Optimizing Package Express Operations in China

Baris Yildiz[1] and Martin Savelsbergh[2]

[1]Department of Industrial Engineering, Koc University, Istanbul, Turkey
[2]H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, USA

September 17, 2019

Abstract

We explore optimization models to support the planning and operations functions at package express carriers in China. The models simultaneously consider ground and air transportation, company-owned and purchased capacity, multiple service products, and shipments becoming available throughout the day. An extensive computational study using real-life data shows the efficacy of the models, provides insights into the structure of effective express service networks, and analyzes the benefits of operational efficiency improvements on performance.

1 Introduction

Package express companies world-wide are facing a rapidly changing environment due to the explosive growth of e-commerce and due to the push by retailers to satisfy their customers' desire for instant gratification by offering faster and faster delivery times. This is especially true in China where the value of B2C e-commerce in 2015 reached $766.5 billion and where more than 430 million people shopped online that year.

To be able to accommodate the expected growth in demand volume and the need for faster service offerings, package express companies are looking for optimization-based tools to support both their planning and operations functions. Motivated by the environment encountered at SF Express, we have designed and implemented optimization models to support express shipment network design which take into account many of the critical features of the environment, such as the integration of ground and air operations, the use of company-owned capacity (in the form of cargo planes) as well as purchased capacity (in the form of belly capacity offered by commercial airlines), the need to offer of multiple service products, and shipments entering the system throughout the day.

This environment is far more complex than those typically presented and analyzed in the literature. However, there is a clear practical need for optimization tools that incorporate these complicating features, and, equally important, that can handle instances of the size encountered in practice. Our efforts are a first step towards providing the industry with such tools and we hope that these efforts stimulate other researchers to do the same.

The term "express shipment" is a generic term to describe time-sensitive freight transportation, and includes the transport of small packages (also referred to as parcels), but also the (sometimes specialized) transport of high-value items/products. Supporting the design of a time- and cost-efficient express shipment network for transporting a large volume of low-valued parcels and for transporting a small volume of high-valued items/products requires different approaches. For example, small package demand typically involves a large number of origin-destination pairs with much of the total volume concentrated in only a small number of origin-destination pairs (representing pairs of large cities). To accommodate such a demand pattern, a combination of purchased capacity, i.e., belly capacity, to serve the small demand origin-destination pairs, and owned capacity, to serve the high demand origin-destination pairs, may be ideal, especially in China where the package express companies have insufficient owned capacity to satisfy all their demand. On the other hand, the demand for transport of high-value items/products tends to occur between a relatively small number of origin-destination pairs and, due to the special packaging and/or handling requirements, purchased capacity may not be an option.

The above discussion reflects that the air service network plays a critical role in the transport of time-sensitive freight. Our research focuses on the design of an air service network, but recognizes that that cannot be done independently of the road service network, as the road service network acts as a feeder system for the air service network. An integrated view of ground and air transport capacity is necessary to expand service coverage in a cost-efficient way.

To accommodate varying demand distributions and densities (due to differences in the markets) we design and implement two optimization models, which can be viewed as being at opposite ends of a spectrum. For the small package market, in which demand is more densely distributed over the service area, and cargo plane capacity is the limiting resource, efficient use of the cargo plane capacity is critical to reducing operational costs. In this case, using cargo planes to serve a limited set of high-volume origin-destination pairs with direct flights is an effective strategy to maximize the utilization of the available transportation capacity, among others because it avoids the use of a "central hub" with time-consuming sorting operations. On the other hand, for the high-value items/products market, in which demand is smaller and less densely (and more evenly) distributed over the service area (compared to the small package market), available flight time is the limiting resource (rather than cargo plane capacity) and the use of transshipment, i.e., allowing planes to meet and exchange cargo at certain locations) is a more effective strategy (which will become clear when we formally introduce our models and present results of our computational experiments). In both cases, we formulate and solve novel mixed integer programming models, where, because of the size of real-life instances, we have to exploit the special structure of the models and, in one case, rely on column generation techniques for their solution.

To summarize, the main contributions of this study are the following.

- The design and implementation of technology to optimize the air operations of an express shipment service network consisting of about 150 airport and 300 ground hubs, with a fleet of about 60 aircraft with capacities ranging from 14 to 42 tons, handling more than 650,000 timed origin-destination demands in four different service classes.

- For the small-package market, we develop a novel direct shipment model that differs substantially from the traditional hub-and-spoke model employed and achieves much higher capacity utilization. For the high-value products market, we develop a novel transshipment model that simultaneously decides the terminals that act as transshipment hubs and the pickup and delivery routes to and from these hubs. Different from what can be found in the literature, our models (1) integrate ground and air transport options, (2) consider company-owned and purchased air capacity, (3) consider multiple service products, and (4) consider realistic order arrival patterns.

- We propose efficient exact algorithms that exploit the specific structure of the two models to solve large-scale real-life instances.

- We conduct an extensive computational study using real-life data to obtain valuable managerial insights. A few insights resulting from our analyses are that:

- Simultaneously considering ground and air transport options is critical to improving operational efficiency and company profits.
- For serving the small package market:
  * Simultaneously considering company-owned and purchased plane capacity is critical to improving service coverage and company profits.
  * In an environment with ample and reasonably-priced capacity on commercial flights, an air service network design based on direct shipments can be more effective than the commonly used hub-and-spoke design.
  * Among possible improvements in operational efficiency, e.g., reducing loading/unloading times at airports, extending the available flight hours for company-owned cargo planes, increasing the efficiency of intra-city operations (the time it takes from the pick up of a package to the delivery of that package to a gateway hub) has the highest impact on company profits.
  * In an environment in which company-owned cargo planes are restricted to fly during the night and in which there is ample and reasonably-priced capacity on commercial flights during the day, offering next-morning service (by noon) in addition to next-day service (by 6pm) does not require significantly more resources, and, therefore, has the potential to increase profits.
  * As long as the per unit capacity operating cost is below some threshold, the use of larger planes (with higher capacities and lower per unit capacity operating cost) is not advantageous.
- For serving the high-value items/products market:
  * The use of transshipments, allows a high number of origin-destination pairs to be served with a relatively small number of cargo planes.
  * Offering next-morning service (by noon) in addition to next-day service (by 6pm) requires significantly more resources, which means that pricing a next-morning service judiciously is critical to ensure it's profitability.

The remainder of the paper is organized as follows. In Section 2, we discuss related literature. In Section 3, we formally define the problem and introduce our two optimization models. In Section 4, we present and analyze the results of an extensive computation study. In section 5, we conclude with some final remarks.

Before continuing, we want to emphasize that the problems and solution approaches considered in this study are motivated by a collaboration with ShunFeng (SF) Express Company (China), but that it is not an exact representation of the operational environment encountered at SF Express as certain aspects, parameters, and constraints have been altered or have not been captured.

2 Relevant literature

Express shipment service network design falls in the broad class of network design problems (Magnanti and Wong, 1984), which have been extensively studied in the literature. We refer the interested reader to comprehensive reviews by Yang and H. Bell (1998); Crainic (2000); Melkote and Daskin (2001), and Agarwal (2002), and more recently by (Prodhon and Prins, 2014). Here, we focus on recent studies involving settings similar to the ones we consider or involving methodologies similar to the ones we propose.

The express shipment service network design is a particularly challenging variant of service network design. One of the earliest studies on the express shipments is Kuby and Gray (1993), who consider the express delivery operations of Federal Express. Barnhart and Schneur (1996) consider a single-hub overnight delivery system and handle sorting, departure slot, circulation, and fleet size constraints. They propose column generation as a solution approach to handle the large-scale instances encountered by the major U.S. express carriers. Removing the single-hub restriction, Kim et al. (1999) consider a multiple-hub system and propose heuristic approaches to solve large-scale instances arising in practice. Armacost et al. (2002) also consider a multi-hub system and introduce an innovative modeling approach based on so-called *composite variables*, which represent groups of routes carrying the shipment volume associated with a subset of the gateway hubs. They show that the use of composite variables provides tight linear programming bounds, which extends the size of instances for which an optimal solution can be found. To deal with the enormous size of the express delivery problems encountered in practice, Barnhart et al. (2002) suggest an iterative solution framework that decomposes the service network design problem into two subproblems: route generation and shipment movement. Using data from a US express delivery company, they demonstrate that their approach produces operating cost savings and provides a valuable tool for evaluating strategic and tactical decisions. Fleuren et al. (2013) also employ a model using composite variables in their study of express shipment service network design for European carrier TNT Express. Louwerse et al. (2014) consider a multi-hub network with multiple service classes and suggest a two-stage heuristic that builds on composite variable based models. Testing their heuristic on modified instances of an express shipment carrier, they demonstrate that their heuristic can produce service network designs that can significantly reduce the transportation costs. Quesada Perez et al. (2016) suggest a new formulation in which they relax the fixed assignment assumption made by most of the previous studies. They present valid inequalities to strengthen their formulation and test their formulation on realistic instances derived from FedEx Europe data to show its efficacy. Because of the tight service guarantees offered in the express shipment industry, air transportation links are a critical component in the service network. Lin and Chen (2004) study a two-echelon (hierarchical) network design problem in which some links represent transport by airplane, which allows meeting tight service time requirements. An implicit enumeration method embedding a least-time path algorithm was developed to solve the resulting network design problem. Meuffels et al. (2015) consider the simultaneous design of the road and air service network, present a two-phase solution approach, and test it on real-world data to show its efficacy. More recently, Martin et al. (2019) consider integrated service selection, pricing, and fulfillment planning for express parcel carriers. That is, they examine the interplay between service pricing and demand, and its impact on the fulfillment process. They propose a genetic algorithm to solve their optimization model. A set of randomly generated test instances is used to assess the efficacy of the proposed methodology. Other extensions of service network design problems that have been explored recently, include stochastic demands (Ng and Lo, 2016), asset balance constraints (Bai et al., 2016), end-to-end service constraints (Balakrishnan et al., 2017), and resource acquisition (Hewitt et al., 2019).

Our study differs from the existing literature in several respects. All of the aforementioned studies focus on hub-and-spoke networks, whereas we do not impose any network structure. In our direct shipment model, we only consider transporting shipments from one airport to another, eliminating the need for sorting and repackaging at a hub airport. Büdenbender et al. (2000) explore the same idea for a direct-flight postal delivery problem and develop a hybrid tabu search/branch-and-bound solution methodology. Seeking an understanding of the impact of different parameter settings for their algorithm on the resulting service networks, they experiment with 19 instances ranging in size from 485 to 3004 requests (origin-destination demands). In our transshipment model, we consider transshipments, i.e., some locations function as "hubs" where planes meet and exchange their cargo, but different from more traditional hub-and-spoke systems, no sorting takes place at the the transshipment hubs. Another important difference is our treatment of ground transportation. In most previous studies, the assignment of stations (cities without airports) to gateway hubs (cities with airports) are fixed. In our models, such assignments are an integral part of the service network design. Contrary to most traditional express shipment service network design settings, our models consider both company-owned capacity as well as purchased capacity (belly capacity available on commercial flights), which is especially important in China. Finally, we consider an environment in which the express carrier offers multiple service classes, and in which shipments enter the air transportation network throughout the day.

As observed above, many express shipment service network design models are solved using column generation and its generalization to integer programs, i.e., branch-and-price (Barnhart et al., 1998), including, among others, Barnhart and Schneur (1996); Barnhart et al. (2000), and Armacost et al. (2002). Yıldız and Karaşan (2017) recently introduced an innovation to path-based formulations that has inspired several of our formulation ideas. They propose a formulation in which partial columns (path segments) are used instead of whole columns (paths). Considering path-segments as variables and concatenating path-segments to construct complete paths makes it easy to consider non-simple path solutions and to include certain types of constraints on paths, which are difficult to incorporate in a standard path-based formulations. Yıldız et al. (2016) and Yıldız et al. (2018) present applications of path-segment formulations and a branch-and-price algorithm to solve large-scale network design problems arising in different transportation and telecommunications applications. In our transshipment model, we also employ the notion of path-segments, as we divide an aircraft's itinerary into a pickup segment and a delivery segment, which enables us to significantly simplify the spatial and temporal coordination.

3 Problem definition

In this section we describe the service network design problem of a large express carrier to find optimal package movements that maximizes the carrier's profit.

We consider a set of inter-city shipments. Each shipment has an origin, a destination, a weight, a pick up time, i.e., the time it is picked up at the origin, and a service class, which determines the due time, i.e., the latest time it can be delivered at its destination. The shipments are grouped into demands based on their origin, destination, and service class. Each demand has an origin city, a destination city, a weight, a ready time, and a due time. The ready time represents the time at which the shipments that constitute the demand are available for transport at the city's outbound gateway hub (the airport in case the city has an airport and a sorting center in case the city does not have an airport). Shipments are grouped into demands using 30-minute time intervals. That is, a demand with ready time of 18:00 constitutes all shipments that become available for transport at the gateway hub between 17:30 and 18:00. The weight of a demand is the sum of the weights of the shipments that constitute the demand. The due time at the destination represents the time at which the shipments need to reach the inbound gateway hub of the destination city to be able to be delivered in time to meet the service specified by the service class. In defining demands, it is assumed that the intra-city time is known, i.e., the time from when a package is picked up to the time it is available for transport at the city's outbound gateway hub as well as the time from when a package arrives at the city's inbound gateway hub to the time at is deliver at its final destination, and that these intra-city times are the same for every package.

The express package carrier owns and operates a number of cargo planes (of various types) for air transportation. Due to government regulations, cargo airplanes can only be operated during a given time window; between 11pm and 8am in China. The express package carrier has the option to purchase air transportation capacity in the form of belly capacity on commercial flights. The carrier can choose from a set of available commercial flights, each with a given origin city, origin departure time, destination city, destination arrival time, available belly capacity, and per-unit cost.

We refer to cities where the express package carrier has airport operations, i.e., package sorting, as hub cities, and all other cities as non-hub cities. Note that a non-hub city can have an airport and that it may be possible to purchase capacity on commercial flights departing from or arriving at that airport.

If beneficial, a combination of ground and air transportation can be used when serving a demand, for example when the origin and/or destination city (or both) are non-hub cities. Furthermore, we note that in some situations, even though ground transportation is cheaper than air transportation, it may still be preferable to serve a demand by air rather than ground, for example to avoid cargo plane capacity to go unused.

The express package carrier earns a revenue from each demand depending on the origin and destination, the size, and the service class. The express package carrier seeks to design a service network that maximizes its profit, i.e., the revenue from demands it serves minus the cost of the belly capacity purchased (i.e., we assume the use of its own cargo planes is "free"). The design of the service network involves determining which demands to serve, determining which routes and schedules to operate with the company-owned cargo planes, and determining what capacity to purchase on commercial flights.

Observe that we are focusing on the design of air service network operations, i.e., the aircraft routes and schedules. The demand can be based on historic data, but may also be adjusted to reflect market trends, or to account for anticipated growth as a result of pricing policy changes, etc. We develop different models for different markets, as the company has to decide how to best use its fleet of aircraft, and this may mean allocating (a number of) aircraft to each of the markets they serve.

Next, we introduce the notation used throughout the paper. Let Q be the set of demands. Each demand $q \in Q$ is characterized by a 6-tuple $\langle o^q, d^q, f^q, r^q, t^q, \bar{t}^q \rangle$, where o^q is the demand's origin, d^q its destination, f^q its weight, t^q its ready time, $\bar{t}^q$ the due time, and r^q the revenue earned if the demand is served. The total weight and the total potential revenue are denoted by F and R, respectively, (i.e., $F = \sum_{q \in Q} f^q$ and $R = \sum_{q \in Q} r^q$). The set of demands with weight less than threshold value $\bar{w}$ is denoted by $Q(\bar{w})$, i.e., $Q(\bar{w}) = \{q \in Q : f^q \leq \bar{w}\}$. Similarly, we define $F(\bar{w}) = \sum_{q \in Q(\bar{w})} f^q$ and $R = \sum_{q \in Q(\bar{w})} r^q$.

