Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

AGILE LOGISTICS SIMULATION AND OPTIMIZATION FOR MANAGING DISASTER RESPONSES

Francisco Barahona
Markus Ettl
Marek Petrik
Peter M. Rimshnick

IBM T.J. Watson Research Center
Yorktown, NY 10520, USA

ABSTRACT

Catastrophic events such as hurricanes, earthquakes or floods require emergency responders to rapidly distribute emergency relief supplies to protect the health and lives of victims. In this paper we develop a simulation and optimization framework for managing the logistics of distributing relief supplies in a multi-tier supply network. The simulation model captures optimized stocking of relief supplies, distribution operations at federal or state-operated staging facilities, demand uncertainty, and the dynamic progression of disaster response operations. We apply robust optimization techniques to develop optimized stocking policies and dispatch of relief supplies between staging facilities and points of distribution. The simulation framework accommodates a wide range of disaster scenarios and stressors, and helps assess the efficacy of response plans and policies for better disaster response.

1 INTRODUCTION

Disaster response requires that emergency supplies are distributed rapidly and widely across the affected area. Many national and regional agencies operate supply chains designed to respond to a large variety of disasters. These supply chains differ significantly from commercial supply chains. First, the delivery of goods is not driven by optimizing profits but instead by fulfilling humanitarian needs. Second, while commercial supply chains can be designed and refined to slowly changing customer demand, the supply chains in disaster response must be set up rapidly with little advance warning and often with little precise information. Because disasters usually result in improvements in infrastructure and response plans, the next disaster typically differs significantly from historical patterns. As a result, there is generally insufficient historical data to construct faithful models of damage and the demand for emergency supplies.

In this paper, we consider the problem of optimizing and simulating inventory levels and deliveries in a supply chain during a disaster response. These are typical logistical challenges faced by government emergency management agencies, such as the Federal Emergency Management Agency (FEMA) in the US or Emergency Management Australia (EMA). Emergency response agencies often rely on supplies that are stored long-term in a small number of large warehouses; we refer to them as *Distribution Centers* (DC). During a response, the response agencies must transport numerous emergency supplies through a number of intermediate processing points to *Points of Distribution* (PODs) where they are distributed to the affected population. We refer to the intermediate processing nodes as *Staging Areas* (SAs). Once established during the disaster response, both PODs and SAs can hold inventories to respond to shifting demand needs. Figure 1 shows an example of a supply chain structure.

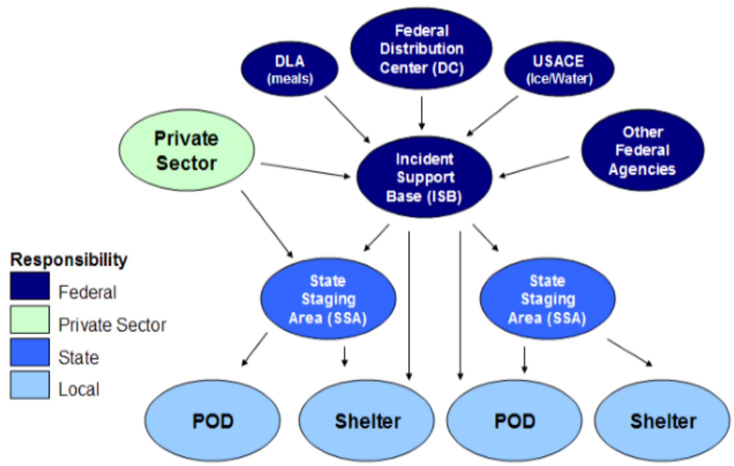

Figure 1: Example of a supply chain of a disaster response agency.

We consider a scenario in which a single commodity is distributed, e.g. bottled water; our models, however, can be extended easily to handle multi-commodity supply chains. Other common commodities that are distributed are, for example, meals ready-to-eat, blankets, temporary housing, or medical supplies.

We focus on the supply chain planning during the initial phase of the response. During this phase, supplies are available only in DCs before the disaster happens. Because there are only a few DCs, it is essential to initiate the transportation of supplies as early as possible. Often, truck services must be procured and the shipments initiated long before the precise extent of the damage—and therefore the demand—is known with any certainty. The challenge from the optimization perspective is that the supplies must be efficiently pre-positioned during the initial stages of the response in SAs in the affected region in order to best respond to unknown demand given unknown damage to the transportation network. In addition, the supplies must be delivered more or less uniformly to the affected population and the optimization models must scale to large areas with thousands of distribution points.

One of the main innovations of our work is the *robust* model of demand uncertainty (Ben-Tal, Ghaoui, and Nemirovski 2009). Robust optimization is an alternative model of uncertainty to the more traditional stochastic optimization (Zipkin 2000, Porteus 2002, Ben-Tal, Ghaoui, and Nemirovski 2009). It computes solutions that are guaranteed to work well in all *plausible* realization of the uncertainty instead of computing solutions that work well in expectation. Robust solution can also be seen as an immunization against the effects of the uncertainty.

Compared with stochastic models, the robust model of uncertainty is simpler in terms of both computational and data complexities (Barbarosoglu and Arda 2004). In particular, robust models are typically easier to solve than stochastic models of uncertainty because their focus is on a single worst case opposed to the average over stochastic samples. Robust optimization problems are also easier to specify based on incomplete data. The uncertainty sets only define plausible scenarios and do not require distributions over realizations. Finally, robust models do not suffer from some of the out of sample extension problems associated with some of the sample-based stochastic optimization algorithms such as approximate dynamic programming (Powell 2007, Ben-Tal, Ghaoui, and Nemirovski 2009).

