Robust Aircraft Routing

Chiwei Yan, Jerry Kung
Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
{chiwei, jkung}@mit.edu

We propose a robust optimization approach to minimize total propagated delay in the aircraft routing problem, a setting first developed by Lan et al. (2006) and then extended by Dunbar et al. (2012, 2014). Instead of minimizing the expected total propagated delay by assuming that flight leg delays follow specific probability distributions, our model minimizes the maximal possible total propagated delay when flight leg delays lie in a pre-specified uncertainty set. We develop exact and tractable solution approaches for our robust model. The major contribution of our model is that it allows us to explicitly model and handle correlation in flight leg delays (e.g., due to weather or various air traffic management initiatives) that existing approaches cannot efficiently incorporate. Using both historical delay data and simulated data, we evaluate the performance of our model and benchmark against the state-of-the-research stochastic approach (Dunbar et al. 2014). In most of the cases, we observe that our model outperforms the existing approach in lowering the mean, reducing volatility, and mitigating extreme values of total propagated delay. In the cases where a deficit in one of the three criteria exists, gains in the other two criteria usually offset this disadvantage. These results suggest that robust optimization approaches can provide promising results for the aircraft routing problem.

Key words: aircraft routing; robust airline planning; delay propagation; robust optimization
History: This version, 05/31/2015

1. Introduction

Airline delays are prevalent and costly. In 2013, almost one out of every four flight legs operated by major U.S. carriers arrived more than 15 minutes late; of these delayed flights, one-third are a result of propagated delay: the late arrival of an aircraft causing a late departure (Bureau of Transportation Statistics 2013). One of the key drivers for these propagated delays is that airlines have in the past employed optimization models in an attempt to maximize profit; this has often led to the creation of tight schedules in an attempt to increase aircraft utilization and decrease crew salaries for time on the ground (Klabjan et al. 2001). As a consequence, delays propagate very rapidly throughout the network. This in turn leads to significant costs for an airline, such as additional pay for the crew, customer dissatisfaction, and further decreased utilization of aircraft due to flight cancellation.

To mitigate the discrepancy between what is planned and what actually happens, the research community has developed robust planning methods to pro-actively consider such delays and dis-

ruptions. These methods create schedules with the objective of constructing plans that have one of two kinds of robustness built in: (1) ease of repair once disrupted or (2) diminished propensity of delay disturbances. Within the first category, Rosenberger et al. (2004) develops a string-based fleet assignment models that embed many short cancellation cycles and limit the number of aircraft that can serve each hub. Such characteristics prevent cancelling a series of flights and isolate the disruption to a particular hub. Smith and Johnson (2006) introduce the concept of "station purity" which limits the number of fleet type that can serve each aiport. This creates additional aircraft swapping opportunities. Shebalov and Klabjan (2006) proposes a robust crew pairing model that maximizes the number of crews that can potentially be swapped in operations. Yen and Birge (2006) develops a two-stage stochastic integer program with recourse for crew scheduling problem. Their model produces robust crew schedule by incorporating disruptions in the evaluation of crew schedules during the long-range planning phase. Recently, Froyland et al. (2013) presents a recoverable robust aircraft routing model as a two-stage stochastic program. They explicitly model aircraft routing deicision in the first stage, and a full set of recovery operations including flight cancellation, delay, and aircraft swapping in the second stage recourse problem.

Our line of research falls into the second category, where we aim to come up with airline schedule that is less susceptible to delay disturbance. One existing idea in this category is to reduce delay propagation by re-timing flight schedule. Ahmadbeygi et al. (2010) uses a propagation tree to minimize propagated delay so that slack in the schedule can be re-allocated to where it is required most. In this research, we are not going to study the re-timing idea. Instead, the focus of this paper is on building robustness into the aircraft routing problem to protect against propagated delay. Aircraft routing is a critical airline planning phase in which the goal is to create a sequence of flight legs to be operated by individual aircraft such that each flight is covered in exactly one routing and all aircraft are properly maintained. As Lan et al. (2006) indicate, this problem can usually be cast as a feasibility problem, thus providing flexibility to achieve desired robustness by designing an appropriate objective function. To illustrate the impact of aircraft routing on delay propagation, suppose two connecting fights are close to each other with respect to timing (the arrival time of the first flight is very close to the departure time of the second flight). In this case, it is not desirable to fly both using a single aircraft, because even a small amount of delay from the first flight will end up propagating to the second flight.

The idea of reducing propagated delay through clever routing is not new. Some of the pioneers include Lan et al. (2006), who develop a stochastic integer program to minimize the total expected propagated delay along aircraft routes, where all the flight leg delays are fitted as independent log-normal distributions using historical data. To solve the program, they design an approximation algorithm based on column generation. To the best of our knowledge, there is no efficient solution

approach existed that can solve this model to optimality, even when flight leg delays are assumed to be independent random variables. This is perhaps due to the nonlinearity in the computation of propagated delay.

Later on, Dunbar et al. (2012) propose an improved solution approach to solve a deterministic integrated aircraft routing and crew pairing problem to minimize total propagated delay, given that flight leg delays are known to be constant. This work was recently extended to incorporate stochastic flight leg delay information (Dunbar et al. 2014). They use historical flight leg delay data to construct random scenarios and develop two algorithms to solve the stochastic program. Borndörfer et al. (2010) develop an alternative robust aircraft routing formulation whose objective is to minimize the sum of the probability of when propagated delay is positive over all flights, assuming that flight leg delays follow certain independent probability distributions. They demonstrate the effectiveness of their model in reducing the amount of propagated delay through computational experiments, but in theory it is unclear how this objective function is related to the actual propagated delay.

In this work, we depart from previous methods (Lan et al. 2006, Dunbar et al. 2014) that use *stochastic optimization* to minimize the expectation of propagated delay. We instead utilize *robust optimization* to minimize the maximal possible propagated delay. In contrast to the previous literature, which models flight leg delays as probability distributions, we utilize uncertainty sets to capture the stochasticity of flight leg delays. The major motivation of applying robust optimization to the aircraft routing problem comes from the ability to handle correlation in flight leg delays. Flights arriving to the same airport at similar times are usually delayed due to various air traffic management initiatives or inclement weather. Incorporating delay correlation into robust aircraft planning could provide additional benefits. As mentioned earlier, no efficient methods capable of dealing with delay correlation data have been explored in the robust aircraft routing literature.

On the other hand, robust optimization has been successfully applied to a number of application areas where the problems are combinatorially hard and involve correlated and complex uncertainty (Bertsimas et al. 2011). Although a major criticism of the robust optimization-based approach is the conservativeness of optimizing against the worst case as opposed to the expectation, under many circumstances, it may still perform better than stochastic optimization in expectation due to proper utilization of correlation data. This is mainly due to the added tractability that robust optimization enjoys, which allows it to handle complex uncertainty. Furthermore, additional protection is built into robust optimization, especially when realized data differs significantly from historical data (Bertsimas and Thiele 2006, Bertsimas et al. 2013b).

Applying robust optimization to the aircraft routing problem is fairly new. Until now, only one study (Marla and Barnhart 2010) has considered this approach. They compare the performance of

three different classes of models: 1) chance-constrained programming, 2) robust optimization, and 3) stochastic optimization (Lan et al. 2006). In their consideration of robust optimization, they model flight leg delays using a budget uncertainty set developed by Bertsimas and Sim (2004). Although their conclusions do not favor the robust optimization approach, we believe that by modeling uncertainty sets differently and solving the resultant robust problem cleverly, we may still be able to improve upon existing methods. Furthermore, we aim to provide comprehensive benchmarks to better evaluate the value of this approach.

The remainder of this paper is organized as follows. In Section 2, we present our robust formulation of the robust aircraft routing problem in detail. We also discuss how to model the uncertainty set that captures flight leg delays. Our uncertainty set is a polyhedral set based on the Central Limit Theorem and incorporates correlation structure in flight leg delays. In Section 3, we develop an exact decomposition solution approach for our robust model based on column-and-row generation. We also discuss detailed computational approaches to solving the resulting separation and pricing sub-problems. In Section 4, we conduct extensive computational experiments to evaluate the benefits of our robust model and benchmark it against a state-of-the-research model. We conclude with a discussion in Section 5.

2. The Robust Aircraft Routing Problem Formulation

In this section, we describe our formulation for the robust aircraft routing problem using the framework of robust optimization. We first outline our mathematical formulation and then discuss how we model and construct uncertainty sets for flight leg delays.

2.1. Robust Aircraft Routing Formulation

The aircraft routing problem can be formally stated as follows: given a set of aircraft of a specific fleet type and a set of flights that must be operated, determine how the fleet can be routed so that each flight leg is included in exactly one aircraft routing, and all aircraft are properly maintained. Among all feasible assignment of aircraft to routes, we seek the one that incurs the least amount of total propagated delay.

We denote by $\mathcal{R}$ the set of all feasible routes, $\mathcal{F}$ the set of flights, and N the total number of available aircraft. For each $r \in \mathcal{R}$, $F(r)$ represents the sequence of flights along route r. All maintenance-feasible routes can be represented by the columns of an $|\mathcal{F}| \times |\mathcal{R}|$ binary matrix A, where $A_{i,j} = 1$ if route j visits flight i, and $A_{i,j} = 0$ otherwise. Binary decision variables x_r denote whether or not route r is chosen in the optimal solution, with $x_r = 1$ if it is selected, and $x_r = 0$ otherwise. As in Lan et al. (2006) and Dunbar et al. (2012), we consider two kinds of delay in our model:

1. *primary delay*, denoted d_j for each flight leg $j \in \mathcal{F}$, which includes en-route delay, passenger connection delay, ground handling delay, and other delays that are not a function of routing.

2. *propagated delay*, denoted p_j^r for flight leg $j \in F(r)$ on route $r \in \mathcal{R}$, represents the amount of delay propagated to flight j caused by the late arrival of the upstream flight.

Given flight leg primary delay $\mathbf{d} \in R_+^{|\mathcal{F}|}$, the propagated delay $\mathbf{p}^r(\mathbf{d}) \in R_+^{|F(r)|}, r \in \mathcal{R}$ can be calculated iteratively as a function of $\mathbf{d}$. Suppose flights (i,j) are two consecutive flights on route r, $\text{slack}_{i,j}$ denotes the slack time for a connection (i,j), which is the difference between the scheduled arrival time of flight i and the scheduled departure time of flight j, minus the mean turnaround time for the corresponding aircraft type. The propagated delay to flight j from flight i then follows the expression

$$p_j^r(\mathbf{d}) = \max\{0, p_i^r(\mathbf{d}) + d_i - \text{slack}_{i,j}\}, \quad \forall (i,j) \in r, \forall r \in \mathcal{R}. \tag{1}$$

With the above notation, given flight leg delays $\mathbf{d}$, we write the deterministic aircraft routing problem (AR) with the objective of minimizing total propagated delay as the following integer program:

$$\begin{aligned}
\textbf{(AR)} \qquad \min_{\mathbf{x}} \quad & \sum_{r\in\mathcal{R}} c_r(\mathbf{d}) x_r = \sum_{r\in\mathcal{R}} \sum_{i\in F(r)} p_i^r(\mathbf{d}) x_r && (2)\\
\text{s.t.} \quad & A\mathbf{x} = \mathbf{e} && (3)\\
& \sum_{r\in\mathcal{R}} x_r \le N && (4)\\
& \mathbf{x} \in \{0,1\}^{|\mathcal{R}|}
\end{aligned}$$

Objective (2) is the sum of propagated delay over all flights, given constant flight leg primary delays $\mathbf{d}$. Constraints (3) are the flight cover constraints that ensure that each flight leg must be covered by only one aircraft routing. Constraints (4) are the fleet count constraints that keep the total number of aircraft used to be less than or equal to the number of available aircraft. To enhance the robustness of this model against all flight leg primary delays lying in a pre-specified uncertainty set $\mathcal{U}$, we consider the following robust aircraft routing (RAR) formulation:

$$\begin{aligned}
\textbf{(RAR)} \qquad \min_{\mathbf{x}} \max_{\mathbf{d}\in\mathcal{U}} \quad & \sum_{r\in\mathcal{R}} c_r(\mathbf{d}) x_r = \sum_{r\in\mathcal{R}} \sum_{i\in F(r)} p_i^r(\mathbf{d}) x_r && (5)\\
\text{s.t.} \quad & A\mathbf{x} = \mathbf{e} && (6)\\
& \sum_{r\in\mathcal{R}} x_r \le N && (7)\\
& \mathbf{x} \in \{0,1\}^{|\mathcal{R}|}
\end{aligned}$$

Objective (5) is to minimize the maximum possible total propagated delay when individual primary flight delays $\mathbf{d}$ are drawn from the uncertainty set $\mathcal{U}$.

2.2. Modeling Uncertainty Set $\mathcal{U}$ for Flight Leg Primary Delays d

In contrast to existing literature (Lan et al. 2006, Borndörfer et al. 2010), in which it is assumed that primary delays on flight legs $\mathbf{d}$ are independent for the purposes of tractability, we believe that flight legs have a clear dependence structure that should be accounted for in order to improve solution quality.