Let the set of cities, the set of cities with an airport, the set of hub cities, and the set of hub cities where transshipment can take place be denoted by C, A, H, and H^T, respectively. Between any two cities $i, j \in C$, let t^g_{ij} denote the ground travel time from i to j. Similarly, between any two cities with an airport $i, j \in A$, let the air travel time from i to j be denoted by t^a_{ij}. The travel times are assumed to include any time required for handling operations at the gateway hubs. The set of cargo plane types owned by the express carrier is denoted by K. For each plane type $k \in K$, the number of cargo planes owned is denoted by η_k. The capacity of a plane of type $k \in K$ is denoted by w_k. The cargo planes are only allowed to operate between $[\underline{\phi}, \overline{\phi}]$. The set of commercial flights with belly capacity is denoted by B. Each commercial flight $b \in B$ is characterized by a 5-tuple $\langle o_b, d_b, c_b, w_b, t_b, \bar{t}_b \rangle$, where o_b is the departure city, d_b is the destination city, c_b is the cost per unit weight, w_b is the available capacity, t_b is the latest cargo accept time, and $\bar{t}_b$ is the arrival time at the destination city.

To accommodate different demand distributions and densities (specific to the markets served), we propose two different models in our study. Next, we present details of each model and their respective solution approaches.

3.1 Direct Shipments Model

In the direct shipments model (DSM), we enforce that any demand transported by air only occupies one flight leg. That is, in DSM, a demand is either transported from its origin city to its destination city by ground transportation, or it is transported in three phases: (1) from its origin city to an airport city using ground transportation, (2) from an airport city to another airport city on a single flight, either on a company-owned cargo plane or on a commercial plane (i.e., using purchased belly capacity), and (3) from an airport city to its destination city using ground transportation. In the latter case, one or both of the ground transportation phases may be "empty". DSM makes the most effective use of the company-owned cargo plane capacity, as any demand transported using the cargo planes uses the capacity for the shortest possible distance.

In a solution to DSM, any demand uses one flight leg of a cargo plane's route. In contrast, in a traditional hub-and-spoke air service network design, a demand typically uses more than one flight leg and thus consumes plane capacity for a longer period of time. At first glance, the capacity utilization in a hub-and-spoke network can look higher, but this does not necessarily imply a more effective air service network, i.e., that more demand is served by the cargo planes. To illustrate this point, consider a hub-and-spoke network in which all demand is first picked up and transported to a hub node, where it is sorted and then transported and delivered at its destination. Assume that each plane executes one pick-up and one delivery path (as is often the case in practice and in most studies in the literature (Kuby and Gray, 1993; Barnhart and Schneur, 1996; Kim et al., 1999; Armacost et al., 2002; Fleuren et al., 2013)). In that case, the maximum demand that can be served is equal to the total capacity of the planes. However, in a solution to DSM, the demand served by cargo planes can be more than their total capacity. In fact, in our computational experiments, we see that it is more than twice their total capacity!

DSM is useful for markets in which demand is densely distributed over the service area, and cargo plane capacity is the limiting resource. Using cargo planes to serve a limited set of high-volume origin-destination pairs with direct flights is an effective strategy to maximize the utilization of the available transportation capacity.

Two major decisions have to be made simultaneously: the routes of each of the company-owned cargo planes and the set of demands to be served using air transportation (either cargo plane or belly capacity). The route of a cargo plane may contain cycles and can start and end at different hub-cities, but it has to start and end within the time window allowed for cargo flights. To ensure a repeatable schedule, we enforce cargo plane balance at each hub-city (the same number of cargo planes of a particular type has to depart and arrive). Note that this does not mean that a cargo plane starts and ends at the same city. If not, a pilot can either stay overnight away from his home base or can be deadheaded back to his home base using commercial flights.

Since each demand represents a set of packages with the same origin, destination, ready and due time, we allow fractional allocations of demands (as done by our industry partner). Next, we introduce the model specific notation.

We discretize time to control the size of the model. Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ be a set of time intervals that span the time window $[\underline{\phi}, \overline{\phi}]$ during which cargo planes are allowed to operate. For each $\theta \in \Theta$, the start time of the interval is denoted by t_θ and the end time of the interval is $\bar{t}_\theta$, i.e., $t_{\theta_1} = \underline{\phi}$, $t_{\theta_i} = \bar{t}_{\theta_{i-1}}$ for $i = 2, \ldots, m$, and $\bar{t}_{\theta_m} = \overline{\phi}$.

Let $\mathcal{L}$ be the set of all flight legs, or lines, that can be executed by the carrier-owned cargo planes. A line $\ell \in L$ is represented by a 3-tuple $\langle o_\ell, d_\ell, \theta_\ell \rangle$, where o_ℓ is the origin, d_ℓ is the destination, and θ_ℓ is the time interval in which the line starts. An itinerary (route) for a company-owned cargo plane is an ordered set of lines $p = \{\ell_1, \ell_2, \ldots, \ell_m\}$, where a route p is *feasible* if the following spatial and temporal consistency conditions are satisfied:

- Except for the first line, each line in p starts where the preceding line ends, i.e., $d_{\ell_{i-1}} = o_{\ell_i}$ for all $i \in \{2, \ldots m\}$.
- Except for the first line, each line in p start after the preceding line's arrival time plus the time needed to complete ground operations τ, i.e., $\bar{t}_{\theta_{\ell_{i-1}}} + \tau \leq t_{\theta_{\ell_i}}$ for all $i \in \{2, \ldots m\}$.

With each route p, we also associate a plane type p_k (i.e., the same ordered set of lines can define different routes by associating a different plane type). Thus, a route p has an associated capacity w_p, which is the capacity of the plane type associated with the route. The set of all feasible routes is denoted by P and the set of feasible routes that include a line $\ell \in L$ is denoted by $P(\ell)$. Similarly, the set of feasible routes that use a plane of type $k \in K$ is denoted by $P(k)$.

Note that a demand $q \in Q$ can be transported by a line $\ell \in \mathcal{L}$, if the demand can reach o_ℓ (using the ground transportation) before the line's departure time t_{θ_ℓ}, and it can reach its destination in time (using ground transportation) after the line's arrival time $\bar{t}_{\theta_\ell}$ at d_ℓ. More formally we say a line $\ell \in \mathcal{L}$ can support a demand $q \in Q$, if $t^q + t^g_{o^q,o_\ell} \leq t_{\theta_\ell}$ and $\bar{t}_{\theta_\ell} + t^g_{d_\ell,d^q} \leq \bar{t}^q$. We denote the set of all demands that can be supported by a line ℓ as $Q(\ell)$ and the set of lines that can support a demand $q \in Q$ by $\mathcal{L}(q)$.

A demand can also use purchased belly capacity. We say that a commercial flight $b \in B$ can support a demand $q \in Q$, if $t^q + t^g_{o^q,o_b} \leq t_b$ and $\bar{t}_b + t^g_{d_b,d^q} \leq \bar{t}^q$. The set of demands that can be supported by a commercial flight $b \in B$ is denoted by $Q(b)$. Similarly, the set of commercial flights that can support a demand $q \in Q$ is denoted by $B(q)$.

We define the following decision variables:

- x^q_ℓ: the amount of demand $q \in Q$ transferred by line $\ell \in L(q)$;
- y^q_b: the amount of demand $q \in Q$ transferred by commercial flight $b \in B(q)$;
- z^q: the amount of demand $q \in Q$ not served; and
- u_p: the number of planes that execute route p.

With these decision variables we define the mixed integer programming formulation of DSM as follows (and in the remainder we denote it by DS).

$$\max \sum_{q \in Q} (f^q - z^q) r^q - \sum_{q \in Q} \sum_{b \in B(q)} c_b y^q_b \tag{1}$$

$$\text{s.t.} \sum_{\ell \in L(q)} x^q_\ell + \sum_{b \in B(q)} y^q_b + z^q = f^q \qquad \forall q \in Q \tag{2}$$

$$\sum_{q \in Q(\ell)} x^q_\ell \leq \sum_{p \in P(\ell)} w_p u_p \qquad \forall \ell \in L \tag{3}$$

$$\sum_{q \in Q(b)} y^q_b \leq w_b \qquad \forall b \in B \tag{4}$$

$$\sum_{p \in P(k)} u_p \leq \eta_k \qquad \forall k \in K \tag{5}$$

$$\sum_{\substack{p \in P(k) \\ p_o = i}} u_p - \sum_{\substack{p \in P(k) \\ p_d = i}} u_p = 0 \qquad \forall k \in K, \forall i \in H \tag{6}$$

$$x^q_\ell \geq 0 \qquad \forall q \in Q, \ell \in L(q) \tag{7}$$

$$y^q_b \geq 0 \qquad \forall q \in Q, b \in B(q) \tag{8}$$

$$u_p \geq 0 \text{ and integer} \qquad \forall p \in P \tag{9}$$

The objective is to maximize profit, which accounts for the revenues earned from served demand and the costs incurred for using belly capacity on commercial flights. Constraints (2) ensure that the amount of demand shipped by cargo planes plus the amount of demand shipped using belly capacity on commercial flights plus the amount of unserved demand equals the total demand. Constraints (3) ensure that the cargo planes assigned to a line provide sufficient capacity. Similarly, Constraints (4) ensure that there is enough belly capacity to serve demand allocated to commercial flights. Constraints (5) ensure that number of cargo planes of type k used does not exceed the number of planes of type k available. Finally, Constraints (6) enforce that, for a given plane type, the number of planes that start their itinerary at a given hub-city is equal to the number of planes that end their itinerary at that hub-city. These constraints ensure that the resulting schedule can be repeated on a daily basis.

In practice, the number of plane types $|K|$ is usually small, however, the number of routes can get very large as the number of hub-cities $|H|$ and number of time intervals $|\Theta|$ increases. For example, assuming 10 hours of flight time, flight legs of two hours, a single plane type ($|K| = 1$), four hub cities ($|H| = 4$), and a one-hour discretization ($|\Theta| = 10$), we have 8208 feasible routes. If instead, we assume two plane types ($|K| = 2$), ten hub cities ($|H| = 8$), and a 30-minute discretization ($|\Theta| = 20$), the number of feasible routes is already more than 4.2 million. In our computational experiments, we have an even larger number of feasible routes as we have an instance with three plane types ($|K = 3|$), more than thirty hub cities ($|H| = 34$) and a 15-minute discretization ($|\Theta| = 40$). Clearly, in an optimal solution only a very small fraction of the feasible routes are selected. Thus, we propose a column generation approach to solve real-life instances (for which solving the mixed integer program directly is impossible).

3.1.1 Column generation

In the column generation algorithm (CG), we start with a finite subset of cargo plane routes, i.e., a restricted formulation, and iteratively add variables as needed to reach and prove optimality (of the linear relaxation). In the remainder, we will denote the linear relaxation by DS_{LP}. At each iteration of the column generation algorithm, we solve a *pricing problem* to identify routes that are not in the restricted formulation, but have a positive reduced cost, or conclude that no such routes exist, in which case the algorithm terminates.

Let LP^n be the restricted DS_{LP} formulation in the n^{th} iteration of algorithm CG. Let $\lambda_\ell, \ell \in L$, $\mu_k, k \in K$ and $\alpha_i^k, i \in A$ be the dual variables associated with the Constraints (3), (5) and (6), respectively, for LP^n. Then the reduced cost π_p of route variable u_p, $p \in P$, is given by

$$\pi_p = \sum_{\ell \in p} w_p \lambda_\ell - \mu_{p_k} + \alpha^k_{o_p} - \alpha^k_{d_p}. \tag{10}$$

Next, we show that the pricing problem can be solved by finding, for each plane type $k \in K$, a longest path in a directed acyclic graph. We define the *pricing graph* $G^k(\bar{L}, A)$ for plane type $k \in K$ as follows:

- The node set $\bar{L}$ contains all the feasible lines and artificial source and sink nodes σ, and $\bar{\sigma}$ (i.e., $\bar{L} = L \cup \{\sigma, \bar{\sigma}\}$),
- The arc set A consists of three sets of arcs A_σ, $A_{\bar{\sigma}}$, and A_L, defined as follows
  - $A_\sigma = \{(\sigma, \ell) : \ell \in L\}$,
  - $A_{\bar{\sigma}} = \{(\ell, \bar{\sigma}) : \ell \in L\}$,
  - $A_L = \{(\ell, \bar{\ell}) : \ell, \bar{\ell} \in L \text{ and } \bar{t}_{\theta_\ell} + \tau \leq t_{\theta_{\bar{\ell}}}\}$,
- We assign a weight κ^k_a to each arc $a \in A$ as follows
  - $\kappa^k_{(\sigma,\ell)} = \alpha^k_{o_\ell}$, for all $a \in A_\sigma$,
  - $\kappa^k_{(\ell,\bar{\sigma})} = w_k \lambda_\ell - \alpha^k_{d_\ell}$, for all $a \in A_{\bar{\sigma}}$, and
  - $\kappa^k_{(\ell,\bar{\ell})} = w_k \lambda_\ell$, for all $a \in A_L$.

The following lemma formally states the relation between solving the pricing problem and finding longest paths in the pricing graphs G^k, $k \in K$.

Lemma 1. *Let $p^* = \{\ell_1, \ell_2, \dots \ell_m\} \in P$ be a route of a plane of type $k \in K$. Then, the variable associated with route p^* has the highest reduced cost in DS_{LP} among the variables associated with routes of a plane of type k if and only if the path $\bar{p} = \{\sigma, \ell_1, \ell_2, \dots \ell_m, \bar{\sigma}\}$ is a longest path in the pricing graph G^k.*

Proof. This follows from the fact that for every feasible route $p = \{\ell_1, \ell_2, \dots \ell_m\}$ in $P(k)$, there exists a path $\bar{p} = \{\sigma, \ell_1, \ell_2, \dots \ell_m, \bar{\sigma}\}$ in the pricing graph G^k with total length $l_p = \sum_{\ell \in P} w_k \lambda_\ell + \alpha^k_{o_p} - \alpha^k_{d_p}$, and that for every path $\bar{p} = \{\sigma, \ell_1, \ell_2, \dots \ell_m, \bar{\sigma}\}$ in the pricing graph G^k, there exists a feasible route in $p = \{\ell_1, \ell_2, \dots \ell_m\}$ with reduced cost $\pi_p = l_p - \mu_k$. □

Lemma 1 implies that the pricing problem can be solved by finding a longest path in G^k from the artificial source σ to the artificial sink $\bar{\sigma}$ for each $k \in K$. Clearly, if for all $k \in K$, the longest path in G^k has length less than μ_k, then one can conclude that there exist no routes not already in the restricted problem that have a positive reduced cost.

Although solving a longest path problem in a general graph is hard, solving a longest path problem in the pricing graph G^k is easy, because it is acyclic. The pseudo code for the longest path algorithm LP can be found in Algorithm 2, in Appendix A.

Finding the highest reduced cost variable associated with a route of a plane of type k for each $k \in K$, and choosing the one with the highest reduced cost overall solves the pricing problem. However, preliminary computations have shown that it is beneficial to simultaneously add multiple variables with positive reduced cost as it tends to reduce the number of column generation iterations.

As the ultimate goal is not only to solve the linear relaxation, but instead to obtain a high-quality feasible solution, choosing which variables (with positive reduced cost) to add to the restricted model is an important decision. In our setting, it is natural to consider routes that serve different lines. So we employ an iterative approach (Algorithm MPCG) to solve the pricing problem and add multiple routes to the restricted model at each column generation iteration. We first find a route with the highest reduced cost. Next, we remove all its lines from the pricing graph and then find a route with the highest reduced cost in the resulting reduced pricing graph. We repeat these steps as long variables with positive reduced cost are found and the total number of such variables does not exceed a limit C. The pseudo code for MPCG can be found in Algorithm 1.