Recently, there has been considerable effort aimed to address optimization problems related to all stages of disaster mitigation, response, and recovery (Altay and Green 2006). We focus on the immediate and short-term response to the disaster and do not address the recovery phase or disaster mitigation. In comparison with previous work, we describe a comprehensive model for immediate disaster response which includes inventory management, transportation, and simulation. The inventory management is based on a network flow model similar to Beamon and Kotleba (2006) but with a more detailed treatment of demand uncertainties. Because the inventory model abstracts away from the details of the transportation, we are able to more faithfully model and optimize demand uncertainties during the response. This model is described in detail in Section 2.

The truck schedules are then computed using a transportation optimization component which models actual truck routes. The truck schedules are subject to realistic constraints on driving time, road damage, and traffic. The truck deliveries are computed to match the inventory model deliveries as closely as possible. Similar logistic models have been used previously but the main difference is that the goal of our model is to match the pre-computed inventory flows (Haghani and Oh 1996, Sheu 2007). This model is described in Section 3. The results of the logistic model are then used in conjunction with a simulator that evaluates the realistic performance of the model. The simulator is described in Section 4 and the results of simulations on a large-scale earthquake response scenario are summarized in Section 5.

2 INVENTORY MODEL

In this section, we describe the inventory model for the disaster response in greater detail. We are assuming a single commodity and the general unit measure is a single truck-load. The inventory is held in multiple nodes with several levels. The shipping model is simplified to a network flow model with a discretized time interval and fractional truck load shipments.

Network flow models of supply chains have been studied extensively in the literature. In comparison with existing work, we focus on a tractable representations of demand uncertainties. There are many uncertainties involved in responding to a disaster: the extent of the demands, travel times, damage to stock nodes, transportation times and capacities. To simplify our model, we focus only on demand uncertainties; We briefly discuss extensions to many other types of uncertainties. Unlike traditional approaches, we use robust optimization models which lead to more tractable models, simpler data requirements, and more robust solutions. While solutions to robust models can be expected to perform worse on average, the worst-case performance may be more appropriate in disaster response (Ben-Tal, Ghaoui, and Nemirovski 2009).

Robust optimization models have been recently proposed as an alternative to stochastic optimization models. Typical stochastic optimization models treat uncertainty as random variables and are formulated as: $\min_{y\in\mathscr{Y}} \mathbf{E}[f(X,y)]$, for a decision variable y, random variable X, and an objective function f. Robust optimization, on the other hand solves a worst case problem: $\min_{y\in\mathscr{Y}} \max_{x\in\mathscr{X}} f(x,y)$, for a plausible set of realizations $\mathscr{X}$. While at first sight such formulations may seem too conservative, they have been shown to work well in many domains when the plausible sets are chosen correctly. The plausible set $\mathscr{X}$ can be either constructed directly from data or from stochastic distributions.

Robust optimization models of inventory management with uncertain demands have been studied previously (Bertsimas and Thiele 2006). This work showed that many properties of stochastic inventory control also hold in robust models. However, the model studied by Bertsimas and Thiele (2006) assumes several important simplifications that limit its applicability to our setting. It assumes that the uncertainty is restricted to be of a rectangular shape which we define formally below (Iyengar 2005). Rectangular uncertainty sets intuitively mean that the demand variations are independent across stock nodes and time. This essentially implies that the plans are computed with respect to the worst case in each node as opposed to a worst case over all nodes.

In addition to rectangular uncertainty sets, Bertsimas and Thiele (2006) consider only static policies. *Static policies*, also known as open-loop control, are easy to compute but do not adjust to new information about demand uncertainty. *Fully adjustable policies*, also known as closed loop control, can on the other hand adjust as new information is update. Adjustable policies are much harder to compute than static ones. *Affinely adjustable policies* or controllers represent a compromise between static and fully adjustable ones (Ben-Tal, Goryashko, Guslitzer, and Nemirovski 2004, See and Sim 2010). Affine controllers represent a compromise between static policies and fully adjustable policies. Truncated affine controllers represent a piecewise extension of affine controllers. These approaches, though, require that the uncertainty is rectangular to make their computation tractable.

Our model lifts the restriction to rectangular uncertainty sets and can be easily adapted to affinely adjustable policies. Non-rectangular uncertainty sets, in general, lead to models that are NP-hard to solve (Ben-Tal, Ghaoui, and Nemirovski 2009). To achieve tractability, we make low-dimensionality assumptions on the uncertainty sets in the form of a factored representation akin to See and Sim (2010). That is, the uncertainty set is defined as a combination of a small number of "extreme" scenarios that introduce limited correlations between plausible demands. Because of the unique needs of disaster response model, we develop new objective functions that measure the fairness of the coverage over multiple stock nodes. These measures enable us to quantitatively evaluate the quality of the solutions in contrast with some previous heuristic approaches (Oh and Haghani 1997).

2.1 Deterministic Formulation

We start by describing a deterministic formulation of the inventory problem. Later we build on this model to introduce the robust model of uncertainty. The planning horizon is discrete and finite with time steps $\mathscr{T} = \{0 \ldots T\}$. The supply chain consists of stock-nodes $\mathscr{W}$. There is initial inventory in each stock node, z_w. There is demand at a subset of the nodes $\mathscr{W}_d \subseteq \mathscr{W}$ at each time step $t \in \mathscr{T}$ and node $w \in \mathscr{W}_D$, $d_{t,w} \in \mathbb{R}_+$. The storage capacity in terms of truck-loads of each non-demand node $w \in \mathscr{W} \setminus \mathscr{W}_D$ is denoted as $s_w \in \mathbb{R}_+$. The set of transportation links between the nodes form a directed acyclic graph $\mathscr{L} \subseteq \mathscr{W} \times \mathscr{W}$. The transportation lead time for an edge $e \in \mathscr{L}$ in discrete time-steps is denoted as $l_e \in \mathbb{N}$. The loading/unloading throughput at each node $w \in \mathscr{W}$ in terms of truck-loads is $q_w \in \mathbb{R}_+$. The costs are defined as follows. The backlogging cost at node $w \in \mathscr{W}$ is denoted as $c_w \in \mathbb{R}_+$. We ignore the shipping costs in describing the formulation to reduce clutter. The shipping costs are simply added to the objective.