As an example, consider two different flights arriving to the same airport within the same hour. Both of these flights will be affected by the same weather conditions – one of the main drivers of primary delay – and may be delayed by similar amounts of time. Furthermore, various air traffic management initiatives are designed to delay flights which share the same characteristics. For instance, ground delay programs (GDPs) assign delays to flights entering into the same destination airport, and airspace flow programs (AFPs) control flights flying through the same flow constrained area (FCA). Moreover, different flow management programs often have complex interactions with each other (Barnhart et al. 2012). Thus, correlation of primary delay between different flights is a natural phenomenon that should be considered when generating aircraft routes.

With this insight in mind, we allow covariance data to play a central role in our uncertainty sets. Using historical schedule data from the Airline Service Quality Performance (ASQP) database, we first compute the primary delay for each flight leg by subtracting the propagated delay from total arrival delay based on the algorithm provided in Lan et al. (2006). We then calculate the sample means $\hat{\mu} \in R_+^{|\mathcal{F}|}$ of primary delays for each flight leg and create a sample covariance matrix $\hat{\Sigma} \succeq 0, \hat{\Sigma} \in R^{|\mathcal{F}|\times|\mathcal{F}|}$. Using this data, we create the following uncertainty sets for flight leg delays $\mathbf{d}$, where $C = \hat{\Sigma}^{-\frac{1}{2}}$, and $\hat{\sigma}_f$ denotes the standard deviation of the primary delay for flight $f \in \mathcal{F}$:

$$\mathcal{U}_p := \left\{ \mathbf{d} \in \mathbb{R}_+^{|\mathcal{F}|} \mid \|\mathbf{C}(\mathbf{d} - \hat{\mu})\|_p \leq \sqrt{|\mathcal{F}|} \cdot \Gamma; \left| \frac{d_f - \hat{\mu}_f}{\hat{\sigma}_f} \right| \leq \Gamma, \forall f \in \mathcal{F} \right\}, \tag{8}$$

where $p \in [0, +\infty]$, $\Gamma \in [0, +\infty)$ are two exogenous parameters. The first constraint in the uncertainty set is inspired by the Central Limit Theorem and is commonly used in the robust optimization literature. To see this, when primary leg delays $\mathbf{d}$ are uncorrelated, $C = \text{diag}(\frac{1}{\sigma_f})_{f\in\mathcal{F}}$, the first constraint is reduced to $\|\text{diag}(\frac{1}{\sigma_f})_{f\in\mathcal{F}}(\mathbf{d} - \mu)\|_p \leq \sqrt{|\mathcal{F}|} \cdot \Gamma$. In the case $p = 2$, this uncertainty set is an ellipsoid centered around μ, where the length of each semi-principal axis is the value of the standard deviation of primary delays for the corresponding particular flight. When the primary delays are correlated, the uncertainty set would correspond to a stretching and rotation of the ellipsoid. For example, when primary delays for two flights are positively correlated, the uncertainty set would capture simultaneous large and small delays for pairs of correlated flights and would lose some regions where one flight delay is large, and the other is small, and vice versa. We also add box constraints for each individual primary delay d_f to control the degree to which each

individual primary delay can vary from its historical mean. The parameter Γ, commonly referred to as the budget of uncertainty, controls the protection level we want: the larger the Γ, the more conservative we are in the sense that we protect against delays that are allowed to lie in a larger set. For example, for this Central Limit Theorem set (8), $\Gamma = 3$ usually gives us a high level of protection when the data follows a normal distribution, because almost all of the probability mass lies in the range of $[\mu - 3\sigma, \mu + 3\sigma]$ for normal random variables. We refer interested readers to Bertsimas et al. (2004) for more details regarding this uncertainty set.

We choose $p = 1$ (L_1 norm) and introduce auxiliary variables $y_i, i \in \{1, \cdots, |\mathcal{F}|\}$ so that $\mathcal{U}$ can be equivalently formulated as a polyhedral set:

$$\mathcal{U} := \left\{ \mathbf{d} \in \mathbb{R}_+^{|\mathcal{F}|} \mid \exists y \in R^{|\mathcal{F}|} \text{ s.t. } \sum_{i=1}^{|\mathcal{F}|} y_i \le \sqrt{|\mathcal{F}|} \cdot \Gamma; \pm c_i^T(\mathbf{d} - \hat{\mu}) \le y_i, \forall i \in \{1, \cdots, |\mathcal{F}|\}; \left| \frac{d_f - \hat{\mu}_f}{\hat{\sigma}_f} \right| \le \Gamma, \forall f \in \mathcal{F} \right\}, \tag{9}$$

where c_i^T is the ith row of matrix C.

3. Solution Approach

In this section, we describe our solution approach via a column-and-row generation framework for the robust aircraft routing model (RAR) introduced in Section 2.1. We also discuss our computational approaches for solving the separation problem and the pricing problem, respectively.

3.1. Column-and-Row Generation Framework

As in the deterministic aircraft routing problem (2)-(4), RAR contains a huge number of potential decision variables (aircraft routes). Because it is impractical to enumerate all feasible routes explicitly, branch-and-price (Barnhart et al. 1998b) is often used to solve such problems. Unfortunately, this means that the traditional dualization approach for obtaining the robust counterpart (RC) introduced by Ben-Tal and Nemirovski (1999) cannot be applied in this setting, because we do not have the full set of decision variables included in the model before starting the solution process. However, we can still apply an iterative cutting-plane method (Bertsimas et al. 2014) to repeatedly solve RAR with a finite subset of the constraints. At each iteration, we check whether any violated constraints can be generated, adding them if they exist, and re-solving until there are no more violated constraints. In order to use the cutting-plane method, we consider the equivalent epigraph reformulation (epi-RAR) for the robust aircraft routing model (RAR),

$$\begin{aligned}
\textbf{(epi-RAR)} \quad \min_{\mathbf{x},z} \quad & z \\
\text{s.t.} \quad & \sum_{r \in \mathcal{R}} \sum_{i \in F(r)} p_i^r(\mathbf{d}) x_r \le z, \quad \forall \mathbf{d} \in \mathcal{U} && (10) \\
& A\mathbf{x} = \mathbf{e} && (11) \\
& \sum_{r \in \mathcal{R}} x_r \le N && (12) \\
& \mathbf{x} \in \{0,1\}^{|\mathcal{R}|}
\end{aligned}$$

Note that *epi-RAR* is a mixed integer linear program with an infinite number of rows (constraint (10)) and a huge number of columns. To address this, we utilize efficient decomposition methods on both rows and columns. We refer to constraints (10) as *robustifying constraints* because they help to protect against all possible flight primary delays in the uncertainty set.

3.1.1. The Separation Problem

For the relaxed problem *epi-RAR* with only a subset of robustifying constraints (10), we let $\mathbf{x}^*$ denote the optimal solution, and we let z^* denote the optimal objective value. We denote by $\mathcal{R}_{\mathbf{x}^*}$ the feasible aircraft routes that are specified by $\mathbf{x}^*$ (i.e., $\mathcal{R}_{\mathbf{x}^*} = \{r \in \mathcal{R} \mid x_r^* = 1\}$). The following separation problem (SEP) checks whether any violated constraints in (10) must be added into the model:

$$\textbf{(SEP)} \qquad z(\mathcal{R}_{x^*}) = \max_{\mathbf{p},\mathbf{d}} \quad \sum_{r \in \mathcal{R}_{x^*}} \sum_{i \in F(r)} p_i^r \tag{13}$$

$$\text{s.t.} \quad p_j^r = \max\{0, p_i^r + d_i - \text{slack}_{i,j}\}, \quad \forall r \in \mathcal{R}_{x^*},\ \forall (i,j) \in r \tag{14}$$

$$\mathbf{d} \in \mathcal{U}$$

Given a specific aircraft routing, *SEP* optimizes the flight leg primary delays $\mathbf{d} \in \mathcal{U}$ to maximize the total propagated delay stated in objective (13). We note here that *SEP* cannot be solved in parallel, since primary delays for flights from different routes could be jointly constrained in uncertainty set $\mathcal{U}$.

Constraint (14) is nonlinear in nature, but we can linearize it through a big-M re-formulation as follows, denoted as *SEP-bigM*:

$$\textbf{(SEP-bigM)} \qquad z(\mathcal{R}_{\mathbf{x}^*}) = \max_{\mathbf{p},\mathbf{d}} \quad \sum_{r \in \mathcal{R}_{\mathbf{x}^*}} \sum_{i \in F(r)} p_i^r \tag{15}$$

$$\text{s.t.} \quad p_j^r \le p_i^r + d_i - \text{slack}_{i,j} + M_j^1 I_j, \quad \forall r \in \mathcal{R}_{\mathbf{x}^*},\ \forall (i,j) \in r \tag{16}$$

$$p_j^r \le 0 + M_j^2 (1 - I_j), \quad \forall r \in \mathcal{R}_{\mathbf{x}^*},\ \forall (i,j) \in r \tag{17}$$

$$\mathbf{d} \in \mathcal{U}$$

$$I \in \{0,1\}^{|\mathcal{F}|}$$

where M_j^1, M_j^2 are sufficiently large constants (big-Ms), and I are auxiliary indicator variables that are explained in more detail below.

Since $\mathcal{U}$ is a polyhedral set in our case, *SEP-bigM* is a mixed integer linear program. Along with the objective, constraints (16) and (17) together restrict that $p_j^r = \max\{0, p_i^r + d_i - \text{slack}_{i,j}\}, (i,j) \in r$, which is the definition of propagated delay. To see this, if auxiliary variable $I_j = 1$, constraints (16) become ineffective, and constraints (17) indicate $p_j^r = 0$ due to the maximization of the objective

function; similarly, if $I_j = 0$, constraints (16) ensure $p_j^r = p_i^r + d_i - \text{slack}_{i,j}$. Thus at the optimal solution, I_j will automatically pick the right value to ensure $p_j^r = \max\{0, p_i^r + d_i - \text{slack}_{i,j}\}$. We solve *SEP-bigM* with route $\mathcal{R}_{x^*}$ and denote the optimal flight leg primary delay allocation by $\mathbf{d}^*$. If $z(\mathcal{R}_{x^*}) \leq z^*$, then no violated constraint need be added; if $z(\mathcal{R}_{x^*}) > z^*$, we add the violated constraint $\sum_{r\in\mathcal{R}} \sum_{i\in F(r)} p_i^r(\mathbf{d}^*) x_r \leq z$ to the relaxed master problem and re-solve.

Solving *SEP-bigM* is not an easy task due to the combinatorial structure of mixed integer programs. However, tightening the big-M constants for each constraint can make the formulation substantially stronger than if big-Ms are assigned arbitrarily large values. We present a method below to determine tightened but sufficiently large big-Ms.

PROPOSITION 1. *Let* $d_j^* = \max_{\mathbf{d}\in\mathcal{U}} d_j, \forall j \in \{1, \cdots, |\mathcal{F}|\}$, *and* $p_j^r(\mathbf{d}^*), \forall j \in \{1, \cdots, |\mathcal{F}|\}$ *be the corresponding propagated delays under flight leg delays* $\mathbf{d}^* = \{d_1^*, \cdots, d_{|\mathcal{F}|}^*\}$. *Then* $M_j^1 = slack_{i,j}$, $M_j^2 = p_j^r(\mathbf{d}^*)$ *are valid values of big-Ms for SEP-bigM.*

Proof. In the optimal solution, indicator variable I_j is 1 if $p_i^r + d_i - \text{slack}_{i,j} \leq 0$, which makes $p_j^r = \max\{0, p_i^r + d_i - \text{slack}_{i,j}\} = 0$. Since constraint (17) becomes $p_j^r \leq 0$ under $I_j = 1$, in order to let $p_j^r = 0$ in the optimal solution, the right hand side of constraint (16) should be greater than or equal to 0, i.e. $p_i^r + d_i - \text{slack}_{i,j} + M_j^1 \geq 0$. Since $p_i^r, d_i \geq 0$, setting $M_j^1 = \text{slack}_{i,j}$ is valid. On the other hand, in the optimal solution, I_j is 0 if $p_i^r + d_i - \text{slack}_{i,j} \geq 0$, thus ensuring $p_j^r = \max\{0, p_i^r + d_i - \text{slack}_{i,j}\} = p_i^r + d_i - \text{slack}_{i,j}$. Since constraint (16) becomes $p_j^r \leq p_i^r + d_i - \text{slack}_{i,j}$ under $I_j = 0$, to make $p_j^r = p_i^r + d_i - \text{slack}_{i,j}$ in the optimal solution, the right hand side of constraint (17) should satisfy $M_j^2 \geq p_i^r + d_i - \text{slack}_{i,j}$. Note that by construction, $d_i \leq d_i^*, \forall i \in \{1, \cdots, |\mathcal{F}|\}, \forall \mathbf{d} \in \mathcal{U}$, thus $p_i^r + d_i - \text{slack}_{i,j} \leq \max\{0, p_i^r + d_i - \text{slack}_{i,j}\} = p_j^r(\mathbf{d}) \leq p_j^r(\mathbf{d}^*), \forall r \in \mathcal{R}_{x^*}, \forall (i,j) \in r$ because $p_j^r(\mathbf{d})$ is a monotonically increasing function. This result shows that $M_j^2 = p_j^r(\mathbf{d}^*)$ is valid. □

Computing tightened big-M according to Proposition 1 requires solving $|\mathcal{F}|$ separate optimization problems. The complexity of these optimization problems depends on how we construct uncertainty set $\mathcal{U}$. Under the set (9) we use, these optimization problems become linear programs, which can be solved very efficiently. With tightened big-M, this formulation works fairly efficiently when we test it for moderately large industry size instances in Section 4. When using it to solve very large instances, careful construction of uncertainty set $\mathcal{U}$ can greatly speed up *SEP-bigM*. For instance, for the uncertainty set described in (8), we could group flights according to their destination airport or pathway and estimate the covariance matrix for each group individually by assuming that primary delays between two flights in two different groups are independent. In that case, $C = \hat{\Sigma}^{-\frac{1}{2}}$ becomes sparse, making *SEP-bigM* easy to solve. Along the same lines, directly estimating a sparse inverse covariance matrix from data is also useful in increasing tractability for large instances (Friedman et al. 2008, Hsieh et al. 2011).

