Scalable Robust and Adaptive Inventory Routing

Dimitris Bertsimas, Swati Gupta, Joel Tay
Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139
dbertsim@mit.edu, swatig@mit.edu, joeltay@mit.edu

We consider the finite horizon inventory routing problem with uncertain demand, where a supplier must deliver a particular commodity to its customers periodically, such that even under uncertain demand the customers do not stock out, e.g. supplying residential heating oil to customers. Current techniques that solve this problem with stochastic demand, robust or adaptive optimization do not scale to real-world data sizes, with the status quo being only able to perform inventory routing for ~100 customers. We propose a scalable approach to solving a robust and adaptive mixed integer optimization formulation that is made tractable with algorithms for generating worst-case demand vectors, heuristic route selection, warm starts and column generation. We demonstrate experimentally a mean reduction in stockouts of over 94% in our robust and adaptive formulations, translating to a cost savings of over 14%. We also show how to modify our model to achieve further cost savings through fleet size reduction. Our robust and adaptive formulations are tractable for ~6000 customers.

Key words: robust optimization, vehicle routing, inventory routing, stock-out, demand, uncertainty

1. Introduction

We consider the rich problem of inventory routing where a supplier has a contract with individual customers to monitor their inventory of a commodity that diminishes over time, and to resupply that commodity to maintain customer stocks above a certain threshold. Some sizeable industries concerned with inventory routing problems of this type are those supplying commodities such as soft drinks in vending machines, portable water in offices, or heating oil in residential areas. In many of these inventory routing applications, the presence of uncertainty in the customers’ demand for the commodity (and other uncertainties in data, e.g. temperature in heating oil usage models) is a critical issue that must be addressed in order to provide solutions that are of practical value in the real world. In this paper, we provide novel scalable and adaptive algorithms to address the inventory routing problem using a robust and adaptive optimization framework.

Given a network of customers spread over a geographic area, the supplier needs to make the following operational decisions:

- **Fleet size**: the supplier needs to decide ahead of the operational period the number of vehicles to be maintained and the crew size required. More crew size and vehicles increase the cost of operation, whereas a reduction in these may impact the quality of service negatively, and require a larger emergency fleet to handle stockouts.
• **Routes and Schedules**: the supplier needs to determine which routes to utilize to visit customers, and when to schedule these routes, while minimizing their cost of operation (thus maximizing their profits).

• **Refuelling quantities**: when a customer is visited, the supplier needs to determine how much of the commodity to resupply. Attempting to resupply all customers to their full capacity might not be feasible for the vehicles’ capacity, or it might limit the number of customers that a vehicle can resupply.

Having defined our key operational decisions, we now consider the key objectives that a supplier is concerned with, namely: (i) **reducing the frequency of stockouts** and (ii) **minimizing the cost of operations**. Reducing the frequency of stockouts is important, as, besides the obvious damage to brand image that results from customers’ stocks being depleted, it is also highly undesirable for suppliers because they have to designate vehicles to make unplanned emergency replenishments of these customers, often at very short notice. Regarding operational cost, much of the inventory routing literature (e.g. [Irnich et al. (2014)](#)) has focused on minimizing the routing cost while maintaining a desired level of service. However, we argue that given the advances in consumer-grade vehicle routing heuristics and technology, reducing the fleet size while ensuring quality of service yields more significant savings. It is thus natural to seek solutions with reduced fleet sizes that are robust to uncertainties in the rate of customers’ demand for the commodity, while minimizing the routing costs.

An important reason that many approaches to this problem do not scale well is that they attempt to solve for the optimal routes. As this requires solving the Travelling Salesman Problem as a subproblem, it becomes difficult to use these approaches to solve problems of the sizes required in real-world applications. Furthermore, as the capital, maintenance and labor cost of the vehicle fleet usually outweigh the fuel cost, the most significant improvement comes from being able to select a schedule for customers that requires fewer vehicles, subject to quality of service requirements.

Current approaches in the literature ([Solyali et al. (2012)](#), [Aghezzaf (2007)](#)) solve only up to around a hundred customers and do not scale to problem sizes that arise in the real world. Our main application throughout the paper is to companies that provide heating oil in residential areas. For example, a typical company of this nature in New England might have a customer base spanning north central Massachusetts and southern New Hampshire with around 10,000 customers. Our key contribution is a robust and adaptive mixed integer optimization (MIO) formulation that scales to large problem sizes, augmented with a demand uncertainty set that varies with temperature and dynamic route generation. Using data sets generated from real temperature data, we demonstrate both the effectiveness and scalability of our approach.
The rate of demand of these commodities has typically been considered in the literature (e.g. Chepuri and Homem-De-Mello (2005), or the survey of Gendreau et al. (1996)) to be either (dynamically) deterministic or stochastic. A deterministic rate of demand, as with many optimization problems, leads to more tractable but less realistic models. A stochastic rate of demand, however, is less tractable for large instances and often leads to heuristic solutions which are sensitive to the assumptions made about the probability distribution of the demand. In contrast, a robust optimization approach combines the tractability of deterministic models with the realism of stochastic approaches by modeling uncertainty in a deterministic manner, and leads to solutions that are less sensitive to the probabilistic assumptions made about the underlying demand.

Our contributions in this section can be summarized as follows:

1) **Robustness.** We present a robust formulation of the uncertainty set for demand that captures, for the case of resupplying heating oil, the dependence of demand on temperature as well as individual customers’ rates of consumption. This results in a novel non-convex uncertainty set which we are able to tractably optimize over, thus generating the critical worst-case demand scenarios.

2) **Adaptability.** For the case where customer demand can be recorded remotely, we present an approach that allows us to adapt our operational decisions according to observed demand. We demonstrate computationally that the adaptive solutions outperform both the deterministic and robust formulations.

3) **Scalability.** By combining novel ways to generate the critical worst-case demand scenarios, automated neighborhood route selection, tabu-search heuristics and generating constraints on the fly for the adaptive formulation, we are able to solve problems with \(~6000\) customers over a time horizon of 151 days, within two hours for both the robust and adaptive formulations.

4) **Quality of solutions.** We demonstrate that the robust solutions of our model materially decrease stockouts and are relatively insensitive to estimation noise in demand and temperature, achieving across a variety of data sets of sizes ranging from 51 to 5915, an average reduction in stockouts of over 94% from a deterministic model. We show that both the robust and adaptive formulations can be used to reduce vehicle fleet size, while still outperforming the deterministic solution. Finally, we demonstrate that the robust and adaptive solutions lead to a decrease in total operational cost for the supplier, when combining routing cost with vehicle fleet cost and the cost of resupplying customers who experience stockouts.

The remainder of this paper is structured as follows: in Section 2, we survey some of the related literature and discuss why current approaches do not scale well. In Section 3, we introduce deterministic, robust and adaptive models for the capacitated inventory routing problem. We define our uncertainty set, and provide an algorithm that maximizes affine functions of demand over the uncertainty set. In Section 4, we discuss our techniques for route generation and some heuristics...
to further improve our routes. We detail our experiments and computational results in Section 5. We finally conclude with overall discussion and some future directions in Section 6.

2. Related Work

Vehicle routing problems (VRPs) arise naturally from many problem contexts, and as such have been extensively studied in many flavors. Beginning with “The Truck Dispatching Problem”, proposed by Dantzig and Ramser (1959), the difficulty of these problems and their relevance to many industries have generated much research over the past few decades.

One of the best-studied formulations of vehicle routing problems is the capacitated vehicle routing problem (CVRP), which in its most basic form describes the problem of determining a minimum-cost set of routes by which a fleet of delivery vehicles with limited capacity delivers quantities of a product or commodity to customers at various locations. When the costs of potential routes and the customer demands are assumed to be fixed and known, this is a deterministic problem. Early approaches for getting exact solutions of the CVRP were for decades dominated by branch-and-bound algorithms (e.g. Christofides and Eilon (1969), Christofides et al. (1981), Laporte et al. (1986)); in addition, branch-and-cut algorithms were later developed with many different families of cuts (Laporte et al. (1985), Augerat (1995), Ralphs et al. (2003), Lysgaard et al. (2004), Baldacci et al. (2004)), often building on research on the Travelling Salesman Problem. More recently, another popular approach is to solve the problem using column generation alongside cut generation (e.g. Fukasawa et al. (2006), Baldacci et al. (2008), Pecin et al. (2014)). We refer the reader to Cordeau et al. (2006), Golden et al. (2008), Laporte (2009), Baldacci et al. (2010), Coelho et al. (2013), and Toth and Vigo (2014) for detailed literature surveys about the CVRP and related vehicle routing problems.