We use the following decision variables. For any link $e \in \mathscr{L}$, the inventory shipped along the link $e \in \mathscr{L}$ during time step $t \in \mathscr{T}$ is denoted as $f_{t,e} \in \mathbb{R}_+$. Alternately, we use $f_{t,v,w}$ for $e = (v,w)$; when there is no such edge the variable is by definition 0. The inventory level at stock node $w \in \mathscr{W}$ is $x_{t,w} \in \mathbb{R}_+$ and the backlog at the node is $b_{t,w} = [\sum_{u=0}^{t} d_{u,w} - x_{t,w}]_+$. Note that for terminal nodes, the value $x_{t,w}$ represents the total deliveries and not the current inventory. We use $[x]_+ = \max\{0, x\}$. The model makes the following general assumptions. Stock nodes are used to model all levels of the supply chain, including the points of demand, but backlogging is allowed only for the demand nodes. Transportation links are unidirectional and the transportation network is *acyclic*.

The network flow between the stock nodes can be readily formulated as a linear program. The constraints in the linear program are as follows for each node $w \in \mathscr{W}$ and time step $t \in \mathscr{T}$. The inventory dynamics of shipments and deliveries are:

$$x_{0,w} = z_w \ , \quad x_{t,w} \le s_w \ , \quad x_{t,w} = x_{t-1,w} + \sum_{v \in \mathscr{W}} \left(f_{t-l_{v,w},v,w} - f_{t,w,v} \right) \ , \quad \sum_{v \in \mathscr{W}} \left(f_{t-l_{v,w},v,w} - f_{t,w,v} \right) \le q_w \ .$$

The last constraint limits the flow into and out of a stock node by its throughput. The constraints above can be abstracted as linear constraints using matrices A_f and A_x and a vector b as $A_f f + A_x x \ge b$, Note that as formulated, the constraints do not include the demands. This is intentional to simplify the introduction of the uncertain demands in the robust formulation.

Assuming that the precise demands d are known, one can formulate the following *linear program*, introduced above, with the objective to maximize the coverage across the stock nodes

$$\min_{x,f} \sum_{\substack{t \in \mathscr{T} \\ w \in \mathscr{W}}} \frac{c_w \cdot [\sum_{u=0}^{t} d_{u,w} - x_{t,w}]_+}{\sum_{u=0}^{t} d_{u,w}} \qquad \text{s.t.} \qquad A_f f + A_x x \ge b \tag{2.1}$$

This optimization problem, however, suffers from several important shortcomings which we address in the remainder of the paper: 1) the formulation does not address demand or lead-time uncertainties, and

2) the coverage that is achieved may be very uneven—easily accessible demand nodes would have much higher coverage.

2.2 Measures of Coverage Quality

Unfortunately, it is often impossible to satisfy all demand in a disaster response. The goal in disaster response is, therefore, to achieve the greatest possible demand coverage given the available resources. This brings up the natural question what is the most fair distribution of the supplies across the region. Simply using the average coverage, or total demand satisfied, leads to very unequal coverage. Another approach would be to consider the worst-case backlog over the demand nodes but this leads to plans that are too conservative and achieve overall low coverage.

To address this issue in a principled way, we describe a new class of *fairness measures* used to evaluate the trade-off between deliveries to multiple nodes. These measures are used to combine the coverage levels or backlog from multiple stock nodes into a single number.

Definition 2.1. A function $\mu : \mathbb{R}_+^{\mathscr{W}} \to \mathbb{R}$ represents a backlog *fairness measure* when it satisfies the following conditions for any two backlog values instantiations $x, y \in \mathbb{R}_+^{\mathscr{W}}$:

1. *Normalization:* $\mu(\mathbf{0}) = 0$.
2. *Monotonicity:* If $x \geq y$, then $\mu(x) \geq \mu(y)$.
3. *Convexity:* $\mu(\alpha \cdot x + (1-\alpha) \cdot y) \leq \alpha \cdot \mu(x) + (1-\alpha) \cdot \mu(y)$ for $\alpha \in [0,1]$.
4. *Uniform indifference:* $\mu(x + c \cdot \mathbf{1}) = \mu(x) + c$ for any $c \in \mathbb{R}$.
5. *Positive homogeneity:* $\mu(c \cdot x) = c \cdot \mu(x)$ for any $c \geq 0$.

This class of fairness measures is inspired by *coherent risk measures* which are used in stochastic finance to model risk-averse decision makers (Follmer and Schied 2011). These functions are attractive both because they have both good properties for the models and are also computationally convenient. Of particular interest is the following alternative representation of fairness measures as a robust value with respect to a set of probability measures.