3.1.2. The Pricing Problem

Each time a new robustifying constraint is added to the relaxed master problem, the resulting program needs to be solved using column generation. Let $G=(\mathcal{N},\mathcal{A})$ be a directed acyclic graph with a single source node s and a single terminal node t. The node set $\mathcal{N}$ represents the set of flights and arc set $\mathcal{A}$ corresponds to feasible connections between flights. For simplicity and testing purposes, the source node and terminal node are dummy nodes that link to all flight nodes. In practice, various operational restrictions could be added (Barnhart et al. 1998a). For example, the source node and terminal node can only be linked to flights departing from and arriving to maintenance-compatible airports, respectively. Each flight node i possesses a weight $-\mu_i$ corresponding to the negative dual price of the i^{th} flight covering constraint (11). We denote by r the dual price of the fleet count constraint (12). Without loss of generality, suppose there are k constraints in the robustifying constraint set (10). Let $\mathbf{d^1},\mathbf{d^2},\cdots,\mathbf{d^k}$ be the corresponding flight primary delays vectors and $s_1,s_2,\cdots,s_k$ be the dual prices for robustifying constraints 1 to k. Thus for the RAR pricing problem, we wish to find an $s-t$ path $\pi^*=\{s,n_1,n_2,\cdots,t\}$ that minimizes reduced cost

$$\pi^* = \underset{\pi \text{ is an } s-t \text{ path}}{\arg\min} \; rc_\pi = \left\{ \left(-r - \sum_{i\in\pi: i\neq s,t} \mu_i \right) - \sum_{j=1}^{k} \left(\sum_{i\in\pi: i\neq s,t} s_j p_i^\pi(\mathbf{d^j}) \right) \right\}. \tag{18}$$

If reduced cost $rc_{\pi^*}<0$, the minimizing route π^* forms a new column of A and is assigned a cost of $\sum_{i\in\pi^*: i\neq s,t} p_i^{\pi^*}(\mathbf{d^j})$ for robustifying constraints $j=1,2,\cdots,k$.

3.1.3. Solving the Pricing Problem

We discuss here in detail the revised label setting algorithm to solve (18), which dynamically calculates the propagated delays using (1) and the reduced cost of the path. As Dunbar et al. (2012) point out, because the propagated delay $p_i^\pi(\cdot)$ is not a simple sum of delays along the path from s to i, problem (18) cannot easily be cast as a minimum cost network flow problem. The label setting algorithm we develop here is modified from Dunbar et al. (2012).

Let π be an $s-t$ path in G (an ordered collection of nodes $\{s,n_1,\cdots,n_q,t\}$ in $\mathcal{N}$ with $(s,n_1),(n_q,t)\in\mathcal{A}$ and $(n_l,n_{l+1})\in\mathcal{A}$ for all $l=1,\cdots,q-1$). For $n\in\pi$, let $\pi(n)$ denote the ordered collection of nodes in the path π truncated so that the final node in the list is n. Define truncated reduced cost $rc_{\pi(n)}=\left(-r-\sum_{i\in\pi(n):i\neq s,t}\mu_i\right)-\sum_{j=1}^{k}\left(\sum_{i\in\pi(n):i\neq s,t} s_j p_i^{\pi(n)}(\mathbf{d^j})\right)$. Denote by $p_n^\pi(\cdot)$ the propagated delay at node n, computed along path $\pi(n)$ using (1). Due to the nonlinear nature of the propagated delay formula (1), our labels must track both the accumulated reduced cost $rc_{\pi(n)}$ at node n along path π and the propagated delay $p_n^\pi(\cdot)$. This motivates the following dominance conditions for paths destined for the same node.

Definition 1 (Path Dominance Condition). Let paths $\pi(n)$, $\eta(n)$ be two different paths destined for the same node n. We say that $\pi(n)$ dominates $\eta(n)$ if $rc_{\pi(n)} \leq rc_{\eta(n)}$, $p_n^{\pi(n)}(\mathbf{d^j}) \leq p_n^{\eta(n)}(\mathbf{d^j}), \forall j \in \{1, \cdots, k\}$ and $(rc_{\pi(n)}, p_n^{\pi(n)}(\mathbf{d^j})) \neq (rc_{\eta(n)}, p_n^{\eta(n)}(\mathbf{d^j}))$ such that $s_j < 0$, where k is the number of robustifying constraints (10) at the current iteration.

Lemma 1. *Let $m \in \mathcal{N}$ such that $(n,m) \in \mathcal{A}$. If $\pi(n)$ dominates $\eta(n)$, then $\{\pi(n), m\}$ dominates $\{\eta(n), m\}$.*

Proof. From (1) we have $\forall j \in \{1, \cdots, k\}$ such that $s_j < 0$,

$$p_m^{\{\pi(n),m\}}(\mathbf{d^j}) = \max\{0, p_n^{\{\pi(n),m\}}(\mathbf{d^j}) + d_n^j - s_{n,m}\}, \tag{19}$$

$$p_m^{\{\eta(n),m\}}(\mathbf{d^j}) = \max\{0, p_n^{\{\eta(n),m\}}(\mathbf{d^j}) + d_n^j - s_{n,m}\}. \tag{20}$$

By Definition 1, (19) and (20),

$$p_n^{\{\pi(n),m\}}(\mathbf{d^j}) = p_n^{\pi(n)}(\mathbf{d^j}) \leq p_n^{\eta(n)}(\mathbf{d^j}) = p_n^{\{\eta(n),m\}}(\mathbf{d^j})$$

Thus we have,

$$p_m^{\{\pi(n),m\}}(\mathbf{d^j}) \leq p_m^{\{\eta(n),m\}}(\mathbf{d^j}),\ \forall 1 \leq j \leq k, s_j < 0 \tag{21}$$

For the reduced cost, linear programming duality tells us that dual prices $s_j \leq 0,\ \forall 1 \leq j \leq k$ for all robustifying constraints (10), so we have:

$$rc_{\{\pi(n),m\}} = rc_{\{\pi(n)\}} - \mu_m - \sum_{j=1}^{k} s_j p_m^{\{\pi(n),m\}}(\mathbf{d^j}) = rc_{\{\pi(n)\}} - \mu_m - \sum_{1 \leq j \leq k: s_j < 0} s_j p_m^{\{\pi(n),m\}}(\mathbf{d^j}), \tag{22}$$

$$rc_{\{\eta(n),m\}} = rc_{\{\eta(n)\}} - \mu_m - \sum_{j=1}^{k} s_j p_m^{\{\eta(n),m\}}(\mathbf{d^j}) = rc_{\{\eta(n)\}} - \mu_m - \sum_{1 \leq j \leq k: s_j < 0} s_j p_m^{\{\eta(n),m\}}(\mathbf{d^j}). \tag{23}$$

Since $rc_{\{\pi(n)\}} \leq rc_{\{\eta(n)\}}$ and (21), we have $rc_{\{\pi(n),m\}} \leq rc_{\{\eta(n),m\}}$. □

By induction, Lemma 1 suggests that if ϖ is a path from m to terminal node t, and $(n,m) \in \mathcal{A}$, we have $rc_{\{\pi(n),\varpi\}} \leq rc_{\{\eta(n),\varpi\}}$. Thus equipped with this definition of path dominance, we can directly apply the label setting algorithm of Dunbar et al. (2012) (Algorithm 1) to solve (18) where at each node, we only create labels for those paths that are not dominated by any other paths at that node.

This label setting algorithm is fairly efficient on the instances we test in Section 4. One major reason for this is that some of the robustifying constraints (10) are not binding in the optimal solution of the relaxed master problem. Thus, some of the dual prices corresponding to those robustifying constraints are zero. This makes the pricing algorithm, or more precisely the dominance condition, easier than it appears to be.

Algorithm 1 (Dunbar et al. 2012) Label Setting Algorithm for Pricing New Columns

Initialize:

Set $I_s = \{s\}$ and $I_i = \emptyset$ for all $i \in \mathcal{N} \backslash \{s\}$.
Set $M_i = \emptyset$ for each $i \in \mathcal{N}$.

loop

if $\bigcup_{i \in \mathcal{N}} (I_i \backslash M_i) = \emptyset$ **then**

return $\arg\min_{\pi(t) \in I_t} rc_{\pi(t)}$

else

choose $i \in \mathcal{N}$ and $\pi(i) \in I_i \backslash M_i$ so that $rc_{\pi(i)}$ is minimal

for $(i, j) \in \mathcal{A}$ **do**

if path $\{\pi(i), j\}$ is not dominated by $\eta(j)$ for any $\eta(j) \in I_j$ **then**

set $I_j = I_j \bigcup \{\pi(i), j\}$

end if

end for

set $M_i = M_i \bigcup \{\pi(i)\}$

end if

end loop

3.2. Overall Algorithm

We summarize the column-and-row generation algorithm that we use to solve *epi-RAR* in Algorithm 2. The algorithm iteratively adds robustifying constraints and new columns by solving the corresponding separation and pricing sub-problems until the upper bound provided by the separation problem matches with the optimal value of the current relaxed *epi-RAR* problem.

4. Computational Experiments

In this section, we compare the performance of our robust model against that of the state-of-the-research approach (Dunbar et al. 2014) on both historical delay data and simulated delay data. All computer programs were written in Java, and linear and integer programs were solved by IBM ILOG CPLEX 12.6 with default settings. We note here that the solution approach (Algorithm 2) for RAR requires branch-and-price at each iteration. Due to implementation difficulties, we only solve relaxed *epi-RAR* at the root node by column generation and do not perform any further branching. However, we observed that the optimal value of the LP relaxation for relaxed *epi-RAR* coincides with that of the IP solution in many cases. All computational tests were performed on a laptop equipped with an Intel Core i7-3520M CPU running at 2.90 GHz and 8 GB memory.

Algorithm 2 Simultaneous Column-and-Row Generation

Given:

Relaxed *epi-RAR* with only constraints (11) and (12)

Initialize:

Add first robustifying constraint $\sum_{r\in\mathcal{R}}\sum_{i\in F(r)} p_i^r(\mathbf{d^0})x_r \le z$ with an arbitrary $\mathbf{d^0}\in\mathcal{U}$.

loop

Solve relaxed *epi-RAR* using branch-and-price, get optimal routes $\mathcal{R}_{x^*}$ and optimal value z^*.

Given $\mathcal{R}_{x^*}$, solve *SEP-bigM*, get optimal flight primary delays $\mathbf{d}^*$ and optimal value $z(\mathcal{R}_{x^*})$.

if $z(\mathcal{R}_{x^*}) > z^*$ **then**

Add robustifying constraint $\sum_{r\in\mathcal{R}}\sum_{i\in F(r)} p_i^r(\mathbf{d}^*)x_r \le z$ to relaxed *epi-RAR*.

else

return $\mathcal{R}_{x^*}$

end if

end loop

4.1. Benchmark Model, Local Approach in Dunbar et al. (2014)

Based on the models and algorithms developed in Dunbar et al. (2012), Dunbar et al. (2014) further incorporate stochastic delay information to minimize expected total propagated delay. They model delay stochasticity by constructing a set of random scenarios Ω, where each scenario $\omega\in\Omega$ corresponds to primary delay values $\mathbf{d}^\omega$ for each flight. Each random scenario is realized with equal probability. They develop two algorithms to tackle this problem: 1) the exact approach and 2) the local approach. Since their approaches were originally designed for integrated aircraft routing and crew scheduling, we simplify here to only aircraft routing. The exact approach enumerates all the feasible aircraft routes $\mathcal{R}$. For each feasible routing $r\in\mathcal{R}$, they calculate its expected total propagated delay $\mathbb{E}_{\mathbf{d}}\left[c_r(\mathbf{d})\right] = \sum_{\omega\in\Omega}\frac{1}{|\Omega|}\left(\sum_{i\in F(r)} p_i^r(\mathbf{d}^\omega)\right)$. They then directly solve the deterministic model in (2)-(4) with all decision variables generated. They change the objective function to be $\sum_{r\in\mathcal{R}}\mathbb{E}_{\mathbf{d}}\left[c_r(\mathbf{d})\right]x_r$, which represents the sum of expected total propogated delay over all selected routings. Although this is an exact solution approach, it is not practical for industry-sized problems due to its lack of an efficient column generation process. As a result, we consider the local approach as our benchmark for comparison. The local approach is an approximation approach that incorporates stochastic delay information from the scenarios within the label-setting algorithm used in the pricing problem. At each step of the label-setting algorithm, it calculates the average delay propagation arriving at each flight node over all scenarios $\omega\in\Omega$. We denote $\hat{p}_j^r$ as the average propagated delay at node j along path r, and $\hat{p}_j^r = \frac{1}{|\Omega|}\sum_{\omega\in\Omega}\max\left\{0, \hat{p}_j^r + d_i^\omega - \text{slack}_{i,j}\right\}$. With this notation, we introduce the path dominance condition used in the local approach in Dunbar et al. (2014).