However, the solutions to deterministic VRPs can be sensitive to errors or uncertainties in the parameters of the problem, becoming suboptimal or even infeasible for real-world actualizations. This has typically been addressed by taking the uncertain parameters as random variables, and utilizing stochastic programming to formulate the model. Assuming a known probability distribution for the uncertain parameters, probabilistic guarantees can then be made (e.g. a chance-constrained VRP). (More generally, the field of stochastic programming is described in much greater detail in Birge and Louveaux (2011) and Shapiro et al. (2014), just to give two examples.) However, stochastic VRPs are much harder to solve than their deterministic counterparts (Dror et al. (1989)). Developing exact algorithms that solve these problems to optimality has been challenging for problems of any realistic size, and much work has been done on heuristics (for a detailed survey, Toth and Vigo (2001)), and recently, metaheuristics (Toth and Vigo (2014)) that work well on VRPs.

One paradigm that has proven useful in approaching problems modeling optimization under uncertainty is Robust Optimization (RO) (for instance Bertsimas and Sim (2003), Ben-Tal et al. (2013)).
This approach leads to solutions that are guaranteed to satisfy the constraints for all uncertain parameters in a chosen uncertainty set, and often leads to tractable models requiring weaker assumptions on the uncertain parameters than stochastic formulations. RO formulations have been found in practice to yield solutions that are competitive with the optimal deterministic solution, and perform significantly better in worst-case scenarios. They also tend to be less affected by errors in parameter estimation or structural assumptions (Goldfarb and Iyengar (2003), Bertsimas and Sim (2004)).

While demand uncertainty has long been considered in its stochastic form (Bertsimas (1992), Bertsimas and Simchi-Levi (1996), Gendreau et al. (1996)), recent works have proven the usefulness of RO in formulating certain varieties of VRPs (Ordóñez (2010)). For instance, Sungur et al. (2008) consider a formulation of the single-stage Robust Capacitated VRP (RCVRP) under demand uncertainty that can be solved deterministically, using the budget-of-uncertainty approach first developed in Bertsimas and Sim (2003), and Gounaris et al. (2013) consider the RCVRP with more general demand uncertainty that can be reformulated to yield numerical solutions.

Solyali et al. (2012) and Aghezzaf (2007) have previously addressed the inventory routing problem within a RO framework. Solyali et al. (2012) report solving instances with a branch-and-cut algorithm, solving a Travelling Salesman problem as a subproblem exactly, for up to 30 customers and a time horizon of seven periods. Aghezzaf (2007) uses a heuristic approach to generate routes, proposing a nonlinear MIO problem, and reports solving for cyclic distribution routes for 50 customers. In contrast, our methods allow us to solve problems with the number of customers two orders of magnitude larger than both of these, over a time horizon which is an order of magnitude larger than Solyali et al. (2012), by solving a deterministic MIO to generate robust solutions, and using a cutting-plane algorithm to generate adaptive solutions.

Finally, we consider the problem of formulating an adaptive multistage robust optimization model. While the fully adaptive robust optimization problem is intractable via a dynamic programming approach, affinely-adaptive robust optimization solutions have been found to perform almost as well, while retaining the tractability of single-stage robust optimization problems (Ben-Tal et al. (2004), Bertsimas et al. (2010)). This approach has recently been applied to the unit commitment problem in power generation (Bertsimas et al. (2013b), Lorca et al. (2016)). Finite adaptability is a different approach that works well for some multistage robust optimization models (Bertsimas and Caramanis (2010), Bertsimas et al. (2011b)), but we chose affine adaptability due to its stronger scalability characteristics.

3. Problem Formulation
To view the problem we address in a concrete context, consider the following inventory routing problem over a finite horizon: A company has customers who consume a homogeneous commodity
over time, and a fleet of vehicles that is used to resupply them. We would like to generate a feasible schedule of routes for the vehicles that satisfies capacity constraints for users and vehicles, and leads to a low likelihood of stockouts for the customers.

A key insight that helps us achieve this is the observation that in practice, customers are often located in small neighborhoods, and that most of the variable cost (i.e., travelling distance) of the routing problem is derived from travel between depots and these small neighborhoods of customers. Within these neighborhoods, then, routes can be optimized sufficiently for industrial purposes by local search algorithms such as 2-opt (Croes (1958)). Therefore, our approach is to think of routes not as a list of customers, but as a neighborhood which a vehicle might travel between in a given time period. Upon selecting a route for a vehicle, a feasible schedule is then one which assigns customers to that vehicle that are on that route, i.e., in the associated selected neighborhood. Correspondingly, we assign costs to routes based on travel between the depot and the customers in the selected neighborhood, bearing in mind that the costs are to be taken as accurate only to the first order. The realized cost will depend on the customers we assign to the vehicle servicing a route.

This has a few key advantages. Firstly, it leverages the current knowledge of the company in the form of extant routes and neighborhoods, driver experience and other geographic and network information. In terms of actual implementation, this allows us to begin with a set of routes that are already known to be feasible, and gradually introduce routes to improve the solution quality of our model as needed. Furthermore, it significantly reduces the solution space of feasible routes, which helps the model to scale to large problem sizes more easily. By varying the sizes and coverage of the set of routes that we optimize over, we can exercise control over the tradeoff between scalability and solution quality, as needed.

For the vehicle routing problem under consideration, our decision-making has to take into account two sources of uncertainty in the demand for the commodity. The more important of these is the uncertainty associated with changes in temperature, which is correlated across all the customers. To a smaller extent, there is also an uncertainty in demand specific to each customer, which we assume is uncorrelated across customers. Using the well-established RO methodology, we define appropriate uncertainty sets (see (10) below) that capture these phenomena. In Section 3.3, we discuss ways to initialize the parameters of this uncertainty set from observations or simulations of the uncertain data.

3.1. Nominal Formulation
We begin by defining the nominal formulation of the inventory routing problem - in other words, we solve the problem for the case where demand is fixed rather than uncertain. Consider $N$ customers
who need to be resupplied over a time horizon $T$. The customers are to be resupplied with a fleet of $M$ vehicles, each of capacity $S$. In a single time period, the vehicles can be assigned to a tour $\theta = 1, \ldots, \Theta$, each which has associated cost $c_\theta$. Each customer $i$ has a maximum capacity of $Q_i$, and we suppose that customer $i$ begins the season with $Z_i$ of the commodity remaining. For the nominal formulation, we assume that demand $d_t^i$ is known for all customers and time periods.

We consider the following decision variables:

- $g_{i,\theta}^t$, the amount of fuel that customer $i$ will be resupplied via route $\theta$ at time $t$,
- $u_t^i$, the total amount of fuel that customer $i$ will be supplied at time $t$,
- Binary variable $v_t^\theta$ which is 1 if and only if tour $\theta$ is selected at time $t$.

Then, the nominal formulation is:

$$\min u, v, g \sum_{t=1}^{T} \sum_{\theta=1}^{\Theta} c_\theta v_t^\theta$$  \hspace{1cm} (1)

$$\text{s.t.} \quad 0 \leq Z_i + \sum_{\tau=1}^{t} u_{\tau}^i - \sum_{\tau=1}^{t} d_{\tau}^i \leq Q_i, \quad \forall i \in [N], \quad \forall t \in [T],$$  \hspace{1cm} (2)

$$\sum_{\theta=1}^{\Theta} v_t^\theta \leq M, \quad \forall t \in [T],$$  \hspace{1cm} (3)

$$u_t^i \leq \sum_{\theta=1}^{\Theta} g_{i,\theta}^t, \quad \forall i \in [N], \quad \forall t \in [T],$$  \hspace{1cm} (4)

$$\sum_{i=1}^{N} g_{i,\theta}^t \leq Sv_t^\theta, \quad \forall \theta \in [\Theta], \quad \forall t \in [T],$$  \hspace{1cm} (5)

$$g_{i,\theta}^t = 0, \quad \forall i \in [N], \quad \forall \theta : i \notin \theta, \quad \forall t \in [T],$$  \hspace{1cm} (6)

$$g_{i,\theta}^t \geq 0, \quad \forall i \in [N], \quad \forall \theta \in [\Theta], \quad \forall t \in [T],$$

$$u_t^i \geq 0, \quad \forall i \in [N], \quad \forall t \in [T],$$

$$v_t^\theta \in \{0, 1\}, \quad \forall \theta \in [\Theta], \quad \forall t \in [T].$$

(1) expresses the cost minimization objective. (2) guarantees that each customer is resupplied so that their supply of the commodity is never depleted. (3) respects the fleet size. (4) ensures that the amount of fuel assigned to refuel a customer is also assigned to some route in the same time period. (5) both allows us to assign fuel to a route only if the route is actually selected, and if so, also enforces vehicle capacity limits. (6) ensures that assignments are only made for customers that are on a given route.