3.1.2 IP Solution

A branch and price (BP) algorithm can be developed to solve DS by embedding the CG algorithm in a branch and bound scheme. However, a much simpler scheme is already able to obtain integer solutions of proven high quality. Preliminary computational experience has shown and that the well-known heuristic approach that solves an IP with the columns generated during the solution of the linear relaxation is very effective and produces near-optimal solutions in almost all cases. We next show that with a slight modification, we can strengthen the DS formulation to improve the computational efficiency and solution qualities of our solution approach.

Note that the main difference between a fractional and an integer solution to DS is the use of "excess capacity". To address this issue, we explicitly consider the capacity required on a given line. Let $\bar{w}_\ell$ be the total volume of the demands that can potentially use ℓ (i.e. $\bar{w}_\ell = \sum_{q \in Q(\ell)} f^q$). For each plane type $k \in K$ and line $\ell \in L$, we define $w^*_{k,\ell} = \min\{\bar{w}_\ell, w_k\}$. Then, replacing (3) with inequality

$$\sum_{q \in Q(\ell)} x^q_\ell \le \sum_{p \in P(\ell)} w^*_{p,\ell} u_p \qquad \forall \ell \in L, \tag{11}$$

Algorithm 1: MPCG Algorithm

Input: $\boldsymbol{\lambda}, \boldsymbol{\mu}, C$
Output: $\mathcal{P}$
Set $\mathcal{P} = \emptyset$;
Order the set of plane types K from smallest to largest capacity ;
foreach $k \in K$ **do**
  Generate G^k;
  Set $\pi = 1$;
  while $|\mathcal{P}| < C$ *and* $\pi > 0$ **do**
    Find longest path p^* in G^k ;
    if $\pi_{p^*} > 0$ **then**
      Add p^* to $\mathcal{P}$;
      Update G^k by removing all arcs used by p^* ;
    Set $\pi = \pi_{p^*}$
return $\mathcal{P}$

we obtain a tighter formulation, which we denote by $\overline{DS}$. Note that replacing (3) with (11) requires a small change in the pricing problem: the arc weights w_p have to be replaced by arc weights $w^*_{p,\ell}$.

3.2 Transshipment Model

In the transshipment model (TSM), a demand transported by air may occupy more than one flight leg, often, but not necessarily, on more than one cargo plane. That is, cargo planes are allowed to meet in a transshipment location to exchange some or all of their cargo. Conceptually, each demand uses one pickup and one delivery flight that meet at a transshipment location. If the transshipment location happens to be the hub-city where the demand enters or exits the air network, then the demand uses only a single flight (either a pickup or a delivery flight). In TSM, each cargo plane's route is composed of two distinct parts: a pickup flight and a delivery flight (where one of them may be the "empty" flight). On the pickup flight, the cargo plane collects demands at various hub-cities to take them to a transshipment location. On the delivery flight, the cargo plane takes demands from the transshipment location and drops them of at various hub-cities. As in the direct shipment model, ground transportation can be used to transfer demand from/to a non-hub city to/from a hub-city. The TSM makes most effective use of the flying time available for company-owned cargo planes to provide broad coverage, i.e., seeks to provide connections between a large number of cities (possibly at the expense of reduced capacity utilization). It is important to note that TSM only considers *transfers*, i.e., cargo can be transferred from one plane to another at certain designated locations, but no sorting takes place at these locations.

The TSM is most useful for a high-value items/products market, in which demand is relatively small and sparsely and fairly evenly distributed over the service area and available flight time is the limiting resource (rather than cargo plane capacity) and where the use of belly capacity on commercial flights is not an option due to special packaging and/or handling requirements. We note that if the use of belly capacity on commercial flights is an option, it is not hard to extend the model to accommodate this possibility; see Appendix B for details.

We assume that there is a single plane type (even though in practice package express carriers may employ multiple plane types). We believe this is a reasonable assumption, since in this setting the demands are assumed to be small compared to cargo plane capacities (and the equipment used for transfer operations at airports is independent of the plane type).

We further assume that both the pickup and the delivery flights are simple paths, i.e., a hub-city is visited at most once on a pickup or delivery path (we will show that this is not a restrictive assumption and that there always exists an optimal solution that only uses simple paths). A pickup path can be represented by a sequence of flight legs, $p = (a_1, a_2, \ldots, a_m)$ with $a_m \in H$ and a delivery path can be represented by a sequence of flight legs $p = (a_1, a_2, \ldots, a_m)$ with $a_1 \in H$. For notational convenience, we denote the origin and destination of a path p with o_p and d_p, respectively. The duration of a path p is denoted by δ_p. We will only consider pickup and delivery paths that start and end in the allowed operating window $[\underline{\phi}, \overline{\phi}]$ and we will refer to such paths as *feasible*. A pickup or delivery path $p = (a_1, a_2, \ldots, a_m)$ is called an *actual-path* if $m \geq 1$ and we denote the set of all feasible actual pickup and delivery paths with P and $\bar{P}$, respectively. For technical reasons that will become clear soon, we also consider *empty paths*, i.e., we allow $m = 0$. The set of empty pickup and delivery paths are denoted by P^o and $\bar{P}^o$, respectively. We define $\mathcal{P} = P \cup P^o$ as the set of all feasible pickup paths and $\bar{\mathcal{P}} = \bar{P} \cup \bar{P}^o$ as the set of all feasible delivery paths.

For each demand $q \in Q$, let e_p^q denote the lower bound on the arrival time of a plane executing pickup path p at the transshipment location d_p if it serves demand q. Note that the calculation of e_p^q involves determining the hub-city from which the demand q is going to be picked up (where the demand enters the air transportation network) and the time required by ground transportation from o^q to that hub-city. Furthermore, let $\bar{e}_{\bar{p}}^q$ denote the time it takes for demand q to reach its destination d^q (including any ground transportation), if it is served on delivery path $\bar{p} \in \bar{P}$. Note that the calculation of $\bar{e}_{\bar{p}}^q$ involves determining the hub-city where demand q will leave the air transportation network to reach its final destination.

Let τ denote the time required to transship packages from one plane to another at a transshipment location. For each demand $q \in Q$, a pickup path p is called a *plausible* pickup path if there exists a delivery path $\bar{p} \in \bar{\mathcal{P}}$ such that $d_p = o_{\bar{p}}$ and $e_p^q + \tau + \bar{e}_{\bar{p}}^q \leq \phi_{d^q}$. A *plausible* delivery paths for a demand $q \in Q$ is defined similarly. The sets of all plausible pickup and delivery paths for demand $q \in Q$ are denoted by $\mathcal{P}(q)$ and $\bar{\mathcal{P}}(q)$, respectively.

In TSM, the movement of a package from its origin to its destination is thought of as consisting of four components. In the first component, ground transportation is used to reach the location where the package enters the air transportation network. In the second component, a pickup path is used to take the package to a transshipment location. In the third component, a delivery path is used to take the package to a location where it exits the air transportation network. In the final component, ground transportation is used to reach the package' final destination. Each component represents a *segment* of the path from origin to destination. TSM seeks to find feasible concatenations of segments for each demand so as to maximize company profit. Concatenation of segments, however, requires both spatial and temporal coordination to ensure that feasible paths are constructed for the demands. The following observations allow us to construct feasible concatenations without explicitly modeling the ground transportation components by incorporating their timing information in the pickup and delivery paths.

Observation 1. *Let $p \in \mathcal{P}$ be a pickup path and let $\hat{Q} \subset Q$ be the set of demands served by p, then the earliest time t^* that a plane executing p can arrive at d_p is $\max_{q \in \hat{Q}}\{e_q^p\}$.*

Observation 2. *Let $\bar{p} \in \bar{\mathcal{P}}$ be a delivery path, let $\hat{Q} \subset Q$ be the set of demands served by $\bar{p}$, and let the earliest departure time of a plane executing delivery path p be t^* (implied by the latest arrival time of a demand in $\hat{Q}$ and the transshipment time τ), then each demand $q \in \hat{Q}$ can arrive at its destination at time $t^* + \bar{e}_{\bar{p}}^q$.*

Observation 3. *Let p be a non-simple pickup path and let $\hat{Q} \subset Q$ be the set of demands served by p. Then there exits a simple path $\hat{p} \in \mathcal{P}$ that visits the same set of locations, picks up the same set of demands, and arrives at the transshipment location at the same time as path p.*

This follows from the fact that whenever a location is visited more than once, it is feasible to pick up all demand at that location at the time of the last visit.

Observation 4. *Let p be a non-simple delivery path and let $\hat{Q} \subset Q$ be the set of demands served by p. Then there exits a simple path $\hat{p} \in \mathcal{P}$ that visits the same set of locations, delivers the same set of demands, and departs the transshipment location at the same time as path p.*

This follows from the fact that whenever a location is visited more than once, it is feasible to deliver all demand at that location at the time of the first visit.

Next, we present a mixed integer programming formulation for TSM. We introduce following decision variables:

- *Time-coordination* variables $t_h, h \in H^T$, indicating the latest arrival of pickup paths at transshipment location $h \in H^T$;
- *Pickup-path* variables $u_p, p \in \mathcal{P}$, indicating whether a pickup path is used or not;
- *Delivery-path* variables $u_{\bar{p}}, \bar{p} \in \bar{\mathcal{P}}$, indicating whether a delivery path is used or not;
- *Demand assignment* variables x_p^q, $p \in \mathcal{P} \cup \bar{\mathcal{P}}$, indicating whether a demand $q \in Q$ is served on path $p \in \mathcal{P} \cup \bar{\mathcal{P}}$ or not;
- *Coverage* variables y^q, $q \in Q$, indicating whether a demand q is served or not; and
- *Transshipment* variables z_h^q, $q \in Q, h \in H^T$, indicating whether a demand q is transferred at transshipment location $h \in H^T$ or not.
- *Empty-path* variables ϵ_i^q/ $\bar{\epsilon}_i^q$, $i \in H^T$, indicting whether a demand $q \in Q$ is served with a empty pickup/delivery path at transshipment location i or not.

With these decision variables formulation TS is defined as follows:

$$\max \sum_{q\in Q} r^q y^q \tag{12}$$

$$\text{s.t.} \sum_{h\in H^T} z_h^q \geq y^q \qquad \forall q \in Q, \tag{13}$$

$$\sum_{\substack{p\in P(q)\\ d_p=h}} x_p^q + \epsilon_h^q \geq z_h^q, \qquad \forall q \in Q, \forall h \in H^T \tag{14}$$

$$\sum_{\substack{\bar{p}\in \bar{P}(q)\\ o_p=h}} x_{\bar{p}}^q + \bar{\epsilon}_h^q \geq z_h^q, \qquad \forall q \in Q, \forall h \in H^T \tag{15}$$

$$t_h \geq \sum_{\substack{p\in P(q)\\ d_p=h}} e_p^q x_p^q + t_{o^q,h}^g \epsilon_h^q \qquad \forall q \in Q, \forall h \in H^T,, \tag{16}$$

$$t_h + \tau + \sum_{\substack{\bar{p}\in \bar{P}(q)\\ s_{\bar{p}}=k}} \bar{e}_{\bar{p}}^q x_{\bar{p}}^q + t_{h,d^q}^g \bar{\epsilon}_h^q \leq \bar{t}^q \qquad \forall h \in H^T, \forall q \in Q(h), \tag{17}$$

$$t_{o_{\bar{p}}} + \tau + \delta_{\bar{p}} u_{\bar{p}} \leq \overline{\phi} \qquad \forall \bar{p} \in \bar{P} \tag{18}$$

$$\sum_{p\in \mathcal{P}} u_p = \rho, \tag{19}$$

$$x_p^q \leq u_p \qquad \forall q \in Q, \forall p \in P(q) \tag{20}$$

$$x_{\bar{p}}^q \leq u_{\bar{p}} \qquad \forall q \in Q, \forall \bar{p} \in \bar{P}(q) \tag{21}$$

$$\sum_{\substack{p\in \mathcal{P}\\ d_p=h}} u_p - \sum_{\substack{\bar{p}\in \bar{\mathcal{P}}\\ s_{\bar{p}}=h}} u_{\bar{p}} = 0 \qquad \forall h \in H^T, \tag{22}$$

$$\sum_{\substack{p\in \mathcal{P}\\ s_p=h}} u_p - \sum_{\substack{\bar{p}\in \bar{\mathcal{P}}\\ d_{\bar{p}}=h}} u_{\bar{p}} = 0 \qquad \forall h \in H^T, \tag{23}$$

$$t_h \geq 0 \qquad \forall h \in H^T, \tag{24}$$

$$u_p \in \{0,1\} \qquad \forall p \in \mathcal{P} \cup \bar{\mathcal{P}} \tag{25}$$

$$x_p^q \in \{0,1\} \qquad \forall q \in Q, \forall p \in \mathcal{P} \cup \bar{\mathcal{P}} \tag{26}$$

$$y^q \in \{0,1\} \qquad \forall q \in Q, \tag{27}$$

$$z_h^q, \epsilon_h^q, \bar{\epsilon}_h^q \in \{0,1\} \qquad \forall q \in Q, h \in H^T \tag{28}$$

The objective is to maximize the profit (total revenue earned from served demand). Constraints (13)-(15) ensure that if a demand $q \in Q$ is served, it is served by a pickup and a delivery path (including empty paths) that arrive at and depart from the same transshipment location. Note that Constraints (13)-(15) ensure the spatial coordination, while Constraints (16)-(17) ensure that the timing of the paths is synchronized and demand q is delivered on time. Constraints (18) ensure that cargo planes respect the operating window. Constraint (19) limits the number of planes used in the air transportation network, and Constraints (20) and (21) ensure that if a pickup or delivery path is used to serve a demand $q \in Q$, there is a plane that executes the path. Finally, Constraints (22) and (23) enforce that the number of arriving and departing planes in a transshipment location match and the resulting schedule can be repeated on a daily bases.

Note that considering the pickup and delivery segments executed by a plane separately, as opposed to considering complete itineraries for a plane, significantly reduces the number of variables. The observations above show that time coordination of pickup and delivery segments can be achieved without introducing timed-copies of path segments (with different visit times at the locations), which also significantly reduces the number of variables. In fact, preliminary experiments have shown that, mostly due to the limited cargo plane operating window (nine hours in practice), it is possible to solve TS directly without the need for more involved approaches, such as column generation or branch and price.

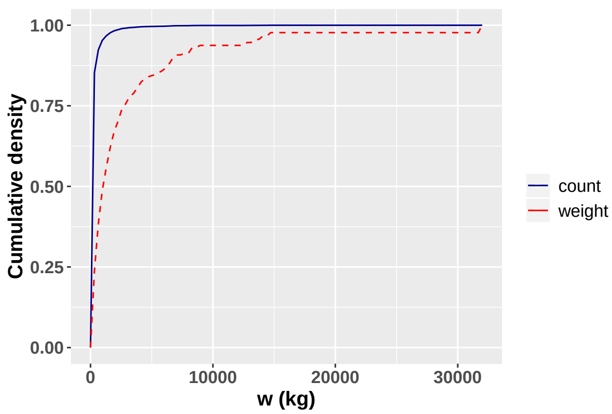
4 Computational Study

In order to evaluate the effectiveness and benefits of our solution methodologies, we have conducted a comprehensive numerical study using data from SF Express (which SF Express uses themselves for strategic network analysis). In the following sections, we discuss these data, how they were used to create instances for the evaluation of the designs produced by our proposed models, and the insights obtained from our computational experiments. Note that we do not include experiments that specifically seek to demonstrate the efficiency of our proposed solution approaches. We believe that being able to solve the large-scale instances used in the experiments is sufficient proof of their efficacy.