Definition 2 (Path Dominance Condition in Dunbar et al. (2014)). Let $\pi(n), \eta(n)$ be two different paths destined for the same node n. We say that $\pi(n)$ dominates $\eta(n)$ if $rc_{\pi(n)} \leq rc_{\eta(n)}$ and $\hat{p}_n^{\pi(n)} \leq \hat{p}_n^{\eta(n)}$.

The optimal solution of the local approach can be computed by column generation using the pricing algorithm (Algorithm 1) with path dominance condition as Definition 2.

4.2. Case Studies Based on Historical Delay Data

To evaluate the effectiveness of our robust model, we apply Algorithm 2 to create routes for two of the largest fleet types operated by a major US airline in the year 2007. The characteristics of the underlying networks are listed in Table 1, and all the flights selected are operated on a daily basis over the testing period. Because in practice the model will be built using historical data and then applied to future operations, all routings were created using ASQP data containing historical flight primary delays for the 31 days of July 2007 (the training set) and then evaluated out-of-sample on the 31 days of August 2007 (the testing set).

Table 1 Characteristics of Two Routing Problems

Network	Number of flight legs	Number of aircraft
N_1	106	24
N_2	117	23

For both sets of data, we benchmark the routes of our robust model against two other routes:

1. The original route in the historical schedule data as the baseline (B):

This is the airline's actual route for the selected flights.

2. The local approach provided in Dunbar et al. (2014) (DFW) as in Section 4.1:

The flight primary delays for the 31 days in July 2007 constitute 31 random delay scenarios. Each random scenario is realized with probability $\frac{1}{31}$.

We evaluate the performance of these three approaches along three criteria: 1) the average total propagated delay, 2) the volatility of total propagated delay (standard deviation), and 3) worst-case total propagated delay (maximum value). Figures 1 and 2 compare the performance of the robust approach over a range of values of the budget of uncertainty $\Gamma \in \{0.2, 0.4, \cdots, 2.8, 3.0\}$ to the DFW method on the training data from July 2007. We compute the ratios: $100 \cdot (\mathrm{B} - \mathrm{DFW})/\mathrm{B}$ and $100 \cdot (\mathrm{B} - \mathrm{RAR})/\mathrm{B}$ in each performance evaluation criterion to measure the improvement over the baseline (B). The long dashed lines represent the relative reduction in percentage in corresponding performance criterion for approach DFW over the baseline case (B), and the solid lines represent the relative reduction in percentage in corresponding performance criterion for approach RAR over the baseline (B) under various values of uncertainty budget Γ. Notice that for the robust approach,

as Γ increases, the performance along the measures of standard deviation and maximum value of propagated delay improves almost monotonically. This is because the routings are created on training data, and larger values of Γ protect against precisely the maximum total propagated delay.

Figure 1 Relative Improvements of the Algorithms RAR over B and DFW over B on Network N_1 (July 2007)

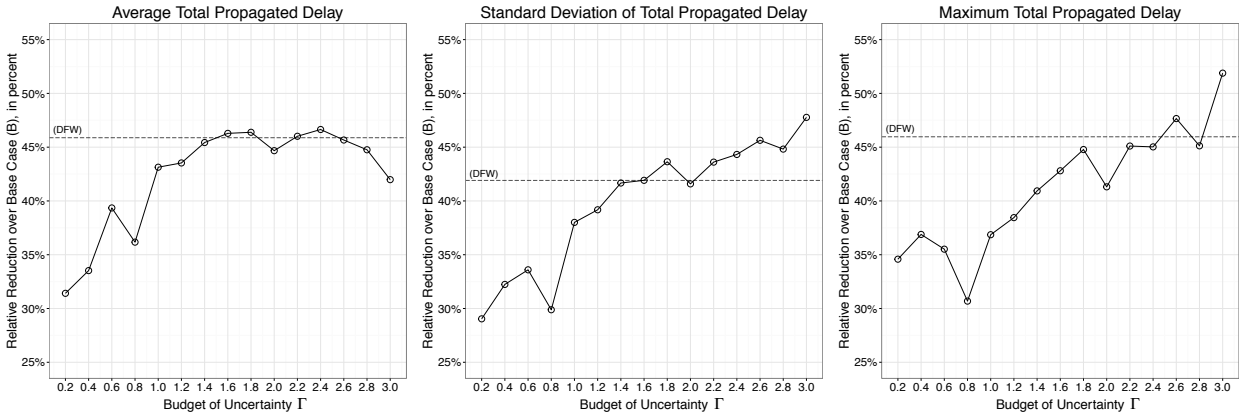
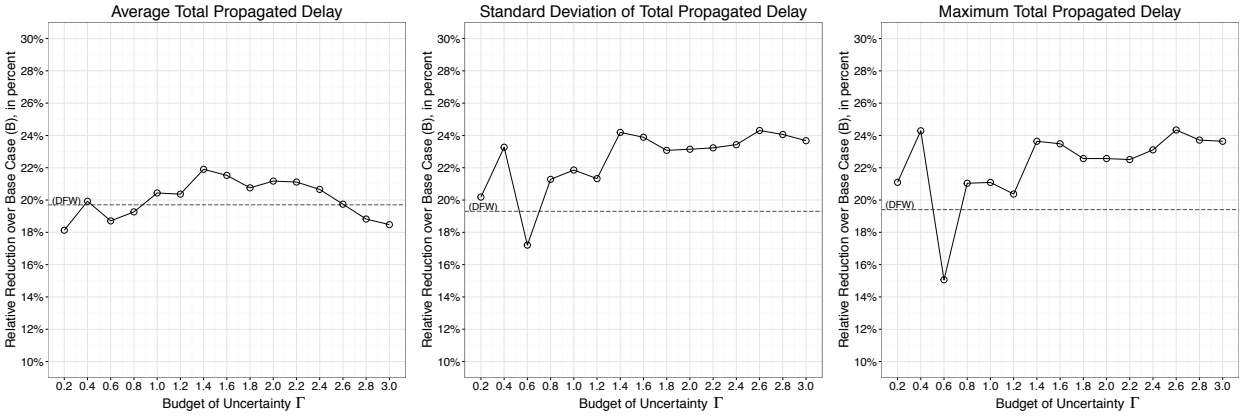


Figure 2 Relative Improvements of the Algorithms RAR over B and DFW over B on Network N_2 (July 2007)



When the optimal routes informed by the robust model formulated using the training data (July 2007) are applied to testing data (August 2007) as in Figures 3 and 4, we no longer see this monotonicity trend. In practice, we would select one particular value of Γ to create future routings. Guidelines for selecting the budget of uncertainty in previous literature (Bertsimas et al. 2011, 2013a) usually focus on probabilistic guarantees, i.e. the uncertainty set can be tuned so that constraints are robustly feasible with at least some probability. However, uncertainty of primary delays in our problem does not affect route feasibility, it only changes the amount of total

propagated delay. Thus, we decide to choose Γ based on the performance of propagated delay on the training set. We prioritize average total propagated delay as our main performance goal, thus based on Figures 1 and 2, we consider $\Gamma = 2.4$ to be appropriate for network N_1 and $\Gamma = 1.4$ to be appropriate for network N_2 because the robust model yields the lowest average total propagated delay under these two Γ in the training set. Table 2 presents summary statistics for all propagated delay performance criteria under the testing set for routes produced by each approach under both networks N_1 and N_2. Rows with values of Γ that are selected based on training set performance are in boldface. Note that generally, the best way to set the parameter Γ would be: (1) divide our historical dataset into training and validation components; (2) select the value of Γ that yields the best performance on the validation set (cross-validation can also be applied); and (3) once that value of Γ is set, we would then use it to generate future aircraft routings. For the sake of simplicity and also due to a dearth of data, we only use training set to select the value of Γ in this paper.

Figure 3 Relative Improvements of the Algorithms RAR over B and DFW over B on Network N_1 (August 2007)

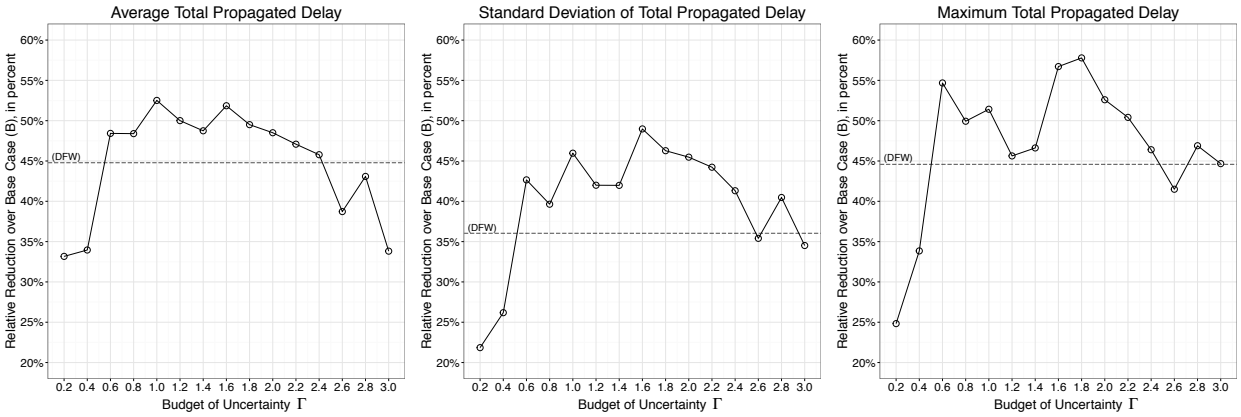
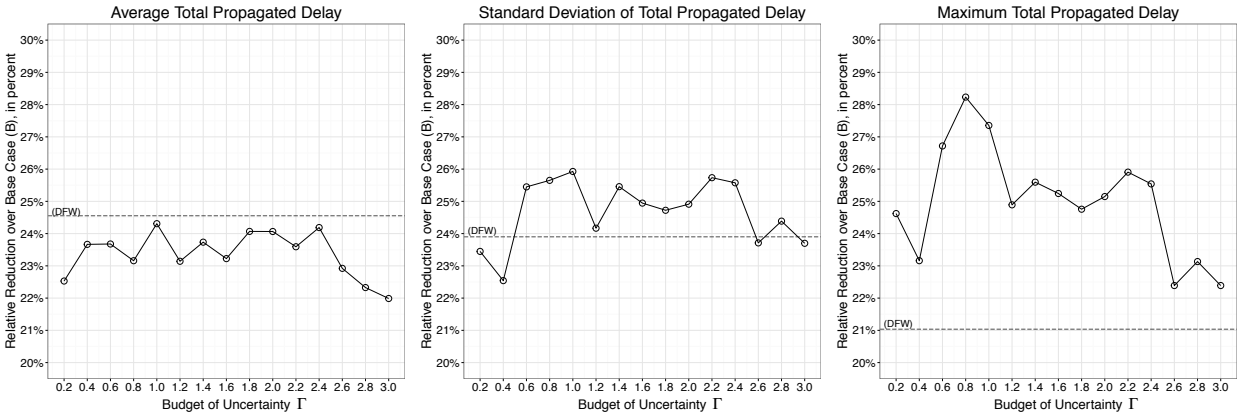


Figure 4 Relative Improvements of the Algorithms RAR over B and DFW over B on Network N_2 (August 2007)



Both approaches improve upon the baseline routing considerably in terms of average, standard deviation, and maximum of total propagated delay. For testing data on network N_1, when $\Gamma = 2.4$, RAR performs better than DFW in all of the performance criteria of total propagated delay that we consider (1.0% larger decrease in average total propagated delay, 6.2% larger decrease in standard deviation, and 1.8% larger decrease in maximum total propagated delay, as compared to baseline). For testing data on network N_2, when $\Gamma = 1.4$, RAR outperforms DFW according to the standard deviation (2.8% larger decrease from baseline) and maximum value of total propagated delay (4.6% larger decrease from baseline), but loses slightly in terms of average total propagated delay (0.8% smaller decrease from baseline).

Aside from the Γ we choose, for both flight networks under testing data, RAR actually outperforms DFW in reducing volatility and extreme delay under a wide range of Γ values $[0.6, 2.4]$ (especially in network N_2, RAR outperforms DFW in reducing extreme delay value under all Γ values). For very small (≤ 0.4) and very large (≥ 2.6) Γ values, RAR loses superiority in at least two evaluation criteria. This is because the uncertainty set $\mathcal{U}$ constructed by these Γ are either too conservative (for large Γ values) or too restrictive (for small Γ values) in representing potential primary delay. In other words, such Γ values do not accurately capture real world data. When $\Gamma \in [0.6, 2.4]$, in network N_1, compared to DFW, RAR can reduce on average 8.7% more of the standard deviation and on average 6.6% more of the maximum delay value from the baseline. For network N_2, RAR can reduce on average 2.6% more of the standard deviation and on average 4.9% more of the maximum delay value compared to DFW from the baseline. The superiority in reducing extreme delay values of RAR intuitively makes sense because of the min-max objective in RAR formulation. The reduction in volatility is a surprising by-product of reduced extreme delay values, even though we do not explicitly include it in the objective for RAR. Most surprisingly, for network N_1, RAR also has better performance in reducing average total propagated delay in the training set. When $\Gamma \in [0.6, 2.4]$, on average 4.3% more of the average total propagated delay can be reduced. In network N_2, RAR underperforms DFW in reducing average delay value, but not by too much (on average only 0.8% when $\Gamma \in [0.6, 2.4]$), especially in light of the gains made in reducing volatile and extreme values.