Proposition 2.2 (e.g. (Follmer and Schied 2011))**.** *For any fairness measure $\mu : \mathbb{R}_+^{\mathscr{W}} \to \mathbb{R}$ there exists a set of probability measures $\mathscr{Q} \subseteq \mathbb{R}_+^{\mathscr{W}}$ such that for any backlog $b \in \mathbb{R}_+^{\mathscr{W}}$:*

$$\mu(b) = \sup_{q \in \mathscr{Q}} \sum_{w \in \mathscr{W}} q_w \cdot b_w.$$

The representation in Proposition 2.2 makes it possible to integrate the measures with out robust optimization framework. We call fairness measures *polyhedral* when the set $\mathscr{Q}$ in Proposition 2.2 is a polytope. Some examples of polyhedral fairness measures are the following. The average backlogs over all demand nodes: $\mu(b) = \mathbf{E}[b] \frac{1}{|\mathscr{W}|} \sum_{w \in \mathscr{W}} b_w$ or the worst-case backlog over the nodes $\mu(b) = \max_{w \in \mathscr{W}} b_w$. Note that a convex combination of any two fairness measures remains a fairness measure. The actual measure that we use is a convex combination of average and AV@R (Follmer and Schied 2011):

$$\mu(b) = (1-\lambda) \cdot \mathbf{E}[b] + \lambda \cdot CV@R(b) = (1-\lambda) \cdot \mathbf{E}[b] + \lambda \min_{\theta \in \mathbb{R}} \left(-\theta + \frac{1}{\alpha} \mathbf{E}\left[[X + \theta]_+ \right] \right) ,$$

where λ and α are parameters that determine importance of uniform coverage.

The fairness measures, as defined above, can be readily extended to measure the coverage across time steps in addition to measure it across inventory nodes. Then, the deterministic optimization in Equation (2.1) using the fairness measures is as follows:

$$\min_{x,f} \quad \max_{q \in \mathscr{Q}} \sum_{\substack{t \in \mathscr{T} \\ w \in \mathscr{W}}} q_{t,w} \cdot \frac{c_w \cdot [\sum_{u=0}^{t} d_{u,w} - x_{t,w}]_+}{\sum_{u=0}^{t} d_{u,w}} \qquad \text{s.t.} \qquad A_f f + A_x x \geq b \tag{2.2}$$

Note that the formulation uses the representation from Proposition 2.2. It can be readily shown that this optimization problem remains a fractional linear program by taking the dual of the inner maximization.

2.3 Uncertain Demands

We are now ready to extend the framework for dynamic inventory optimization problems, which can be used to model supply chains in the disaster response. We will need some additional notation. The demand is forecasted at the beginning of the optimization. The forecast at each time step t for a node w is denoted as $\hat{d}_{t,w}$. The actual realization of the demand is $d_{t,w}$, and we use $\hat{d}_{t,w} = \hat{d}_{t,w} - d_{t,w}$ to denote the demand deviation from the forecast.

The robust optimization model assumes a set of plausible demands $\mathscr{D} \subseteq \mathbb{R}^{\mathscr{T}\times\mathscr{W}}$. We describe how these sets are constructed below. Each element of the set represents a possible realization of the demands across all nodes and all time steps. We assume that these precise demands are unknown initially and the precise demand $d_{t,w}$ is observed at time t. The static, or open loop, robust optimization version of Eq. (2.2) is as follows:

$$\min_{x,f} \quad \max_{d\in\mathscr{D}} \max_{q\in\mathscr{Q}} \sum_{\substack{t\in\mathscr{T}\\ w\in\mathscr{W}}} q_{t,w} \cdot \frac{c_w \cdot \left[\sum_{u=0}^{t} d_{u,w} - x_{t,w}\right]_+}{\sum_{u=0}^{t} \bar{d}_{u,w}} \qquad \text{s.t.} \qquad A_f f + A_x x \geq b \tag{2.3}$$

Notice that the optimization is over the worst case possible realization of the demand, but the normalization to compute the coverage is in terms of the forecasted demand. The motivation for using $\bar{d}$ in the normalization is to allow for a convex formulation.

2.4 Factored Representation of Uncertainty

Unfortunately, the problem Equation (2.3) is non-convex and may be NP-hard to compute in general (Bertsimas and Goyal 2011). In particular, the optimization over the worst-case demand realization d leads to non-convexities. This is because the function $[\cdot]_+$ is convex in d. Bertsimas and Thiele (2006) address this difficulty by relaxing the problem to rectangular uncertainties, which may lead to solutions that are too conservative in our setting. This approach is also known as a safe approximation (Ben-Tal, Ghaoui, and Nemirovski 2009). We take an alternative approach and use uncertainty sets that are non-rectangular but are simplified by having a small number of extreme points and small dimensionality.

We use a causal representation of the uncertainty. Even though the uncertainty has an effect on a large number of stock nodes, it is usually caused by a small number of causes which are typically easier to define than their impacts. These causes result in significant correlations between demand deviations across the individual nodes. Assume a given set of factors h^j_w for $j = 1\ldots m$ and $w \in \mathscr{W}$. That is, each h^j represents a single cause, such as the location of an earthquake's epicenter, and the corresponding set of demand nodes affected by this cause. It is plausible for each of the events to happen, but it is very unlikely for several of them to influence the demand simultaneously. The uncertainty can be represented:

$$\mathscr{D} = \left\{ \bar{d} + \sum_j h^j \cdot \xi_j \ : \ \sum_j \xi_j = 1,\ \xi \geq \mathbf{0} \right\}.$$

This assumes that the factors modify the demand with respect to the forecast. A more general formulation of this set is $\mathscr{D} = \bar{d} + \operatorname{conv}\left(\{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_n\}\right)$, with the deviations corresponding to $\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_n$. The representation of the uncertainty set in terms of its extreme points above has several advantages in our setting. First, it is often easier to describe most extreme plausible samples than to derive the appropriate linear inequalities. Second, we propose algorithms that have a polynomial running time in the number of extreme points. The approach for using the extreme points of the uncertainty set relies on the following well-known result.