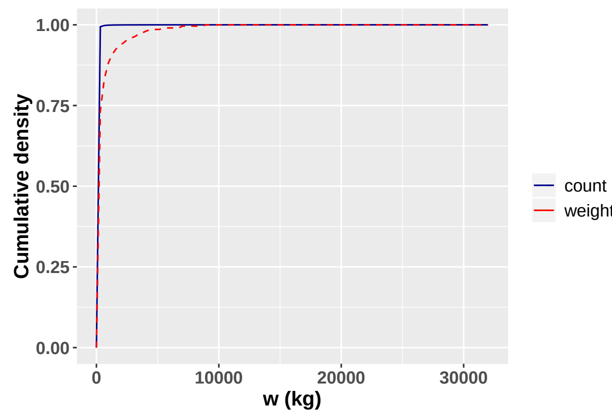
4.1 Data used in our analysis

The data related to demand, owned capacity, and cost and availability of capacity on commercial flights represents the data SF Express uses internally for strategic network analysis.

- The demand represents shipments requiring "next-day service", i.e., the shipments need to be delivered by 6 pm the next day.
- There are more than 100,000 demands, each representing a set of shipments from a specific origin to a specific destination available at a particular time; the weight of a demand represents the total weight of the shipments that make up the demand.
- In Figure 1, we show the cumulative distributions of demand counts (number of origin-destination pairs) and demand weight (weight associated with origin-destination pairs) for demand between hub-cities and demand between cities where in at least one of the cities the carrier does not have an airport operations. For a given value w, the solid line shows the fraction of (the number of) demands with weight at most w (i.e., $\frac{|Q(w)|}{|Q|}$), and the dashed line shows the fraction of the total weight of the demand represented by demands with weight at most w (i.e., $\frac{F(w)}{F}$). We note that in terms of weight, 37.1% of the demand is between hub-cities cities (and, thus, 62.9% of the demand is between cities where in at least one of the cities the carrier does not have airport operations). In terms of counts, 4.2 % of the demands are between hub-cities (95.8% of the demand is between cities where in at least one of the cities the carrier does not have airport operations).
- In Figure 2, we show the ready time distribution of demands between hub-cities cities and of demands where in at least one of the origin and destination city the carrier does not have airport operations.
- In Figure 3, we illustrate the spatial distribution of demand over the service region (i.e., China). Each demand is represented by a line between its origin and its destination with the thickness of the line being proportional to the weight of the demand.
- There are 34 hub-cities, shown as red dots in Figure 3.

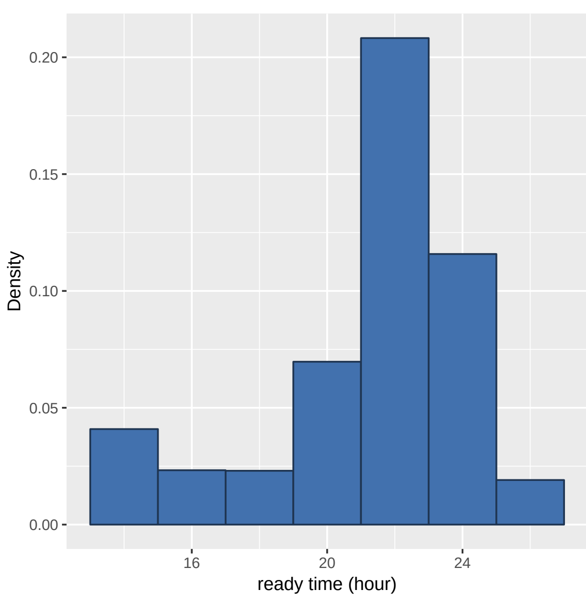


(a) Demand between hub-cities

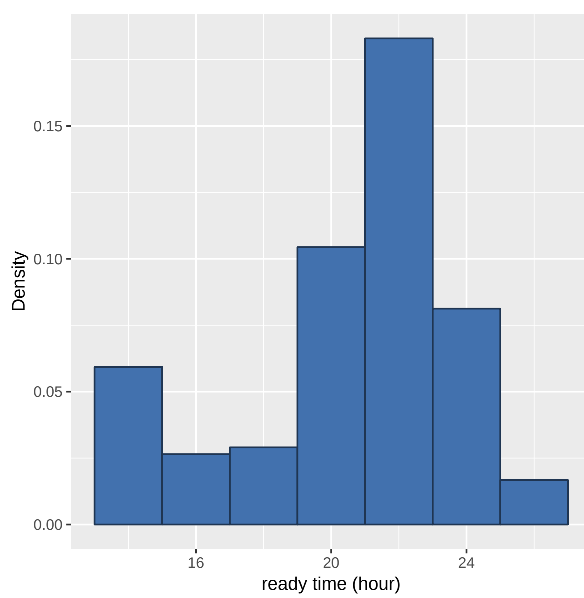


(b) Demand between cities where in at least one city the carrier does not have airport operations

Figure 1: The cumulative density functions of express package demand.



(a) Demand between hub-cities



(b) Demand between cities where in at least one city the carrier does not have airport operations

Figure 2: Distribution of demand ready times.

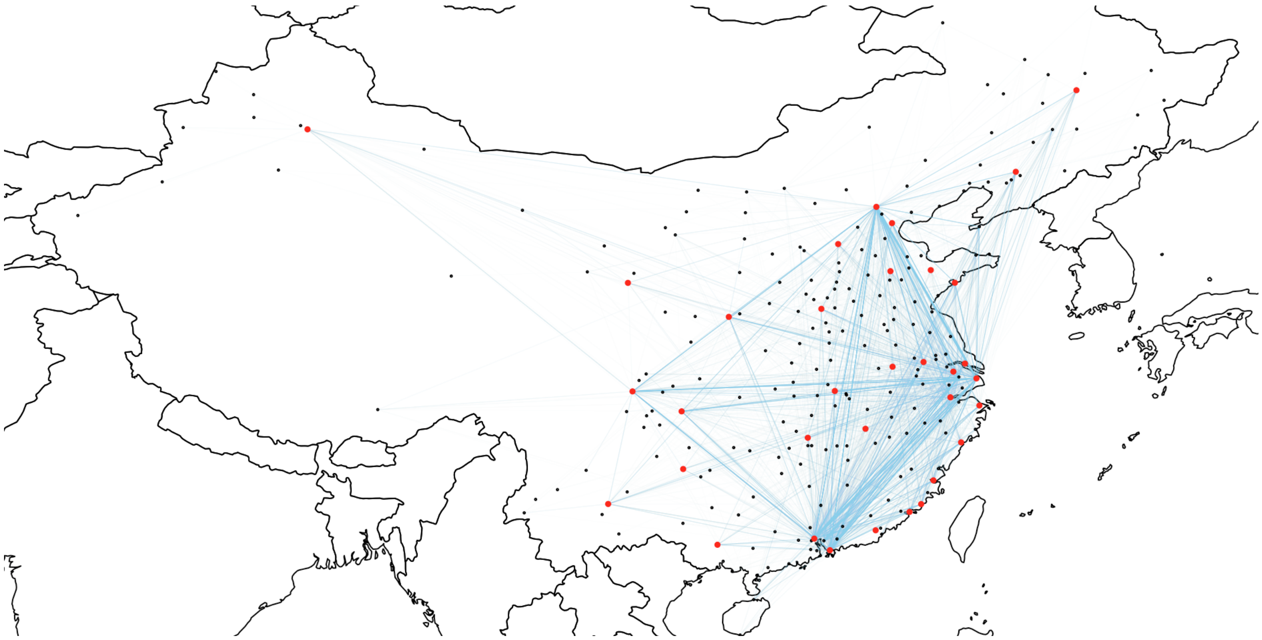


Figure 3: The spatial distribution of express package demand.

- Government regulations impose that company-owned cargo planes can only operate between 11 pm and 8 am.
- There are three types of cargo planes with capacities 14, 28, and 42 tons, respectively, which we will refer to as Type 1, Type 2, and Type 3. The fleet of cargo planes consists of 32 Type 1 planes, 19 Type 2 planes, and 5 Type 3 planes.
- There are more than 6000 commercial flights (from 86 origins to 109 destinations) on which belly capacity is available. The distribution of the cost of belly capacity on these commercial flights is shown in Figure 4. The average cost of belly capacity is \$3.25 per kg.
- Ground transportation is available between almost 2500 city pairs. We assume that ground transportation is flexible and can be "aligned" with the express shipment air network (i.e., ground transportation between a non-airport city and an airport city can be scheduled so that packages reach the gateway hub at the airport city in time to be loaded onto departing flights, and, similarly, ground transportation between an airport city and a non-airport city can be scheduled so that packages arriving at the airport city reach the gateway hub in their non-airport city in time to make service).
- We assume that intra-city operations (both from a shipment's pickup location to the gateway hub and from the gateway hub to the shipment's delivery location) require 5.5 hours and that these intra-city operation times are identical for all cities.
- We assume that handling operations at airports take one hour and that these handling times are identical for all airports.
- We assume that the revenue for demand q is equal to $r^q = 5 + 0.23 t^a_{o_q,d_q}$, where we want to stress that this revenue function does not reflect or represent the prices charged by SF Express. The parameters of the revenue function are obtained by first fitting a linear regression model to the

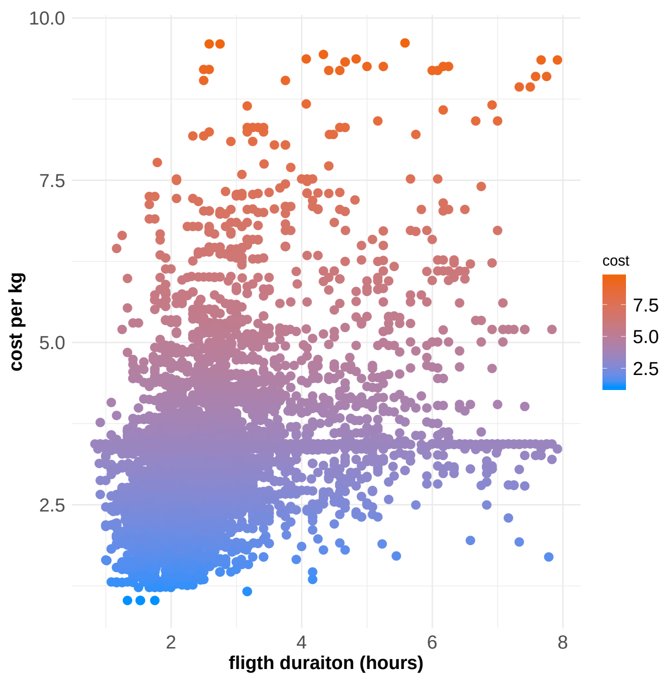


Figure 4: The cost of belly capacity as a function of flight duration.

belly costs (taking the air travel distance as the independent variable) and adding a margin of \$2.73. The revenue function is set up in this way to ensure that, in almost all cases, it is beneficial (i.e., profitable) to serve demand.

- When we consider next-morning deliveries (i.e., shipments are guaranteed to be delivered before noon the next day), we assume the revenue is 50% more than next-day delivery (for the same origin destination pair).

4.2 The Direct Shipment Model

In this section, we seek to answer, among others, the following questions regarding the benefits of using the direct shipment model to analyze and plan express package deliveries.

- What are the characteristics of high quality express shipment air network designs (i.e., company-owned cargo plane routes, purchased capacity, ground transportation, etc.)?
- What operational efficiency improvements have the greatest potential to improve profitability (i.e., improving efficiency of intra-city operations, improving efficiency of airport handling operations, modifying the cargo plane operation window, etc.)?
- How does the composition of the company-owned cargo planes affect the express shipment air network design and profit?
- Can multiple service classes be supported effectively by a single express shipment air network design?
- Does the model effectively integrate the use of ground and air transportation?
- Does the model effectively integrate the use of company-owned and purchased air capacity?

To try and answer these questions, we have conducted five groups of experiments. The groups of experiments aim to investigate: (1) the benefits of multi-modal transportation (i.e., ground and air), (2) the benefits of improvements in the operational parameters, (3) the impact of changes to availability and cost of belly-capacity, (4) the impact of changes in the demand composition (i.e., service class variations), and (5) the impact of changes in fleet composition. Details of each group of experiments are given below.

In the first set of experiments (E1), we consider three settings: the base case (BC), a setting in which no capacity can be purchased on commercial flights, i.e., no belly capacity (NB), and a setting in which no ground transportation is available to take demand to originating at or destined for a non-airport city to/from an airport-city (NG).

In the second set of experiments (E2), we consider twelve variants. We study variants generated by a combination of the following three basic operational parameters.

- Cargo plane operation time window: $[\underline{\phi}, \overline{\phi}] \in \{[23:00, 08:00], [23:00, 09:00], [22:00, 08:00]\}$.
- Intra-city operations time: $\phi_c \in \{4.5, 5.5\}$.
- Airport operations time: $\tau \in \{0.5, 1\}$.

In the third set of experiments (E3), we vary the available belly capacity as well as the cost of using belly capacity on commercial flights. We consider the following variants:

- Available belly capacity is multiplied by $\beta^w = \{0.5, 0.75, 1, 1.25, 1.5\}$; and
- Cost of belly capacity is multiplied by $\beta^c = \{0.5, 0.75, 1, 1.25, 1.5\}$.

In the forth set of experiments (E4), we consider four variants by assuming that $\sigma = \{0, 10\%, 20\%, 30\%\}$ of demand represents shipments with service class next-morning.

Finally, in the fifth set of experiments (E5), we consider five levels of total capacity: 252, 504, 756, 1008, 1260 and 1512 tons. For each total capacity level, we consider three variants in which the fleet of cargo planes is composed of a single plane type with 14, 28, or 42 tons of capacity.

4.2.1 Implementation Details

As shown in Figure 1, the bulk of the total demand (by weight) is composed of relatively small number of demands (which also happen to have origins and destinations that are geographically close and in the center of the service area). Clearly, these large demands are natural candidates for transportation by cargo planes as it leads to high capacity utilizations. Consequently, the many demands with small weight do not play a significant role in determining the cargo plane routes; belly capacity on commercial flights is more suitable for these demands (which are also more likely to occur near the periphery of the service area). Indeed, preliminary experiments have revealed that the cargo plane routes are (mostly) determined by demands that are larger than some threshold w. So, from a computational efficiency perspective, it is a sensible approach to solve DS in two steps. In the first step, DS is solved with demands $Q(\bar{w})$ to determine the cargo plane routes. Then, in the second step, the problem is resolved with the entire set of demands Q, but with the cargo plane routes fixed. In Table 1, we show the solutions that we obtain for the base case when we apply this two-step process, for various threshold values $\bar{w}$. The column headings CP, $Belly$ and NS indicate the total weight of the demand shipped by the company-owned cargo planes, the demand shipped on commercial flights, i.e., using belly capacity, and the demand that is not served. The optimality gap, calculated using the value of the solution to the linear relaxation of the full problem, i.e., with $\bar{w} = 0$, and the value of the best integer solution found by the two-step approach, is reported in the column with heading *gap*. We see that for values smaller than $\bar{w} = 100$, the distribution of the demand among the three categories does not change (for all the practical purposes) and the optimality gap does not improve (it even gets worse for $\bar{w} = 25$, where the model gets so large that finding high-quality feasible solutions becomes very difficult). Therefore, in all further experiments, we consider $\bar{w} = 100$ and solve the instances with the two-step approach outlined above.

In all our experiments, we consider a 15-minutes time discretization. Our algorithms are implemented in Java. All experiments were run on 64-bit machine with an Intel Xeon E5-2650 v3 processor at 2.30 GHz running Linux and using CPLEX 12.6 for solving LPs and IPs. The time limit for solving the IP (after the LP relaxation has been solved) is set to 20 hours.