3.2. Robust Formulation

Now we move to the robust formulation of the inventory routing problem. Here, rather than assume we know what the demand $d$ is, we assume rather that it lies within an uncertainty set $U$ which
we have constructed beforehand. We discuss the construction of $\mathcal{U}$ in more detail in the next subsection. We also assume that the amounts of fuel that customers start with, $Z_i$, take values in the interval $[\bar{Z}_i, \bar{Z}_i]$.

As before, we consider the same variables $g_{t,\theta}^i, u_t^i$ and $v_{\theta}^t$. Then the robust formulation is:

$$\min_{u,v,g} \sum_{t=1}^{T} \sum_{\theta=1}^{\Theta} c_{\theta} v_{\theta}^t$$

s.t.

$$0 \leq Z_i + \sum_{\tau=1}^{t} u_{\tau}^i - \sum_{\tau=1}^{t} d_{\tau}^i \leq Q_i, \; \forall i \in [N], \; \forall t \in [T], \; \forall d \in \mathcal{U}, \; \forall Z_i \in [\bar{Z}_i, \bar{Z}_i],$$

(7)

$$\sum_{\theta=1}^{\Theta} v_{\theta}^t \leq M, \; \forall t \in [T],$$

$$u_t^i \leq \sum_{\theta=1}^{\Theta} g_{t,\theta}^i, \; \forall i \in [N], \; \forall t \in [T],$$

$$\sum_{i=1}^{N} g_{t,\theta}^i \leq S v_{\theta}^t, \; \forall \theta \in [\Theta], \; \forall t \in [T],$$

$$g_{t,\theta}^i = 0, \; \forall i \in [N], \; \forall \theta : i \notin \theta, \; \forall t \in [T],$$

$$g_{t,\theta}^i \geq 0, \; \forall i \in [N], \; \forall \theta \in [\Theta], \; \forall t \in [T],$$

$$u_t^i \geq 0, \; \forall i \in [N], \; \forall t \in [T], v_{\theta}^t \in \{0, 1\}, \; \forall \theta \in [\Theta], \; \forall t \in [T].$$

The only difference with the nominal formulation is (7), which guarantees that each customer is resupplied so that their supply of the commodity is never depleted for all realizations of the demand and initial amounts.

### 3.3. Constructing $\mathcal{U}$

We describe here one method of constructing $\mathcal{U}$ based on insights from the Central Limit Theorem (see Bandi and Bertsimas (2012)), particularly applicable to the scenario of supplying heating oil to residences during winter. To do this, we assume that for any given customer, expected demand is constant above a certain temperature and increases linearly as the temperature decreases below that point. Specifically, for customer $i$, we assume that there exists a breakpoint $\Psi_i$ above which expected demand is $B_0^i$, and that if the temperature decreases below $\Psi_i$, the expected demand increases with a slope (against temperature) of $B_1^i$. We operate with the supposition that $\Psi_i, B_0^i$ and $B_1^i$ have been estimated for each customer from historical data.

We now assume that for each time period $t$, the temperature $\tau_t$ is subject to i.i.d. variation, and thus construct a CLT-style uncertainty set $\mathcal{U}_\tau$ for the temperature,
\[ U_\tau = \left\{ \tau : \left| \sum_{t=1}^{T} (\tau_t - \bar{\tau}_t) \right| / \sigma_\tau \sqrt{T} \leq \Gamma_\tau, \quad \bar{\tau}_t - 3\sigma_\tau \leq \tau_t \leq \bar{\tau}_t + 3\sigma_\tau \quad \forall t \in [T] \right\}. \quad (8) \]

Here \( \bar{\tau}_t \) and \( \sigma_\tau \) are the mean and standard deviation of the temperatures respectively, and \( \Gamma_\tau \) is a robust parameter that we are free to select, which we discuss below. We refer to the value \( \sqrt{T} \sigma_\tau \Gamma_\tau \) as the budget of variation in temperature, i.e., the net amount our temperatures are allowed to vary from their means.

We next consider the additional noise in the demand. For simplicity, we assume the demand is subject to additional zero-mean noise that has the same distribution for each time period, but is i.i.d. across customers, and thus construct a CLT-style uncertainty set \( U_\epsilon \) for the noise in demand,

\[ U_\epsilon = \left\{ \epsilon : \left| \sum_{i=1}^{N} \epsilon_i / \sigma_\epsilon \sqrt{N} \right| \leq \Gamma_\epsilon, \quad -3\sigma_\epsilon \leq \epsilon_i \leq 3\sigma_\epsilon \quad \forall i \in [N] \right\}, \quad (9) \]

where \( \sigma_\epsilon \) is the standard deviation of the demand noise.

This gives us our uncertainty set for demand, \( U \), which, as described above, consists of all demand vectors for which the corresponding temperature and demand noise simultaneously lie within the uncertainty sets \( U_\tau \) and \( U_\epsilon \), respectively.

\[ U = \left\{ d : d_t^i = B_0^i + B_1^i \max(0, \Psi_i - \tau_t) + \epsilon_i, \quad \tau \in U_\tau, \quad \epsilon \in U_\epsilon \right\}, \quad (10) \]

where \( B_0^i, B_1^i \) and and \( \Psi_i \) are all parameters estimated from data.

**Selecting robust parameters** The uncertainty sets \( U_\tau \) and \( U_\epsilon \) involve the parameters \( \Gamma_\tau \) and \( \Gamma_\epsilon \) that represent the planner’s desired balance between optimality and robustness. We next outline our approach for selecting these parameters. Assuming that temperatures \( \tau_t \) are independent for each time period \( t \), with mean \( \bar{\tau}_t \) and variance \( \sigma_\tau^2 \) from an otherwise unknown distribution, we select \( \Gamma_\tau \) such that \( U_\tau \) contains the realized temperature with probability 99% for large \( T \). Specifically, from the Central Limit Theorem,

\[ \lim_{T \to \infty} \mathbb{P} \left( \left| \sum_{t=1}^{T} (\tau_t - \bar{\tau}_t) \right| / \sigma_\tau \sqrt{T} \leq \Phi^{-1}(0.99) \right) = 0.99, \quad (11) \]

where \( \Phi \) is the cdf of the standard normal distribution, and so we select \( \Gamma_\tau = \Phi^{-1}(0.99) \). A similar approach is used for selecting \( \Gamma_\epsilon \). For other possible approaches to selecting the robust parameters,
For a given planning horizon \( T \) and \( N \) customers, let the demands \( d \in \mathbb{R}^{N \times T} \) lie in the uncertainty set \( U \) given in (10), which is non-convex, necessitating a novel approach to generate critical worst-case scenarios.

Note that the only robust constraint in our formulation is constraint (7), which requires us to protect against the maximum and minimum values of \( \sum_{\tau=1}^t d_{i\tau} \) over \( U \) for each customer \( i \) in \([N]\) and each day \( t \) in \([T]\). We next give algorithm Opt-Temp that allows us to optimize over \( U \) an affine combination of convex non-increasing functions of temperature. Note that demand without customer-specific noise is a convex non-increasing function of temperature in our model. In addition, as each robust constraint only involves one customer, the worst-case \( \epsilon_i \) can always be taken to be \( 3\sigma_i \) for maxima, and \(-3\sigma_i \) for minima. Given a customer \( i \) and day \( t \), we can use these to construct a demand vector \( d_i \in \mathbb{R}^T \) that maximizes the sum \( \sum_{\tau=1}^t d_{i\tau} \). This enables us to solve the robust formulation as a deterministic problem, vastly improving computational performance. For notational convenience, we refer to the natural projection of \( U \) onto the set of demand vectors for customer \( i \) as \( U_{[i]} \).

**Summary of algorithm:** To maximize the sum \( \sum_{t=1}^T a_t d_t(\tau_t) \) for \( \tau \in U_{[i]} \) (8), where \( d_t(\tau_t) \), for each \( t \), is a convex non-increasing function of \( \tau \), we let the set of days with non-negative affine coefficients, i.e., \( a_t \geq 0 \), be \( T_1 \), and those with negative affine coefficients, i.e., \( a_t < 0 \), be \( T_2 \). In algorithm Opt-Temp, we consider two cases: (i) \( |T_1| \geq |T_2| \) and (ii) \( |T_1| < |T_2| \). For the first case, we set all temperatures to be at their upper bounds, i.e., \( \tau_t = \bar{\tau}_t + 3\sigma_{\tau} \). We then greedily choose the days in \( T_1 \) and for each such \( t \), decrease its corresponding temperature as far as possible. In the second case, we set all the temperatures to be at their lower bounds, i.e., \( \tau_t = \bar{\tau}_t - 3\sigma_{\tau} \). We then optimize the restricted objective function over the days \( T_2 \) using standard convex optimization techniques. In both cases, we ensure that the temperatures selected respect the bound \( |\sum_{t=1}^T \tau_t - \bar{\tau}_t| \leq \Gamma_{\tau} \sqrt{T} \sigma_{\tau} \), where \( \Gamma_{\tau} \) is a robust parameter. To prove optimality, we show that there exists an optimal solution with at most one temperature not attaining one of its bounds, and that our algorithm finds such a solution.