Theorem 1 Let $\mathscr{X}$ be the feasible set defined by the constraints in Equation (2.3). Then, the following equality holds:

$$\min_{x,f\in\mathscr{X}} \max_{d\in\mathscr{D}} \max_{q\in\mathscr{Q}} \sum_{\substack{t\in\mathscr{T}\\ w\in\mathscr{W}}} q_{t,w}\cdot\frac{c_w\cdot[\sum_{u=0}^{t} d_{u,w}-x_{t,w}]_+}{\sum_{u=0}^{t}\bar{d}_{u,w}} = \min_{x,f\in\mathscr{X}} \max_{d\in\operatorname{ext}\mathscr{D}} \max_{q\in\mathscr{Q}} \sum_{\substack{t\in\mathscr{T}\\ w\in\mathscr{W}}} q_{t,w}\cdot\frac{c_w\cdot[\sum_{u=0}^{t} d_{u,w}-x_{t,w}]_+}{\sum_{u=0}^{t}\bar{d}_{u,w}}\;,$$

where ext represents the extreme points of the polytope.

The proposition follows readily from the convexity of the inner optimization problem in d. Therefore, the optimal solution will be in one of the extreme points; the restriction to these values therefore does not influence the optimal objective. Using the extreme points, the optimization can be formulated readily as a linear optimization problem by taking the dual of the inner optimization problem. The solution of this linear program is then used in the transportation model to compute actual solutions.

3 TRANSPORTATION MODEL

In this section, we describe a transportation model that builds on the inventory optimization model to compute truck routes and schedules. The objective of the transportation model is to use a hierarchical approach to solve the dynamic vehicle routing problem (DVRP) in multi-stage distribution networks with multiple sources and delivery constraints. The objective of DVRP is to maximize the demand fulfillment in delivery locations with fairness. This problem can be categorized as a pickup and delivery vehicle routing problem, where goods are transported between pickup and delivery locations. We schedule and assign multiple types of vehicles to pick up and deliver commodities among locations in different levels of distribution networks.

3.1 Problem Description

As with the inventory optimization model, the total time horizon for DVRP is discretized and denoted by T. We take four-level networks as an example to demonstrate the whole system of DVRP as follows. The top level locations, which correspond to DCs, have an available inventory for each time t. The second level locations have commodity requests as the demand, for each time t. Vehicles located at vehicle depots corresponding to the first level locations deliver available supplies from the first level locations to satisfy demands at the second level locations. The arrived commodities to a second level location at time period t and remaining inventory in this location are then available for further shipment from the location at time period t. The routing is similar for the third and the fourth level. The vehicles starting at a location in level k only travel between locations in level k and locations in level $k+1$. Each time they pick up one truck load at a location in level k and deliver the entire truck load to a location in level $k+1$ and return back to k. Therefore, a route for one vehicle consists of a sequence $(l_1,t_1),(l_2,t_2),\cdots,(l_{2p+1},t_{2p+1})$, where l_{2i+1} is a location in level k, $0\le i\le p$; l_{2j} is a location in level $k+1$, $1\le j\le 2p$; and t_i is the arrival time at location l_i. The demand satisfaction is evaluated at the bottom level on the network which corresponds to PODs.

Each location from the lower levels has a demand for each time t. These demands are produced by the delivery schedule based on the inventory model optimization. Demands at time t can be satisfied by delivered commodities at time t or earlier. Just as in Section 4 it is important that we consider the fairness of the distribution. Because the transportation scheduling problem is more intricate than the inventory management problem, we focus on a specific instance of a fairness measure, as defined in Definition 2.1:

$$\min \sum_{w;t\in\mathscr{T}} \left(\alpha\cdot[b_{tw}-d_{tw}/\beta]^+ + b_{tw}\right). \tag{3.1}$$

An example of a sensible values of the parameters are $\alpha=0.2$ and $\beta=2$. Here b_{wt} and d_{tw} represent the unsatisfied demand and total demand respectively at location w at time t. Objective Equation (3.1) provides a linear fairness judgment, if the unsatisfied demand is more than half of the demand, we add a penalty of $\alpha \cdot b_{tw}$; otherwise, it will be just the value b_{tw}. The purpose of this penalty is to avoid concentrating the unsatisfied demand in any particular location.

We describe two algorithms for DVRP, a greedy algorithm, and a column generation refinement. Each of these algorithms deals with two-level networks, if the network has k levels, we have to apply a two-level algorithm $k-1$ times.

3.2 Greedy Algorithm

Assume that we have a two-level network—if there are multiple levels, each pair of levels can be optimized independently. The greedy algorithm simply assigns trucks loads to trucks based on which is the earliest available one. At each iteration we send a truck from an upper-level location to a low-level location, and then back to an upper-level location. To choose which truck to send, we compute for each truck the shortest two-step trip, and then choose a truck that after this two-step trip would have the minimum total traveling time. At each iteration we update the inventory and demand of the chosen locations. The procedure iterates until all demand is assigned or all trucks have reached a time limit.

3.3 Column Generation

Here we develop a column generation approach that builds on the greedy algorithm to compute better routes at the expense of a higher computing time. We use the greedy algorithm as a subroutine that generates possible routings, then these routings are used in an integer program that chooses the best combination.

In order to generate routes, we modify the greedy algorithm as follows. At each iteration of the greedy algorithm, instead of choosing a truck with minimum total traveling time, we choose at random a truck among the k lowest traveling times. Here k is a parameter that takes the values 3 or 4. This randomization allows us to generate many different good routings that will be given to the column generation procedure.