Table 1: Comparison of the solutions for various weight thresholds ($\bar{w}$)

			Weight (%)		
$\bar{w}$	$\frac{F(\bar{w})}{F}$	gap	SF	Belly	NS
400	0.42	0.15	65.5	24.9	9.6
200	0.54	0.10	72.9	19.4	7.7
100	0.65	0.09	73.9	19.2	6.9
50	0.76	0.09	73.3	19.0	7.7
25	0.84	0.10	72.9	19.2	7.9

Preliminary computational experiments also revealed that a fairly large number of columns is generated when solving the linear relaxation and that solving an integer program over the final set of columns is computationally challenging. Therefore, we solve the integer program in two stages. In the first stage, we only consider the columns that have positive values in the solution to the linear relaxation and solve the integer program over this smaller set of columns. (We allocate half of the time available for solving the integer program to this first phase.) Then, in the second stage, we expand the set of columns with any column that has been in the basis during at least one of the column generation iterations. The solution to the resulting integer program will be our final solution (where, of course, we initialize the integer program with the solution found in the first stage).

In Table 2, we report statistics regarding the solution process and the solutions obtained for all instances in our test set. The first 12 columns in the table define the instance. The column $MIP.t$ shows the time it takes (in seconds) to solve the IP, where TL indicates that the time limit of 72000 seconds was reached before proving optimality. The column $LP.t$ shows the time it takes to solve the linear relaxation and the column $cg.it$ shows the number of column generation iterations. The total number of routes generated by the column generation algorithm can be found in the column *#paths*. Finally, the last column (with heading *gap*) gives the optimality gap, i.e., relates the value of the best integer solution obtained to the value of the linear relaxation (where we recall that the column generation algorithm uses $\bar{w} = 100$).

We see that our algorithm performs quite well, as it terminates with small integrality gaps, 5% on average. It demonstrates that considering only a restricted set of columns, in this setting, results in high-quality solutions. We also note that to solve the linear relaxation only a very small fraction of the feasible routes has to be generated and that this is done in a relatively small number of column generation iterations, 77.3 on average.

4.2.2 Analysis

In this section, we analyze the results of the numerical experiments (with DSM) and discuss their practical implications.

Table 2: Computational performance of the DS solutions ($\bar{w} = 100$)

	Instance Parameters											Solution Statistics				
E	$\underline{\phi}$	$\overline{\phi}$	ϕ_c	τ	mode	β^w	β^c	%NM	T1	T2	T3	t.time	l.time	cg.it	#paths	gap
E1	23	32	5.5	1	M	1	1	0	32	19	5	TL	2187	87	6864	0.05
	23	32	5.5	1	M	0	0	0	32	19	5	TL	4380	65	5972	0.09
	23	32	5.5	1	U	1	1	0	32	19	5	39886	518	87	7993	0.03
E2	23	32	5.5	0.5	M	1	1	0	32	19	5	TL	3843	94	7991	0.05
	23	32	4.5	1	M	1	1	0	32	19	5	TL	2138	68	5886	0.03
	23	32	4.5	0.5	M	1	1	0	32	19	5	TL	4858	93	8502	0.04
	23	33	5.5	1	M	1	1	0	32	19	5	TL	4457	118	10058	0.05
	23	33	5.5	0.5	M	1	1	0	32	19	5	TL	12299	147	12885	0.04
	23	33	4.5	1	M	1	1	0	32	19	5	TL	4614	122	10906	0.03
	23	33	4.5	0.5	M	1	1	0	32	19	5	TL	16467	140	12733	0.03
	22	32	5.5	1	M	1	1	0	32	19	5	TL	3741	101	8230	0.07
	22	32	5.5	0.5	M	1	1	0	32	19	5	TL	12113	124	11715	0.05
	22	32	4.5	1	M	1	1	0	32	19	5	TL	5823	125	11410	0.04
	22	32	4.5	0.5	M	1	1	0	32	19	5	TL	15044	159	13938	0.04
E3	23	32	5.5	1	M	0.5	1	0	32	19	5	TL	2597	79	6790	0.06
	23	32	5.5	1	M	0.75	1	0	32	19	5	TL	2728	102	9437	0.05
	23	32	5.5	1	M	1.25	1	0	32	19	5	TL	5313	101	8076	0.05
	23	32	5.5	1	M	1.5	1	0	32	19	5	TL	5313	101	8076	0.05
	23	32	5.5	1	M	1	0.5	0	32	19	5	TL	4747	85	7412	0.03
	23	32	5.5	1	M	1	0.75	0	32	19	5	TL	2362	90	7425	0.04
	23	32	5.5	1	M	1	1.25	0	32	19	5	TL	2202	68	6067	0.06
	23	32	5.5	1	M	1	1.5	0	32	19	5	TL	4123	133	12753	0.07
E4	23	32	5.5	1	M	1	1	10	32	19	5	TL	2192	73	6278	0.05
	23	32	5.5	1	M	1	1	20	32	19	5	TL	2318	73	6393	0.05
	23	32	5.5	1	M	1	1	30	32	19	5	TL	3131	116	11096	0.05
E5	23	32	5.5	1	M	1	1	0	108	0	0	TL	933	56	3859	0.01
	23	32	5.5	1	M	1	1	0	90	0	0	TL	924	69	4056	0.02
	23	32	5.5	1	M	1	1	0	72	0	0	TL	954	55	2845	0.03
	23	32	5.5	1	M	1	1	0	54	0	0	TL	559	34	2286	0.03
	23	32	5.5	1	M	1	1	0	36	0	0	TL	447	36	1962	0.03
	23	32	5.5	1	M	1	1	0	18	0	0	TL	264	24	1355	0.02
	23	32	5.5	1	M	1	1	0	0	54	0	TL	997	71	4021	0.04
	23	32	5.5	1	M	1	1	0	0	45	0	TL	933	68	3516	0.05
	23	32	5.5	1	M	1	1	0	0	36	0	TL	921	55	3043	0.06
	23	32	5.5	1	M	1	1	0	0	27	0	TL	501	34	2065	0.06
	23	32	5.5	1	M	1	1	0	0	18	0	TL	332	35	1577	0.05
	23	32	5.5	1	M	1	1	0	0	9	0	38639	186	21	998	0.03
	23	32	5.5	1	M	1	1	0	0	0	36	TL	916	70	3676	0.06
	23	32	5.5	1	M	1	1	0	0	0	30	TL	869	59	3435	0.06
	23	32	5.5	1	M	1	1	0	0	0	24	40411	773	36	2408	0.07
	23	32	5.5	1	M	1	1	0	0	0	18	TL	491	31	1998	0.07
	23	32	5.5	1	M	1	1	0	0	0	12	66182	352	23	1356	0.05
	23	32	5.5	1	M	1	1	0	0	0	6	2183	113	12	523	0.03

Detailed results of the E1 experiments can be found in Table 5 in Appendix C, but the critical information is captured in Figure 5, which gives the distribution of the demand, i.e., demand served using cargo planes, demand served using belly capacity on commercial flights, and demand not served (both in terms of weight and in terms of counts) as well as the cost of the purchased belly capacity for the three scenarios investigated.

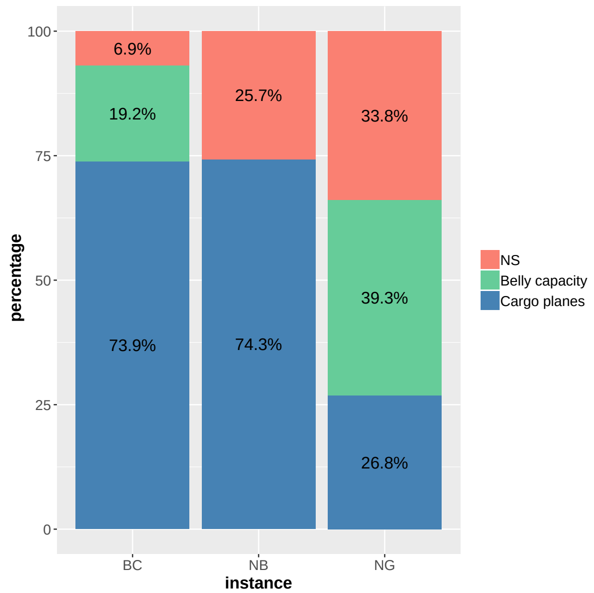
The first thing one notices when looking at Figure 5 is the importance of simultaneously considering both company-owned and purchased capacity and of simultaneously considering road and air transportation to ensure an acceptable level of demand coverage. Without ground transportation only 60.1% of demand, in terms of weight, is served, i.e., 26.8% of demand is covered by company-owned cargo planes and 39.3% is covered by belly capacity of commercial flights (1% and 60.7% in terms of counts). The impact on profit is even greater, because the money spent on belly capacity on commercial flights is much higher (as shown in Part (c)). Without the option to use belly capacity on commercial flights, 74.3% of the demand is served (in terms of weight). However, the impact on profit is small. In fact, the use of belly capacity on commercial flights mostly ensures a high coverage of demand; market share, of course, has many intangible benefits, which is why the use of belly capacity on commercial flights is still enormously valuable.

In Figure 6, we depict the routes for each of the types of cargo planes (in the base case solution). Recalling the demand distribution, see Figure 3, we observe that the routes produced by the direct shipment model align with the underlying demand distribution. There are basically three groups of demands, i.e., city-to-city pairs with high, medium, and low volumes. Type 3 planes, i.e., the large capacity planes of which there are only a few, are mostly used between large cities with high demand volumes. Type 2 planes serve large demand OD pairs too, but their routes also include lines between city pairs with medium demand volumes. Type 3 planes, i.e., the small capacity of which there are many, are used to expand the service coverage to periphery.

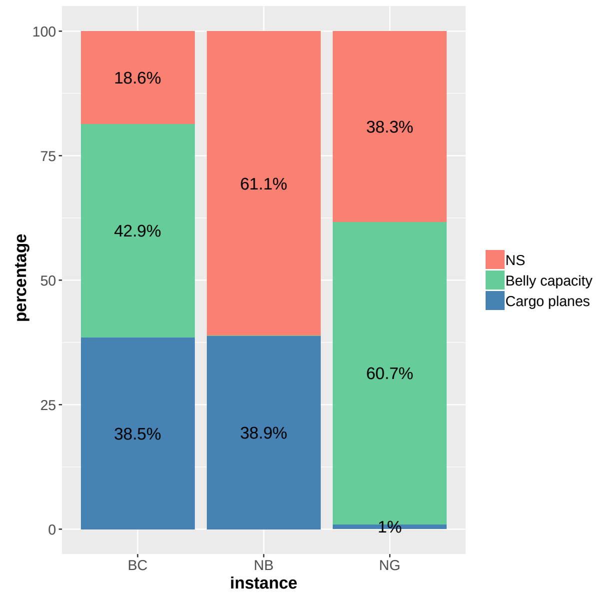
As we observed earlier, the cost of belly capacity varies significantly, see Figure 4, even for city-to-city pairs for which the flight duration is the same. An analysis of the solution to the base case reveals that the belly capacity on commercial flights, which is used to complement the company-owned capacity, is carefully selected: the average cost paid for belly capacity is \$2.68 per kg, which is significantly smaller than the general average of \$3.25 per kg (see Table 5 in the appendix for additional details).

Looking at the company-owned cargo plane utilization is also interesting. We see that the average capacity utilization is slightly over 60%, which, at first glance, does not seem very high. However, as mentioned before, solely looking at capacity utilization values is deceiving when trying to judge the effectiveness of the air service network. One of the primary goals of the direct shipment model is to minimize the time that a demand occupies cargo plane capacity. In a solution to the direct shipment model, any demand uses just one flight leg of a cargo plane's route. In contrast, in a solution to the transshipment model (and also in a hub-and-spoke air service network), a demand typically uses more than one flight leg and thus consumes plane capacity for a longer period of time. Consequently, the capacity utilization tends to be higher, but this does not necessarily imply a more effective air service network in which more demand is served by company-owned cargo planes. Recall that the maximum demand that can be served in a hub-and-spoke network where each plane performs one pickup path and one delivery path is at most the total capacity of the planes. In our base case solution, the demand served by the company-owned cargo planes is more than two times their total capacity!

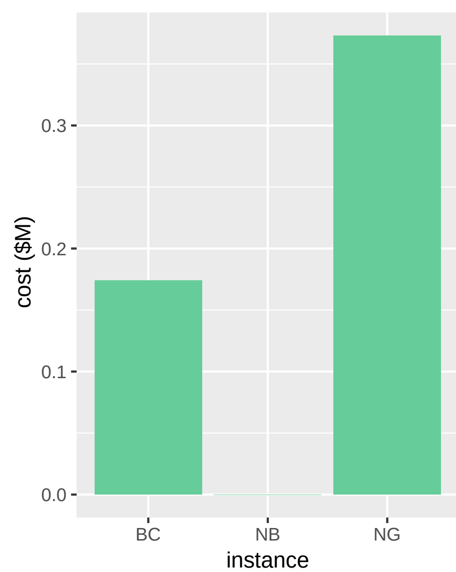
Given that in a solution to the direct shipment model a demand uses just one flight leg of a cargo plane's route, the number of flight legs in the cargo planes' routes is an important indicator of the effectiveness of the air service network. The results in Table 5 show that the number of flight legs in the cargo planes' routes is quite high, especially considering the fact that each flight also implies an hour of ground time for handling operations, which means that for a cargo plane with a route with three flight



(a) Demand served (by weight)

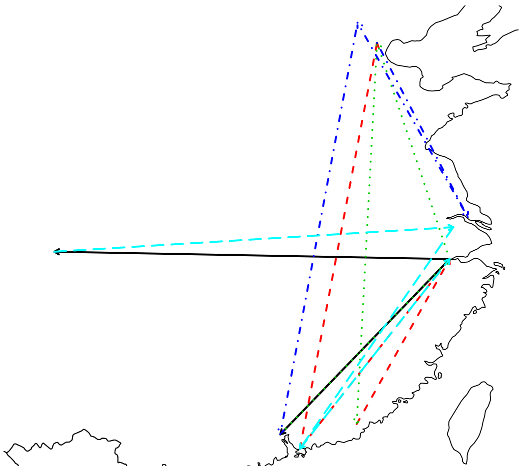


(b) Demand served (by count)

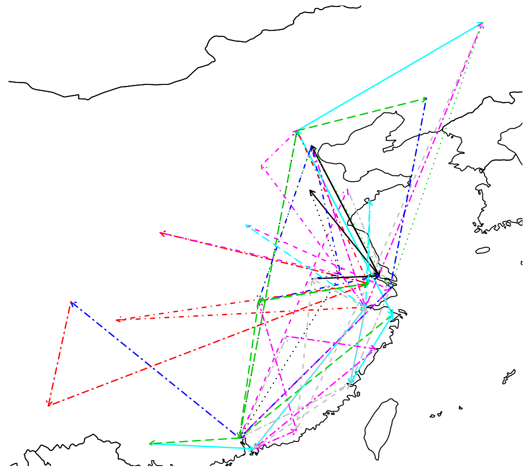


(c) Cost of the purchased capacity

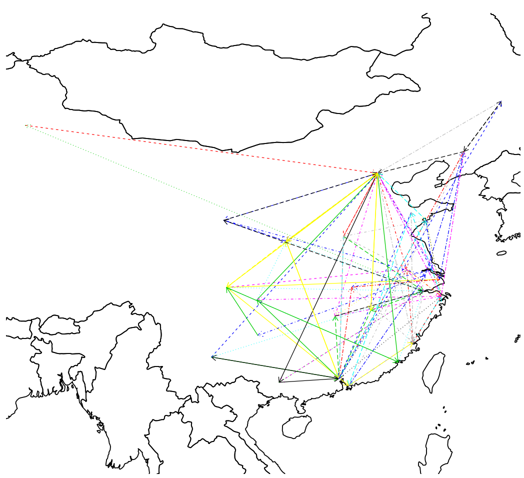
Figure 5: The demand served and the cost of purchased capacity for different multi-modal service network configurations.



(a) Type 1 planes (42 tons capacity)



(b) Type 2 planes (28 tons capacity)



(c) Type 3 planes (14 tons capacity)

Figure 6: The cargo plane routes in the base case solution.

legs 33% of the available flight time (nine hours) is taken up by ground time.