Formally, we present in Algorithm 1 an algorithm Opt-Temp for maximizing an affine combination of convex non-increasing functions over \( U_{[i]} \). The algorithm finds, for convex non-increasing functions \( d_t(\tau) \) and coefficients \( a_t \), a temperature vector yielding \( \max_{\tau \in U_{[i]}} \sum_{t=1}^T a_t d_t(\tau_t) \). In our presentation of the algorithm we use a sorting function Sort\((R)\), which sorts the set of days \( R \) in descending order of the difference in the objective function when the temperature is changed from \( \bar{\tau}_t + 3\sigma_{\tau} \) to \( \bar{\tau}_t - 3\sigma_{\tau} \), i.e., \( \text{Sort}(R) = \{t_1, t_2, \ldots, t_{|R|}\} \) such that \( \Delta(t_x) \geq \Delta(t_y) \) whenever \( x < y \), where:

\[
\Delta(t) = a_t(d_t(\bar{\tau}_t - 3\sigma_{\tau}) - d_t(\bar{\tau}_t + 3\sigma_{\tau})).
\]
\textbf{Algorithm 1: Opt-Temp} \\
\textbf{Input:} $\Gamma > 0$, $\sigma$, $\bar{\tau} \in \mathbb{R}^T$, $a \in \mathbb{R}^T$, $d_i : \mathbb{R} \rightarrow \mathbb{R}$ $\forall t \in [T]$ \\
\textbf{Output:} $\tau \in \arg \max_{\tau \in \mathcal{U}_\tau} \sum_{t=1}^{T} a_t d_i(\tau_t)$ \\
$T_1 = \{t \in T : a_t \geq 0\}$, $T_2 = T \setminus T_1$, $k = l = m = 1$; \\
\textbf{if} $|T_1| \geq |T_2|$ \textbf{then} \\
$\tau_t = \bar{\tau}_t + 3\sigma \ \forall t \in [T]$; \\
$F = 3T\sigma$; \\
$\{t_1, t_2, \ldots, t_{|T_1|}\} = \text{Sort}(T_1)$; \\
\textbf{while} $F \geq 6\sigma - \Gamma \sqrt{T}\sigma$ and $k < |T_1|$ \textbf{do} \\
$$(\tau_{t_k}, F) \leftarrow (\tau_{t_k} - 6\sigma, F - 6\sigma);$$ \\
$k \leftarrow k + 1;$ \\
\textbf{end} \\
\textbf{if} $F > -\Gamma \sqrt{T}\sigma$ and $k < |T_1|$ \textbf{then} \\
$q^* = \arg \max_{q \in T_1} D(q, k, F);$ \\
\textbf{if} $q^* \leq k - 1$ \textbf{then} \\
$$(\tau_{t_k}, \tau_{q^*}, F) \leftarrow (\tau_{t_k} - 6\sigma, \tau_{q^*} + 6\sigma - F - \Gamma \sqrt{T}\sigma, -\Gamma \sqrt{T}\sigma);$ \\
\textbf{end} \\
\textbf{else} \\
$$(\tau_{q^*}, F) \leftarrow (\tau_{q^*} - F - \Gamma \sqrt{T}\sigma, -\Gamma \sqrt{T}\sigma)$; \\
\textbf{end} \\
\textbf{end} \\
\textbf{else} \\
$$\tau = \arg \max_{\tau \in \mathcal{U}'} \sum_{t=1}^{T} a_t d_i(\tau)$$ for $\mathcal{U}' = \mathcal{U}_\tau \cap \{\tau : \tau_t = \bar{\tau}_t - 3\sigma \ \forall t \in T_1\};$ \\
\textbf{end} \\
where \\
$$D(q, k, F) = \\
\begin{cases} \\
a_q d_q(\bar{\tau}_q + 3\sigma - F) - a_q d_q(\bar{\tau}_k + 3\sigma) & \text{if } q > k, \\
a_q d_q(\bar{\tau}_q + 3\sigma - F) + a_{k+1} d_{k+1}(\bar{\tau}_{k+1} - 3\sigma) - a_q d_q(\bar{\tau}_q - 3\sigma) - a_{k+1} d_{k+1}(\bar{\tau}_{k+1} + 3\sigma) & \text{if } q \leq k. \\
\end{cases}$$

\textbf{Theorem 1.} The temperature vector $\tau^* \in \mathbb{R}^T$ output by the Algorithm 1 maximizes $\sum_{t=1}^{T} a_t d_i(\tau_t)$ over $\mathcal{U}_\tau$.

\textbf{Proof:} We first show that $\tau^*$ is feasible. For the case where $|T_1| < |T_2|$, the temperatures are guaranteed to be feasible by definition of the optimization subproblem that we solve (Note that as we only optimize for $T_2$, this is a convex optimization problem and so tractable). To show feasibility for the case where $|T_1| \geq |T_2|$, we consider the bookkeeping variable $F$, which tracks the value of $\sum_{t=1}^{T}(\tau_t - \bar{\tau}_t)$. Before we update a temperature, we check that $F$ will not exceed the CLT-type bounds $-\Gamma \sqrt{T}\sigma \leq F \leq \Gamma \sqrt{T}\sigma$, and limit the magnitude of our updates accordingly. Similarly, the
temperatures are initialized at their upper bounds and never decreased by more than $6\sigma$, the width of the feasible interval for a single temperature. Also, note that we are assured of the existence of a feasible solution (e.g., setting the temperatures to their mean values). Thus, $\tau^*$ is feasible.

Next we prove that $\tau^*$ is optimal. Suppose we had a feasible temperature vector where for some $r \in T_1$, $\tau_r > \bar{\tau}_r - 3\sigma$, and for some $s \in T_2$, $\tau_s < \bar{\tau}_s + 3\sigma$. Then we could decrease $\tau_r$ and increase $\tau_s$ by some small $\epsilon$, while not decreasing the objective function. This means that we can limit ourself to optimal solutions where either the temperatures in $T_1$ all attain their lower bounds $\bar{\tau}_r - 3\sigma$, or the temperatures in $T_2$ all attain their upper bounds $\bar{\tau}_s + 3\sigma$. (We will show that the smaller set attains its bounds.)

Case (i) $|T_1| < |T_2|$: We show that in this case, there exists at least one optimal temperature vector $\tau^*$ such that $\tau^*_r = \bar{\tau}_r - 3\sigma$ for all days in $T_1$. (Note that such an optimal temperature vector is easy to find: for days in $T_1$, all the temperatures are at their lower bounds, and temperatures for days in $T_2$ can be found using linear optimization). Consider any optimal temperature vector $\tau^{opt}$ that maximizes $\sum_{t=1}^{T} a_t d_t(\tau_t)$ such that all the temperatures in $T_2$ attain their upper bounds, i.e. $\tau^{opt}_t = \bar{\tau}_t + 3\sigma$ for $t \in T_2$ (if not, as argued above, all the temperatures in $T_1$ must be at their lower bounds, thus proving our claim). Let $F^{opt} = \sum_{t=1}^{T} (\tau^{opt}_t - \bar{\tau}_t) = \sum_{t \in T_1} (\tau^{opt}_t - \bar{\tau}_t) + 3\sigma|T_2|$. Note that $F^{opt} \leq \Gamma\sqrt{T}\sigma$ since $\tau^{opt}$ is feasible. Now, consider a temperature vector $\tau'$ such that $\tau'_r = \bar{\tau}_r - 3\sigma$ for $t \in T_1$ and $\tau'_t = \bar{\tau}_t + 3\sigma$ for $t \in T_2$. Let $F' = \sum_{t \in T_1} (\tau'_t - \bar{\tau}_t) + \sum_{t \in T_2} (\tau'_t - \bar{\tau}_t) = 3(|T_2| - |T_1|) \geq 0$. Also, note that $F' \leq F^{opt} \leq \Gamma\sqrt{T}\sigma$. Thus, $\tau'$ is feasible and its function value is no worse than $\tau^{opt}$. Hence, we have proved that there exists an optimal temperature vector which attains the lower bounds for temperatures in $T_1$.

In this case, optimality follows from the definition of the optimization subproblem that we solve, restricted to $T_2$.

