



Models for relief routing: Equity, efficiency and efficacy

Michael Huang^a, Karen Smilowitz^{a,*}, Burcu Balcik^b

^a Industrial Engineering and Management Sciences, Northwestern University 2145 Sheridan Road, Evanston, IL 60208, USA

^b Industrial Engineering Department, Kadir Has University, Cibali, Istanbul 34083, Turkey

ARTICLE INFO

Keywords:

Vehicle routing
Supply allocation
Equity
Humanitarian relief

ABSTRACT

In humanitarian relief operations, vehicle routing and supply allocation decisions are critically important. Similar routing and allocation decisions are studied for commercial settings where efficiency, in terms of minimizing cost, is the primary objective. Humanitarian relief is complicated by the presence of multiple objectives beyond minimizing cost. Routing and allocation decisions should result in quick and sufficient distribution of relief supplies, with a focus on equitable service to all aid recipients. However, quantifying such goals can be challenging. In this paper, we define and formulate performance metrics in relief distribution. We focus on efficacy (i.e., the extent to which the goals of quick and sufficient distribution are met) and equity (i.e., the extent to which all recipients receive comparable service). We explore how efficiency, efficacy, and equity influence the structure of vehicle routes and the distribution of resources. We identify trends and routing principles for humanitarian relief based on the analytical properties of the resulting problems and a series of computational tests.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Recent disasters have increased attention on the effectiveness of humanitarian aid; there is increased pressure from donors on relief agencies to show that pledged aid and goods are reaching those in need quickly (Van Wassenhove, 2006). This has led to a larger focus on improving humanitarian relief logistics. The recent large scale disasters and relief efforts (e.g., 2004 Asian tsunami, 2005 Pakistan earthquake, 2010 Haiti earthquake) highlight the need for improved logistics in the field. For example, following the Tsunami, the amount of pledged relief overwhelmed the relief agencies' ability to properly store and distribute the aid (Russell, 2005). Given the trends in the impact of disasters on vulnerable populations and economies worldwide and the criticality of logistics in humanitarian relief operations, it has been increasingly highlighted that Operations Research (OR) can help improve the logistics of humanitarian aid (Van Wassenhove, 2006).

This paper focuses on last-mile distribution in a humanitarian relief chain from a distribution center to beneficiaries (Balcik et al., 2008). Relief supplies must be delivered quickly in sufficient amounts. Given limitations in transportation resources and relief supplies, and damaged infrastructure, it is challenging to plan last mile operations (Balcik et al., 2008). Distribution decisions are often made ad hoc, which may lead to inefficient use of resources, slow response, and inadequate or inequitable relief deliveries.

The classical vehicle routing problem (VRP) minimizes total transportation costs; in humanitarian relief, one is primarily concerned with whether the routes are able to deliver the aid quickly. Focusing on relief response times, Campbell et al. (2008) show that the choice of objective affects how aid is distributed. With alternative objectives of minimizing the last arrival time and minimizing the sum of arrival times, the authors demonstrate that superior service times may be achieved

* Corresponding author. Address: 2145 Sheridan Road, TECH, M233 Evanston, IL 60208, USA. Tel.: +1 847 491 4693; fax: +1 847 491 8005.

E-mail addresses: Michael-Huang@u.northwestern.edu (M. Huang), ksmilowitz@northwestern.edu (K. Smilowitz), burcu.balcik@khas.edu.tr (B. Balcik).

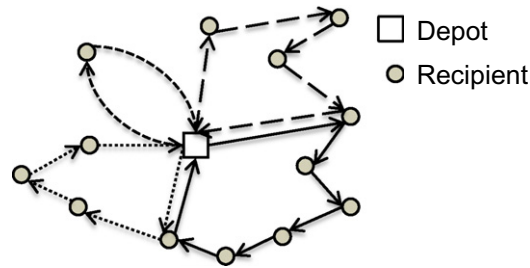


Fig. 1. A set of routes in LMDP.

than those resulting from a traditional VRP objective. They show that minimizing total routing costs results in solutions with longer response times.

Equitable aid distribution among recipients is also a critical consideration in humanitarian relief (Beamon and Balcik, 2008). However, equity may be hard to define both in practice and for modeling purposes. In this paper, we characterize equity in terms of the disparity between service levels among aid recipients, where service levels are characterized by delivery speed and amount. Even so defined, it is not clear how equity should be modeled. Furthermore, it is not obvious how different and potentially conflicting objectives of equity, efficacy and efficiency affect the route structures in the relief context.

Balcik et al. (2008) present an analysis of last mile operations and show how modeling these operations with all the associated complexities can make it difficult to study the underlying differences in route design that occur. Our goal in this paper is to simplify the modeling of last mile distribution to a more stylized setting where we can more easily gain insights into the effects of equity and other considerations in relief distribution. We refer to this problem setting as the last-mile delivery problem (LMDP). The LMDP designs vehicle routes and distribution schedules for a fleet of vehicles delivering supplies from a distribution center, yet unlike Balcik et al. (2008), we focus on a single period problem where each vehicle performs at most one trip to deliver one commodity type. Although more restricted than Balcik et al. (2008), our problem setting still captures many critical characteristics of the last mile environment; in particular, the LMDP in this paper would well address relief operations in rural areas where each vehicle can perform only one trip per period and supplies are distributed in the form of standard packages/pallets. Relief organizations often make daily distribution plans rather than considering multiple days ahead, especially during the initial phases of the emergency due to the chaotic environment and lack of reliable information (Manttyvaara, 2010). An example of LMDP routes is presented in Fig. 1. The demand of a site can be satisfied by more than one vehicle in the LMDP, similar to the Split Delivery Vehicle Routing Problem (SDVRP). The SDVRP is motivated by the potential savings in costs due to splitting demand among multiple vehicles (Dror and Trudeau, 1989). Split deliveries in last mile relief distribution can allow one to quickly serve large demands using limited resources. Our problem setting allows us to analyze the impact of different objectives on route structures and the performance of aid distribution, in terms of (i) efficiency (transportation costs), (ii) efficacy (quick and adequate response), and (iii) equity (fairness as measured by deviations between recipients in efficacy). The incorporation of equity in particular leads to unique challenges.

We develop a set partitioning model for the LMDP to incorporate alternative objectives, based on efficiency, efficacy and equity metrics. Analytical properties of the LMDP models are examined. A numerical study of small instances, in which it is possible to obtain optimal solutions for each of the objectives, is conducted to develop insights into route characteristics. Our analysis shows that there exist substantial differences among solutions that attempt to minimize efficiency compared to those that are concerned with efficacy and equity. Heuristics are developed to solve large instances of the LMDP variations.

This paper is organized as follows: in Section 2, we review the relevant literature. Section 3 gives a detailed description and analysis of the LMDP variations. Section 4 proposes heuristics for the LMDP variations. Section 5 discusses practical implications of this work and future directions.

2. Literature review

There is a growing literature that addresses relief supply distribution in humanitarian logistics. One of the earliest studies that addresses distribution of relief supplies from a distribution center to several camps is Knott (1987). The author formulates a linear programming model that finds the number of trips each vehicle makes to maximize the amount of deliveries while minimizing transportation costs and discusses how considering additional aspects might make the problem very difficult to solve. More recently, the literature addresses more complicated relief supply distribution problems and captures the inherent complexities of the humanitarian relief environment, including multiple commodities, multiple transportation modes or vehicle types, multiple periods, uncertain or varying demand and supply levels and transportation network conditions, and delivery time windows. Most studies develop heuristic algorithms to solve the proposed models (e.g., Haghani and Oh, 1996; Barbarosoglu et al., 2002; Ozdamar et al., 2004; Lin et al., 2009; Shen et al., 2009; Van Hentenryck et al., 2010) or use commercial solvers to test the models for small to moderate problem instances (e.g., Barbarosoglu and Arda, 2004;

Tzeng et al., 2007; Balcik et al., 2008). Studies that focus on relief supply distribution consider a number of different objectives, including minimizing logistics costs (e.g., Haghani and Oh, 1996), minimizing the amount of unsatisfied demand (e.g., Ozdamar et al., 2004), and minimizing service completion or delays (e.g., Barbarosoglu et al., 2002; Yi and Ozdamar, 2007). Equity, although critical, is considered by only a few studies. Tzeng et al. (2007) consider fairness in the problem of choosing transfer point locations in distributing supplies from a set of supply points to demand points. A multi-period multi-objective model is developed that considers, in addition to minimizing total cost and travel time, fairness by maximizing the minimum service satisfaction among demand points. In Balcik et al. (2008), routing and supply allocation decisions are considered jointly in distributing multiple types of relief supplies. The authors minimize the maximum unsatisfied demand percentage over demand locations along a planning horizon. Delivery scheduling decisions are driven by costs due to supply allocation and transportation over multiple days. Lin et al. (2009) study a similar problem that considers delivery of prioritized items in a relief scenario; the authors develop a multiple objective model to minimize unsatisfied demand and travel costs and ensure equity.

Campbell et al. (2008) focus on service time and evaluate two objective functions: minimizing the maximum arrival time of supplies (minmax), and minimizing the sum of arrival times of supplies (minsum) – thereby minimizing the average arrival time. As each demand location is visited exactly once, and full satisfaction of demand is ensured, supply equity is not a concern in the study. The authors explore the impact of different objective functions. For the traveling salesman problem and the VRP, solutions obtained from minmax and minsum objectives are compared with solutions obtained by minimizing total travel time. It is shown that the minmax and minsum objectives ensure better response times for the demand locations, at the expense of reduced efficiency (i.e., increased total route length or costs). Our paper extends this work by considering demands at each recipient which necessitates alternate equity metrics.