As for computational tractability, the efficiency of the RAR model depends heavily on the tractability of the separation problem. When the separation problem is easy to solve (e.g., as in network N_2), the RAR model is more efficiently solved for all values of Γ when compared to DFW. However, when the separation problem is relatively hard to solve (e.g., for network N_1 and larger Γ values), RAR is less efficient to solve. Overall, for both models, two moderately-large industry size instances can be solved to optimal in 20 minutes. Further computational time reduction for RAR could be achieved by performing matrix sparsifying tricks suggested in Section 3.1.1. We also point

Table 2 Comparative Algorithmic Performance on Two Test Networks

Approach	AVG. Total Propagated Delay (minute)	STD. Total Propagated Delay (minute)	MAX Total Propagated Delay (minute)	Total CPU Time (second)	*SEP-bigM* CPU Time (second)	Number of Robustifying Cuts Added
Flight Network N_1, 106 flights / 24 aircraft						
B	1,042.6	1,409.8	5,856	–	–	–
DFW	575.7	901.8	3,245	108.35	–	–
RAR ($\Gamma = 0.2$)	696.8	1,083.9	4,402	11.56	5.22	3
RAR ($\Gamma = 0.4$)	688.5	1,023.7	3,874	41.86	34.92	4
RAR ($\Gamma = 0.6$)	537.7	795.3	2,654	55.72	46.79	4
RAR ($\Gamma = 0.8$)	538.1	837.1	2,932	91.37	81.72	4
RAR ($\Gamma = 1.0$)	495.1	749.3	2,845	51.83	40.34	4
RAR ($\Gamma = 1.2$)	521.2	804.5	3,184	168.11	145.00	8
RAR ($\Gamma = 1.4$)	534.4	804.7	3,126	229.39	200.75	6
RAR ($\Gamma = 1.6$)	502.0	707.6	2,535	574.65	543.75	6
RAR ($\Gamma = 1.8$)	526.5	745.1	2,472	895.88	851.19	6
RAR ($\Gamma = 2.0$)	536.9	756.2	2,776	810.67	764.12	7
RAR ($\Gamma = 2.2$)	551.7	773.5	2,904	784.55	722.94	7
RAR ($\Gamma = 2.4$)	**565.3**	**813.9**	**3,139**	**857.48**	**781.74**	**7**
RAR ($\Gamma = 2.6$)	638.9	895.8	3,426	752.75	672.25	4
RAR ($\Gamma = 2.8$)	593.4	825.5	3,110	966.78	869.10	9
RAR ($\Gamma = 3.0$)	690.0	908.2	3,240	367.71	299.19	5
Flight Network N_2, 117 flights / 23 aircraft						
B	1,131.9	1,651.4	7,392	–	–	–
DFW	854.0	1,256.7	5,837	55.54	–	–
RAR ($\Gamma = 0.2$)	876.9	1,243.6	5,572	17.24	3.80	2
RAR ($\Gamma = 0.4$)	864.0	1,258.4	5,680	18.08	6.83	2
RAR ($\Gamma = 0.6$)	863.9	1,211.1	5,417	25.36	7.44	3
RAR ($\Gamma = 0.8$)	869.7	1,207.9	5,305	18.82	7.10	2
RAR ($\Gamma = 1.0$)	856.7	1,203.4	5,370	48.43	18.85	4
RAR ($\Gamma = 1.2$)	869.9	1,232.0	5,552	27.32	9.61	2
RAR ($\Gamma = 1.4$)	**863.2**	**1,211.0**	**5,500**	**44.56**	**12.99**	**4**
RAR ($\Gamma = 1.6$)	869.0	1,219.3	5,526	51.30	14.42	4
RAR ($\Gamma = 1.8$)	859.5	1,222.9	5,562	45.75	11.09	3
RAR ($\Gamma = 2.0$)	859.5	1,219.9	5,533	40.58	13.19	4
RAR ($\Gamma = 2.2$)	864.8	1,206.5	5,477	32.65	9.08	3
RAR ($\Gamma = 2.4$)	858.1	1,209.1	5,504	48.56	16.05	4
RAR ($\Gamma = 2.6$)	872.4	1,239.4	5,737	46.29	13.17	3
RAR ($\Gamma = 2.8$)	879.2	1,228.4	5,682	38.22	5.38	2
RAR ($\Gamma = 3.0$)	883.0	1,239.6	5,737	33.52	5.85	2

out here that the solution time of RAR does not depend on the size of the historical delay data for training. No matter how many days of historical flight delay we use, we can transform all of them into a single uncertainty set of the same size. On the other hand, the solution time of DFW grows approximately linearly with the number of days involved in the training set because the algorithm needs to average over all training days for each flight leg during the column generation process. This suggests RAR might require less computational effort when we want to use more historical delay data to train our model.

In this case study, careful modeling of the uncertainty set $\mathcal{U}$ allows us to reduce the volatility of solution performance. In some cases, average performance also benefits when compared to existing state-of-the-research stochastic optimization approach. We suspect that these benefits come from

the following: (1) the robust approach provides additional robustness when the realized primary delays are slightly different from the estimated distribution in the training data set; and (2) the robust model provides a tractable and exact approach to deal with correlated primary delays, and incorporating delay correlation improves the robustness of resultant routings.

4.3. Case Studies Based on Simulated Delay Data

To provide more insight into the relative performance of RAR over DFW, we conduct an additional round of computational experiments where the flight primary delay data are generated from simulated probability distributions instead of historical data. We seek to quantify the robustness of both methods with respect to changes in probability distributions. These case studies are motivated by the idea that primary flight leg delay rarely corresponds to a specific distribution, but rather a composite of several types of delays, each with differing individual distributions that may vary throughout different times of the day, month, and year (Tu et al. 2008). Because of this distribution ambiguity and variation over time, it is an important feature of robust operations planning methods to be able to protect against ambiguity of future delays when trained on historical data.

Previous literature suggests several classes of candidate probability distributions that are commonly used to model flight leg delay (Mueller and Chatterji 2002, Tu et al. 2008, Bai 2006). Out of these, we pick three representative ones: normal (truncated), gamma, and log-normal. The computational experiment is set up as follows. For the training data set, we use historical flight delay data in July 2007 (the training set we use in Section 4.2) to compute the observed mean and variance of each flight leg and the Spearman's rank correlation coefficient matrix of all flight legs. We pick Spearman's rho instead of the widely-used Pearson's rho because Spearman's rho preserves order when nonlinear transformations are applied to random variables. This is convenient for generating random variables with different distributions, but the same correlation structure (MathWorks 2014). We then generate the set of training data with 1,000 samples for each flight leg such that the marginal distributions of flight legs follow a truncated normal distribution with the same mean and variance calculated from the July 2007 data. We also fix the Spearman's rank correlation coefficient matrix of the simulated data to be the same as the one calculated from July 2007 data. For testing data sets, we create three different groups of testing sets where the marginal distributions of flight leg delay are distributed as truncated normal, gamma, and log-normal respectively. For each group, we create testing sets in the following way:

- *Deviation in mean*

We keep the standard deviation and Spearman's rank correlation coefficient matrix of the testing set the same as the training set. We then generate testing data by setting the mean to be 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2 times the mean of the training set.

- *Deviation in standard deviation*

We keep the mean and Spearman's rank correlation coefficient matrix of the testing set the same as the training set. We then generate testing data by setting the standard deviation of the testing set to be 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2 times the standard deviation of the training set.

- *Deviation in correlation structure*

We keep the mean and standard deviation of the testing set the same as the training set. We then generate testing data with different Spearman's rank correlation coefficient matrix. We define the multiplier of correlation $\alpha \in (-\infty, 1]$ as a parameter to control how the correlation structure deviates from the training set. $\alpha = 0$ means there is no change in correlation structure. As α increases to 1, the primary delays become more independent, with full independence at $\alpha = 1$. As α decreases to $-\infty$, the primary delay data becomes more correlated, with perfect correlation at $\alpha = -\infty$ (positive or negative correlations are preserved from the training data). We pick $\alpha = \ln(3/4), \ln(5/6), \ln(11/12), 0, \ln(1+(e-1)/12), \ln(1+(e-1)/6), \ln(1+(e-1)/4)$ to generate testing data with various correlation structures. The detailed meaning of α and the correlation perturbing procedures are provided in Appendix B.

In total, we have 19 testing sets for each distribution and each flight network. We apply the above generation procedure to both networks N_1 and N_2 that we used in Section 4.2. We use the training set to create routings using both RAR and DFW and then test the performance of the generated routings under the testing sets created using the procedure above. Note that in DFW, there are 1000 random scenarios and each realizing with probability $\frac{1}{1000}$. Similar to Section 4.2, for both networks, we select uncertainty budget Γ which has the best performance in reducing average total propagated delay in the training set. For network N_1, $\Gamma = 2.0$; for network N_2, $\Gamma = 1.0$. Since we do not have a baseline routing anymore, we calculate the relative performance ratio of RAR over DFW as $100 \cdot (\text{DFW} - \text{RAR})/\text{DFW}$ for each performance criteria. The results are summarized in Tables 3, 4 and 5, where each table corresponds to a specific testing data delay distribution. Complete results with all values of $\Gamma \in \{0.2, 0.4, \cdots, 2.8, 3.0\}$ are presented in Figures 5 - 22 in Appendix A.

Overall, in all 114 testing sets, RAR outperforms DFW in reducing average value in 84% (96/114) of the cases, reducing standard deviation in 99% (113/114) of the cases, and reducing extreme value in 97% (111/114) of the cases.

When testing data is distributed in the same way as the training data (truncated normal, Table 3), for both networks, RAR consistently outperforms DFW in reducing standard deviation and extreme value, especially when testing data has a larger mean or standard deviation than the training set. The only exception is the case when standard deviation multiplier is 0.5 in network N_1. On the other hand, in many cases, DFW outperforms RAR in reducing average propagated delay,

especially when the testing data has a smaller mean or variance compared to the training data and when the tesing data has a different correlation structure. Overall, similar to what we observe in the computational results of network N_2 in Section 4.2, the improvement in reducing volatility and extreme value comes at the cost of a diminished performance with regard to decreasing the average value. However, this cost is small in comparison to the reductions in volatility and extreme value that we observe.

When the testing set is distributed as gamma or log-normal (i.e., different from the training set, Tables 4 and 5), in most of the cases when only mean and standard deviation are varying, RAR consistently results in a route which has substantially smaller standard deviation, extreme value and even average value. Exception occurs only when the standard deviation of the testing set is 50% that of the training set in network N_1. But again, the increase in average total propagated delay is relatively small. When correlation structure deviates in the testing data, we see that the impact of such deviation on RAR's performance is much larger than the impact of deviation in mean or standard deviation. Under almost all correlation multipliers, all three performance criteria (especially for the extreme value reduction criterion) are inferior to the cases when correlation structure is invariable. Deviation in correlation changes the shape of the uncertainty set, which might cause this substantial downgrade in RAR's performance. Though RAR is less effective in these correlation-varying cases, it still outperforms DFW in at least two criteria. Moreover, the small inferiority in one criterion, if exists, is usually offset by the superiority of the other two criteria.

In a word, compared to the existing stochastic optimization approach, the RAR model provides additional benefit when realized flight leg delay distributionally differs from historical data.

5. Concluding Remarks and Future Research Directions

In this paper, we propose a robust optimization-based formulation for the aircraft routing problem to minimize the maximal possible total propagated delay, assuming flight primary leg delays live in a pre-specified uncertainty set. We provide data-driven methods to construct a flight leg delay uncertainty set that not only includes the uncertainty of individual flights, but also the correlations among different flights. We then reformulate the robust problem as an integer program. We propose an exact decomposition solution approach under a column-and-row generation framework. Importantly, this solution method can be applied to general robust optimization problems where the nominal problem is solved through branch-and-price.