Case (ii) $|T_1| \geq |T_2|$: Similar to the previous case, we can assume that the temperatures in $T_2$ all attain their upper bounds, i.e. for all $t \in T_2$, we have $\tau_t = \bar{\tau}_t + 3\sigma$. We next show that there exists such an optimal solution where at most one temperature $\tau_t$ for a $t \in T_1$ is neither at $\bar{\tau}_r - 3\sigma$ nor $\bar{\tau}_t + 3\sigma$.

Suppose we had some feasible solution with $r, s \in T_1$, $\tau_r \neq \bar{\tau}_r \pm 3\sigma$, $\tau_s \neq \bar{\tau}_s \pm 3\sigma$. We want to adjust these temperatures so that one attains its bound, without decreasing the objective function. Let $a = \min(\tau_r - (\bar{\tau}_r - 3\sigma), \bar{\tau}_s + 3\sigma - \tau_r)$, $b = \min(\bar{\tau}_r + 3\sigma - \tau_r, \tau_s - (\bar{\tau}_s - 3\sigma))$. By the convexity of $d_r$ and $d_s$, we use Jensen’s inequality to get:

$$\frac{b}{a+b} d_r(\tau_r - a) + \frac{a}{a+b} d_r(\tau_r + b) \geq d_r(\tau_r), \quad (12)$$

$$\frac{a}{a+b} d_s(\tau_s - b) + \frac{b}{a+b} d_s(\tau_s + a) \geq d_s(\tau_s). \quad (13)$$
Adding these inequalities implies that either $d_r(\tau_r - a) + d_s(\tau_s + a)$ or $d_r(\tau_r + b) + d_s(\tau_s - b)$ must be at least $d_r(\tau_r) + d_s(\tau_s)$, and so we can adjust $\tau_r$ and $\tau_s$ as desired. We thus can limit ourselves to considering temperature vectors with at most one temperature not attaining either of its 3-sigma bounds.

Finally, suppose we knew that $\tau_t$ was the temperature not attaining its bounds. Then, a simple greedy algorithm for the temperature values at lower and upper bounds would give the optimal temperature vector.

In our algorithm, we iterate over all the choices for the day with the temperature not attaining its bounds, and select the one with the best objective value. The remaining temperatures are set to their upper or lower bounds, sorted so that they have the same output a greedy algorithm would have. Therefore, we obtain a temperature vector that maximizes the objective function over both sets of days, $T_1$ and $T_2$.

We now can explicitly find the minima and maxima over $U$ for the sums of demand seen in the robust constraints. For the maximum demand, we construct a worst-case temperature vector for $\sum_{r=1}^t d_i^r$ using the above algorithm. As mentioned above, the robust constraints each involve just a single customer and so $\epsilon_i$ can be taken to be $3\sigma_i$.

For the case of minimum demand, although the above ideas apply, the problem is much easier since we are required to solve a convex optimization subproblem. In fact, for each $s \in [T]$ and $i \in [N]$, we can compute the minimum value of $\sum_{t=1}^s d_i^t$ by solving the following linear optimization problem:

$$\min \sum_{t=1}^s d_i^t$$
\[ \text{s.t. } \]
- $B \leq \sum_{t=1}^T (\bar{\tau}_t - \tau_t) \leq B,$
- $\bar{\tau}_t - 3\sigma_t \leq \tau_t \leq \bar{\tau}_t + 3\sigma_t, \ \forall t \in [T],$
- $d_i^t \geq B_i^0 + B_i^1 x_t - 3\sigma, \ \forall i \in [N], \ \forall t \in [T],$
- $x_t \geq \Psi_t - \tau_t, \ \forall t \in [T],$
- $x \geq 0, \ d_i \geq 0.$

Similar to before, $\epsilon_i$ can be taken to be $-3\sigma_i$. This allows us to replace our robust constraints with $2NT$ deterministic constraints, in each case picking the appropriate endpoint of the interval $[\bar{Z}_i, \bar{Z}_i]$ to robustify against (i.e. $\bar{Z}_i$ for lower bounds and $\bar{Z}_i$ for upper bounds).

In practice, we observed that as the robust constraints for time $t$ do not involve customer demands for time periods beyond that, it improved the performance of our algorithm to project $U$ onto
the first $t$ time periods and find the worst-case vector corresponding to $\Gamma_t \sqrt{t/T}$. This weakens the theoretical probabilistic guarantees that we can make, because the Central Limit Theorem might not be a good approximation when $N$ and/or $T$ are small. However, in our experiments this adaptation did not result in a significant increase in stockouts, but it did produce a significant decrease in the cost (and conservativeness) of the models. Note that the protection is weakest against the earlier time periods at the very start of the planning horizon, when a customer is less likely to stockout anyway.

3.4. Affine Adaptive Robust Formulation

With the advancement of technology, it is becoming increasingly feasible for companies to install sensors in customers’ buildings. This might allow them, for instance, to track the daily consumption of their customers, improving the solution quality of their planning models. While it may be impractical to alter the fleet and crew schedule on short notice, we adapt our formulation so that the quantity of fuel resupplied will now be partially responsive to the actual demand observed. Without this new information from sensors, a company is limited to observations made during scheduled deliveries, i.e., the aggregated demand between refuelling decisions, which is much less informative.

We now define an affine adaptive robust formulation. Instead of having the model decide on exact amounts to refuel each customer daily, we set the quantities refuelled to be affine functions of the demand in the previous days, and solve for the coefficients of these affine functions.

To make the formulation adaptive, we substitute each $u^i_t$ with an affine function of previous days’ demands: $u^i_t = b^i_{0,t} + \sum_{j=1}^{t-1} b^i_{j,t} d^i_j$ (remember that consumption for a day occurs after any refuelling on that day), where the various $b^i_{j,t}$ are now variables we are solving for. Similarly, we substitute each $g^i_t,\theta$ with $g^i_t,\theta = a^i_{0,t} + \sum_{j=1}^{t-1} a^i_{j,t} d^i_j$, where $a^i_{j,t,\theta}$ are variables.

This leads to the following formulation:

$$\min_{a,b,v,g} \sum_{t=1}^{T} \sum_{\theta=1}^{\Theta} c^i v^i_\theta$$

s.t.  
\[0 \leq Z_i + \sum_{\tau=1}^{t} (b^i_{0,\tau} + \sum_{j=1}^{\tau-1} b^i_{j,\tau} d^i_j) - \sum_{\tau=1}^{t} d^i_\tau \leq Q_i, \quad \forall i \in [N], \quad \forall t \in [T], \quad \forall d \in U, \quad (14)\]  
\[\sum_{\theta=1}^{\Theta} v^i_\theta \leq M, \quad \forall t \in [T], \quad (15)\]  
\[b^i_{0,t} + \sum_{j=1}^{t-1} b^i_{j,t} d^i_j \leq \sum_{\theta=1}^{\Theta} (a^i_{0,t} + \sum_{j=1}^{t-1} a^i_{j,t} d^i_j), \quad \forall i \in [N], \quad \forall t \in [T], \quad \forall d \in U, \quad (16)\]
\[ a_{i,t}^{0,t} = 0, \forall i \in [N], \forall \theta : i \notin \theta, \forall t \in [T], \forall j \in \{0, \ldots, t-1\}, \]
\[ a_{i,t}^{0,t} \geq 0, \forall i \in [N], \forall \theta \in [\Theta], \forall t \in [T], \forall j \in \{0, \ldots, t-1\}, \]
\[ b_{i,t}^{0,t} \geq 0, \forall i \in [N], \forall t \in [T], \forall j \in \{0, \ldots, t-1\}, \]
\[ v_{\theta}^t \in \{0,1\}, \forall \theta \in [\Theta], \forall t \in [T]. \]

Note that all the constraints in the adaptive robust formulation have the same interpretation as their counterparts in the robust formulation, although fuel supplied is now adaptive in that it is an affine function of demand. Furthermore, the starting quantities, \( Z_t \), are no longer taken to be uncertain, as we would expect real-time measurements of demand to also yield exact information about the customers’ remaining fuel.

The number of variables in the adaptive formulation is an order of magnitude greater than the nominal or robust case. Thus, it is impractical to solve it using a deterministic linear MIO, as we did for the nominal formulation. In addition, the constraints (14), (15) and (16) involve products of our decision variables and the uncertain demand. This means that to separate over these constraints one would need to solve a quadratic optimization problem over a non-convex set.