We now proceed to analyze how changes in operational efficiency affect the performance measures of interest. We present the parameters configurations for the instances used in this part of the computational study in Table 3. The detailed results regarding the solutions obtained can be found in Table 6 in Appendix D, but the critical information is captured in Figure 7, which gives the change relative to the base case, i.e., configuration A in Table 3, in demand served by company-owned cargo planes and using commercial flights (in terms of weight). We see that not all operational efficiency improvements have the same impact (some are clearly more impactful).

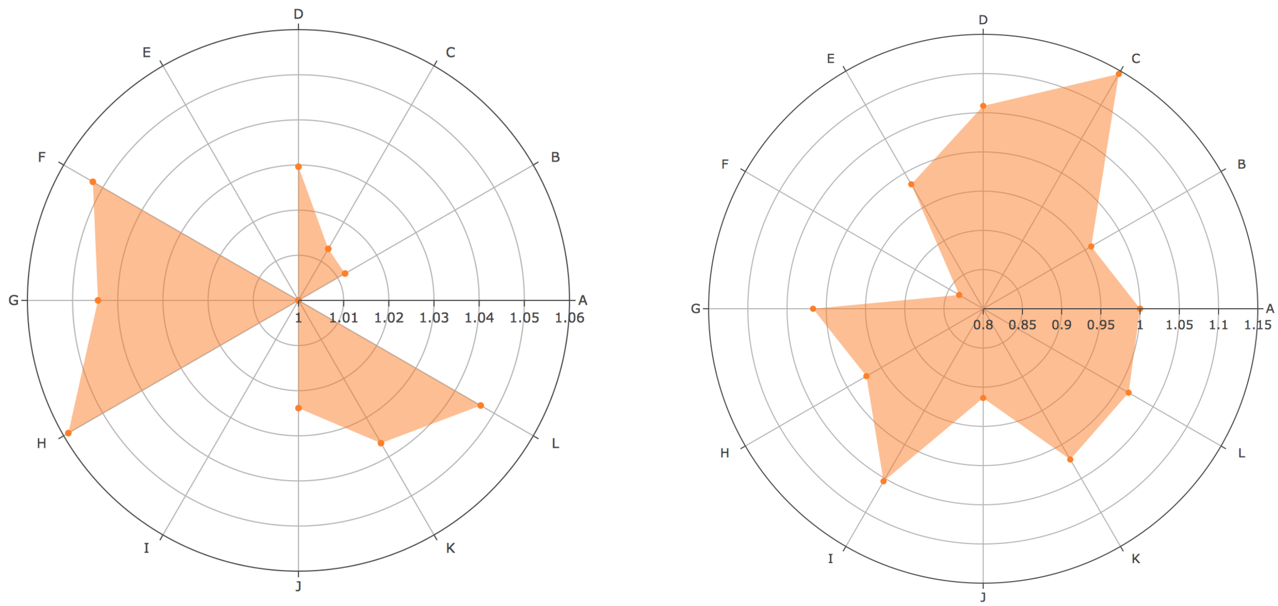
Table 3: Instance configurations

Configuration				
ID	$\underline{\phi}$	$\overline{\phi}$	ϕ_c	τ
A	23	32	5.5	1
B	23	32	5.5	0.5
C	23	32	4.5	1
D	23	32	4.5	0.5
E	23	33	5.5	1
F	23	33	5.5	0.5
G	23	33	4.5	1
H	23	33	4.5	0.5
I	22	32	5.5	1
J	22	32	5.5	0.5
K	22	32	4.5	1
L	22	32	4.5	0.5

We observe that improving the efficiency of handling operations at hubs and of intra-city operations have the potential, by themselves, to significantly increase the demand served (by allowing a more efficient use of the company-owned cargo planes). As expected, when we can increase the demand served by company-owned cargo planes there tends to be a commensurate decrease in the demand served by belly capacity on commercial flights (see for example Configuration F) When examining the impact of extending the operating window for the cargo planes, we see that extending the operating window by one hour does, in and of itself, not impact the demand served by company-owned cargo planes (Configuration E and I). However, extending the operating window along with other efficiency improvements has a synergistic effect, and amplifies the benefits of efficiency gains in airport handling and intra-city operations. This highlights the interactions between various operational aspects of the transport system. For example, an earlier start of a cargo plane (in case the operating window is extended at the front end) can only lead to improvements if there is demand that is ready to be picked up. Thus, allowing the cargo planes to start on hour earlier or end an hour later may not be helpful if demand ready times do not change.

Next, we explore how sensitive the air service network design and its associated performance metrics are to changes in the belly capacity parameters, i.e., the availability and the cost of belly capacity on commercial flights. The detailed results of the experiments can be found in Table 7 in Appendix E, but the critical information is captured in Figure 8, which shows the change in the demand served by company-owned cargo planes and the change in belly capacity spend as the available of belly capacity and the cost of belly capacity is modified. In the figures, a solid line reflects the change as a result of modifying the availability of belly capacity and a dashed line reflects the change as a result of modifying the cost of belly capacity. We vary the available belly capacity as well as the cost of belly capacity by multiplying the original values by 0.5, 0.75, 1, 1.25, and 1.5.

We see that the impact of an increase in available belly capacity or a reduction in the cost of belly capacity on the demand served is much smaller than the impact of a decrease in available belly capacity or a rise in the cost of belly capacity. This suggests that the current level of available belly capacity as



(a) Demand served with owned capacity

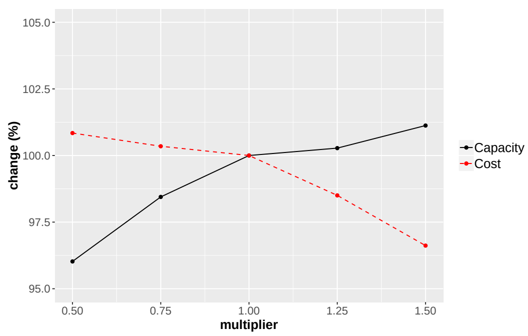
(b) Demand served with purchased capacity

Figure 7: Impact of operational improvements on the demand served by owned and purchased capacity.

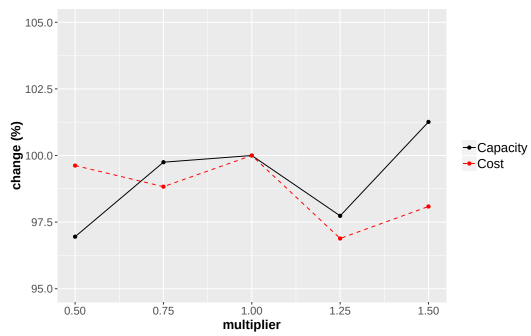
well as the current cost of belly capacity is "well-aligned" with the set of demands.

It also interesting to observe that the demand served by the cargo planes mostly decreases when the available belly capacity or the cost of the belly capacity changes. When more belly capacity becomes available (or is less expensive), then some of the demand that before was served by cargo planes is shifted to belly capacity so as to allow the cargo planes to serve "hard to reach" demands that were not served before, even though this results in lower capacity utilization. Similarly, when less belly capacity is available (or is more expensive), then some of the demand that before was served (profitably) by belly capacity is shifted to cargo planes, even though this means cargo planes have to serve more hard to reach demand and sacrifice capacity utilization.

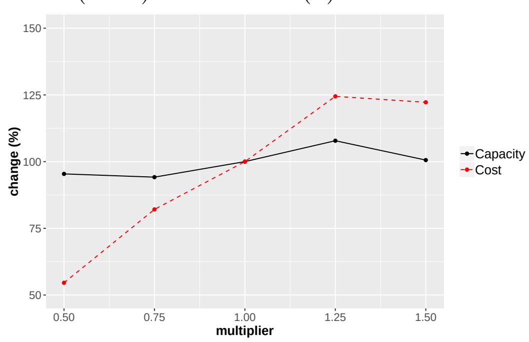
In the last decade, express delivery companies have aggressively introduce new service offerings. In the analyses carried out up to now, demand represented only packages that needed to be delivered by 6pm the next day. Next, we investigate the impact on the air service network, when a fraction of the demand represents packages that need to be delivered by noon the next day. The detailed results of the experiments can be found in Table 8 in Appendix F, but the critical information is captured in Figures 9 and 10. In Figure 9a, we show the change in total demand served when the fraction of demand requiring next morning delivery increases from zero to 0.3, where we assume that this fraction is the same across all demands. (The latter assumption is not necessarily realistic as it is more likely that demand requiring next-morning-am service are available for transportation earlier in the day.). Maybe somewhat surprisingly, the impact is relatively small, especially in terms of demand weight. This is due, primarily, to the fact that the operating window for cargo planes ends at 8am (which implies that demands that are not on the final flight legs arrive much earlier), and most of the demand transported by the cargo planes can naturally be delivered before noon that day. Figure 9b, which shows the fraction of the next-morning demand that is served, reveals that almost all next-morning demand is served (again, especially in terms of demand weight). Clearly, the optimization recognizes and takes advantage of the higher revenue of next-morning demand.



(a) Demand served (total)



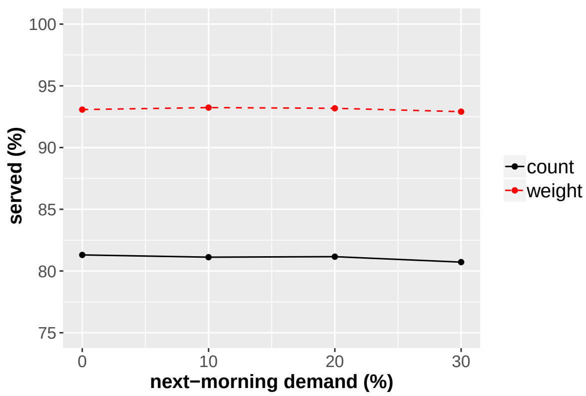
(b) Demand served (by the cargo planes)



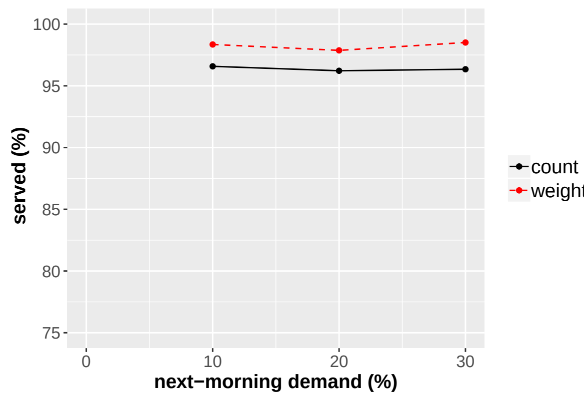
(c) Amount spent for the purchased capacity

Figure 8: Change in the carriers profit and amount of demand supported by the company owned aircraft with the belly capacity and cost variations.

Figure 10 provides a more in-depth analysis and shows the percentage of served next-day and next-morning demand as a function of demand ready times (for $\sigma = 0.3$). As expected, almost all demand that is ready early in the day is served, while the percentage of demand served decreases the later in the day demand is ready. Interestingly, even though the time available to serve next-morning demand is less, a higher percentage of that demand is served. Again, the optimization exploits the higher revenues received from next-morning deliveries.



(a) Total demand served as a function of the next-morning demand fraction.



(b) Next-morning demand served as a function of the next-morning demand fraction.

Figure 9: The impact on demand served as a function of the service class distribution.

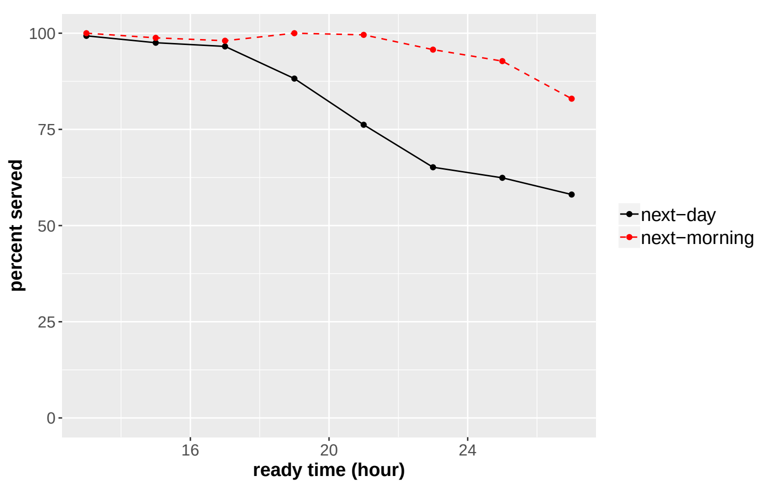
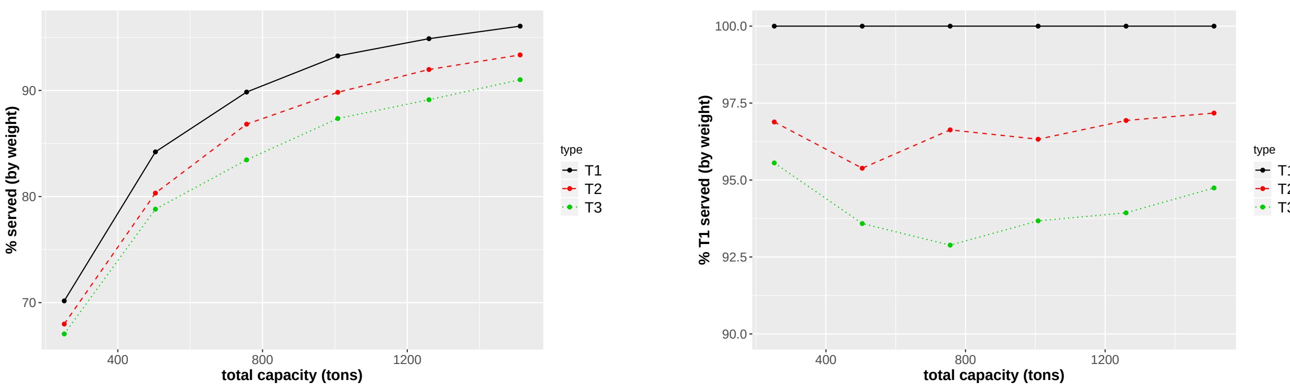


Figure 10: Percentage of the served demand (weight) by the demand ready times $\sigma = 0.3$.

In the final set of experiments, we examine the impact of the cargo plane fleet composition on demand served, i.e., investigate whether it is better to have a few high capacity planes (with lower operating costs per unit capacity) or many small capacity planes (with higher operating costs per unit capacity). The detailed results of the experiments can be found in Table 9 in Appendix G, but the critical information is captured in Figures 11a and 11b. In Figure 11a, we show the demand served for different fleet compositions. Specifically, we compare three fleet compositions, all with the same total capacity, but the first consisting of only Type 1 planes, the second consisting of only Type 2 planes, and the third consisting of only Type 3 planes. As expected, for all three fleet compositions, we see diminishing returns as the total capacity increases. What may be somewhat surprising is that the difference in demand served between the fleet compositions is not very pronounced. Note that a fleet composed of only Type 1 planes (each with a capacity of 14 tons) is always preferable from an air service network design perspective, if the costs of operating the cargo planes and the capital outlays are ignore (which is what we do in our model). However, those costs can significantly affect the long-term profits of a package express company and have to be considered when strategic decisions about fleet composition are made. Larger planes typically require less capital outlay and have smaller operating costs per unit capacity, and, from that perspective, are therefore more attractive. Thus, a trade-off has to be made between operational efficiency (in terms of demand served) and capital layout and operating costs (in terms of fleet composition). In this regard, Figure 11b, which depicts operational efficiency in a different way, is of interest. It shows what fraction of the demand served by a fleet composed of only Type 1 planes can be achieved by fleets of only Type 2 and of only Type 3 planes. We see that the difference is never more than 7.5% and much smaller when the total capacity is relatively small. Considering the fact that Type 3 planes have three times more capacity than Type 1 planes, such a small difference is somewhat surprising and suggests that careful capacity planning is important.



