stable demand.
6.4 Discussion

Overall, the results in Sections 6.2-6.3 are similar both to each other and to our earlier results for FC. When vehicle type 2 is used (Section 6.2, 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).

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.