Using real-world instances with actual schedule and delay data, along with simulated delay data, we demonstrate the effectiveness and efficiency of our method. We show that by incorporating delay correlation, our robust model outperforms the state-of-the-research stochastic optimization

Table 3 Relative Performance Ratio (RPR) under Different Mean / Standard Deviation / Correlation Multiplier of Testing Data Following Truncated Normal Distribution

Multiplier of mean	Multiplier of std	Multiplier of correlation	RPR of mean reduction (%)	RPR of std reduction (%)	RPR of extreme value reduction (%)
Flight Network N_1, 106 flights and 24 aircraft, best $\Gamma = 2.0$					
0.50	1.00	0.00	-1.0	6.5	10.0
0.75	1.00	0.00	-0.4	6.9	9.0
1.00	1.00	0.00	0.2	7.3	8.9
1.25	1.00	0.00	0.8	7.7	8.9
1.50	1.00	0.00	1.4	8.1	9.5
1.75	1.00	0.00	1.9	8.4	9.5
2.00	1.00	0.00	2.5	8.7	9.4
1.00	0.50	0.00	-12.4	-0.7	6.7
1.00	0.75	0.00	-4.1	5.5	11.8
1.00	1.25	0.00	2.1	7.4	8.8
1.00	1.50	0.00	3.1	7.2	6.9
1.00	1.75	0.00	3.6	7.0	6.0
1.00	2.00	0.00	3.9	6.9	5.3
1.00	1.00	$\ln(3/4)$	0.2	8.4	10.3
1.00	1.00	$\ln(5/6)$	0.6	8.0	10.2
1.00	1.00	$\ln(11/12)$	-0.1	6.2	4.4
1.00	1.00	$\ln(1+(e-1)/12)$	-0.8	6.4	11.0
1.00	1.00	$\ln(1+(e-1)/6)$	-0.6	5.9	1.5
1.00	1.00	$\ln(1+(e-1)/4)$	-0.7	6.8	9.5
Flight Network N_2, 117 flights and 23 aircraft, best $\Gamma = 1.0$					
0.50	1.00	0.00	0.3	1.5	2.1
0.75	1.00	0.00	0.3	1.6	2.2
1.00	1.00	0.00	0.4	1.7	1.9
1.25	1.00	0.00	0.4	1.9	1.9
1.50	1.00	0.00	0.5	2.0	1.7
1.75	1.00	0.00	0.5	2.1	1.8
2.00	1.00	0.00	0.6	2.3	1.7
1.00	0.50	0.00	-0.1	1.1	2.9
1.00	0.75	0.00	0.2	1.5	2.1
1.00	1.25	0.00	0.5	1.9	1.6
1.00	1.50	0.00	0.6	2.0	0.9
1.00	1.75	0.00	0.7	2.2	0.4
1.00	2.00	0.00	0.9	2.4	0.6
1.00	1.00	$\ln(3/4)$	-0.3	1.1	0.1
1.00	1.00	$\ln(5/6)$	-0.4	0.7	2.9
1.00	1.00	$\ln(11/12)$	-0.3	1.1	0.6
1.00	1.00	$\ln(1+(e-1)/12)$	-0.1	0.7	0.9
1.00	1.00	$\ln(1+(e-1)/6)$	0.1	1.7	1.1
1.00	1.00	$\ln(1+(e-1)/4)$	-0.4	0.5	2.4

approach in reducing all three performance criteria considered: average value, standard deviation and maximum value of total propagated delay in most of the cases. In the cases when deficit in one criterion exists, such inferiority is usually offset by gains in the other two criteria.

We attribute these benefits to characteristics of our solution method: (1) the robust approach provides a tractable and exact method for dealing with correlated uncertainty, thus providing additional robustness against propagated delay; and (2) the robust approach is less vulnerable

Table 4 Relative Performance Ratio (RPR) under Different Mean / Standard Deviation / Correlation Multiplier of Testing Data Following Gamma Distribution

Multiplier of mean	Multiplier of std	Multiplier of correlation	RPR of mean reduction (%)	RPR of std reduction (%)	RPR of extreme value reduction (%)
Flight Network N_1, 106 flights and 24 aircraft, best $\Gamma = 2.0$					
0.50	1.00	0.00	4.6	9.5	17.3
0.75	1.00	0.00	3.6	8.3	14.5
1.00	1.00	0.00	3.0	7.7	12.4
1.25	1.00	0.00	2.6	7.5	11.2
1.50	1.00	0.00	2.5	7.5	10.3
1.75	1.00	0.00	2.5	7.5	9.5
2.00	1.00	0.00	2.6	7.6	8.6
1.00	0.50	0.00	-5.4	6.5	11.1
1.00	0.75	0.00	0.8	7.7	12.2
1.00	1.25	0.00	3.9	7.8	13.3
1.00	1.50	0.00	4.6	8.1	14.3
1.00	1.75	0.00	5.0	8.5	15.4
1.00	2.00	0.00	5.4	8.8	16.4
1.00	1.00	$\ln(3/4)$	2.9	7.9	10.6
1.00	1.00	$\ln(5/6)$	3.2	7.6	11.2
1.00	1.00	$\ln(11/12)$	1.0	3.6	5.8
1.00	1.00	$\ln(1+(e-1)/12)$	1.8	6.9	7.0
1.00	1.00	$\ln(1+(e-1)/6)$	2.1	5.2	-2.5
1.00	1.00	$\ln(1+(e-1)/4)$	2.9	8.2	8.6
Flight Network N_2, 117 flights and 23 aircraft, best $\Gamma = 1.0$					
0.50	1.00	0.00	3.7	9.4	27.9
0.75	1.00	0.00	2.6	7.1	26.9
1.00	1.00	0.00	2.0	5.7	18.9
1.25	1.00	0.00	1.7	4.9	16.8
1.50	1.00	0.00	1.4	4.4	15.1
1.75	1.00	0.00	1.3	4.1	13.9
2.00	1.00	0.00	1.2	3.9	13.0
1.00	0.50	0.00	0.0	2.1	5.1
1.00	0.75	0.00	1.1	4.2	16.5
1.00	1.25	0.00	2.7	6.9	23.4
1.00	1.50	0.00	3.3	7.8	26.9
1.00	1.75	0.00	3.8	8.6	27.3
1.00	2.00	0.00	4.2	9.3	27.5
1.00	1.00	$\ln(3/4)$	0.8	3.4	2.3
1.00	1.00	$\ln(5/6)$	0.2	2.9	7.8
1.00	1.00	$\ln(11/12)$	1.3	4.2	3.8
1.00	1.00	$\ln(1+(e-1)/12)$	0.8	2.7	0.9
1.00	1.00	$\ln(1+(e-1)/6)$	0.9	4.0	1.7
1.00	1.00	$\ln(1+(e-1)/4)$	0.6	2.3	4.8

to the variations in the input data when compared to stochastic optimization approaches (i.e., the robust approach provides additional robustness when the realized flight delays distributionally differ from historical data).

The idea of applying robust optimization to robust airline planning is fairly new. There are many possible future research directions in this area. For example, a direct extension of this work could be to apply robust optimization to the integrated aircraft routing and crew pairing problem.

Table 5 Relative Performance Ratio (RPR) under Different Mean / Standard Deviation / Correlation Multiplier of Testing Data Following Log-normal Distribution

Multiplier of mean	Multiplier of std	Multiplier of correlation	RPR of mean reduction (%)	RPR of std reduction (%)	RPR of extreme value reduction (%)
Flight Network N_1, 106 flights and 24 aircraft, best $\Gamma = 2.0$					
0.50	1.00	0.00	4.8	12.5	23.1
0.75	1.00	0.00	3.5	11.3	21.9
1.00	1.00	0.00	2.6	10.5	20.5
1.25	1.00	0.00	2.1	9.9	19.2
1.50	1.00	0.00	1.8	9.3	17.7
1.75	1.00	0.00	1.7	8.9	16.4
2.00	1.00	0.00	1.8	8.7	15.3
1.00	0.50	0.00	-4.8	10.0	18.2
1.00	0.75	0.00	0.4	10.3	19.7
1.00	1.25	0.00	3.7	10.6	21.1
1.00	1.50	0.00	4.4	10.7	21.5
1.00	1.75	0.00	4.9	10.9	21.9
1.00	2.00	0.00	5.3	11.1	22.2
1.00	1.00	$\ln(3/4)$	0.9	6.9	8.7
1.00	1.00	$\ln(5/6)$	2.2	7.9	9.3
1.00	1.00	$\ln(11/12)$	-1.1	1.1	3.1
1.00	1.00	$\ln(1+(e-1)/12)$	0.3	6.5	4.8
1.00	1.00	$\ln(1+(e-1)/6)$	0.0	3.2	-3.2
1.00	1.00	$\ln(1+(e-1)/4)$	1.7	7.0	2.9
Flight Network N_2, 117 flights and 23 aircraft, best $\Gamma = 1.0$					
0.50	1.00	0.00	4.1	13.7	28.9
0.75	1.00	0.00	3.2	11.7	28.5
1.00	1.00	0.00	2.5	9.9	28.0
1.25	1.00	0.00	2.1	8.6	27.4
1.50	1.00	0.00	1.8	7.6	26.8
1.75	1.00	0.00	1.6	6.9	26.1
2.00	1.00	0.00	1.5	6.4	25.4
1.00	0.50	0.00	0.7	5.9	25.3
1.00	0.75	0.00	1.7	8.5	27.3
1.00	1.25	0.00	3.0	10.9	28.3
1.00	1.50	0.00	3.5	11.6	28.4
1.00	1.75	0.00	3.9	12.2	28.5
1.00	2.00	0.00	4.2	12.7	28.6
1.00	1.00	$\ln(3/4)$	0.5	3.2	1.6
1.00	1.00	$\ln(5/6)$	-0.3	2.4	6.6
1.00	1.00	$\ln(11/12)$	1.2	3.9	3.2
1.00	1.00	$\ln(1+(e-1)/12)$	0.7	2.3	1.1
1.00	1.00	$\ln(1+(e-1)/6)$	0.7	3.6	-0.3
1.00	1.00	$\ln(1+(e-1)/4)$	0.4	2.4	5.4

It may also be interesting to investigate tailored approaches to modeling flight delay uncertainty sets, so as to achieve greater tractability as well as probabilistic performance guarantees. We believe that robust optimization, with its merits of tractability and attractive solution quality, represents a promising direction for dealing with uncertainty in airline scheduling problems.

Acknowledgments

The authors want to thank Professors Cynthia Barnhart and Dimitris Bertsimas at MIT, Professor Vikrant Vaze at Dartmouth for their valuable comments and suggestions. The authors also thank the referees for several suggestions that improved the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship [Grant No. 1122374] (J. Kung). Any opinion, findings, and conclusions or recommendations expressed in this material are ours and do not necessarily reflect the views of the National Science Foundation.

Appendix A: Detailed Computational Results in Section 4.3

Figures 5 - 13 depict the performance in flight network N_1, and Figures 14 - 22 depict the performance in flight network N_2. Specifically, Figures 5 - 7, 14 - 16 present the relative performance ratio $100 \cdot (\text{DFW} - \text{RAR})/\text{DFW}$ when mean of the testing data under three different distributions deviate from the training data for two flight networks. Figures 8 - 10, 17 - 19 show the relative performance ratio $100 \cdot (\text{DFW} - \text{RAR})/\text{DFW}$ when standard deviation of the testing data under three different distributions deviate from the training data for two flight networks. Figures 11 - 13, 20 - 22 show the relative performance ratio $100 \cdot (\text{DFW} - \text{RAR})/\text{DFW}$ when correlation structure of the testing data under three different distributions deviate from the training data for two flight networks. Performance is evaluated in three criteria: (1) average total propagated delay, (2) standard deviation of total propagated delay, (3) maximum total propagated delay. The long dashed line indicates 0% relative performance ratio, which marks the place indicating that DFW and RAR have the same performance.

Figure 5 Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_1)