(a) Effect of fleet composition on total demand served.

(b) Comparison of demand served for different fleet compositions.

Figure 11: Effect of cargo plane fleet composition.

4.3 Experiments with Transshipment Model

In this section, we seek to answer, among others, the following questions regarding the benefits of using the transshipment model to analyze and plan high-value items/products deliveries.

- What are the characteristics of high-quality transshipment air network designs (i.e., number and locations of airports where transshipments take place, pickup routes, delivery routes, ground transportation, etc.)?
- What operational efficiency improvements have the greatest potential to improve profitability (i.e., improving efficiency of intra-city operations, improving efficiency of airport handling operations, modifying the cargo plane operation window, etc.)?
- Does the model effectively integrate the use of ground and air transportation?

To analyze the value of the TSM, we have modified the demand data so that it is more representative of the setting we have in mind. Below, we list the modifications; all other aspects have remained the same.

- Only demands with weight between 1000 and 1250kg are considered. In Figure 12, we illustrate the spatial distribution of the demand over the service region; 58 of the demands are between hub-cities, and 42 of the demands require ground transportation (i.e., either the origin or the destination is not a hub-city).
- The fleet consists of four cargo planes.
- Ground transportation is available between every origin-destination pair with (ground) travel time less then 8 hours.
- The per-unit revenue for demand q is $r^q = 100 + 100(t^a_{o_q,d_q})^{1.25} + 100I_{q_o} + 100I_{q_d}$, where indicator I_c $(c \in C)$ is 1 if c is a hub-city, and 0 otherwise. Note that per-unit revenue is a super-linear function of distance, i.e., as the flexibility (to serve the demand) decreases, the revenue increases super-linearly. The fact that the flexibility also decreases when the origin and/or destination is not a hub-city (as ground transportation takes time) is also reflected in the per-unit revenue.

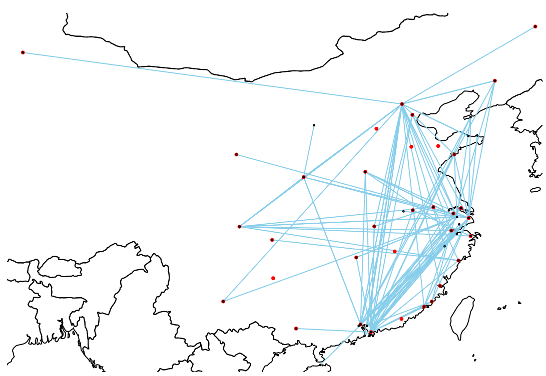


Figure 12: The spatial distribution of high-value items/products demand.

To try and answer the questions listed above, we have conducted four groups of experiments. The groups of experiments aim to investigate: (1) the benefits of multi-modal transportation (i.e., ground and air), (2) the benefits of improvements in the operational parameters, (3) the impact of changes in the demand composition, and (4) the impact of changes in the fleet size. Details of each group of experiments are given below.

In the first set of experiments (TS-E1), we consider two settings: the base case (BC) and a setting in which no-ground transportation is available to take demand to originating at or destined for a non-hub city to/from a hub-city (NG).

In the second set of experiments (TS-E2), we consider 12 variants. We study variants generated by a combinations of the following three basic operational parameters.

- Cargo aircraft operation time windows: $[\underline{\phi}, \overline{\phi}] \in \{[23:00, 08:00], [23:00, 09:00], [22:00, 08:00]\}$.
- Intra-city operations time: $\phi_c \in \{4.5, 5.5\}$.
- Airport operations time: $\tau \in \{0.5, 1\}$.

In the third set of experiments (TS-E3), we consider four variants by assuming that $\sigma = \{0, 10\%, 20\%, 30\%\}$ of the demand represents shipments that need to be delivered before noon the next day.

In the last set of experiments (TS-E4), we consider four larger fleet sizes: 5, 6, 7 and 8.

As preliminary computational experiments revealed that the pickup and delivery paths used in high-quality air service network designs typically consist of two flight legs and have a duration of less than six hours, in subsequent experiments we restricted ourselves to pickup and delivery paths with these characteristics. In all our experiments we set the time limit to 48 hours.

4.3.1 Analysis

In this section, we analyze the results of the numerical experiments (with TSM) and discuss their practical implications.

In Table 4 we present the results for the TS-E1 problem instance solutions, where we compare the performance of the TS solutions with and without ground transportation (denoted as BC and NG in the table). In Table 4, the columns *count* indicates the number of demands (out of 100) that have been served in the respective solution. Similarly, the column *value* presents the percent of the total demand value served (total revenue captured) in the optimal solution. The column $\#T$ indicates the number of airports used as transshipment locations. The column *t.util* indicates the time utilization of the planes (including the ground handling operations). The last two columns *time* and *gap* shows the cpu time (seconds) used by the algorithm and the integrality gap, respectively.

Table 4: The importance of integrating ground and air transportation.

ID	count	value	#T	t.util	time	gap
BC	78	77.6	2	94.1	27124	0
NG	31	29.1	1	100	4439	0

The results in Table 4 clearly demonstrate the critical importance of ground transportation. The fraction of demand served more than doubles when ground transportation is integrated with air transportation (and the increase in the fraction of captured revenue is even greater). However, not surprisingly, determining a high-quality integrated (multi-modal) service network is much harder computationally. When ground transportation is considered, the number of demand assignment variables increases significantly (from about 200,000 to 800,000), which results in a much larger and more difficult integer program.

We now proceed to analyze how changes in operational efficiency affect the performance measures of interest. Detailed results regarding the solutions obtained can be found in Table 10 in Appendix H, but the critical information is captured in Figure 13, which shows the change relative to the base case (the instance IDs again refer to the descriptions in Table 3).

Figure 13 reveals that improving the efficiency of intra-city operations gives the highest increase in demand served. Interestingly, improving the efficiency of ground operations does not have a significant impact (in contrast to what we saw for the direct shipment model). This observation highlights the different characteristics of the direct shipment and transshipment models. In the direct shipment model, the number of flight legs in a cargo planes itinerary is of critical importance, and this number strongly depends on the time time spend on ground operations. Even moderate time savings in ground operations may allow the addition of one (or more) legs to an itinerary, which, in turn, increases the demand served. However, such efficiency improvements cannot easily be converted to an increase in coverage in the transshipment model, where demand is much more spread out. We also see that the timing of the flight hour extension is critical. Our results show that, starting the cargo flights one hour earlier (configurations

I, J, K, L) does not make a difference (compared to A, B, C and D) in the amount of demands served, However, extending the flight time by one hour have a significant impact (compare configurations A and E).

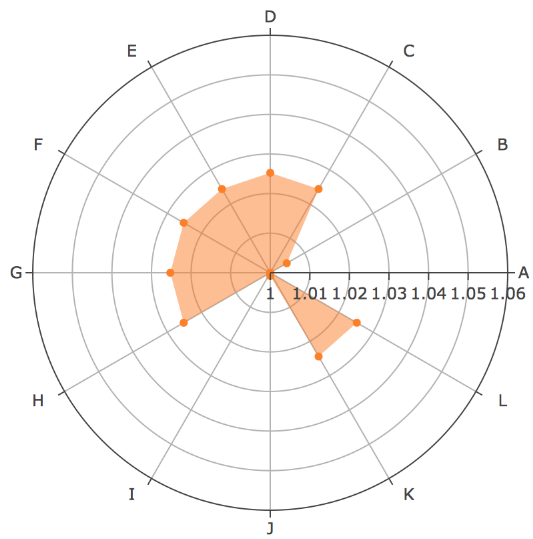
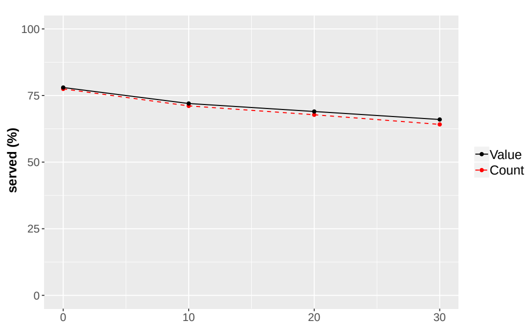
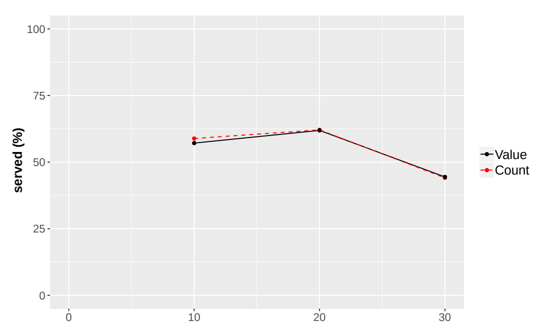


Figure 13: The impact of operational improvements on demand served.

Next, we explore the effect of changes in the service class distribution of the demand. Detailed results can be found in Table 11 in Appendix I, but the critical information is captured in Figure 14, which shows the total demand served (by value and count) for different service class distributions. We observe that in this setting (high-value items/products) the effect is much more pronounced than in the previous setting (express packages). Because it is more challenging to serve demand, the impact of differences in the available delivery time has a greater impact. As the revenue for shipments that have to delivered before noon the next day is 50% higher, TSM ends up serving fewer (a reduction of about 15%), but higher revenue demands (the net loss in value is reduced to 7%). These results underline the time-constrained nature of a high-value items/products delivery system, and the impact of the pricing of different service classes on the air network system design.



(a) Total demand served as a function of the next-morning demand fraction.



(b) Next-morning demand served as a function of the next-morning demand fraction.

Figure 14: Effect of service class distribution on the carriers optimal profit and served demand count.

Finally, we explore the impact of an increase in fleet size. Detailed results can be found in Table 12 in Appendix J, but the critical information is captured in Figure 15, which shows the effect of increasing the fleet size on demand served.

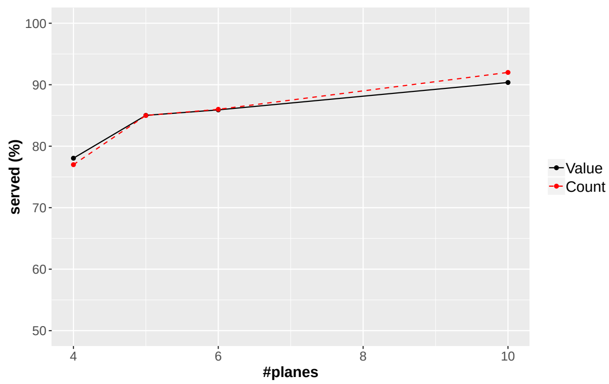


Figure 15: Impact of increasing fleet size.

The results are as expected, initially an increase in fleet size results in size able jump in demand served, but at some point it becomes more challenging to increase demand served as the only unserved demand left are the "hard to reach" demands. Therefore, the service coverage (which demands to serve) and capacity planning (how much planes to allocate in this service) decisions have to be made very carefully, considering the distribution of the demand over the service region.

5 Final remarks

Motivated by the environment encountered at SF Express, we have developed two optimization-based tools to support the planning and operations functions of package express carriers in China. The two tools focus on markets with characteristics that are at opposite ends of a spectrum.

One aspect of our research is worth reiterating. One of our optimization models only considers direct shipments. As we have shown, in some settings an air express shipment network based on direct shipments can have significant advantages over a hub-and-spoke network, even though the research community and most practitioners have focused on the latter. Considering only direct shipments offers operational efficiency, but can limit coverage. Our research demonstrates, however, that simultaneously considering ground and air and simulataneously considering owned and purchased capacity can mitigate the limitations of considering only direct shipments. The presence of (sufficient) belly capacity on commercial flights allows for the "cherry picking" of city pairs that provide high demand for direct transport using owned cargo planes, without sacrificing service coverage. (Note that a portion of this high demand almost certainly represents demand arriving and/or departing with ground transportation.

Our work is only a first step towards developing a broad suite of tools to support the planning and operations functions of package express carriers in China. Many other challenges that can be addressed using optimization models/tools remain, and we plan to pursue these in the near future. Here, we mention only one. The models presented in this paper, may suggest changes to the existing air service network. Such changes cannot be implemented overnight (as they impact Chinese airspace, which is controlled by the Civic Aviation Administration of China) and thus such changes need to be introduced gradually. This leads to interesting incremental network design problems.

Acknowledgment

Baris Yildiz was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the grant number 2219. Furthermore, this work would not have been possible without the assistance of the air service network planning team at SF Express. We thank them not only for providing us with real-life data, but also for the many informative and insightful discussions.

References

Y. K. Agarwal. Design of capacitated multicommodity networks with multiple facilities. *Operations Research*, 50(2):333–344, 2002.

A. P. Armacost, C. Barnhart, and K. A. Ware. Composite variable formulations for express shipment service network design. *Transportation science*, 36(1):1–20, 2002.

R. Bai, J. R. Woodward, and N. Subramanian. A new fast large neighbourhood search for service network design with asset balance constraints. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2016.

A. Balakrishnan, G. Li, and P. Mirchandani. Optimal network design with end-to-end service requirements. *Operations Research*, 65(3):729–750, 2017.

C. Barnhart and R. R. Schneur. Air network design for express shipment service. *Operations Research*, 44(6):852–863, 1996.

C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations research*, 46(3):316–329, 1998.

C. Barnhart, C. A. Hane, and P. H. Vance. Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems. *Operations Research*, 48(2):318–326, 2000.

C. Barnhart, N. Krishnan, D. Kim, and K. Ware. Network design for express shipment delivery. *Computational Optimization and Applications*, 21(3):239–262, 2002.

K. Büdenbender, T. Grünert, and H.-J. Sebastian. A hybrid tabu search/branch-and-bound algorithm for the direct flight network design problem. *Transportation Science*, 34(4):364–380, 2000.

T. G. Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

H. Fleuren, C. Goossens, M. Hendriks, M.-C. Lombard, I. Meuffels, and J. Poppelaars. Supply chain–wide optimization at tnt express. *Interfaces*, 43(1):5–20, 2013.

M. Hewitt, T. G. Crainic, M. Nowak, and W. Rei. Scheduled service network design with resource acquisition and management under uncertainty. *Transportation Research Part B: Methodological*, 128: 324–343, 2019.

D. Kim, C. Barnhart, K. Ware, and G. Reinhardt. Multimodal express package delivery: A service network design application. *Transportation Science*, 33(4):391–407, 1999.

M. J. Kuby and R. G. Gray. The hub network design problem with stopovers and feeders: The case of federal express. *Transportation Research Part A: Policy and Practice*, 27(1):1–12, 1993.

C.-C. Lin and S.-H. Chen. The hierarchical network design problem for time-definite express common carriers. *Transportation research part B: methodological*, 38(3):271–283, 2004.

I. Louwerse, J. Mijnarends, I. Meuffels, D. Huisman, and H. Fleuren. Scheduling movements in the network of an express service provider. *Flexible Services and Manufacturing Journal*, 26(4):565–584, 2014.

T. L. Magnanti and R. T. Wong. Network design and transportation planning: Models and algorithms. *Transportation science*, 18(1):1–55, 1984.

F. Martin, V. Hemmelmayr, and T. Wakolbinger. Integrated service selection, pricing and fullfillment planning for express parcel carriers-enriching service network design with customer choice and endogenous delivery time restrictions. Technical Report 6986, WU Vienna University of Economics and Business, 2019.

S. Melkote and M. S. Daskin. Capacitated facility location/network design problems. *European journal of operational research*, 129(3):481–495, 2001.

W. J. M. Meuffels et al. *The design of road and air networks for express service providers*. CentER, Tilburg University, 2015.

M. Ng and H. K. Lo. Robust models for transportation service network design. *Transportation Research Part B: Methodological*, 94:378–386, 2016.