Outside of humanitarian relief, equity-related issues have been widely studied using OR models in areas such as public facility location; see, for example, Erkut (1993), Mulligan (1991), and Marsh and Schilling (1994). Although routing decisions affect equitable service provision, most of the routing studies in the literature address only efficacy and efficiency objectives. In routing, equity-related issues have only recently received attention; and then, only limited to few applications mostly addressing public/nonprofit sector problems (see Balcik et al., 2011). As stressed by Marsh and Schilling (1994), since there may not be a single equity metric for a specific application, analyses that explore the tradeoffs between equity and other objectives can be helpful in understanding the use of different equity policies and metrics. As such, this paper defines and formulates different equity metrics focusing on last mile operations in humanitarian relief.

3. LMDP modeling and analysis

This section describes the LMDP in further detail. Section 3.1 presents the general problem setting. Section 3.2 presents an integer programming formulation of the problem. Section 3.3 introduces the performance metrics for efficiency, efficacy and equity that define the LMDP model variations; Section 3.4 discusses analytical properties of the LMDP models. Section 3.5 presents a computational study of the different models.

3.1. Problem setting

As mentioned earlier, the goal of this paper is to study a stylized version of the LMDP to gain insights into the underlying route structures. The LMDP is defined on a graph $G = (N^0, A)$ which we assume is complete and symmetric. The node set N^0 includes the depot ($i = 0$) and the demand locations ($i \in N: i, \dots, N$). Travel costs (in hours), $c_{ij} \forall (i, j) \in A$, are non-negative and satisfy the triangle inequality. Routes begin at time $t = 0$. Demand is measured in units of full pallets of relief supplies. Pallets are modeled as a single commodity, but each pallet may be composed of different types of supplies. The set $K = \{1, \dots, K\}$ represents a fleet of vehicles operating from the depot to serve the aid recipients and C is the capacity of each vehicle measured in pallets. Demand may be split across multiple vehicles. We assume that the demand of any recipient is less than the capacity of a vehicle; $d_i \leq C \forall i \in N$ and sufficient supplies are accumulated at the depot which can be delivered to recipients ($KC \geq \sum_{i \in N} d_i$). Vehicles may complete at most one trip and thus a vehicle visits a node at most once and may not return to the depot to refill its capacity.

The LMDP involves two types of decisions. Routing decisions determine for each vehicle a set of nodes and visit sequence (and hence the arrival times). Distribution decisions determine the number of pallets delivered at each visit. We develop performance metrics to measure the effects of these decisions and use the metrics as objective functions of different models of the LMDP. Efficiency (Z_1) is measured by the total travel costs. Efficacy (Z_2) is measured by the sum of each pallet's arrival time. Equity (Z_3) is measured by several functions which capture the disparity in efficacy among nodes.

3.2. Integer program formulation

The LMDP is formulated with an extension of the set partitioning formulation of the VRP. The set partitioning formulation of the VRP assumes that the set $R = \{1, \dots, R\}$ of feasible routes is given. Each route $r \in R$ represents a set of nodes and an order in which the nodes are visited. From this information, the following parameters can be obtained for each route. The travel cost c_r is the time required to travel the length of the route. The parameter b_{ir} denotes the presence of node i on route

r : $b_{ir} = 1$ if node i is visited by route r , and 0 otherwise. The parameter a_{ir} denotes the arrival time at node i on route r , if route r visits node i (i.e., $b_{ir} = 1$). If route r does not visit node i (i.e., $b_{ir} = 0$), a_{ir} equals T , an upper bound on the time horizon.

We assume that each candidate route in \mathbf{R} is performed by at most one vehicle. In Section 3.4, we show that there exists an optimal solution where each route is run at most once. The binary decision variables x_r determine whether route r is selected; $x_r = 1$ if route r is selected, and 0 otherwise. The integer decision variables y_{ir} indicate the number of pallets delivered to node i by route r .

We define the LMDP(Z_f) as the model of the LMDP with the objective function of minimizing the generic metric Z_f ; specific metrics are introduced in Section 3.3. The LMDP(Z_f) is formulated as:

$$\text{Min } Z_f(x, y) \quad (1a)$$

subject to

$$\sum_{r \in \mathbf{R}} x_r \leq K \quad (1b)$$

$$y_{ir} \leq d_i b_{ir} x_r \quad \forall i \in \mathbf{N} \quad \forall r \in \mathbf{R}, \quad (1c)$$

$$\sum_{i \in \mathbf{N}} y_{ir} \leq C \quad \forall r \in \mathbf{R}, \quad (1d)$$

$$\sum_{r \in \mathbf{R}} y_{ir} = d_i \quad \forall i \in \mathbf{N}, \quad (1e)$$

$$y_{ir} \text{ integer}, \quad x_r \text{ binary} \quad (1f)$$

Constraint (1b) guarantees that the number of routes does not exceed the fleet size. Constraints (1c) ensure that a node is delivered supplies along a route only if it is visited on the route. Constraints (1c) also set an upper bound on the delivery amounts equal to the demand for that location. Constraints (1d) limit the amount delivered on each route to the vehicle capacity. Constraints (1e) ensure each recipient node receives precisely the number of pallets required. Constraints (1f) specify that the delivery variables are integer and the routing variables are binary.

3.3. Performance metrics

In this section, metrics are formulated to characterize the objective functions based on efficiency, efficacy, and equity. An example is presented to illustrate each objective's impact on routing solutions.

3.3.1. Efficiency

Let Z_1 denote the measure of efficiency which calculates total travel time for selected routes:

$$Z_1(x, y) = \sum_{r \in \mathbf{R}} c_r x_r \quad (2)$$

Because the Z_1 -based objective is the traditional VRP objective, the LMDP(Z_1) is equivalent to the SDVRP. The SDVRP reduces costs by allowing vehicles to fully use capacity (see Archetti et al., 2006; Archetti et al., 2008).

3.3.2. Efficacy

Let Z_2 denote the measure of efficacy which calculates the speed and sufficiency of deliveries.

$$Z_2(x, y) = \sum_{i \in \mathbf{N}} \sum_{r \in \mathbf{R}} a_{ir} y_{ir} \quad (3)$$

The arrival time at node i on route r , a_{ir} , is multiplied by the amount of pallets delivered to determine the sum of arrival times in that delivery. Eq. (3) sums the arrival times across all pallets to yield the *demand-weighted arrival time*. The amount delivered, y_{ir} , represents the change in cumulative unsatisfied demand and the demand-weighted arrival time is calculated by summing the product of arrival time and change in cumulative demand. The equivalent value of the *time-weighted unsatisfied demand* can be found by summing the product of time between deliveries and cumulative unsatisfied demand.

The objective of minimizing Z_2 is an extension of the objective of minimizing the sum of arrival times for unit demand from Campbell et al. (2008). The unit demand LMDP(Z_2) has been given several names in the literature including the delivery man problem, the traveling repairman problem, and the minimum latency problem.

3.3.3. Equity

We introduce three metrics to measure equity. The first two, Z_3^a and Z_3^b , are deviation-type equity metrics that measure the spread in service level across nodes. While Z_3^a and Z_3^b are intuitive, they are not (as shown in Section 3.3.4) suitable as an objective in the LMDP model; therefore, we use Z_3^a and Z_3^b metrics for evaluation purposes. The third equity metric Z_3 cap-

tures response time and equity. To calculate the equity in delivery times, we disaggregate the Z_2 -based objective by node. Let s_i be the service level of node i measured by the demand weighted arrival time normalized by the demand of node i and let \bar{s} denote the average service level across nodes:

$$S_i = \frac{1}{d_i} \sum_{r \in R} a_{ir} y_{ir} \quad (4)$$

$$\bar{s} = \frac{1}{N} \sum_{i \in N} s_i \quad (5)$$

Metrics Z_3^a and Z_3^b address the disparity in the s_i values. Specifically, Z_3^a captures the maximum pairwise difference and Z_3^b captures the standard deviation.

$$Z_3^a(x, y) = \max_{i \in N} s_i - \min_{i \in N} s_i \quad (6)$$

$$Z_3^b(x, y) = \sqrt{\sum_{i \in N} \frac{(s_i - \bar{s})^2}{N}} \quad (7)$$

The Z_3 metric is a compromise between efficacy and equity. Similar to the Z_2 metric, Z_3 minimizes waiting time and serves demand quickly, but also incorporates equity by weighing units of demand differently. To incorporate equity, we consider a convex disutility function $f(w)$, where w is the fraction of unsatisfied demand. Fig. 2 plots an example function. Because the rate of change of the disutility is greater when the percentage of unsatisfied demand is closer to 1 (i.e., the demand served is low), the first pallets delivered receive greater benefit than the last.

Similar to how the Z_2 metric measures the *demand-weighted arrival time*, the Z_3 metric minimizes the *disutility-weighted arrival time*. The disutility function encourages a vehicle not to necessarily satisfy a node's entire demand but rather to save supply to serve another node. Unlike Z_2 where the change between delivered demand is accessible from information in the model (y_{ir}), the change in disutility requires a calculation dependent on many variables. It is easier to formulate the equivalent value of *time-weighted disutility*. We discretize time horizon into hours. For each hour $t \in T = 1, \dots, T$ and each node $i \in N$, the percentage of unsatisfied demand of node i at time t , w_{it} , is calculated and a penalty equal to $f(w_{it})$ is accrued. Thus, Z_3 , the time-weighted disutility is calculated as follows:

$$w_{it} = \left(1 - \frac{1}{d_i} \sum_{r: a_{ir} < t} y_{ir} \right) \quad (8)$$

$$Z_3(x, y) = \sum_{t=1}^T \sum_{i=1}^N f(w_{it}) \quad (9)$$

Eq. (9) sums the penalties accrued due to the percentage of unsatisfied demands of each node $i \in N$ over each hour $t \in T$. In practice, $f(w_{it})$ does not need to be calculated for every hour in T , but only when $f(w_{it})$ changes. This can be found by sorting the arrival times of deliveries to node i and determining the time between successive deliveries. The following piecewise linear disutility function is used in our calculations:

$$f(w) = \begin{cases} \frac{4w}{13}, & w < 0.25 \\ \frac{8w-1}{13}, & 0.25 \leq w < 0.5 \\ \frac{16w-5}{13}, & 0.5 \leq w < 0.75 \\ \frac{24w-11}{13}, & w \geq 0.75 \end{cases} \quad (10)$$

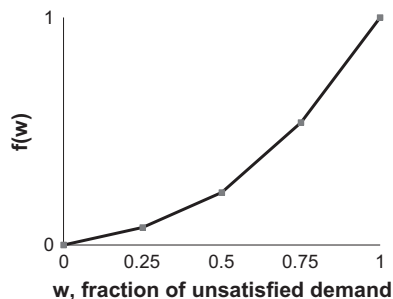


Fig. 2. An example piecewise-linear disutility function for unsatisfied demand.