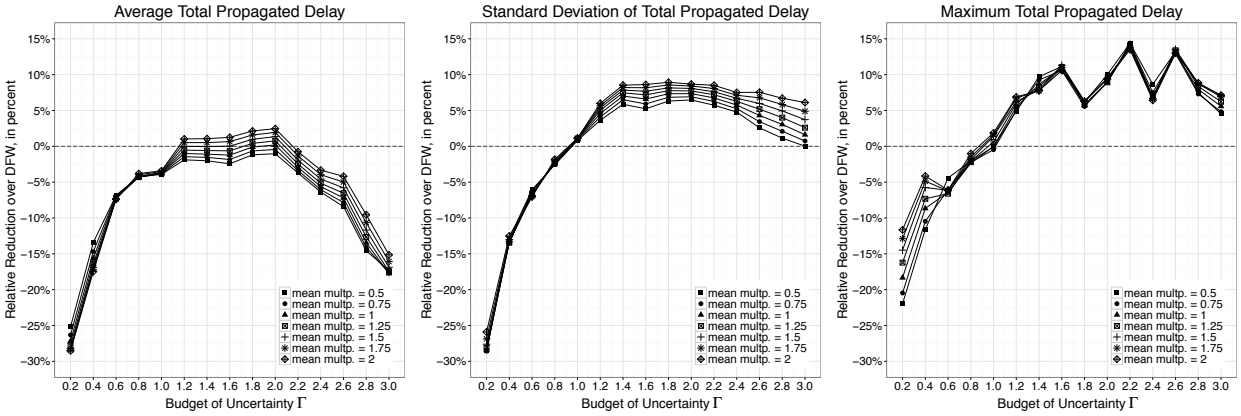


Figure 6 **Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_1)**

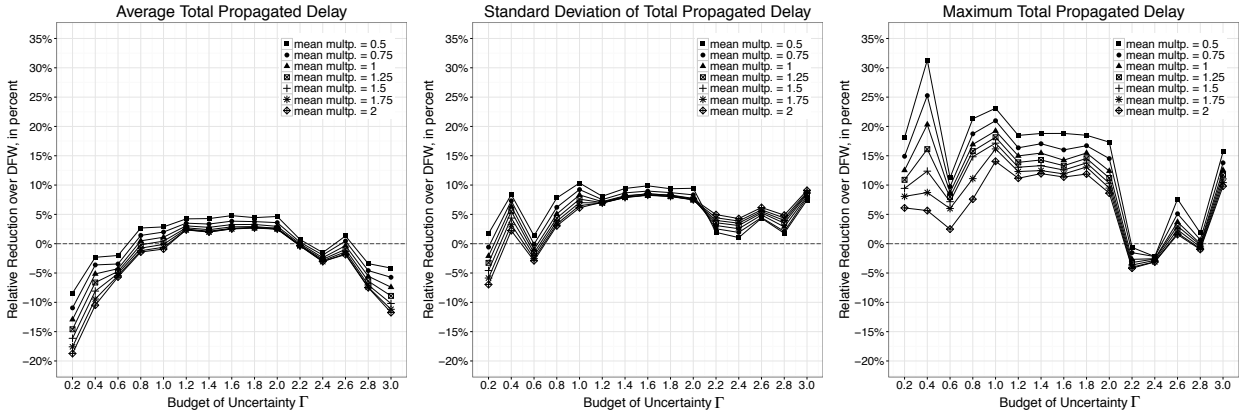


Figure 7 **Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_1)**

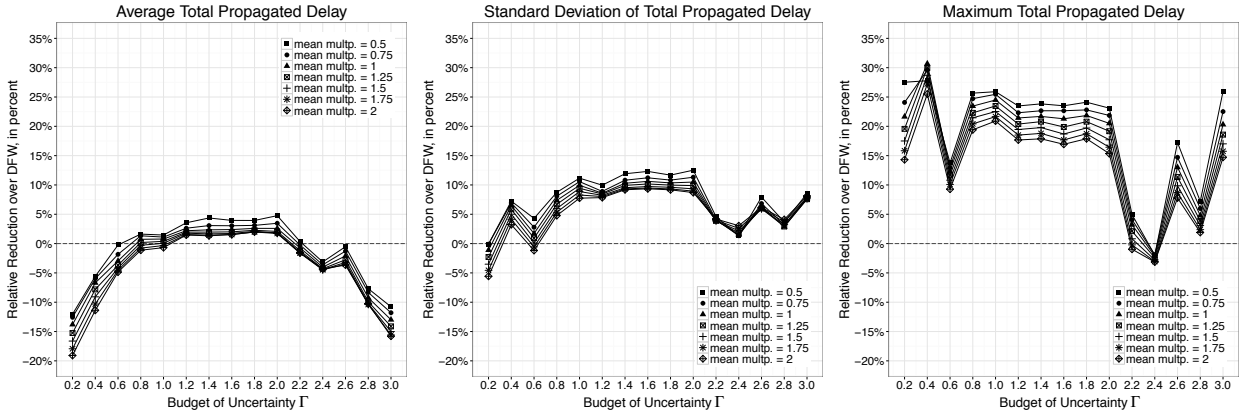


Figure 8 **Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_1)**

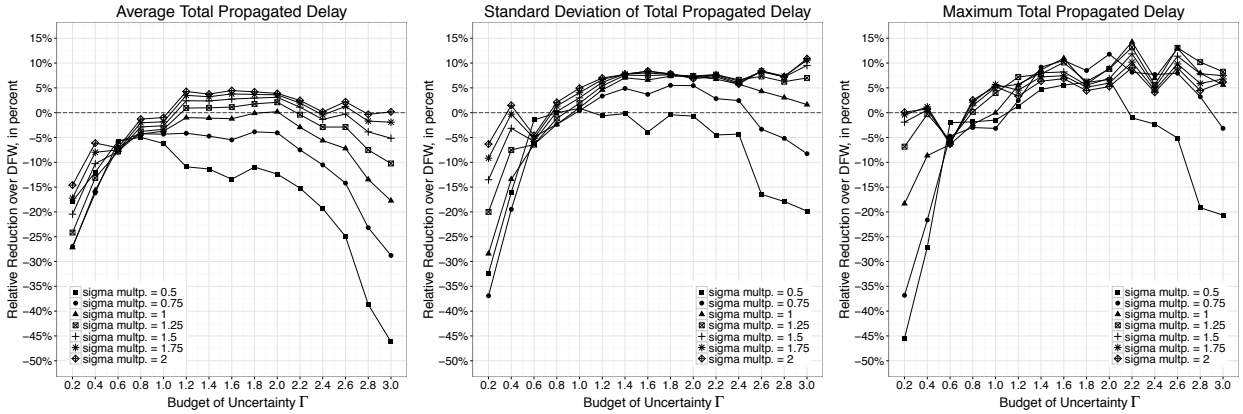


Figure 9 Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_1)

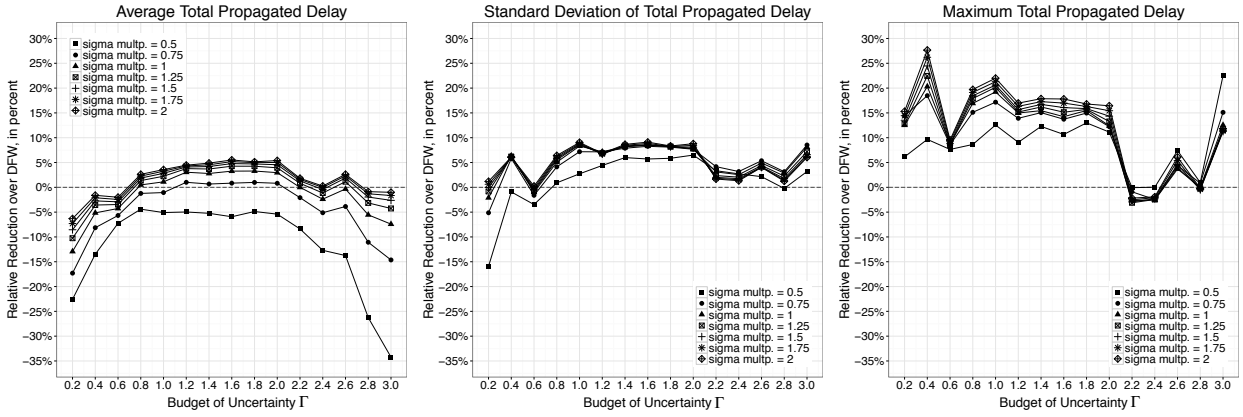


Figure 10 Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_1)

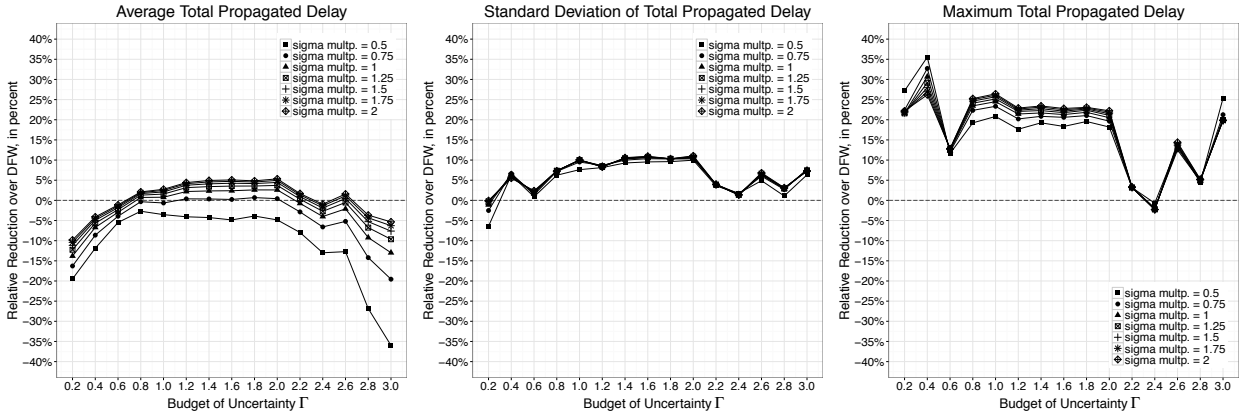


Figure 11 Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_1)

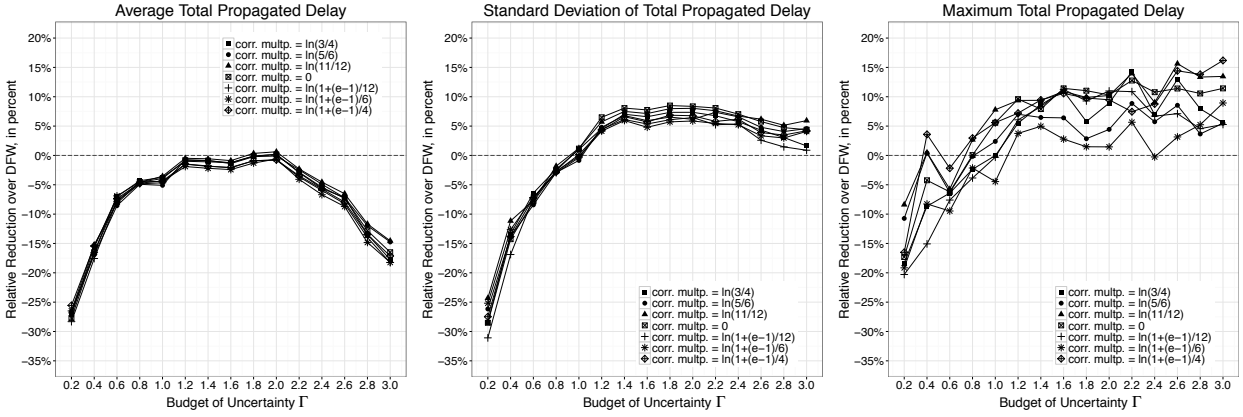


Figure 12 **Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_1)**

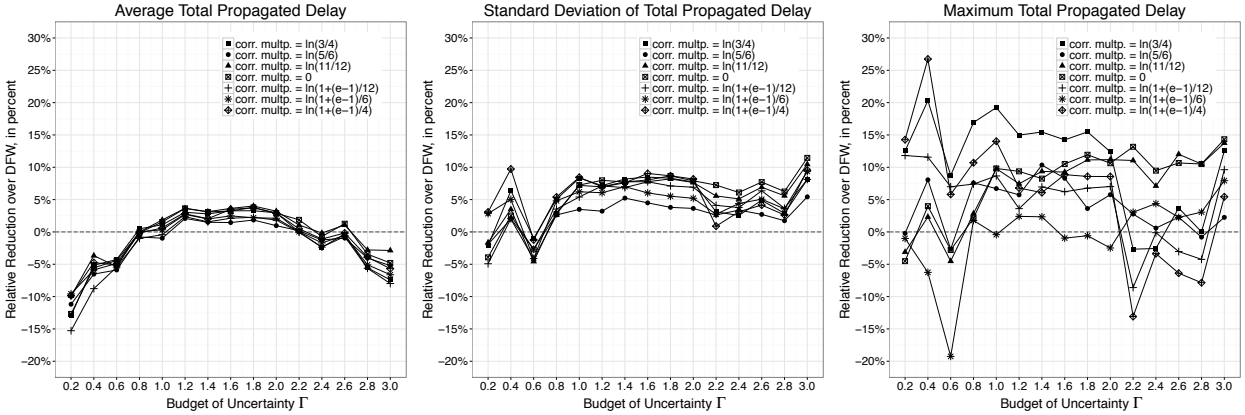


Figure 13 **Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_1)**

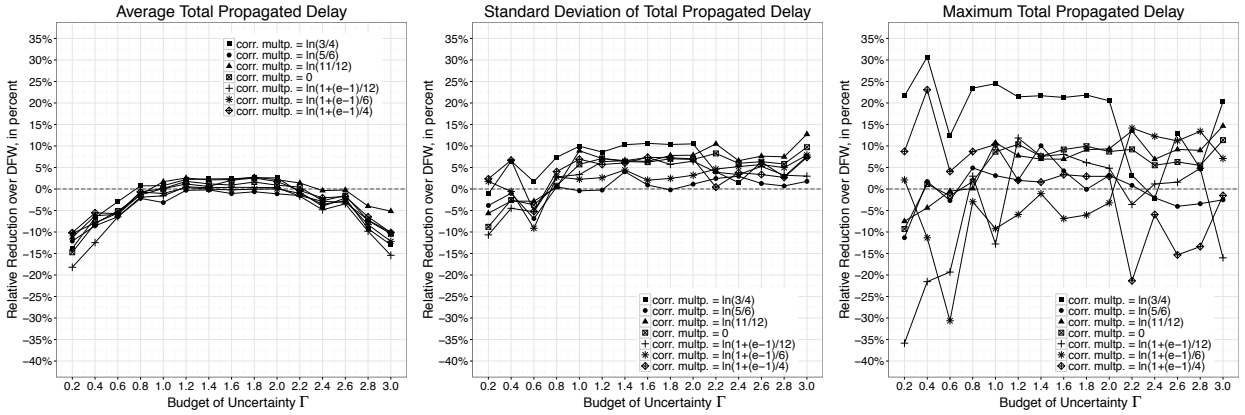


Figure 14 **Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_2)**

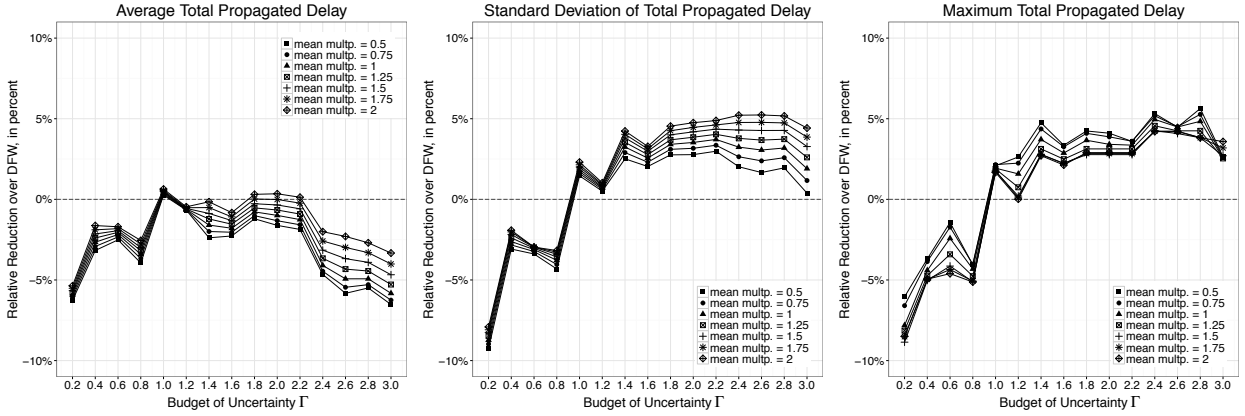


Figure 15 Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_2)