C. Prodhon and C. Prins. A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1):1–17, 2014.

J. M. Quesada Perez, J.-S. Tancrez, and J.-C. Lange. " a multi-hub express shipment service network design model with flexible hub assignment. In *EURO 2016: 28th European Conference on Operational Research*, 2016.

H. Yang and M. G. H. Bell. Models and algorithms for road network design: a review and some new developments. *Transport Reviews*, 18(3):257–278, 1998.

B. Yıldız and O. E. Karaşan. Regenerator location problem in flexible optical networks. *Operations Research*, 65(3):595–620, 2017.

B. Yıldız, O. Arslan, and O. E. Karaşan. A branch and price approach for routing and refueling station location model. *European Journal of Operational Research*, 248(3):815–826, 2016.

B. Yıldız, O. E. Karaşan, and H. Yaman. Branch-and-price approaches for the network design problem with relays. *Computers & Operations Research*, 2018.

Appendix A Finding a longest path in an directed acyclic graph

Algorithm 2: Longest Path Algorithm

Input: G^k
Output: p^*

Find the topological order of nodes O in G^k;
Set $v_\sigma = 0$ and $v_\ell = -\infty$ for all $\ell \in \bar{L}$;
Set $pred_\ell = null$ for all $\ell \in \bar{L}$;
foreach $\ell \in O$ **do**
    foreach $(\ell, \bar{\ell}) \in A$ **do**
        if $v_{\bar{\ell}} < v_\ell + w_{(\ell,\bar{\ell})}$ **then**
            $v_{\bar{\ell}} = v_\ell + w_{(\ell,\bar{\ell})}$;
            $pred_{\bar{\ell}} = \ell$
$p^* = \emptyset$;
$\ell = \bar{\sigma}$;
while $\ell \neq \sigma$ **do**
    if $pred_\ell \neq \sigma$ **then**
        add $(pred_\ell, \ell)$ to beginning of p^*;
    $\ell = pred_\ell$
return p^*

Appendix B Extending TSM to consider purchased capacity

To accommodate purchased capacity, we introduce variable $\beta_b^q, q \in Q, b \in B(q)$ indicating whether demand q is served by the commercial flight $b \in B$ ($\beta_b^q = 1$) or not ($\beta_b^q = 0$).

In order to account for the cost of the belly capacity, objective (12) has to be changed to

$$\max \sum_{q \in Q} r^q y^q - c_b \beta_b^q \tag{29}$$

To ensure that a demand is either served by the purchased belly capacity or with the owned cargo planes, and to ensure weight limit on purchased belly capacity are respected, the following constraints have to be added to TS:

$$\sum_{b \in B(q)} \beta_b^q + y^q \leq 1 \qquad \forall q \in Q, \tag{30}$$

$$\sum_{q \in Q(b)} f^q \beta_b^q \leq w_b \qquad \forall b \in B. \tag{31}$$

Appendix C E1 Experiment results

In Table 5, the column ID indicates the incidence identification code (BC: base case, NB: no belly capacity, NG: no ground transportation). In the columns $Value$, we show what percentage of the total demand value (R) is captured as profit (SF), paid for the belly cost (B) and left unserved (NS). Similarly, the columns $Weight$ and $Count$ indicates what percent of the total demand weight and count is transferred by the company owned aircrafts (SF), transferred by the rented belly capacity (B) and left unserved (NS). For the belly capacity utilization statistics, the column *cost* indicates the average cost paid (per kg) for the belly usage. The column $m.w$ shows the mean weight of demands transferred by the belly capacity. The column *#used* indicates the number of commercial flights that are used in the solution to transfer some demand. The column $\#full$ indicates the how many of the commercial flights are used to their full capacity in the respective solution. For the carrier's cargo aircraft utilization stats, the column $m.w$ indicates the mean demand weight for those transported by the owned airplanes. Columns *v.util* and *t.util* indicate the average capacity and flight time utilizations. Similarly the column *#legs* indicates the average number of flight legs (on the average) executed by the cargo aircraft. Finally, the last three columns present statistics about the solution procedure.

Table 5: Solution results for the E1 problem instances

	Value (%)			Weight (%)			Count (%)			Belly Use				Cargo Aircraft Usage			
ID	SF	B	NS	SF	B	NS	SF	B	NS	m.c	m.w	#used	#full	m.w	v.util	t.util	#legs
BC	83.8	9.2	7.0	73.9	19.2	6.9	38.5	42.9	18.6	2.68	14.5	2009	402	62.2	0.64	0.98	3.23
NB	74.2	0.0	25.8	74.3	0.0	25.7	38.9	0.0	61.1	0.00	0.0	0	0	62.0	0.64	0.99	3.25
NG	46.5	19.7	33.7	26.8	39.3	33.8	1.0	60.7	38.3	2.80	21.0	2238	1051	886.5	0.24	0.98	3.27

Appendix D E2 Experiment results

Table 6 presents the results for the E2 instance solutions. In Table 6 the column ID indicates the incidence identification code (as presented in Table 3). The rest of the columns present the same statistics as defined in Appendix C

Table 6: Solution results for the E2 problem instances

	Value (%)			Weight (%)			Count (%)			Belly Use				Cargo Aircraft Usage			
ID	SF	B	NS	SF	B	NS	SF	B	NS	m.c	m.w	#used	#full	m.w	v.util	t.util	#legs
A	83.8	9.2	7.0	73.9	19.2	6.9	38.5	42.9	18.6	2.68	14.5	2009	402	62.2	0.64	0.98	3.23
B	84.4	8.8	6.9	74.8	18.4	6.8	42.7	39.6	17.7	2.67	15.0	2000	377	56.8	0.58	0.99	3.59
C	86.8	10.0	3.2	74.9	22.0	3.2	40.0	50.9	9.2	2.55	14.0	2145	500	60.7	0.67	0.98	3.13
D	87.0	9.3	3.6	76.1	20.3	3.6	43.6	47.7	8.8	2.57	13.8	2160	442	56.6	0.59	0.98	3.55
E	83.7	9.0	7.3	73.8	18.9	7.3	39.4	41.6	19.0	2.68	14.7	1969	393	60.7	0.59	0.98	3.48
F	86.1	7.7	6.2	77.8	16.0	6.2	44.9	38.0	17.1	2.69	13.7	1982	316	56.2	0.54	0.98	4.00
G	87.3	9.3	3.3	77.2	19.5	3.3	44.0	47.1	8.8	2.68	13.4	2166	431	56.8	0.62	0.98	3.45
H	88.3	8.6	3.1	78.2	18.6	3.1	48.3	43.5	8.2	2.57	13.9	2098	391	52.5	0.55	0.99	3.95
I	82.2	9.8	8.0	71.8	20.2	8.0	38.2	42.5	19.3	2.71	15.4	1985	441	61.0	0.59	0.94	3.32
J	84.6	8.5	6.9	75.7	17.5	6.8	41.7	39.8	18.6	2.72	14.3	1936	378	58.8	0.54	0.96	3.91
K	86.8	9.3	3.8	76.6	19.6	3.8	41.9	48.5	9.6	2.66	13.1	2170	426	59.2	0.61	0.98	3.50
L	87.7	9.1	3.2	77.3	19.5	3.2	45.1	46.1	8.9	2.61	13.7	2134	418	55.6	0.55	0.99	3.91

Appendix E E3 Experiment results

Table 7 presents the results for the E3 instance solutions. The columns B^w and B^c indicates the belly capacity and cost multipliers for each problem instances (i.e., the case $B^w = 1$ and $B^c = 1.00$, corresponds to the base case). The rest of the columns preset the same information as in the previous tables.

.

Table 7: Solution results for the E3 problem instances

		Value (%)			Weight (%)			Count (%)			Belly Use				Cargo Aircraft Usage			
β^w	β^c	SF	B	NS	SF	B	NS	SF	B	NS	m.c	m.w	#used	#full	m.w	v.util	t.util	#legs
0.5	1	80.7	8.8	10.6	71.6	17.7	10.6	36.6	41.0	22.4	2.77	14.0	2209	930	63.4	0.63	0.98	3.16
0.75	1	82.9	8.7	8.4	73.7	17.9	8.4	37.9	41.8	20.4	0.00	0.0	2054	0	63.1	0.63	0.99	3.29
1.25	1	83.4	9.9	6.7	72.2	21.1	6.7	38.6	43.2	18.2	2.63	15.8	1984	336	60.7	0.62	0.98	3.25
1.5	1	84.8	9.2	6.0	74.8	19.3	5.9	40.4	41.9	17.6	2.68	14.9	1960	211	60.0	0.64	0.99	3.25
1	0.5	88.8	5.0	6.2	73.6	20.2	6.1	38.4	44.3	17.3	1.39	14.8	2091	425	62.2	0.64	0.98	3.25
1	0.75	85.8	7.6	6.7	73.0	20.4	6.6	37.8	44.0	18.2	2.07	15.0	2070	442	62.6	0.63	0.98	3.23
1	1.25	80.2	11.4	8.4	71.6	20.1	8.3	37.3	42.1	20.6	3.18	15.5	1942	432	62.2	0.63	0.98	3.21
1	1.5	78.6	11.2	10.1	72.5	17.5	10.1	37.8	39.5	22.7	3.60	14.3	1799	387	62.1	0.63	0.98	3.20
1	1	83.8	9.2	7.0	73.9	19.2	6.9	38.5	42.9	18.6	2.68	14.5	2009	402	62.2	0.64	0.98	3.23

Appendix F E4 Experiment results

Table 8 presents the results for the E4 instance solutions. The column σ indicates the fraction of next-morning demand (i.e., the case $\sigma = 0$, corresponds to the base case). The rest of the columns preset the same information as in the previous tables. .

Table 8: Solution results for the E4 problem instances

	Value (%)			Weight (%)			Count (%)			Belly Use				Cargo Aircraft Usage			
σ	SF	B	NS	SF	B	NS	SF	B	NS	m.c	m.w	#used	#full	m.w	v.util	t.util	#legs
0.0	83.8	9.2	7.0	73.9	19.2	6.9	38.5	42.9	18.6	2.68	14.5	2009	402	62.2	0.64	0.98	3.23
0.1	84.2	9.1	6.7	73.4	19.9	6.8	39.1	42.1	18.8	2.67	15.3	2042	0	60.8	0.64	0.98	3.23
0.2	84.0	9.5	6.6	71.8	21.4	6.8	37.3	44.0	18.7	2.65	15.8	2035	479	62.3	0.62	0.98	3.27
0.3	84.1	9.3	6.6	71.1	21.8	7.1	37.5	43.4	19.2	2.67	16.3	2060	473	61.5	0.63	0.98	3.20

Appendix G E5 Experiment results

Table 9 presents the results for the E5 instance solutions. In Table 9 the column θ indicate the plane types and W indicates the total capacity (i.e., for each instance the total number of the cargo planes is W/w_θ). The rest of the columns preset the same information as in the previous tables. .

Table 9: Solution results for the E5 problem instances

	Value (%)			Weight (%)			Count (%)			Belly Use				Cargo Aircraft Usage			
θ	SF	B	NS	SF	B	NS	SF	B	NS	m.c	m.w	#used	#full	m.w	v.util	t.util	#legs
T1	89.4	6.6	4.0	81.9	14.1	3.9	52.0	35.1	12.9	2.63	13.1	1951	271	51.1	0.57	0.99	3.21
	87.2	7.6	5.2	79.0	15.8	5.1	45.1	39.1	15.7	2.68	13.1	1979	306	56.8	0.65	0.99	3.24
	82.6	10.6	6.8	70.7	22.6	6.7	37.4	44.9	17.7	2.62	16.3	2030	512	61.2	0.74	0.99	3.19
	76.0	13.7	10.3	61.4	28.5	10.2	28.7	49.9	21.4	2.70	18.5	2164	699	69.5	0.83	0.99	3.31
	64.8	19.3	15.9	46.0	38.2	15.8	18.8	54.7	26.5	2.82	22.7	2235	991	79.3	0.89	0.99	3.47
	48.0	22.1	29.9	26.2	43.9	29.8	8.6	58.2	33.2	2.81	24.5	2261	1180	99.1	0.98	0.99	3.61
T2	85.1	8.2	6.7	76.2	17.1	6.7	39.1	42.0	18.9	2.67	13.2	2027	351	63.1	0.53	0.99	3.24
	82.3	9.6	8.1	71.7	20.3	8.0	35.2	44.5	20.4	2.65	14.8	2018	424	66.1	0.59	0.99	3.24
	77.9	11.8	10.3	65.3	24.5	10.2	30.8	47.1	22.1	2.68	16.9	2078	562	68.7	0.65	0.99	3.36
	72.4	14.3	13.3	57.4	29.4	13.2	23.4	51.5	25.2	2.71	18.5	2134	698	79.7	0.75	0.99	3.41
	61.4	18.9	19.8	42.6	37.7	19.7	15.4	54.7	30.0	2.79	22.4	2189	976	89.8	0.83	0.99	3.44
	46.1	21.8	32.0	24.4	43.6	32.0	5.7	59.2	35.2	2.79	23.9	2266	1178	139.8	0.92	0.98	3.56
T3	80.7	10.2	9.1	70.2	20.9	9.0	33.9	44.8	21.4	2.73	15.1	1990	458	67.1	0.48	0.98	3.25
	77.9	11.1	11.0	66.1	23.0	10.9	29.9	46.4	23.7	2.70	16.1	2053	523	71.7	0.53	0.98	3.33
	73.9	13.3	12.8	60.4	26.9	12.6	26.1	48.9	25.0	2.77	17.9	2086	639	74.9	0.61	0.98	3.33
	67.6	15.7	16.7	51.9	31.5	16.5	19.9	52.8	27.3	2.79	19.4	2138	791	84.6	0.67	0.98	3.44
	59.2	19.6	21.3	39.9	38.9	21.2	13.6	56.2	30.2	2.81	22.5	2214	992	94.8	0.80	0.97	3.33
	45.2	21.8	33.0	24.0	43.0	33.0	5.2	58.9	35.9	2.82	23.7	2269	1164	149.5	0.92	0.97	3.50

Appendix H TS-E2 Experiment results

Table 10 presents the results for the TS-E2 instance solutions. The instance IDs are named as described in Table 3 and the column headings have the same meaning has described for Table 4.

Table 10: Solution Results for the TS-E2 problem instances

ID	count	value	#T	t.util	time	gap
A	77	78.05	2	90.09	89829	0
B	77	78.43	2	88.47	126716	0
C	79	79.96	2	89.46	79785	0
D	79	80.02	2	93.21	52848	0
E	79	79.96	2	80.38	57434	0
F	79	80.02	2	88.88	50687	0
G	79	80.02	2	92.73	44283	0
H	79	80.02	2	99.99	100058	0
I	76	76.89	2	79.51	172808	0.03
J	77	78.11	2	90.39	172847	0.01
K	79	79.96	2	80.35	167618	0
L	79	80.02	2	91.25	89947	0

Appendix I TS-E3 Experiment results

Table 11 presents the results for the TS-E3 instance solutions. The column heading σ indicates the next-morning demand fraction and the rest of the column headings have the same meaning has decried for Table 4.

Table 11: Solution results for the TS-E3 problem instances

σ	count	value	#T	t.util	time	gap
0	77	78.1	2	90.09	89829	0.00
0.1	71	70.0	3	97.43	172858	0.09
0.2	67	67.1	3	95.57	172853	0.14
0.3	65	63.5	3	91.14	172841	0.18

Appendix J TS-E4 Experiment results

Table 12 presents the results for the TS-E4 instance solutions. The column heading ρ indicates the available number of planes and the rest of the column headings have the same meaning has decried for Table 4.

Table 12: Solution results for the TS-E4 problem instances

ρ	count	value	#T	t.util	time	gap
4	77	78.1	2	90.09	89829	0.00
5	85	85.0	3	90.40	172808	0.02
6	86	85.9	2	90.56	172807	0.06
10	92	90.4	2	97.08	172808	0.04