The function is chosen to mimic the shape of a convex polynomial function. As more pallets are delivered, w decreases; the progressively increasing slope ensures that the last pallets delivered are weighted less than the first.

3.3.4. An illustrative comparison of objectives

Fig. 3 presents a network with three recipient nodes ($i = 1, 2, 3$) each with a demand of four pallets. Consider an instance with two vehicles, each with capacity of six pallets. Fig. 4 presents the optimal solutions for each LMDP variation. The figure shows the route path and delivery times and amounts at each node. Let $\pi(k)$ denote the route that is run by vehicle k . The routes obtained by the model variations differ in terms of delivery times and amounts. Table 1 displays the service level for each node (i.e., s_i values) and the metric values for each solution. Naturally, each solution optimizes its respective metric at the expense of other metrics.

This simple example illustrates initial trends that guide our subsequent analysis. While all equity metrics have a common goal of reducing the disparity in service times among recipients, the model variations can yield differences in solutions. In the solution displayed in Fig. 4c, the two vehicles travel in opposite directions, each dropping off two units of demand at each of the nodes. The service level for all nodes is 9.5, yet nodes 1 and 3 can improve their service level without lowering the service level at node 2 if the vehicles that visit them first deliver their full demand. In such a solution, service levels of the nodes are $s_1 = 6$, $s_2 = 9.5$ and $s_3 = 4$. Thus in the LMDP(Z_3^a) and LMDP(Z_3^b) solutions, nodes 1 and 3 are penalized to match the service level at node 2 to minimize the disparity between service levels. While this is only one example, one can imagine

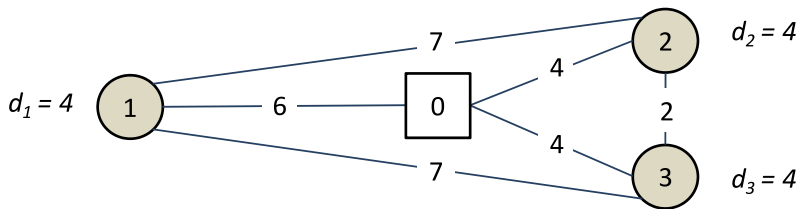


Fig. 3. A sample distribution network with three recipient nodes. Arc lengths are shown on each arc.

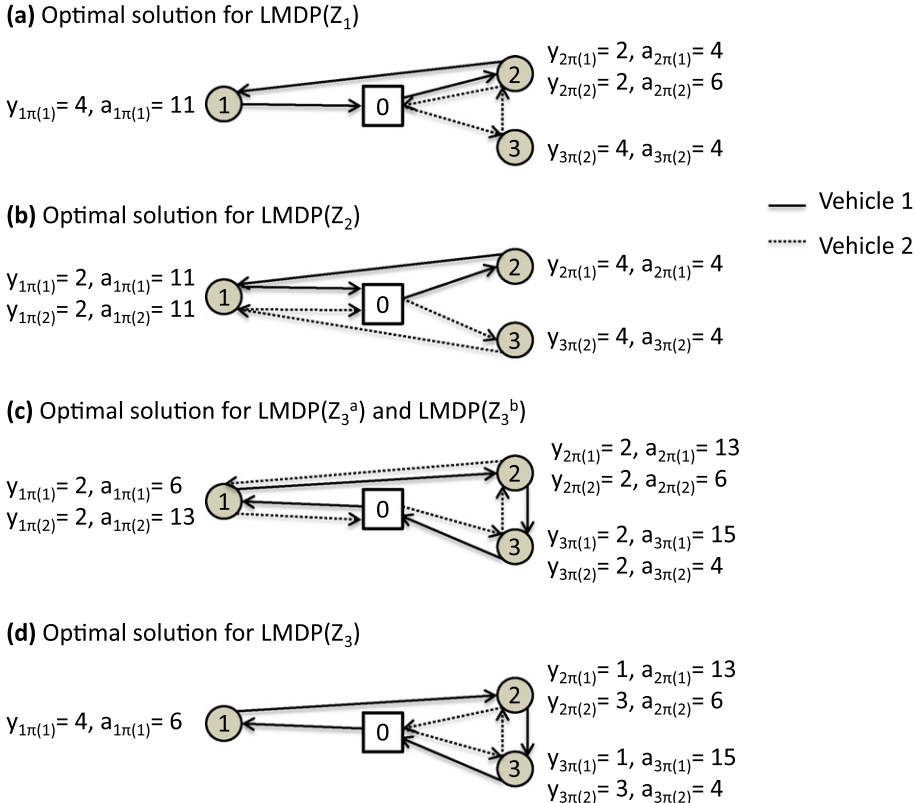


Fig. 4. Optimal solutions for the LMDP variations.

Table 1
Performance of optimal solutions under different objectives.

Solution for	Node service level			Metric				
	s_1	s_2	s_3	Z_1	Z_2	Z_3^a	Z_3^b	Z_3
LMDP(Z_1)	11	5	4	27	80	7	3.09	19.46
LMDP(Z_2)	11	4	4	34	76	7	3.30	19.00
LMDP(Z_3^a)	9.5	9.5	9.5	38	114	0	0	21.77
LMDP(Z_3^b)	9.5	9.5	9.5	38	114	0	0	21.77
LMDP(Z_3)	6	7.75	6.75	29	82	1.75	0.72	17.38

in a large network, every route taking detours so that all nodes receive equally poor service. At a nominal cost to efficiency and efficacy (7% and 8% increase, respectively), the LMDP(Z_3) solution greatly outperforms the LMDP(Z_1) and LMDP(Z_2) solutions with respect to the Z_3^a and Z_3^b metrics. The large improvement in equity can be attributed to the allocation of demand at node 3. Rather than completely serving node 3, vehicle 2 delivers an extra unit to node 2. The solution does not suffer large increases in Z_1 and Z_2 because there is still an incentive to serve demand quickly.

Additionally, this example illustrates how multiple optima to the LMDP models may produce unfavorable results. In the solution displayed in Fig. 4a, the solution can improve its Z_2 , Z_3^a , Z_3^b and Z_3 without worsening its Z_1 metric value by reversing the route run by vehicle 1. In the computational studies, we utilize weighted secondary objectives to select preferred options among multiple optima, as discussed in Section 3.5.1.

3.4. Analytical properties

In this section, we explore route structures for the LMDP models. In an early SDVRP paper, Dror and Trudeau (1990) prove Property 3.1 regarding splitting in optimal SDVRP routes. We present the related Properties 3.2 and 3.3 for the LMDP(Z_2) and LMDP(Z_3). Property 3.4 validates our modeling assumption that each route is run at most once. The proofs of Properties 3.2–3.4 are presented in Appendix A.

Property 3.1 (Dror and Trudeau). For LMDP(Z_1), there exists an optimal solution where for every pair of nodes, at most one vehicle visit both nodes.

Property 3.2. The result from Property 3.1 holds for LMDP(Z_2).

Property 3.3. The result from Property 3.1 does not hold for LMDP(Z_3).

Property 3.4. For each of the objective functions of the LMDP based on Z_1 , Z_2 and Z_3 , there exists an optimal solution where each route is run at most once.

Properties 3.1 and 3.2 imply that routes cannot share a subset of nodes with more than one recipient node. This suggests that for the LMDP(Z_1) and LMDP(Z_2), splitting is only beneficial for reasons linked to capacity. For the LMDP(Z_1), splitting of demand allows for fewer vehicles to be used. With the triangle inequality, a reduction in routes run translates to a reduction in travel costs. In contrast, in the LMDP(Z_2) where response time is minimized, it is desirable to utilize the entire fleet of vehicles. Although the entire fleet is used, splitting may occur if a vehicle can reach a recipient earlier than the other vehicles but has insufficient supply to serve the recipient's full demand. Properties 3.1 and 3.2 can be used to add cuts to the formulation for their respective problems. The constraints $\sum_{r:(b_{ir}>0)\&(b_{jr}>0)} x_r \leq 1 \quad \forall i, j \in N$ may be added to the formulations of the LMDP(Z_1) and LMDP(Z_2). Because splitting a node's demand among multiple vehicles can improve the equity-based Z_3 objective, Property 3.1 does not hold for LMDP(Z_3), as shown in Fig. 4d. Splitting occurs so that a node may receive some of its initial supply earlier at the expense of another node receiving its remaining supply later. We show how the differences between the properties regarding Z_1 and Z_2 with the property regarding Z_3 impact route structure in the following section.

3.5. Computational study of small instances

In this section, small instances of the LMDP models are solved using a commercial solver and solutions under the different objectives are compared. Section 3.5.1 introduces the test cases. Sections 3.5.2 and 3.5.3 present observations on route structure and performance metrics respectively.

3.5.1. Test cases

Small instances are composed of an 8-node set and a 10-node set. For each set, five location distributions are paired with five demand distributions. The location distributions locate the nodes on a 100×100 square with the depot at the center; recipient nodes are either (1) clustered (C), (2) evenly distributed (E), or (3) randomly distributed (R1, R2, R3). The

demand distributions allow demands to range between 1 and 4 pallets. Demand distributions are either (1) clustered by location creating low and high demand regions (C), (2) evenly distributed (E), or (3) randomly distributed (R1, R2, R3). Detailed descriptions of the instances are available in Appendix B of the expanded version of this paper (Huang et al., 2010). Vehicles have a capacity of 10 pallets ($C = 10$); and the fleet size is one more than the minimum for feasibility ($K = \lceil \sum_i d_i / C \rceil + 1$). Travel costs are set to be the Euclidean distance between two nodes rounded to the nearest integer.