We instead use a cutting-plane algorithm that exploits the structure of the uncertainty set \( U \), to tractably solve the adaptive formulation. Given a candidate solution, one can now optimize for affine functions of demand (possibly a different one for each constraint) as before, to generate a feasible cutting plane. Thus, we can first use OPT-TEMP to give us a worst-case temperature vector for a potential constraint-candidate pair. As noise for each customer is constant across time periods, the noise \( \epsilon_i \) for a worst-case demand vector for that constraint-candidate pair is given by a greedy algorithm sorting on its coefficient, \( \sum_{i=1}^{N} a_{i,t}^{0,t} \sum_{j=1}^{t-1} a_{i,t}^{j,t} \). This allows us to add new deterministic constraints from demand vectors for the violated constraints, reoptimize the model, and generate a new candidate solution.

4. Route Generation

Route generation is a widely studied problem, especially given its importance in various vehicle and inventory routing problems (for example, see Laporte (1992), Francis et al. (2008), Golden et al. (2008), applied to a plethora of real-world applications such as routing for bakery companies (Pacheco et al., 2012), blood product distribution (Hemmelmayr et al. 2009a, Hemmelmayr et al. 2009b), grocery industry (Semet and Taillard 1993), ship-routing (Gunnarsson et al. 2006)). The literature is ripe with a number of exact (Baldacci et al. 2011, Laporte 1992, Laporte et al. 1986) and heuristic (Hemmelmayr et al. 2009b, Vidal et al. 2012, Liu 1997, Semet and Taillard 1993) approaches for route generation. Since tractability is a major concern with exact approaches, we employ heuristic
methods to generate a feasible set of routes. We would however like to emphasize that exploring route generation techniques is not the main focus of this work.

In this work, we consider routes to be not just feasible tours, but be feasible neighborhoods where a schedule specifies which subset of neighbours to serve. Our formulation operates under the approximation that there is a fixed cost to ‘visit’ a neighborhood, irrespective of how many houses are actually resupplied with the commodity. This is a reasonable approximation in our problem context since the cost of routing is a second order cost, compared to the cost of customer stockouts (affecting customer satisfaction and reputation of the firm) and the cost of maintaining a fleet of vehicles.

We generate an initial set of feasible neighborhoods for our datasets in two ways: (a) using a user-operated GUI where the supplier can manually select neighborhoods that are typically served together, (b) using an automated sweep of the customer locations. The first approach is preferred when the supplier would like to utilize prior knowledge and accumulated expertise about different neighborhoods. Our second approach is an automated sweep of the geographic area under consideration. Our algorithm creates a cover of the entire space with ‘neighborhoods’ or boxes so that the number of customers in each box lies in an interval. This interval is selected so that a vehicle is able to resupply about half the customers in the neighborhood to maximum capacity. Experimentally, these sweeps generate a good first set of feasible routes that make the problem scalable.

We further improve the quality of the routes using a set-cover formulation inspired by the work of \cite{Cacchiani2014} augmented with well-studied tabu-search heuristics for improving routes (for e.g. in \cite{Cordeau1997, Gendreau1994}). We consider a set of possible schedules for the customers, covered using feasible routes such that on any day at most $M$ vehicles are used. However, we deviate from their work by assuming that the cost, $c_\theta$, of a route $\theta$ is given by the Euclidean distance of the tour suggested by the 2-opt (TSP) heuristic from the depot that a customer is served from (which is usually good enough in practice). We construct the following input from a pre-computed solution of the nominal problem.

- $T$, the total number of days in the planning horizon,
- $M$, the maximum number of vehicles in the fleet,
- $N$, the number of customers,
- $S_i$, the set of valid schedules that a customer $i$ could be visited at. In order to construct this set, we consider the service schedule suggested by the nominal solution, and shift it by allowing each customer to be visited up to three days before or after the scheduled delivery.
- $\Omega$, the set of feasible routes. We initialize $\Omega$ with the set of routes obtained by either neighborhood selection or automatic sweep. We will now describe how we use the following formulation to improve the quality of the set of routes.
We next generate routes, using the following set-covering-like formulation. Let $a^i_\theta$ be a constant equal to 1 if customer $i$ is on route $\theta$, and 0, otherwise, for all $i \in [N], \theta \in \Omega$. Let $b^t_p$ be equal to 1 if service schedule $p$ is feasible for customer $i$, i.e., $p \in S_i$ and $b^t_p$ is 0, otherwise. We use two sets of binary variables: $x^t_\theta$ and $y^i_p$. Here $x^t_\theta$ is 1 if and only if route $\theta$ is selected on day $t$ and 0, otherwise. Finally, $y^i_p$ is 1 if and only if combination $p$ is selected for customer $i$ and 0, otherwise. We now formulate a binary integer program as follows:

\[
SC_{MIO} = \min \sum_{\theta \in \Omega} \sum_{t \in [T]} x^t_\theta c_\theta \tag{17}
\]

s.t. $\sum_{p \in S_i} y^i_p = 1, \quad i \in [N], \tag{18}$

$\sum_{\theta \in \Omega} x^t_\theta a^i_\theta - \sum_{p \in S_i} y^i_p b^t_p \geq 0, \quad i \in [N], \quad t \in [T], \tag{19}$

$\sum_{\theta \in \Omega} x^t_\theta \leq M, \quad \forall t \in [T], \tag{20}$

$x^t_\theta \in \{0, 1\}, \quad r \in \Omega, \quad t \in [T],$

$y^i_p \in \{0, 1\}, \quad p \in S_i, \quad i \in [N].$

The objective function aims at minimizing the cost of the routes selected. Constraints (18) guarantee that exactly one feasible service schedule is selected for a customer. Constraints (19) guarantee that if the selected service schedule for customer $i$ requires service on day $t$, then there must be a route selected on day $t$ with customer $i$ on the route. Constraints (20) ensure that at most $M$ vehicles are used on any day, thereby respecting the fleet size.

We relax the above integer optimization problem and do column generation on the resulting optimization relaxation. We generate candidate routes using the following heuristic operations:

- **Insert** a customer into an existing neighbouring route,
- **Swap** two customers from neighboring routes,
- **Remove** customers from an existing route,
- **Construct** new routes for each day $t$ by considering customers $i$ such that their dual variables $p_{i,t}$ take large values.

For each candidate route $\theta$ for each day $t$, we compute its reduced cost as follows: $\bar{c}_\theta = c_\theta - \sum_{i \in \theta} p_{i,t} - p_M$ where $p_{i,t}$ is the optimal dual variable corresponding to constraints (19) and $p_M$ is the dual variable corresponding to the constraint (20). We add the route that has the most negative reduced cost out of a set of candidate solutions, until the relative decrease in the objective value (17) of the relaxed linear optimization problem $SC_{LO}$ (corresponding to $SC_{MIO}$) is less than 1%.
5. Computational Experiments

To test the scalability of our problem formulations and the quality of our solutions, we generated a number of datasets based on real-world problems. We imported customer locations from a few instances in the TSPLIB, the standard test bed of the Traveling Salesman Problem, with the size of these instances (i.e. the number of customers) ranging from 51 to 5915 (data instances eil51, rat99, kroB200, rat575, pcb1173, d2103, r15915). For simplicity, we assumed a homogenous fleet of vehicles (in particular, with identical maximum capacity), operating from a single depot located at the centroid of the users.

We assumed all the customers to have homogenous heating oil tanks with identical maximum capacity. For each customer, we generated a base temperature above which their expected demand was near-zero and constant, and below which it increased linearly as temperature decreased. We randomly generated family sizes for each customer ranging from 2 to 4, and scaled the mean demand accordingly, adding noise in both the temperature and for each user’s demand as described in Section 3.2. To generate temperatures, we used actual data for Boston for the months of November 2013 to March 2014, representing a full season of heating oil consumption (Weather Underground (2014)).

We further generated estimated initial amounts for each customer, and subjected these to further zero-mean uncertainty proportional to the difference of these amounts to the full customer capacity, encapsulating the principle that a customer with higher usage or a customer who was serviced a longer time ago should have more uncertainty in their starting amounts. To be precise, if the estimated initial amount for a customer was \( z_{i}^{est} \), the noisy initial amount was

\[
z_{i} = z_{i}^{est} + (Q - z_{i}^{est}) \times U_{i}, \quad U_{i} \sim U(-1/2, 1/2).
\]

Family sizes and uncertain initial amounts were randomized separately to get training and testing datasets. We used the training set to tune our robust parameters \( \Gamma_{r} \) and \( \Gamma_{e} \), and set these parameters correspondingly in the testing set to test the performance of our approach in terms of running time and effectiveness. We assumed in our experiments that a centralized depot serves all the customers, although the formulation generalizes easily to applications with multiple depots, each with their own fleet of vehicles. We construct routes (with 30-40 customers on each route) for each dataset such that any customer lies on approximately 3-4 routes. We further augment these routes with dynamically searched routes such that the covering cost of the customers with respect to feasible schedules (heuristically generated from the past demands) decreases. We detail various parameters of the dataset in Appendix [A].