We let N be the total number of vehicles which we can use in this two-level problem. $V(h)$ is the set of routes that is generated for vehicle h by the greedy algorithm. $\mathscr{U}$ is the index set of upper level locations. $\mathscr{L}$ is the index set of lower level locations. The inventory arriving at location w at period t is ℓ_{it} for $w \in \mathscr{U}$. The demand at location w at period t is denoted by d_{tw} for $w \in \mathscr{L}$.

The variables of the integer program are as follows. For each route r we have a binary variable x_r. It takes the value 1 if the route is chosen, and 0 otherwise. The variable o_{tw} represents the inventory at location w and period t that is left at this same location for period $t+1$. The variable y_{tw} is the inventory at location w and period t left at this location to satisfy demand at later periods. We use b^1_{tw} and b^2_{tw} to represent the unsatisfied demand at location w at time t. The second variable b^2_{tw} has a higher cost, and $b^1_{tw} \leq d_{tw}/\beta$. This is to impose fairness in the objective. In each iteration we then solve the following mixed integer linear program:

$$
\begin{aligned}
\max_{x,y,b,o} \quad & \sum_{w\in\mathscr{L}}\sum_{t\in\mathscr{T}} b^1_{tw} + (1+\alpha)\cdot b^2_{tw} \\
\text{s.t.} \quad & \sum_{r\in V(h)} x_r = 1, \\
& \sum_{r:(i,t)\in r} x_r + o_{tw} = \ell_{tw} + o_{t-1,w}, \quad w\in\mathscr{U},\ t\in\mathscr{T} \\
& \sum_{r:(j,t)\in r} x_r + b^1_{tw} + b^2_{tw} + y_{t-1,w} = d_{tw} + y_{tw}, \quad w\in\mathscr{L},\ t\in\mathscr{T} \\
& b^1_{tw} \leq d_{tw}/2 \quad w\in\mathscr{L},\ t\in\mathscr{T},
\end{aligned}
$$

In addition, the variables are further constrained as $x_r \in \{0,1\}, y_{tw}, b^1_{tw}, b^2_{tw}, o_{tw} \geq 0$. The variables x_r that take the value one give the desired routing.

4 SIMULATION

In this section, we describe a simulation framework that we used to evaluate the inventory and optimization models. The simulation evaluates the optimized inventory and transportation schedules in a complex, life-like environment under various scenarios and assumptions. The platform used for the simulation was AnyLogic 6, a Java based simulation tool that supports discrete event, agent based, and system dynamics type models. The simulator models numerous types of stochasticity and is more granular than the optimization models. These simulation features can be broken down into three basic parts: the demand model, the supply model, and the consumption model. As we describe in Section 5, we estimate the simulation parameters from existing disaster studies.

4.1 Demand Model

The average demand μ_{tC} is modeled at the county c level as a fraction of the affected population. The actual demand is then a random variable with a triangular distribution: $\delta_{tc} \sim \text{Triangular}(\mu_{tc}(1-\alpha), \mu_{tc}, \mu_{tc}(1+\alpha))$ where α is an adjustable input parameter. We assume that μ_{tc} is quadratic function of the time.

To compute the demand at an individual POD, we consider the affected population close to the POD compared to other PODs nearby. When there is one POD in the county c, we assume that all of the demand is channeled to this POD. The demand (in person-units) at POD w at time t originating from county c is $d_{tw}^{c} = \delta_{tc} \cdot N_c / |\mathscr{W}_c|$ where $\mathscr{W}_c$ represent the set of PODs in the county, and N_c represents the total population of the county. In counties with no PODs, the contribution of the demand on PODs in other counties is distributed based on the distance $d_{tw}^{c} = \delta_{tc} \cdot N_c / \left(r_w^c \cdot \sum_{u \in \mathscr{W}_c} \frac{1}{r_u^c} \right)$, where r_w^c represents the distance from the center of county c to POD w. Finally, the total demand at POD w at time t given by $d_{tw} = \sum_{c \in Counties} d_{tw}^{c}$.

4.2 Supply Model

As described above, the distribution network consists of a hierarchy, with DCs at the top, going down to PODs at the bottom. Inventory is transported between nodes in the network via trucks. As such, trucks represent the key agents in the agent-based model of supply distribution.

Trucks attempt to proceed according to the schedule provided by the transportation optimizer. This schedule can be provided once at the beginning of the simulation, or multiple times throughout the simulation by having AnyLogic, via a custom Java API, call the optimizers at set intervals through the planning horizon. As would be the case in a real disaster situation, we model two main elements that may cause a truck to deviate from its initial schedule.

First, the schedules are conditional on *loading and unloading constraints*. Each site w in the network has a loading dock of size n_w. This dock is used both for loading and unloading. Second, the travel time is *stochastic*, generated according to travel times in ESRI ArcGIS . One of the features of ArcGIS is that various types of barriers can be added as layers to a given map. This allowed us to model granular road closures, such as bridges being out, as well as decreased transportation time over wide areas of a map. In the simulation model, travel times between nodes u and w was thus modeled as: $t_{uw} = e_{uw} + \rho$ where $\rho \sim \text{Triangular}(-c, 0, c)$ with e_{uw} representing the travel time given by ESRI, and c a parameter.

4.3 Consumption Model

The third component within the simulation defines how the commodity is consumed by victims at the PODs. In that sense, it forms the interface between the demand model and the supply model. We assume the commodity is consumed at the beginning of each period. Let $s_{tw}i$ represent the available supply at POD w at the end of period t , and c_{tw} the consumption. Formally, $c_{tw} = \min\left\{s_{t-1,w}, \hat{d}_{t,w}\right\}$ and $s_{tw} = s_{t-1,w} + a_{tw} - c_{tw}$ where a_{tw} is the amount of delivered commodity in period t and $\hat{d}_{tw} = d_{tw} + \gamma \cdot \left[\hat{d_{t-1,w}} - c_{t-1,w}\right]_+$. In other words, in addition to the demand coming from the demand model, we allow unsatisfied demand from the previous period to be rolled over into the new period, based on an input rollover parameter γ .