To choose preferred options among multiple optima, the following objectives are used: $Z_1 + \alpha Z_2$ for LMDP(Z_1), $Z_2 + \alpha Z_1$ for LMDP(Z_2), and $Z_3 + \alpha Z_1$ for LMDP(Z_3), where α is chosen to be small enough to ensure that the primary objective is minimized. The metric Z_3 is not used as a secondary objective due its complexity and the associated computational time. For the LMDP(Z_3), Z_1 is used as the secondary objective because efficacy (as measured by Z_2) is already partially addressed in the Z_3 metric.

A C++ program using the C++ API of ILOG Concert Technology (ILOG, 2008) is developed; the mixed integer programs are solved using Concert 2.7 and CPLEX 11.2. The program is run on 2.4 GHz 64-bit (4 MB L2 cache) CPU machine with 8 GB of RAM. To ensure computational tractability, we reduce the size of the candidate route set R , by restricting R to only the feasible routes that do not visit more than five nodes. The restriction should not have a large impact on LMDP(Z_2) and LMDP(Z_3) solutions as the solutions aim to balance the demand roughly equally; however, the restriction may affect LMDP(Z_1) solutions where longer routes reduce the fleet’s travel time. Therefore, although these solutions are optimal for the given setting, we should note that the setting is restricted by route length. Lastly, a 60-h limit is imposed on the solver for each LMDP instance.

3.5.2. Route structure

Several observations can be made regarding the general nature of the optimal solutions for the different models of the LMDP. To illustrate these observations we provide a set of solutions for one 10-node instance of the LMDP variations in Fig. 5. For each of the diagrams, with the exception of the depot at the center, the size of the node is representative of the amount of demand requested by the node.

In Table 2, we present aggregate results for all instances. Columns 3 and 4 present the percentage of solutions which use the minimum number of vehicles and the percentage of solutions that use all available vehicles (one more than minimum, in this case). Column 5 displays the percentage of routes in which the last node visited coincides with the node on the route that is farthest from the depot. Columns 6 through 9 display the average arrival time of nodes for the different levels of node demand.

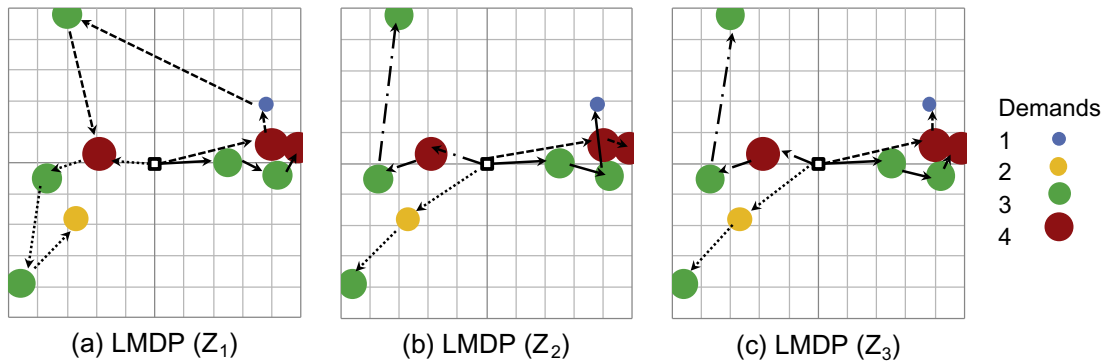


Fig. 5. Optimal solutions for Z_1, Z_2 and Z_3 for a 10-node instance.

Table 2
Summary results for route structure across 8-node and 10-node instances.

Number of nodes	Model	Vehicles used		% Last node visited is farthest	Average arrival time by node demand			
		Min (%)	Min + 1 (%)		1	2	3	4
8-Nodes	LMDP(Z_1)	88	12	49	85.0	70.2	58.8	49.1
	LMDP(Z_2)	0	100	89	60.5	52.6	47.5	42.4
	LMDP(Z_3)	0	100	93	53.3	51.9	50.9	46.3
10-Nodes	LMDP(Z_1)	80	20	38	88.3	72.0	56.3	42.7
	LMDP(Z_2)	0	100	79	62.7	53.1	45.5	39.4
	LMDP(Z_3)	0	100	86	55.8	53.5	48.4	41.2

Observation 1. Number of vehicles

With only a few exceptions, the optimal LMDP(Z_1) solution uses the minimum number of vehicles (see columns 3 and 4 of Table 2). This point is made in Campbell et al. (2008); given the triangle inequality, minimizing travel costs results in the use of as few vehicles as possible. Instances where the LMDP(Z_1) does not use the minimum number of vehicles occur when multiple optima exist and the use of an extra vehicle can reduce service times without adding travel time. These exceptions occur in 12% of the 8-node instances and 20% of the 10-node instances. Optimal solutions to the LMDP(Z_2) and LMDP(Z_3) use all available vehicles which reduces the service time of the nodes at the cost of increased travel time. In Fig. 5a, the LMDP(Z_1) solution splits delivery to one node with demand of four pallets to use a total of three vehicles instead of four.

Observation 2. Route shape

Routes in optimal LMDP(Z_1) solutions tend to be hooked-shaped, which is consistent with earlier studies of the VRP (Daganzo, 2005). By curving back toward the depot, the routes reduce the distance from the final recipient back to the depot. In contrast, in the optimal solutions to the LMDP(Z_2) and LMDP(Z_3), where the line haul from the final recipient back to the depot is not measured in the objective, optimal routes tend to be more ray-shaped, moving away from the depot (as seen in Fig. 5b and c). We quantify these trends over all instances by measuring the percent of solutions in which the last node visited on a route is the farthest node from the depot on that route. For the 8-node instances, the last node visited is the farthest from the depot 49% of the time in the LMDP(Z_1) solutions, 89% of the time in the LMDP(Z_2) solutions, and 93% of the time in the LMDP(Z_3) solutions. For the 10-node instances, these percentages are 38%, 79%, and 86%, respectively.

Observation 3. Impact of node demand

In the LMDP(Z_2), where service time is measured in the objective, the routes are highly demand driven. Nodes that have high demand are served at a premium, often at the expense of recipients with low demand. Nodes with demand 1 are often served last on their respective routes. In Fig. 5b, the LMDP(Z_2) solution crosses two routes so that a high-demand node may be served earlier. In the LMDP(Z_3), as long as the nodes are visited by a single vehicle, the nodes are served without regard to their demand level. Thus, the smallest nodes are treated more equally; they are served by the closest route and are not bypassed for nodes with higher demands. To quantify this observation we evaluate the average arrival time of nodes based on the demand of the node. This highlights two interesting trends. The LMDP(Z_1) solution has longer arrival times for all nodes because LMDP(Z_1) solutions use fewer vehicles. The use of fewer vehicles means that the nodes on average receive their demand at later times. Further, nodes with lower demand often have longer arrival times as they are more likely to share their route with more nodes; again longer routes translate to longer arrival times. While each model exhibits a higher average arrival time for low-demand nodes, the deviation in arrival time is much lower in the LMDP(Z_3) solutions than the other models. Coincidentally, the hook-like characteristics of the routes in the LMDP(Z_2) solution result from node demand as the routes curve back to serve the smallest nodes that had been ignored earlier.

3.5.3. Metric performance

Each model (LMDP(Z_1), LMDP(Z_2) and LMDP(Z_3)) is evaluated relative to the five metrics (Z_1 , Z_2 , Z_3^a , Z_3^b and Z_3). To evaluate a solution with respect to metrics Z_1 , Z_2 , and Z_3 , we calculate the percentage difference between the metric value of the solution and the metric value of the optimal solution for that metric. For example, let $Z_j^{LMDP(Z_1)}$ be the Z_j metric value of the optimal LMDP(Z_k) solution and $Z_j^* = Z_j^{LMDP(Z_j)}$, then to evaluate the optimal LMDP(Z_1) solution with respect to Z_2 we calculate $\Delta_2^{LMDP(Z_1)} = (Z_2^{LMDP(Z_1)} - Z_2^*)/Z_2^*$. For the metrics Z_3^a and Z_3^b , the percent difference is taken from the best solution among the LMDP(Z_1), LMDP(Z_2) and LMDP(Z_3) solutions.

Table 3 summarizes the results of the 8-node and 10-node instances by aggregating the 25 instances with different demand and location distributions. Several observations may be made from Table 3.

Table 3

Summary of deviation in metric values for LMDP models.

Model		Optimality gap by metric									
		8-Node instances					10-Node instances				
		Δ_1	Δ_2	Δ_3	Δ_3^a	Δ_3^b	Δ_1	Δ_2	Δ_3	Δ_3^a	Δ_3^b
LMDP(Z_1)	Avg	–	29%	33%	65%	64%	–	29%	34%	72%	68%
	Max	–	65%	74%	135%	119%	–	61%	68%	332%	222%
	Min	–	7%	7%	0%	0%	–	3%	2%	0%	0%
LMDP(Z_2)	Avg	18%	–	1%	2%	2%	17%	–	2%	7%	4%
	Max	39%	–	11%	15%	13%	39%	–	14%	49%	30%
	Min	5%	–	0%	0%	0%	2%	–	0%	0%	0%
LMDP(Z_3)	Avg	17%	2%	–	0%	0%	17%	2%	–	2%	1%
	Max	38%	33%	–	5%	2%	39%	13%	–	23%	18%
	Min	8%	0%	–	0%	0%	1%	0%	–	0%	0%

Table 4

Summary of deviation in metric values for LMDP models for 8-node instances with minimum number of vehicles.