For our computational experiments, we let the nominal, robust and adaptive formulations solve for two hours each, using the nominal solution as a warm start (though infeasible) to the robust model, and the robust solution as a warm start to the adaptive model.
While the adaptive model we presented in Section 3.4 schedules a customer with refuelling quantities that are affine in all of that customer’s observed demand, we improved the tractability of our implementation of the adaptive model by relaxing the number of terms of the adaptability. Specifically, we limited the refuelling quantity for a customer for time period $t$ to be a base amount $b_{i,t}$, plus a term linear in that customer’s demand during time period $t-1$ (i.e., $d_{i,t-1}$), a term linear in the total demand of that customer during time periods $t-3$ and $t-2$ (i.e., $d_{i,t-3} + d_{i,t-2}$), and a term linear in the total demand of that customer during time periods $t-7$ to $t-4$ (i.e., $d_{i,t-7} + \cdots + d_{i,t-4}$). We also solved for non-adaptive quantities across different vehicles, i.e. for $g_{i,\theta}$.

We used the robust solution as a warm start to the base amount, and initialized the other affine coefficients as zero. Note that this is, by design, already a feasible solution to the adaptive model, ensuring that the adaptive model always gave feasible output.

Each dataset (both training and testing) contained fifty generated scenarios for any computational experiment. All our instances were solved with Gurobi 6.0.0 on a Intel Xeon E5687W (3.1 GHz) processor with 16 cores and 128 GB of RAM.

To investigate the quality of the solutions resulting from our models, we ask the questions of a) whether our robust and adaptive models lead to fewer stockouts, b) what effect this has on service cost, and (c) whether this suggests a possible reduction of the vehicular fleet size.

### 5.1. Stockouts

We first investigate whether our robust and adaptive inventory routing models lead to a reduction in stockouts. As mentioned above, the three formulations were solved sequentially (nominal-robust-adaptive) on problems of a fixed vehicle fleet size and a decision horizon of 151 days. Figures 1, 2 and 3 show the average percentage of customers who experienced stockouts for the nominal, robust and adaptive formulations respectively.

We observe that across data sets, while the nominal formulation had between 160%-225% of customers stocking out (some customers experienced stockouts multiple times), the robust formulation decreased this to below 9% of all customers, and in most cases half of that or even less. Stockouts decreased further by 0.5%-1% of all customers from the robust to the adaptive formulation for all the data sets. We also notice that the robust and adaptive formulations were less sensitive than the nominal formulation to increasing variance in the noise (i.e., errors in tuning the robust parameters).

We explore how the reduction in stockouts was distributed across the fifty scenarios for each data point in the above experiment. Figure 4 shows the standard box plot for the reduction in robust model stockouts as a fraction of nominal model stockouts, for the uncertainty regime of the base level of variance in the noise. We observe that every scenario generated had at least 94% relative reduction in stockouts from the nominal to robust models, with an average of over 96% relative reduction for each dataset.
Figure 1  Average stockout percentages for the nominal solutions for data sets of different sizes.

Figure 2  Average stockout percentages for the robust solutions for data sets of different sizes.
Figure 3  Average stockout percentages for the adaptive solutions for data sets of different sizes.

Figure 4  25%-75% quantiles and extreme values for the reduction in robust model stockouts as a fraction of nominal model stockouts.
5.2. Service cost

We next consider the effect on the cost of servicing the customers with the different formulations. As before, we solve problems of a fixed vehicle fleet size. To get the combined cost of the problem, we consider costs from two sources, namely: (a) the variable cost from the routes, which is the objective function of the optimization model, and (b) the cost of refuelling a customer who experiences a stockout. Because these stockouts occur randomly throughout the course of a time period, and must be addressed urgently, the planner must send an emergency refuelling vehicle out each time a customer stocks out. We assume that due to the reduced efficiency of the smaller emergency refuelling vehicle, its cost per unit distance is twice that of the usual refuelling vehicle fleet. Thus, for an instance with variable cost $c_{\text{var}}$ and refuelling cost $c_{\text{re}}$, the combined cost is $c_{\text{var}} + 2c_{\text{re}}$.

Table 1 compares the combined cost for the respective models, along with the percentage gap compared to the best lower bound the solver could find within the time limits we set. $C_N$, $C_R$ and $C_A$ are the combined costs of the nominal, robust and adaptive models respectively, while $G_N$, $G_R$ and $G_A$ are the respective provable duality gaps output by the Gurobi solver.

In all cases, the robust model had a combined cost no higher than 86% of the nominal model’s. With larger data sets of over a hundred customers, the cost savings were 44% or more of the combined cost of the nominal model. Furthermore, the adaptive model had a combined cost that was 0.2%-0.3% lower than that of the robust model, i.e. 0.1%-0.3% lower than the combined cost of the nominal model.

<table>
<thead>
<tr>
<th>Customers</th>
<th>$C_N$</th>
<th>$G_N$</th>
<th>$C_R$</th>
<th>$G_R$</th>
<th>$C_A$</th>
<th>$G_A$</th>
<th>$C_R/C_N$</th>
<th>$C_A/C_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>16032</td>
<td>26.9</td>
<td>13791</td>
<td>37.2</td>
<td>13768</td>
<td>56.6</td>
<td>0.860</td>
<td>0.859</td>
</tr>
<tr>
<td>99</td>
<td>83527</td>
<td>45.7</td>
<td>66894</td>
<td>65.1</td>
<td>66684</td>
<td>66.5</td>
<td>0.801</td>
<td>0.798</td>
</tr>
<tr>
<td>200</td>
<td>5.295e6</td>
<td>77.3</td>
<td>2.966e6</td>
<td>74.3</td>
<td>2.959e6</td>
<td>75.5</td>
<td>0.560</td>
<td>0.559</td>
</tr>
<tr>
<td>575</td>
<td>1.001e6</td>
<td>1.61</td>
<td>459040</td>
<td>16.4</td>
<td>457547</td>
<td>18.9</td>
<td>0.459</td>
<td>0.457</td>
</tr>
<tr>
<td>1173</td>
<td>1.474e7</td>
<td>36.2</td>
<td>8.003e6</td>
<td>46.1</td>
<td>7.986e6</td>
<td>47.9</td>
<td>0.543</td>
<td>0.542</td>
</tr>
<tr>
<td>2103</td>
<td>3.190e7</td>
<td>31.8</td>
<td>1.373e7</td>
<td>38.0</td>
<td>1.370e7</td>
<td>39.7</td>
<td>0.431</td>
<td>0.429</td>
</tr>
<tr>
<td>5915</td>
<td>3.946e8</td>
<td>35.6</td>
<td>1.676e8</td>
<td>43.1</td>
<td>1.670e8</td>
<td>44.7</td>
<td>0.425</td>
<td>0.423</td>
</tr>
</tbody>
</table>

Table 1   Costs and solver gaps for data sets of different sizes.

5.3. Fleet reduction

Finally, whereas in the previous subsections, the vehicle fleet size was constant for each data set, here we investigate the tradeoffs of reducing the vehicle fleet size. We focus on a single data set with $N = 575$, for which our previous experiments used a fleet of 11 vehicles.

To allow the models to output a solution even with an infeasibly small fleet size, we introduce slack variables into our model that allow the demand constraints to be relaxed for a steep penalty (we took this to be $10^7$ times the amount of violation). Taking the combined cost introduced in
Section 5.2 we now further add to this the fixed cost of a vehicle fleet of a given size, where we assume that the fixed cost of operating a vehicle is equivalent to the variable cost from travelling a distance of 10,000 units. Thus, for an instance with variable cost $c_{\text{var}}$ and refuelling cost $c_{\text{re}}$, utilizing a fleet of size $N$, the combined cost is $c_{\text{var}} + 2c_{\text{re}} + 10,000N$. Table 2 compares the new combined cost for the best solutions for vehicle fleets of different sizes. $C_N$, $C_R$ and $C_A$ are the combined costs of the nominal, robust and adaptive models respectively, while $S_N$, $S_R$ and $S_A$ are the average number of customers who stock out.