5 Experimental Results

In this section, we overview the result for an application of the proposed model to a disaster caused by a hypothetical earthquake in the New Madrid Seismic Zone of the southern and midwestern United States (Elnashai, Cleveland, Jefferson, and Harrald 2008). This scenario was developed by the Central United States Earthquake Consortium (CUSEC). Figure 2 depicts the fraction of the population affected, based on CUSEC models. The two rectangular inner regions show the areas in which the travel time is increased. In total, the earthquake was projected to affect 6 states, 141 counties (73 of which had some water infrastructure damage) and 2.7 million people. We evaluated our models using hypothetical but realistic projections of available response assets and resources. In all three scenarios, we estimated the following available assets: 8 regional DCs, 30 intermediate staging areas, 366 PODs, 900 top level trucks (DC to staging area), and unlimited lower level trucks (staging area to PODs).

First, we show the overall coverage achieved for three different scenarios in Figure 3. Since this disaster has not yet occurred, we do not have a baseline. The baseline, therefore, is based on the described inventory and transportation models with fixed location of the intermediate SAs. In the optimized asset deployment scenario, the inventory model is extended to a mixed integer model which can also optimize which set of SAs to open from a larger subset. Finally, the last model also optimizes available processing rate at staging areas. As demonstrated in the figure, using optimal asset and staff deployments allows for significant increases in coverage rates throughout the time horizon. Under these assumptions, increasing staffing resources seems to have the most impact on achievable coverage.

Next, we evaluate the benefits and losses due to using robust optimization. The more traditional approach would be to model the demand uncertainty using a stochastic distribution. To make the comparison, we assume that the demand is distributed according to the assumptions described in Section 4 with the difference that we assume three uncertainty factors based on the epi-center of the earthquake. We generate the robust set based on sampled scenarios generated from the assumed distribution. The full planning problem is too large to solve optimally for the stochastic model when using sample average approximation (SAA) (Shapiro, Dentcheva, and Ruszczynski 2009). Therefore, we make the comparison on 10% of the staging areas from a single state (Arkansas) and the demand in that state only. The solution quality is then evaluated with respect to the *mean* performance with the assumed distribution. The results in Figure 4 and Figure 5 show that, at least in our application, the robust optimization achieves comparable solution quality to the stochastic model with a dramatic reduction in computation time.

We also evaluate how the solution quality depends on the fairness measure used in the inventory optimization model. The optimization using worst-case coverage over PODs leads to unacceptably low coverage. Figure 6 compares the coverage for the average objective and the fair objective described in Section 2.

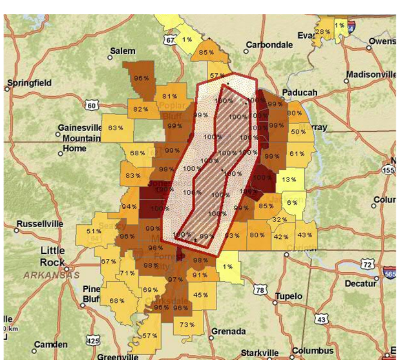

Figure 2: Projected damage intensity in the disaster scenario simulation.

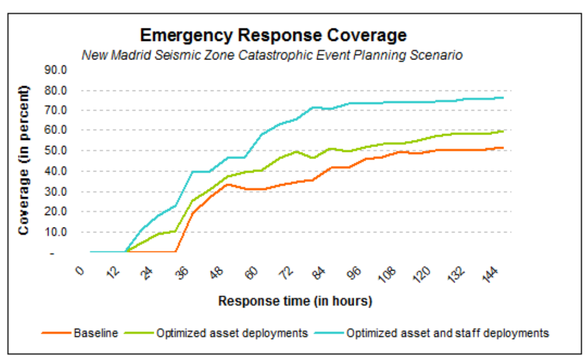

Figure 3: Response demand coverage for three simulated scenarios.

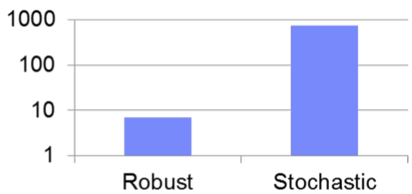

Figure 4: Computation time in seconds.

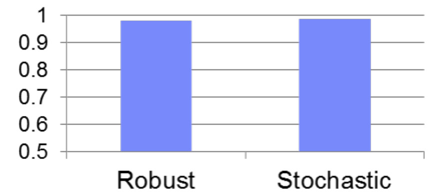

Figure 5: Relative solution quality with stochastic quality being 1.

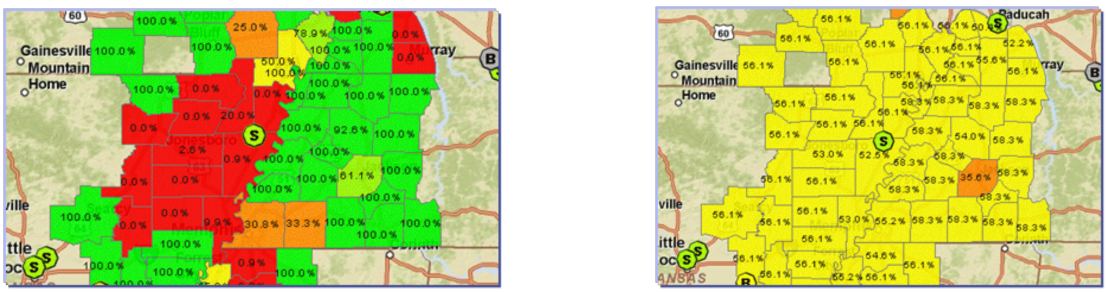

Figure 6: Coverage comparison for uniform and fairness measures.