Model		Optimality gap by metric				
		Δ_1	Δ_2	Δ_3	Δ_3^a	Δ_3^b
LMDP(Z_1)	Avg	–	8%	12%	17%	15%
	Max	–	32%	56%	109%	74%
	Min	–	0%	0%	0%	0%
LMDP(Z_2)	Avg	8%	–	5%	10%	11%
	Max	26%	–	32%	54%	38%
	Min	0%	–	0%	0%	0%
LMDP(Z_3)	Avg	7%	4%	–	1%	1%
	Max	29%	21%	–	23%	18%
	Min	0%	0%	–	0%	0%

Observation 4. The differences in route structure between the efficiency model (LMDP(Z_1)) solutions and solutions of the other variations translate into large differences in metric values.

By not considering travel from the last recipient to the depot, travel costs increase by 17% to 18%, on average, when Z_1 is not the primary objective. The LMDP(Z_2) and LMDP(Z_3), by minimizing arrival times, implicitly reduce distance, yet the LMDP(Z_1) does not consider efficacy and equity, and these metrics are far from optimal when Z_1 is the primary objective, with average gaps of 29% for Z_2 , 65–72% for Z_3^a , and 64–68% for Z_3^b .

Observation 5. The similarities in route structure of efficacy and equity (LMDP(Z_2) and LMDP(Z_3)) solutions yield similar metric values. However, differences in equity among the solutions increase when capacity is reduced

In contrast to the large difference observed between the LMDP(Z_1) solutions and the other models, the optimal LMDP(Z_2) and LMDP(Z_3) solutions have similar metric values; in many instances, the solutions are identical.

In terms of Z_3^a and Z_3^b , the LMDP(Z_3) solutions perform slightly better than the LMDP(Z_2) solutions. In the 8-node instances, the LMDP(Z_2) averages 2% worse than the best solution while the LMDP(Z_3) averages less than 1% worse. In the results from the 10-node instances there is a larger gap. For Z_3^a and Z_3^b , the LMDP(Z_2) is 7% and 4% worse than the best solutions, respectively, while the LMDP(Z_3) is only 2% and 1% worse. We conjecture that the difference is due to a reduction in the amount of slack capacity that is available for the 10-node instances. This is supported by results with fewer vehicles as a reduction in number of vehicles results in a reduction of capacity.

Table 4 presents results for the 8-node instances when the number of available vehicles is reduced to the necessary minimum. We could not obtain LMDP(Z_3) solutions for many 10-node instances in the time limit.

Observation 6. When the three models are constrained to use the minimum number of vehicles, the differences in metric values between the LMDP(Z_1) solutions and the other two LMDP models decreases.

The loss of the extra vehicle results in closer metric values between the optimal solutions of the LMDP(Z_1) and the optimal solutions of the LMDP(Z_2) and LMDP(Z_3); in one instance, the solutions are identical for the three models. We note the decrease in optimality gap when comparing Table 4 with Table 3. With respect to Z_1 , the optimal LMDP(Z_2) and LMDP(Z_3) solutions average only 8% and 7% higher than the optimal LMDP(Z_1) solutions, respectively, and the LMDP(Z_1) has Z_2 and Z_3 metric values averaging only 8% and 12% higher than the optimal solutions for the LMDP(Z_2) and LMDP(Z_3), respectively.

Reducing the number of vehicles effectively reduces capacity, and consistent with Observation 5, the differences between the LMDP(Z_2) model and the LMDP(Z_3) model become more prominent. With respect to Z_2 , the LMDP(Z_3) performs 4% worse and with respect to Z_3 , the LMDP(Z_2) performs 5% worse. Most notable is that with respect to the equity metric Z_3^b , the optimal LMDP(Z_2) solutions average 10% worse than the best solution compared to 1% of the optimal LMDP(Z_3) solutions. We posit that as total demand approaches the total capacity of the fleet, the percent difference in solutions becomes similar to the example in Section 3.3, where the LMDP(Z_3) exhibits a large improvement in equity at a nominal cost to efficiency and efficacy.

3.5.4. Computational effort

Table 5 reports the time taken in CPU minutes. The LMDP(Z_2) requires the least amount of time to solve. The LMDP(Z_3) requires the most time, occasionally reaching the 60-h limit. For larger instances, heuristics are needed to solve the LMDP; therefore we propose heuristics in Section 4.

4. Large scale LMDP instances

The previous section shows how route structure and metric values depend on the choice of objective function for the LMDP. Importantly, the results also show that the solution times for the LMDP(Z_3) grow quickly. This section proposes heuristics to solve larger instances of the LMDP. A greedy random adaptive search procedure (GRASP) framework is introduced in Section 4.1 to solve the LMDP(Z_1) and LMDP(Z_2). In Section 4.2, we present a new approach for LMDP(Z_3) since the

Table 5
Solution times in CPU minutes.

Model	Demand distribution	Time in CPU minutes									
		Location distribution									
		C	E	R1	R2	R3	C	E	R1	R2	R3
8-Node instances					10-Node instances						
LMDP(Z_1)	C	3	1	1	1	1	23	30	26	21	100
	E	1	2	1	1	1	12	12	19	12	20
	R1	1	14	6	6	1	24	40	23	88	31
	R2	720	10	1	6	1	722	50	57	66	57
	R3	1	1	1	2	1	21	20	27	35	23
LMDP(Z_2)	C	0	0	0	0	0	1	1	1	3	1
	E	2	0	0	0	0	5	23	6	3	3
	R1	0	0	0	0	0	2	2	2	1	1
	R2	0	0	0	0	1	21	2	2	3	3
	R3	0	0	0	0	0	1	2	1	1	1
LMDP(Z_3)	C	48	1974	10	130	13	117	3619	3541	1014	3614
	E	176	3613	40	223	97	3611	3607	3616	766	3620
	R1	15	191	138	982	840	32	193	1098	897	3618
	R2	6	369	153	2053	964	113	507	2275	1082	3615
	R3	62	1390	8	66	13	15	29	695	96	3631

methods used for the LMDP(Z_1) and LMDP(Z_2) cannot address the complexities of this model. Section 4.3 presents a computational study of the large instances.

4.1. Heuristics for the LMDP(Z_1) and LMDP(Z_2)

To solve the LMDP(Z_1) and LMDP(Z_2), we utilize a GRASP framework (Feo and Resende, 1995). GRASP is a metaheuristic in which a set number of repetitions is executed. Each repetition consists of a random constructive phase and a local improvement phase. In the constructive phase, a feasible solution is iteratively constructed. At each iteration, an insertion of a recipient into an existing route is selected randomly from a restricted candidate list of insertions which pass an evaluation function. The constructive phase is followed by an improvement phase where the constructive solution is improved via a deterministic local search to find a local optima. The best solution among all the repetitions is kept. Further details including a flowchart of the heuristic is provided in Huang et al. (2010).

4.2. Heuristic for the LMDP(Z_3)

The heuristic for the LMDP(Z_3) shares some similarities to the GRASP framework in that the heuristic employs a set number of repetitions. Each repetition is evaluated with respect to Z_3 and the best solution is kept. However, the constructive and improvement methods of the LMDP(Z_1) and LMDP(Z_2) do not adapt well to the LMDP(Z_3) due to factors described in the following paragraph. Therefore, we pattern a heuristic for the LMDP(Z_3) using an alternative two-phase approach to incorporate the goals of equity, efficiency and efficacy. A detailed description of the heuristic (particularly the 2nd phase) is available in Huang et al. (2010).

In the LMDP(Z_3), the dependency between routes prevents the decomposition of the problem by route; it is difficult to determine the contribution towards the Z_3 value from a single route. In addition, the value of Z_3 is very sensitive to the allocation decisions; the swapping of a single pallet from one delivery to another may have a large impact on the objective value. For these reasons, it is difficult to construct a solution or modify a solution to improve the objective as is done for the LMDP(Z_1) and LMDP(Z_2).

We draw on the insight that the optimal solution for the LMDP(Z_3) delivers a portion of demand to each node quickly and then distributes the remainder. The two-phase approach comprises of (1) an initial phase to deliver a fraction, β , of the demand to each node within a time τ and (2) a second phase to satisfy the remainder of the demand either by increasing the initial delivery amounts or by appending additional stops onto the routes. Note that the deliveries in the latter option are appended to the end of the routes rather than inserted into a route to ensure that the initial phase deliveries are performed before τ . Equity is incorporated by ensuring that each node receives a portion of its demand initially limiting the deviation in service level across nodes; efficacy and efficiency are incorporated via effective solutions for the two phases. The approach may be generalized to use node-specific β_i values to represent different urgency levels among the nodes.

The resulting solution is dependent on the parameter values of β and τ . A large β value and a small τ value correspond to serving most of the demand quickly; however, this is often infeasible. For a given β , if τ is too small, the instance is infeasible as it is not possible to serve β of the demand within τ . While one might expect that a large value of β would result in more equitable solutions, this is not always the case. A large value for β often requires a large value of τ to ensure feasibility. When τ is large, significant disparity in service levels may appear in the initial phase. A small value for β leaves opportunity for the

service level to deviate in the second phase. Fortunately, as shown in Section 4.3.4, we find that the two-phase approach is fairly robust with respect to the choice of β .

In the initial phase, to ensure feasibility, the value of β is set *a priori* and the value of τ is set after the initial phase. To deliver the initial demand, we perform one repetition of the LMDP(Z_2) heuristic on the graph $G = (\mathbf{N}^0, \mathbf{A})$ with reduced demands $\lceil \beta d_i \rceil$ for all $i \in \mathbf{N}$. The solution to the LMDP(Z_2) for the reduced demand problem produces feasible routes that quickly deliver the initial demand; τ is set to be the latest arrival time in this solution. Using the example from Section 3.3.4, Fig. 6a shows the solution after the initial phase when $\beta = 0.75$. This solution yields a value of $\tau = 6$.

The second phase is initialized with routes and deliveries from the initial phase; node i begins with an unsatisfied demand equal to $d_i - \lceil \beta d_i \rceil$. We refer to this demand as “unassigned demand” since these units are not associated with routes at the start of the second phase. The heuristic designates each vehicle as capacity feasible or capacity infeasible based on a conservative estimate, which calculates the load on the vehicle plus the load from unassigned demand of nodes visited by the vehicle minus the vehicle capacity. A positive value represents a capacity infeasible vehicle and a negative value represents a capacity feasible vehicle. Delivery amounts are increased by $d_i - \lceil \beta d_i \rceil$ for all nodes on capacity feasible vehicles. For nodes visited by multiple vehicles, the heuristic increases the delivery amount to the node on the first vehicle found; for other vehicles, the delivery amount is not increased.