<table>
<thead>
<tr>
<th>Vehicles</th>
<th>$C_N$</th>
<th>$S_N$</th>
<th>$C_R$</th>
<th>$S_R$</th>
<th>$C_A$</th>
<th>$S_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>959490</td>
<td>852.9</td>
<td>529377</td>
<td>15.12</td>
<td>527747</td>
<td>12.38</td>
</tr>
<tr>
<td>10</td>
<td>1013552</td>
<td>932.54</td>
<td>519377</td>
<td>15.12</td>
<td>517747</td>
<td>12.38</td>
</tr>
<tr>
<td>9</td>
<td>1003552</td>
<td>932.54</td>
<td>509377</td>
<td>15.12</td>
<td>507747</td>
<td>12.38</td>
</tr>
<tr>
<td>8</td>
<td>993552</td>
<td>932.54</td>
<td>499377</td>
<td>15.12</td>
<td>497747</td>
<td>12.38</td>
</tr>
<tr>
<td>7</td>
<td>983552</td>
<td>932.54</td>
<td>556791</td>
<td>110.00</td>
<td>548587</td>
<td>95.04</td>
</tr>
<tr>
<td>6</td>
<td>991370</td>
<td>976.52</td>
<td>546791</td>
<td>110.00</td>
<td>538587</td>
<td>95.04</td>
</tr>
<tr>
<td>5</td>
<td>981370</td>
<td>976.52</td>
<td>1493560</td>
<td>1888.88</td>
<td>1493049</td>
<td>1860.24</td>
</tr>
<tr>
<td>4</td>
<td>1458104</td>
<td>1823.82</td>
<td>1483560</td>
<td>1888.88</td>
<td>1482509</td>
<td>1885.62</td>
</tr>
<tr>
<td>3</td>
<td>2157939</td>
<td>3233.10</td>
<td>2040526</td>
<td>2998.22</td>
<td>2038394</td>
<td>2994.58</td>
</tr>
</tbody>
</table>

Table 2 Stockout percentages for $N = 575$ with different fleet sizes.

We observe that the robust and adaptive solutions allow us to decrease the fleet size to 8 without increasing stockouts, thereby decreasing the combined cost. Decreasing the fleet size below 8 leads to an increase in combined cost for the robust and adaptive models, as demand is shifted from scheduled refuellings to emergency refuellings. On the other hand, with the nominal model, removing even one vehicle leads to an increased combined cost, as the savings from the smaller fleet size are lost to increased refuelling costs from the increased numbers of customers who stock out.

6. Conclusion

We consider the finite horizon inventory routing problem with uncertain demand, which do not scale well with current techniques. We present robust and adaptive formulations for this problem that are tractable for ~6000 customers. For an uncertainty set where customers’ demands demonstrate limited dependence, for which standard methods of robust optimization are insufficient, we construct an algorithm that allows us to find worst-case scenarios deterministically (robust formulation) or relative to a candidate solution (adaptive formulation). We show a significant decrease in stockouts (over 94% in all test cases) for our models, translating to a 14% decrease in cost for the supplier. In addition, we show that our models, with slack variables, are capable of providing further cost savings through a reduction in the vehicle fleet size.
In this paper, we give a proof of concept that scales an inventory routing problem to a large number of customers for the heating oil industry. However, the solution techniques are worth exploring further for a variety of problems, utilizing well-constructed uncertainty sets that model the natural uncertainties in the problem data. We would reduce the complexity of the solution space through careful examination of the problem domain - for instance, in the inventory routing problem we consider, through the observation that routing costs are only of second-order importance to the planners. Another interesting direction might be to explore more sophisticated ways of modeling emergency refueling routing decisions, and smarter ways of managing fuel quantities dynamically.

A natural question that arises out of this work would be to attempt to understand why, although the solutions to our models that we find in reasonable time give us much of the benefit of robust optimization (i.e. experimentally reducing stockouts by over 94%), the provable gaps in the output of the Gurobi solver are still relatively large. One possible approach to this end might be to reduce these gaps through further theoretical improvements, for example, along the lines of [Bertsimas and de Ruiter (2015)].

Acknowledgments
This material is based upon work supported partly by NSF Grant Nos. CCF-1115849 and OR-1029603, ONR Grant Nos. N00014-14-1-0072, N00014-12-1-0999, N00014-15-1-2083. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and the Office of Naval Research.

Appendix A: Data
Let the number of customers in the data set be $N$ and let each customer have a capacity of $Q$ (set to 20). We consider the length of the planning horizon, $T$, to be 151 time units, but in our experiments although we solve the model for $T = 160$, we only calculate costs for the first 141 time units to account for end of horizon effects. We also assume a homogenous fleet of vehicles each with capacity $S$ (set to 200).

1. Estimated initial amounts: We generate the estimated initial level of oil for each customer $i \in N$ using the following formula:
   
   $$z^e_{i} = Q \times (1 - \min(0.9, |X_i|/3)),$$

   where $X_i$ are i.i.d. standard normal random variables sampled once for each customer. We generate $z^e_{i}$ once for each customer for all the training scenarios, and once for each customer for all the testing scenarios.

2. Realized initial amounts: Once the estimated initial amounts are fixed, we generate actual customer levels at the start of the horizon, called $z_i$. These are generated with randomness proportional to the difference of the estimated from $Q$. More precisely, for each scenario, we generate the initial customer level using:

   $$z_i = z^e_i + (Q - z^e_i) \times U_i,$$

   where $U_i$ are i.i.d. uniform random variables distributed as $U_i \sim U(-1/2, 1/2)$. We finally clip the $z_i$ within the interval $[0.5, Q]$. 


3. **Estimated Temperature**: We set the estimated temperatures $T_{est}^t$ according to Weather Underground Data for the temperatures in Boston for the months of November 2013 to March 2014, representing a full season of heating oil consumption (Weather Underground (2014)).

4. **Realized Temperature**: We consider different scenarios with the noise in temperature, $\delta_t$, varying in the set $\{0.02, 0.04, 0.06, \ldots, 1.0\}$. For each value for $\delta_t$, we create instances with temperature generated using the following relation:

$$T_t = T_{est}^t + \max(-3 \cdot \delta_t, \min(3 \cdot \delta_t, \delta_t \cdot X_t)).$$

5. **Fleet Size**: For datasets with less than 1000 customers, we assume a fleet with approximately $\sqrt{N}/2$ vehicles. Otherwise, we assume approximately $3D/ST$ vehicles, where $D$ is the total mean demand across all the customers over the entire planning period, $S$ is the vehicle capacity and $T$ is the planning period. For data sets of size 51, 99, 200, 575, 1173, 2103 and 5915, this means a fleet size of 3, 4, 7, 11, 24, 42 and 117 vehicles respectively.

6. **Routes**: We start with a rectangular sweep of the customer locations and generate a set of routes such that smaller datasets have an average of around 15 customers, and larger datasets have an average of around 25 customers. The characteristics of these automatically generated routes are included in Table 3 showing the number of routes (Num Routes), Minimum cost of the routes (Min cost), Maximum cost of the routes (Max cost), Average cost of the routes (Avg cost), Minimum route size (Min size), Maximum route size (Max size) and Average route size (Avg size). We experimented with increasing the number of routes covering each customer, and augmenting the set of routes by tabu-search. Neither of these improved the cost of the model by more than 2%, and in fact the increased complexity rendered the robust and adaptive models insoluble in under twelve hours for for datasets of size larger than 200.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Num routes</th>
<th>Min cost</th>
<th>Max cost</th>
<th>Avg cost</th>
<th>Min size</th>
<th>Max size</th>
<th>Avg size</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>15</td>
<td>113.3</td>
<td>234.1</td>
<td>153.6</td>
<td>10</td>
<td>20</td>
<td>13.6</td>
</tr>
<tr>
<td>99</td>
<td>27</td>
<td>198.2</td>
<td>521.1</td>
<td>297.9</td>
<td>11</td>
<td>23</td>
<td>14.67</td>
</tr>
<tr>
<td>200</td>
<td>55</td>
<td>2454.2</td>
<td>10200.7</td>
<td>5216.8</td>
<td>11</td>
<td>24</td>
<td>14.54</td>
</tr>
<tr>
<td>575</td>
<td>49</td>
<td>280.1</td>
<td>1169.9</td>
<td>568.34</td>
<td>21</td>
<td>29</td>
<td>23.47</td>
</tr>
<tr>
<td>1173</td>
<td>98</td>
<td>1184.8</td>
<td>5973.7</td>
<td>3162.2</td>
<td>20</td>
<td>32</td>
<td>23.93</td>
</tr>
<tr>
<td>2103</td>
<td>152</td>
<td>1174.0</td>
<td>10653.8</td>
<td>3553.7</td>
<td>21</td>
<td>41</td>
<td>27.58</td>
</tr>
<tr>
<td>5915</td>
<td>469</td>
<td>2504.7</td>
<td>40715.4</td>
<td>14526.0</td>
<td>21</td>
<td>37</td>
<td>25.19</td>
</tr>
</tbody>
</table>

Table 3 Characteristics and coverage of the initial automatically generated routes.

**References**


Toth, Paolo, Daniele Vigo. 2001. The Vehicle Routing Problem. SIAM.