6 CONCLUSION

We described a simulation and optimization framework for managing the logistics of distributing relief supplies in a multi-tier supply network. The simulation model captures optimized stocking of relief supplies, distribution operations at federal or state-operated staging facilities, demand uncertainty, and the dynamic progression of disaster response operations. We apply robust optimization techniques to develop optimized stocking policies and dispatch of relief supplies between staging facilities and points of distribution. The simulation framework accommodates a wide range of disaster scenarios and stressors, and helps assess the efficacy of response plans and policies for better disaster response.

Our results on a hypothetical, but realistic, projections for damage during an earthquake at New Madrid Fault zone indicate that our models represent a viable approach for disaster response. The models scale well even to a large disaster, covering a population of several million people. The scalability is partially achieved by relying on robust models of uncertainty. The new class of fairness measures that we propose also offers a flexible modeling framework which preserves the tractability of our approach.

ACKNOWLEDGMENTS

We thank Rajesh Ravi and Yichong Yi for technical support and for implementing visualization tools.

REFERENCES

Altay, N., and W. G. Green. 2006. “OR/MS research in disaster operations management”. *European Journal of Operational Research* 175:475–493.

Barbarosoglu, G., and Y. Arda. 2004. “A two-stage stochastic programming framework for transportation planning in disaster response”. *Journal of the Operational Research Society* 55:43–53.

Beamon, B. M., and S. A. Kotleba. 2006. “Inventory modelling for complex emergencies in humanitarian relief operations”. *International Journal of Logistics: Research and Applications* 9:1–18.

Ben-Tal, A., L. E. Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski. 2004. "Anjustable robust solutions of uncertain linear programs". *Mathematical Programming* 99:351–376.

Bertsimas, D., and V. Goyal. 2011. "On the power and limitations of affine policies in two-stage adaptive optimization". *Mathematical Programming* 134:491–531.

Bertsimas, D., and A. Thiele. 2006. "A Robust Optimization Approach to Inventory Theory". *Operations Research* 54:150–168.

Elnashai, A. S., L. J. Cleveland, T. Jefferson, and J. Harrald. 2008. "New Madrid Earthquake Impact Analysis - Planning Scenario". Technical report, Institute for Crisis, Disaster, and Risk Management.

Follmer, H., and A. Schied. 2011. *Stochastic Finance: An introduction in discrete time*. Walter de Gruyter.

Haghani, A., and S. Oh. 1996. "Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations". *Transportation Research A: Policy and Practice* 30:231–250.

Iyengar, G. N. 2005. "Robust Dynamic Programming". *Mathematics of Operations Research* 30:257–280.

Oh, S.-C., and A. Haghani. 1997. "Testing and evaluation of a multi-commodity multi-modal network flow model for disaster relief management". *Journal of Advanced Transportation* 31:249–282.

Porteus, E. L. 2002. *Foundations of Stochastic Inventory Theory*. Stanford Business Books.

Powell, W. B. 2007. *Approximate Dynamic Programming*. Wiley-Interscience.

See, C.-T., and M. Sim. 2010. "Robust Approximation to Multiperiod Inventory Management". *Operations Research* 58 (3): 583–596.

Shapiro, A., D. Dentcheva, and A. Ruszczynski. 2009. *Lecture Notes on Stochastic Programming: Modeling and Theory*. SIAM.

Sheu, J.-B. 2007. "An emergency logistics distribution approach for quick response to urgent relief demand in disasters". *Transportation Research E: Logistics and Transportation Review* 43 (6): 687–709.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw Hill/Irwin.

AUTHOR BIOGRAPHIES

FRANCISCO BARAHONA is a Research Scientist at the IBM T. J. Watson Research Center specializing on optimization. He has over 20 years of experience in optimization projects dealing with airlines optimization, transportation planning, facility location and network design. Before joining IBM he was a professor at the University of Waterloo in Canada. He has a Ph. D. in Operations Research. Dr. Barahona has been an Associate Editor for SIAM Journal on Optimization, Operations Research and Management Science, and EURO Journal on Computational Optimization. His email address is barahon@us.ibm.com.

MARKUS ETTL is a manager in the Business Analytics and Mathematical Sciences Department at IBMs T.J. Watson Research Center. His current research focuses on advanced analytics for supply chain, pricing, revenue management, and procurement. He is a frequent speaker at conferences, universities, and customer events. His email is msettl@us.ibm.com.

MAREK PETRIK is a Research Scientist at the IBM T. J. Watson Research Center. He received a Ph.D. in Computer Science from University of Massachusetts Amherst in 2010. His research interest is in dynamic programming and robust optimization. He works on applications in inventory management and recommender systems. His email and web addresses are mpetrik@us.ibm.com and http://people.cs.umass.edu/~petrik.

PETER RIMSHNICK is an Advisory Software Engineer at the IBM T.J. Watson Research Center. His interests include applied optimization, simulation, and analytics technology. Since joining IBM in 2010, he has worked on projects involving portfolio optimization, commodity transportation planning, manufacturing simulation and incentive compensation design. He holds an M.A. in Economics from Princeton University and an M.Eng. in Operations Research from Cornell University. His email address is pmrimshn@us.ibm.com.