The heuristic then iteratively serves the unassigned demand. In each iteration, for each node with unassigned demand, the heuristic finds the cheapest way (in terms of travel time) to append the node to a capacity feasible vehicle. The cheapest option among all such nodes is executed. The feasibility status is updated for vehicles: if a vehicle becomes capacity feasible (i.e., the remaining nodes with unassigned demand on that route may be now served by the vehicle), the delivery amounts are increased to fully satisfy demand. The loop is repeated until all demand is assigned. Like its LMDP(Z_1) and LMDP(Z_2) counterparts, the heuristic runs for 5000 iterations and the best solution retained.

Fig. 6b shows the final solution for the example using this approach. The route used by vehicle 1 is extended to deliver one pallet to nodes 2 and 3. The amount delivered to node 1 by vehicle 1 is increased to four pallets. The route used by vehicle 2 does not change. In this example, the solution is the optimal LMDP(Z_3) solution.

4.3. Computational study

In Section 4.3.1, we evaluate the ability of the heuristic to achieve high quality solutions for the small instances. For the large scale experiments, instances developed from the 100-node test cases from (Solomon, 1987) for VRP with Time Windows (VRPTW) (available at <http://neo.lcc.uma.es/radi-aeb/WebVRP/>) are used. In the original data set, there are four unique location configurations; instances differ by demands and time windows. Configurations C1 and C2 correspond to data sets with clustered locations; configuration R has random locations and configuration RC is a combination of random and clustered locations. For each C1 and C2 configuration, 10 instances are created by assigning nodes a random demand between 1 and 5; 20 instances are created for each R and RC configuration.

4.3.1. Performance against optimal restricted solutions

We compare the heuristic solutions against the route-length restricted optimal solutions for the LMDP(Z_1), LMDP(Z_2) and LMDP(Z_3) for the 8-node and 10-node instances. Overall the heuristics perform very well. The LMDP(Z_1) heuristic matches or performs better than the solver solution in all but one instance with a gap of 3.8%. Because the heuristics are not restricted to routes that visit at most five nodes, the LMDP(Z_1) heuristic does on occasion result in lower travel times than the solution found with CPLEX. The LMDP(Z_2) heuristic finds optimal solutions for every instance. The LMDP(Z_3) heuristic achieves the optimal Z_3 value in all but two of the 10-node instances; in those cases, the gaps are 4.1% and 1.2%. Run times for the heuristics for the 8-node and 10-node instances for all models are less than a second of CPU time.

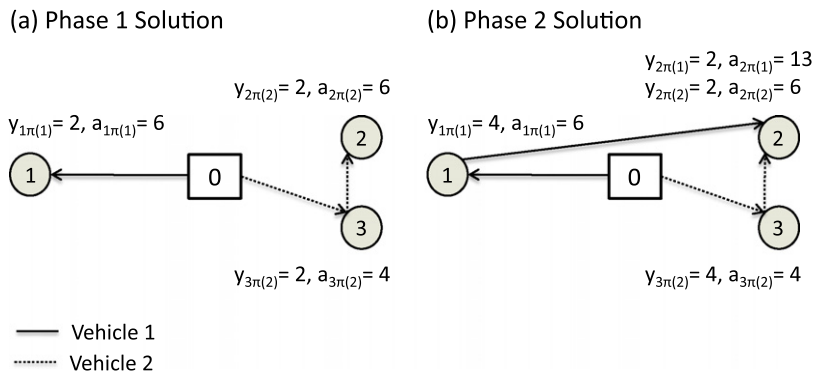


Fig. 6. Solution using two-phase approach with $\beta = 0.75$.

Table 6
Summary of deviation in metric values for LMDP heuristics for large instances.

Heuristic for		Gap by metric				
		Δ_1 (%)	Δ_2 (%)	Δ_3 (%)	Δ_3^a (%)	Δ_3^b (%)
LMDP(Z_1)	Avg	0	27	31	63	37
	Max	0	41	43	115	63
	Min	0	15	14	6	9
LMDP(Z_2)	Avg	16	1	4	8	4
	Max	26	3	9	50	21
	Min	9	0	1	0	0
LMDP(Z_3)	Avg	16	1	0	1	0
	Max	26	11	0	18	5
	Min	6	0	0	0	0

4.3.2. Metric performance for large instances

Since there is no guarantee of optimality in the heuristic solutions, in the large instances, we evaluate the solutions with respect to each metric by calculating the percent difference between the metric value for that solution from the best solution for that metric among the three heuristic solutions. Table 6 aggregates the 60 large instances and presents the percent difference between each heuristic and the best solution among the heuristics. Results for each location configuration is provided in Huang et al. (2010), but the three location types do not exhibit a large difference; hence, results presented here are aggregated over the 60 instances.

While it is intuitive that each heuristic solution should perform its best on the model it attempts to solve, the LMDP(Z_3) heuristic solution often outperforms the LMDP(Z_2) heuristic solution with respect to Z_2 . The trends in relative metric values are similar to the observations seen in Section 3.5.3 in that the metric values for the solutions for the LMDP(Z_2) model and LMDP(Z_3) model are similar to each other and differ more greatly with the solutions for the LMDP(Z_1); furthermore, while the LMDP(Z_2) and LMDP(Z_3) exhibit similar solutions, the LMDP(Z_3) in general yields solutions that are more equitable.

Averaging over all the instances, for the Z_2 and Z_3 metrics, the LMDP(Z_1) solutions average 27% and 31% respectively higher than the best heuristic solution for the respective metrics. With regards to the Z_1 metric, the heuristic LMDP(Z_2) and LMDP(Z_3) solutions both average 16% higher than the best heuristic Z_1 solutions attained by the LMDP(Z_1) heuristic. With respect to the equity metric Z_3^b , the optimal LMDP(Z_2) solutions are on average 4% higher than the best heuristic solution while the LMDP(Z_3) solutions are on average less than 1% higher.

To examine the effect of excess capacity, we alter the number of vehicles and study its effect. Let M denote the minimum number of vehicles necessary for feasibility ($M = \lceil \frac{1}{c} \sum_i d_i \rceil$). Table 7 presents the percentage differences with M , $M + 1$, $M + 2$, and $M + 3$ vehicles.

As the number of available vehicles increases, the percent difference between the heuristic LMDP(Z_1) solution and the best Z_2 and Z_3 heuristic solution increases. This occurs because it is preferable to use the minimum number of vehicles in the LMDP(Z_1), while the opposite is true for the LMDP(Z_2) and LMDP(Z_3). We also note that as more vehicles become available, the LMDP(Z_2) and LMDP(Z_3) metric values become closer; the extra capacity of the vehicles allows for greater parallel delivery reducing the impact of node demand.

4.3.3. Route structure for large instances

Unlike the small instances investigated in Section 3.5, the solutions in the large instances exhibit demand splitting among multiple vehicles. The choice of model also influences the degree to which visits to a node are split among multiple vehicles. For instance, the LMDP(Z_3) solution should have a higher number of nodes visited by multiple vehicles due to splitting for both capacity and equity reasons.

Table 7
Average deviation in metric values for LMDP model heuristics solutions with different numbers of vehicles.

Heuristic for	$K = M$					$K = M + 1$				
	Δ_1 (%)	Δ_2 (%)	Δ_3 (%)	Δ_3^a (%)	Δ_3^b (%)	Δ_1 (%)	Δ_2 (%)	Δ_3 (%)	Δ_3^a (%)	Δ_3^b (%)
LMDP(Z_1)	0	24	29	55	33	0	27	31	63	37
LMDP(Z_2)	19	4	11	34	18	16	1	4	8	4
LMDP(Z_3)	15	0	0	1	0	16	1	0	1	0
Heuristic for	$K = M + 2$					$K = M + 3$				
	Δ_1 (%)	Δ_2 (%)	Δ_3 (%)	Δ_3^a (%)	Δ_3^b (%)	Δ_1 (%)	Δ_2 (%)	Δ_3 (%)	Δ_3^a (%)	Δ_3^b (%)
LMDP(Z_1)	0	32	35	73	44	0	35	38	79	51
LMDP(Z_2)	20	1	3	7	3	23	0	3	7	3
LMDP(Z_3)	19	1	0	1	1	23	0	0	1	0

Table 8
Average number of split deliveries in 100-node instances.

Heuristic for	Average number of nodes visited by:		
	1 vehicle	2 vehicles	3 + vehicles
LMDP(Z_1)	93.38	6.60	0.02
LMDP(Z_2)	90.65	8.87	0.48
LMDP(Z_3)	89.52	9.73	0.75

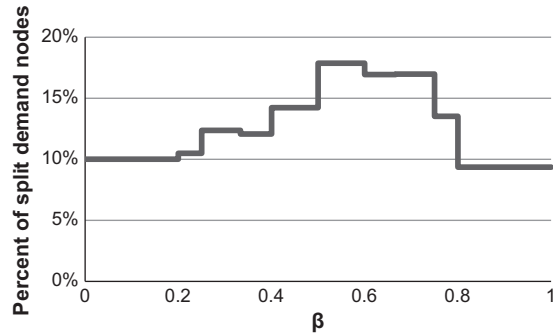


Fig. 7. Percentage of split-demand nodes relative to β .

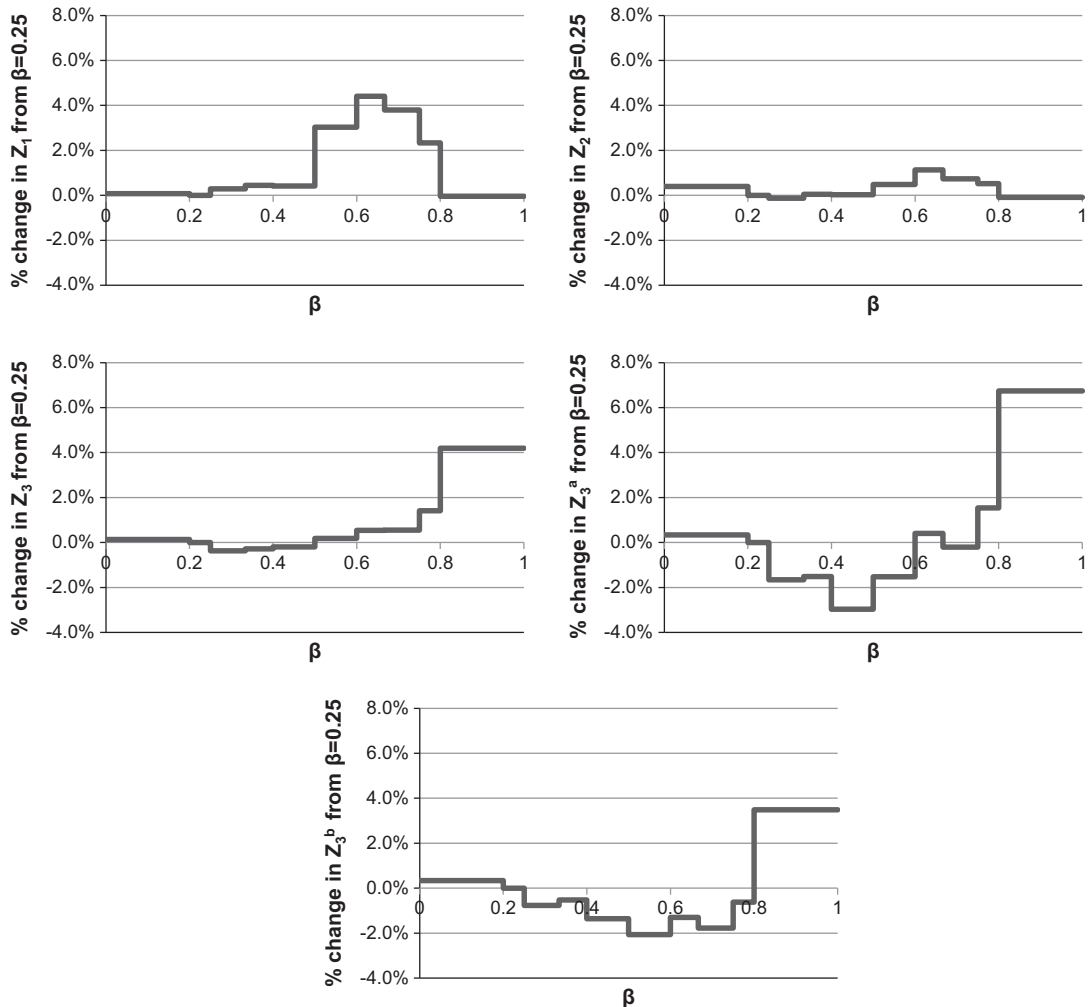


Fig. 8. Percent change of heuristic LMDP(Z_3) metric values from base case relative to β .

Table 8 shows the number of vehicles visiting each node for each of the heuristic solutions to the LMDP(Z_1), LMDP(Z_2) and LMDP(Z_3). As expected, the heuristic for the LMDP(Z_3) exhibits more splitting than the other two models albeit by a slim margin. Further, the splitting is not excessive as the large majority of the nodes are served by only one vehicle. The emergence of splitting in the solutions impacts route structure; because the LMDP(Z_3) heuristic attempts to deliver to each node a fraction of its demand, some routes must double back to deliver to nodes whose demand cannot be satisfied before time τ . As a result, the ray-like structure seen in the routes of the LMDP(Z_3) for small problems is no longer seen in the larger cases.

4.3.4. Choice of β for LMDP(Z_3) Heuristic

This section investigates how the value of β , the fraction of demand that is served in the initial phase of the LMDP(Z_3) heuristic, impacts the LMDP(Z_3) solutions. Specifically, we examine how the choice of β affects the degree of splitting and metric values of the LMDP(Z_3) solutions.

Fig. 7 shows the effect of β on splitting. The curve shows the percentage of nodes that are visited by multiple vehicles in the LMDP(Z_3) solutions with respect to β aggregated across the 60 large instances. The discrete nature of the curve is due to the ceiling function when delivering $\lceil \beta d_i \rceil$.

Splitting is minimal when β is close to 0 or close to 1. When $\beta = 0$, the LMDP(Z_3) heuristic reduces to its second phase in which splitting for equity does not occur. At this point 10% of the nodes are visited by multiple vehicles due to vehicle capacity. As β increases, the number of split nodes increases peaking at 18% in the interval between $\beta = 0.5$ and $\beta = 0.6$. As β increases further, the amount of nodes split drops off rapidly. When $\beta = 1$, the LMDP(Z_3) heuristic reduces to the initial phase of the heuristic and thus is equivalent to the LMDP(Z_2) heuristic; therefore no splitting is done for equity reasons.

Fig. 8 shows how the metric values of the LMDP(Z_3) change as β changes. For each metric, we plot the percent difference in metric value with respect to β compared to the metric value when $\beta = 0.25$, averaged over all 60 instances.

Overall, the choice of β does not have a large impact on the metric values of solutions. With the exception of the Z_3^a metric, the values do not change by more than 5%. However, small trends are observed. The Z_1 and Z_2 metric values increase (decrease) with increases (decreases) in node splitting observed in Fig. 7. As more nodes are visited by multiple vehicles to improve equity, the distance traveled and total wait times increase. Conversely, the metric values for the equity metrics Z_3^a , Z_3^b and Z_3 improve when β takes values that encourage more splitting. Combining these results and factoring the rate of change in the curves in Fig. 8, a small value for β achieves good improvements in efficiency and efficacy, with a modest loss of equity.

One can set β considering the tradeoffs among metrics and characteristics of the relief setting. For instance, if efficiency and efficacy are emphasized, Fig. 8 suggests a low value of β is preferable; if equity is more of a concern, a higher value of β may be more appropriate.

5. Concluding remarks

This paper formulates efficacy and equity metrics and shows that there is a significant difference between solutions that focus on both efficacy and equity and solutions that focus on the traditional commercial concern of efficiency. In the following sections, we discuss the implication of our findings.

5.1. Practical implications

The results from this paper provide several insights for relief operations. In practice, relief organizations tend to utilize a vehicle fleet consisting of many small vehicles, unlike in commercial settings where large vehicles are available and may be preferable to achieve economies of scale. The organizations acquire transportation resources locally as maintaining and shipping large vehicles long distances, especially overseas, can be very expensive and time consuming. Further, local drivers are more knowledgeable of the regions' geography and customs. Characteristics of the post-disaster transportation network affect vehicle selection decisions; for instance, remote locations in a rural area may only be accessed by small trucks. Observation 1 suggests that operating a large number of small vehicles has the additional benefit of resulting in more effective and equitable aid distribution at the expense of a modest increase to traveling and operating costs (e.g., costs of fuel, drivers, etc.).

While having a large fleet of small vehicles is beneficial in improving the efficacy and equity of relief deliveries, operating and coordinating a large fleet might be difficult for relief logisticians in the field. For instance, the logistics coordinator of the Finnish Red Cross was responsible for coordinating a fleet of 150 trucks from a single warehouse during the relief operations after the Pakistan earthquake in 2005 (Mantyaara, 2010). While there is an increasing effort to develop and improve the logistics software in the relief sector, the current software is mostly used for tracking, monitoring and reporting purposes and lacks modules that would support operational decisions. Indeed, routing and delivery scheduling decisions are made mostly according to the insights and experiences of the logisticians. Our heuristics described in Section 4 are efficient and easy to understand, and hence could be a valuable decision support tool to manage large scale last mile operations, if embedded in such software. Observations 2 and 3 focusing on route shape dependent on objective and node demand represent a first step to developing even simpler rules-of-thumb so that a coordinator may make decisions without advanced computing resources.

In humanitarian relief environments, there may be large differences between the needs of populations at different locations. Although Formulation (1) does not explicitly differentiate populations in terms of their vulnerabilities, our heuristic

approach can easily incorporate the criticality and urgency of the needs at different locations. Specifically, relief supply deliveries can be prioritized by modifying β_i and τ values so that vehicles can reach more vulnerable locations first.

While setting the β values, it is also important to evaluate the characteristics of the relief network. As discussed earlier, while splitting may help achieve more equitable and effective solutions, it also increases travel costs. Moreover, in regions where the transportation network is damaged and security is a concern, or when access to certain locations is problematic for other reasons such as poor weather conditions, relief organizations may prefer satisfying the demand of a location in a single delivery if possible. In such cases, a small β value leading to few splits may be more reasonable.

5.2. Future work

One avenue of extension to this work is to analyze the LMDP variations under uncertainty, both in relaxing demands and travel times. In the chaotic environment that follows the immediate aftermath of a disaster there is significant uncertainty. For instance, demand is not immediately known and can only be roughly estimated. Also, travel times between two locations can be uncertain due to damaged infrastructure and unreliable information regarding the road conditions. Therefore, insights into the behavior of efficiency, efficacy, and equity under a stochastic setting would be valuable.

Another avenue of focus is to use the insights gained to generate rule-of-thumb solutions that are easy to find and simple to implement. Because relief agencies often lack technology and computational resources, solutions which may be implemented by hand would be valuable. Rule of thumb solutions would also be quick to calculate and in planning stages could serve as an approximation for costs until more sophisticated solutions are found. Our analysis of stylized distribution problems is an important first step in this area.

To gather insight on how to address concerns of efficiency, efficacy and equity, the LMDP has been separated into three variations. However, in practice, agencies must handle these concerns concurrently. Therefore, it may be important to consider a multi-objective variation of the LMDP which combines the three factors into a single objective.

Acknowledgements

This research has been supported by the National Science Foundation, CMMI 0348622 and CMMI 0654398, as well as the Northwestern University Transportation Center. We also would like to thank the many relief organizations that provided insights for the paper, including Ari Mantyvaara.

Appendix A. Proofs of the analytical properties

Property 3.2. The result from Property 3.1 holds for LMDP(Z_2).

Proof. Properties 3.1 and 3.2 share a similar argument: for every solution in which two nodes i and j are visited by two vehicles 1 and 2, we show that there exists a solution in which only one of the nodes is visited by both vehicles that performs, if not better, equally as well.

Let $y_{i\pi}(k)$ be the amount delivered to node i by vehicle k and $a_{i\pi}(k)$ to be the arrival time of the delivery. Without loss of generality, define node i and vehicle 1 so that $y_{i\pi(1)} = \min \{y_{i\pi(1)}, y_{i\pi(2)}, y_{j\pi(1)}, y_{j\pi(2)}\}$. Consider the following three cases and their corresponding reallocation of the delivery amounts such that there exists only one route that serves both nodes:

Case 1: $(a_{i\pi(1)} - a_{i\pi(2)}) \geq (a_{i\pi(1)} - a_{i\pi(2)})$.

In this case, we use the following reallocation:

$$y_{i\pi(1)}^{new} = 0 \quad y_{i\pi(2)}^{new} = y_{i\pi(2)} + y_{i\pi(1)} \quad y_{i\pi(1)}^{new} = y_{i\pi(1)} + y_{i\pi(1)} \quad y_{j\pi(2)}^{new} = y_{j\pi(2)} - y_{i\pi(1)}$$

For each nodes i and j , the reallocation delivers the same amount as the initial allocation. Also, each vehicle devotes the same amount of capacity. The new allocation is an improvement over the old one:

$$Z_2(x, y) - Z_2(x, y^{new}) = a_{i\pi(1)}y_{i\pi(1)} - a_{i\pi(2)}y_{i\pi(1)} - a_{j\pi(1)}y_{i\pi(1)} + a_{j\pi(2)}y_{i\pi(1)} \geq 0$$

Case 2: $(a_{i\pi(1)} - a_{i\pi(2)}) < (a_{i\pi(1)} - a_{i\pi(2)})$ and $y_{i\pi(2)} \geq y_{i\pi(1)}$

We use the following reallocation:

$$y_{i\pi(1)}^{new} = y_{i\pi(1)} + y_{j\pi(1)} \quad y_{i\pi(2)}^{new} = y_{i\pi(2)} - y_{j\pi(1)} \quad y_{j\pi(1)}^{new} = 0 \quad y_{j\pi(2)}^{new} = y_{j\pi(2)} + y_{j\pi(1)}$$

$$Z_2(x, y) - Z_2(x, y^{new}) = -a_{i\pi(1)}y_{j\pi(1)} + a_{i\pi(2)}y_{j\pi(1)} + a_{j\pi(1)}y_{j\pi(1)} + a_{j\pi(2)}y_{i\pi(1)} > 0$$

Case 3: $(a_{i\pi(1)} - a_{i\pi(2)}) < (a_{i\pi(1)} - a_{i\pi(2)})$ and $y_{i\pi(2)} < y_{i\pi(1)}$

$$y_{i\pi(1)}^{new} = y_{i\pi(1)} + y_{i\pi(2)} \quad y_{i\pi(2)}^{new} = 0 \quad y_{j\pi(1)}^{new} = y_{j\pi(1)} - y_{i\pi(2)} \quad y_{j\pi(2)}^{new} = y_{j\pi(2)} + y_{i\pi(2)}$$

$$Z_2(x, y) - Z_2(x, y^{new}) = -a_{i\pi(1)}y_{j\pi(2)} + a_{i\pi(2)}y_{j\pi(1)} + a_{j\pi(2)}y_{j\pi(2)} + a_{j\pi(2)}y_{i\pi(2)} > 0$$

In each of the cases, the Z_2 value is maintained or improved by a reallocation that sets the amount delivered by a vehicle to a node to 0. The solution can be further improved by deleting the visit. Under the triangle inequality, the arrival time of all nodes that were originally visited after the deleted node on the same route are reduced. Thus there exists an equal, if not a better, solution where only one of the nodes is visited by both vehicles. \square

Property 3.3. The result from Property 3.1 does not hold for $\text{LMDP}(Z_3)$.

Proof. Fig. 4d is a counterexample. Nodes 2 and 3 are visited by both routes. The demand at node 3 is split because vehicle 2 withholds one pallet so that node 2 may receive the pallet instead at time 6. Due to the convex disutility function in Z_3 , serving some demand early at all nodes is more important. \square

Property 3.4. For each of the objective functions of the LMDP based on Z_1 , Z_2 and Z_3 , there exists an optimal solution where each route is run at most once.

Proof. Given that $d_i \leq C \forall i \in N$, demand at a node can be fully satisfied through a direct delivery route. Therefore, a direct delivery route is taken only once. To prove the property for more general cases, we focus on routes that visit more than one node. Suppose there are two vehicles that run the same route that visits nodes i and j . However, such a route would contradict Properties 3.1 and 3.2, which imply that for optimal solutions for $\text{LMDP}(Z_1)$ and $\text{LMDP}(Z_2)$ any subset of recipient nodes with cardinality greater than 1 is served by at most one vehicle. Therefore, a particular route is taken at most once under Z_1 and Z_2 . For Z_3 , consider a reallocation used in the first case of the proof to Property 3.2. Because the vehicles are performing the same route, the arrival time at the nodes are the same for the two vehicles. Therefore, skipping a node only serves demand earlier and does not result in the demand of another node being served later. Hence the new route improves Z_3 . \square

References

- Archetti, C., Savelsbergh, M., Speranza, M., 2006. Worst-case analysis for split delivery vehicle routing problems. *Transportation Science* 40 (2), 226–234.
- Archetti, C., Savelsbergh, M., Speranza, M., 2008. To split or not to split: that is the question. *Transportation Research Part E* 44, 114–123.
- Balcik, B., Beamon, B., Smilowitz, K., 2008. Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems* 12 (2), 51–63.
- Balcik, B., Irvani, S., Smilowitz, K., 2011. A Review of Equity in Nonprofit and Public Sector: A Vehicle Routing Perspective. *Encyclopedia of Operations Research and Management Science*.
- Barbarosoglu, G., Arda, Y., 2004. A two-stage stochastic programming framework for transportation planning in disaster response. *Journal of the Operational Research Society* 55 (1), 43–53.
- Barbarosoglu, G., Ozdamar, L., Cevik, A., 2002. An interactive approach for hierarchical analysis of helicopter. *European Journal of Operational Research* 140 (1), 118–133.
- Beamon, B., Balcik, B., 2008. Performance measurement in humanitarian relief chains. *International Journal of Public Sector Management* 21 (1), 4–25.
- Campbell, A., Vandenbussche, D., Hermann, W., 2008. Routing for relief efforts. *Transportation Science* 42 (2), 127–145.
- Daganzo, C., 2005. *Logistics Systems Analysis*. Springer, Berlin.
- Dror, M., Trudeau, G., 1989. Savings by split delivery routing. *Transportation Science* 23, 141–145.
- Dror, M., Trudeau, G., 1990. Split delivery routing. *Naval Research Logistics* 37, 383–402.
- Erkut, E., 1993. Inequality measures for location problems. *Location Science* 1 (3), 199–217.
- Feo, T., Resende, M., 1995. Greedy randomized adaptive search procedures. *Journal of Global Optimization* 6, 109–133.
- Haghani, A., Oh, S., 1996. Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations. *Transportation Research Part A* 30 (3), 231–250.
- Huang, M., Smilowitz, K., Balcik, B., 2010. Models for Relief Routing: Equity, Efficiency and Efficacy (Expanded Version). <http://www.iems.northwestern.edu/docs/working_papers/michael_huang_dec10.pdf>.
- ILOG, 2008. ILOG CPLEX 11.2 User's Manual.
- Knott, R., 1987. The logistics of bulk relief supplies. *Disasters* 11 (2), 113–115.
- Lin, Y.-H., Batta, R., Rogerson, P., Blatt, A., Flanigan, M., 2009. A logistics model for delivery of prioritized items: application to a disaster relief effort. Department of Industrial and Systems Engineering, University at Buffalo (SUNY), Buffalo, Working Paper.
- Mantyyvaara, A., 2010. Logistics Coordinator. Personal Communication.
- Marsh, M., Schilling, D., 1994. Equity measurement in facility location analysis: a review and framework. *European Journal of Operational Research* 74 (1), 1–17.
- Mulligan, G., 1991. Equality measures and facility location. *Papers in Regional Science* 70 (4), 345–365.
- Ozdamar, L., Ekin, E., Kucukyazici, B., 2004. Emergency logistics planning in natural disasters. *Annals of Operations Research* 129, 217–245.
- Russell, T., 2005. The Humanitarian Relief Supply Chain: Analysis of the 2004 South East Asia Earthquake and Tsunami. Massachusetts Institute of Technology.
- Shen, Z., Dessouky, M., Ordóñez, F., 2009. A two-stage vehicle routing model for large-scale bioterrorism emergencies. *Networks* 54, 255–269.
- Solomon, M., 1987. Algorithms for the vehicle routing and scheduling problem with time windows. *Operations Research* 35, 254–265.
- Tzeng, G., Cheng, H., Huang, T., 2007. Multi-objective optimal planning for designing relief delivery systems. *Transportation Research Part E* 43 (6), 673–686.
- Van Hentenryck, P., Bent, R., Coffrin, C., 2010. Strategic planning for disaster recovery with stochastic last mile distribution. In: Lodi, A., Milano, M., Toth, P. (Eds.), *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Springer, Berlin/Heidelberg, pp. 318–333.
- Van Wassenhove, L., 2006. Blackett Memorial Lecture Humanitarian Aid Logistics: supply chain management in high gear. *Journal of the Operational Research Society* 57, 475–489.
- Yi, W., Ozdamar, L., 2007. A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research* 179, 1177–1193.