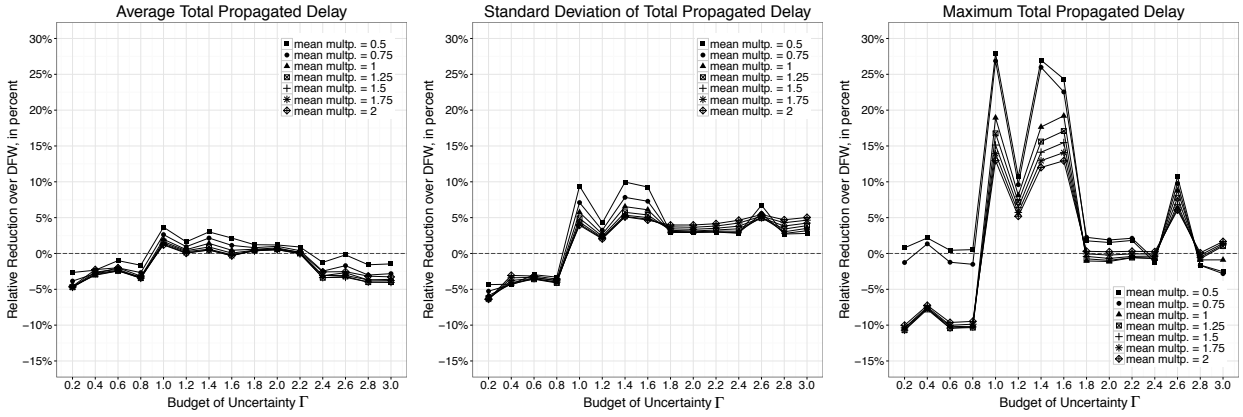


Figure 16 Impact of Deviation in Mean (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_2)

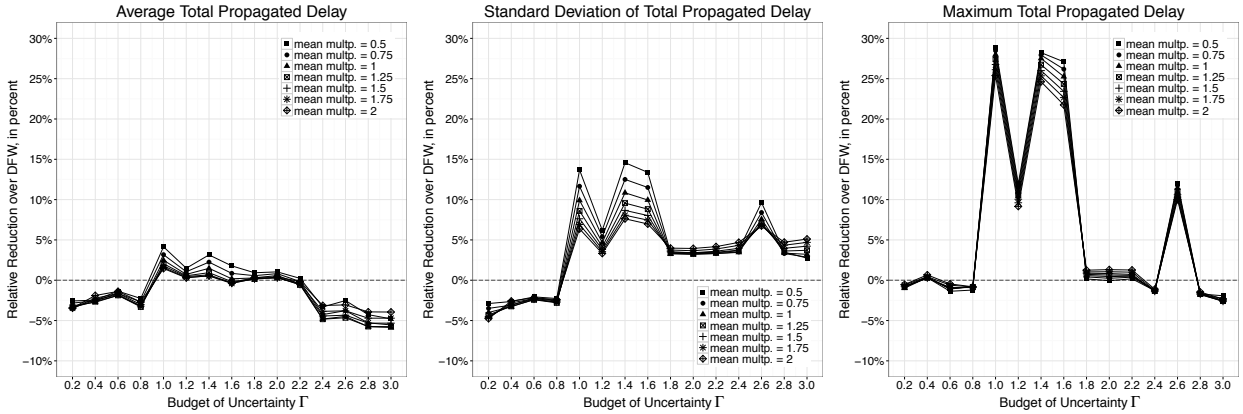


Figure 17 Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_2)

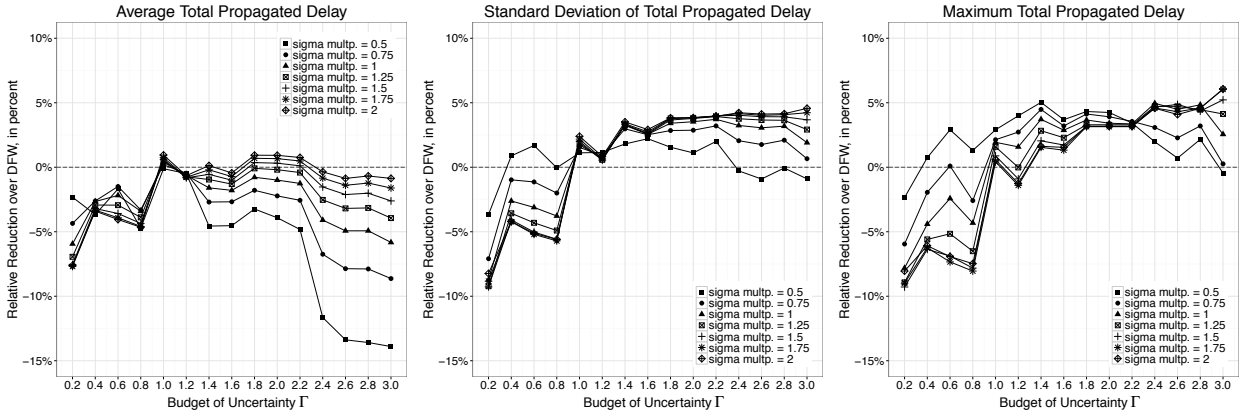


Figure 18 **Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_2)**

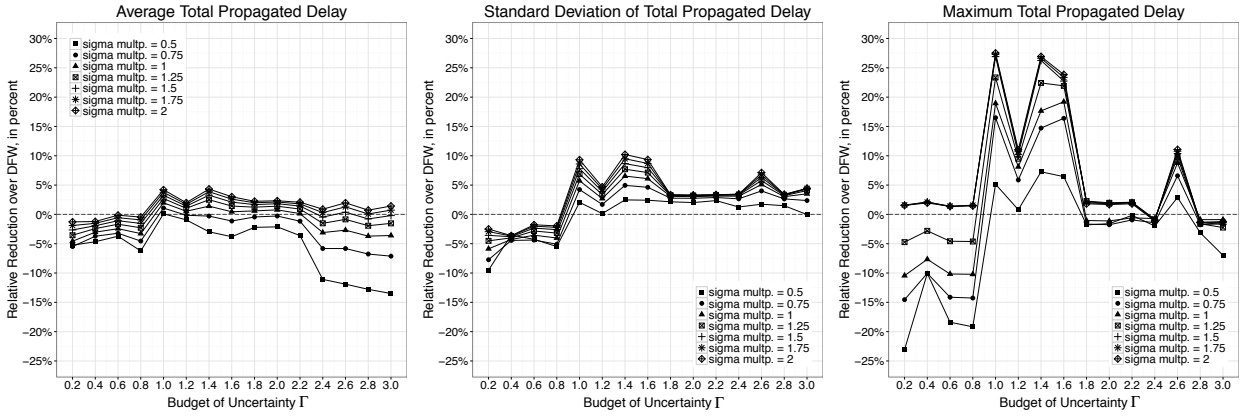


Figure 19 **Impact of Deviation in Standard Deviation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_2)**

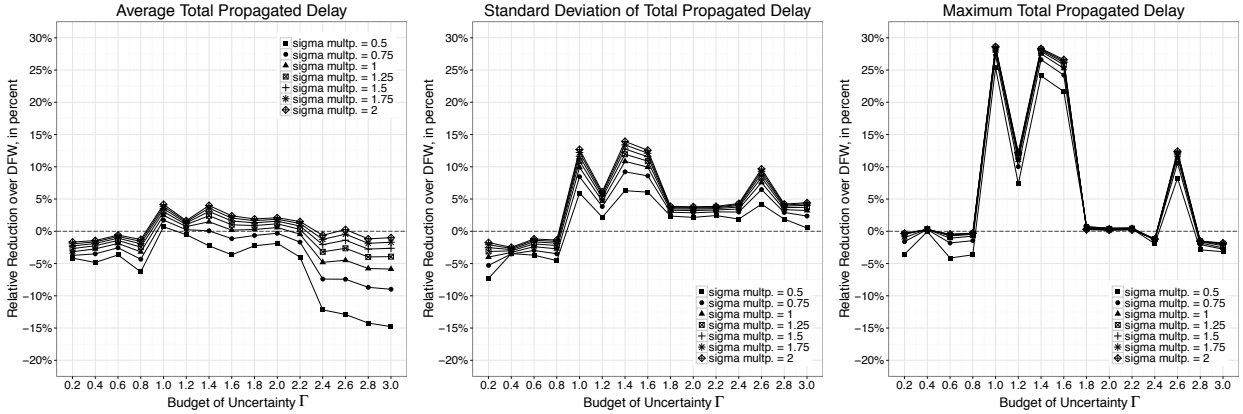


Figure 20 **Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Truncated Normal Distribution / Testing Flight Network: N_2)**

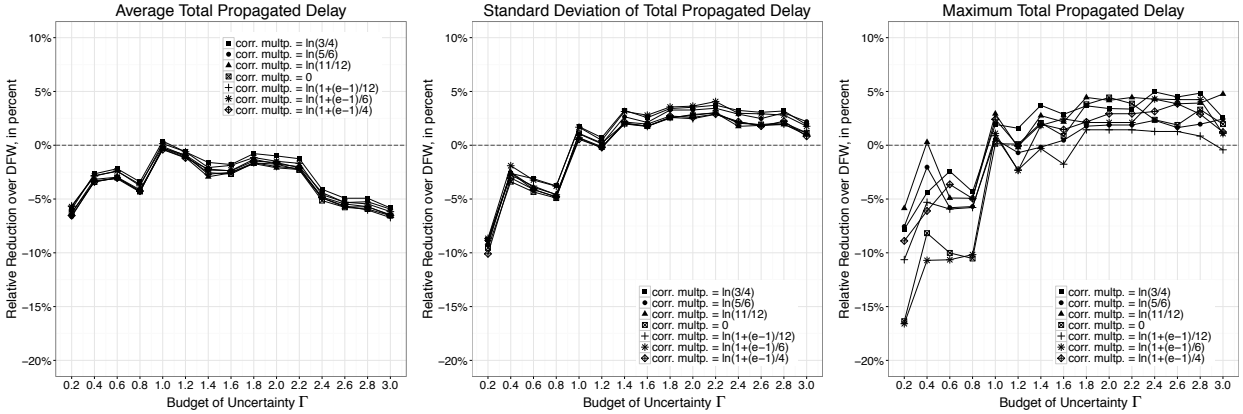


Figure 21 Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Gamma Distribution / Testing Flight Network: N_2)

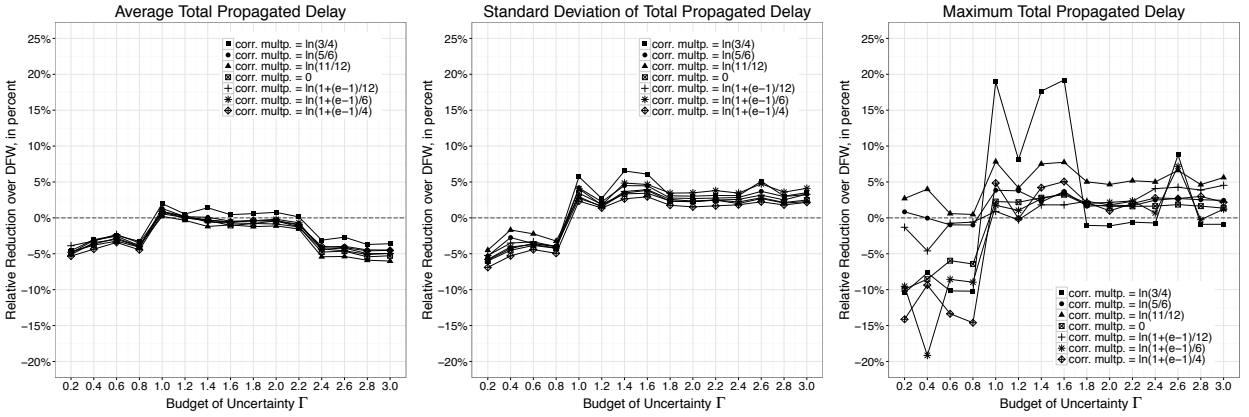
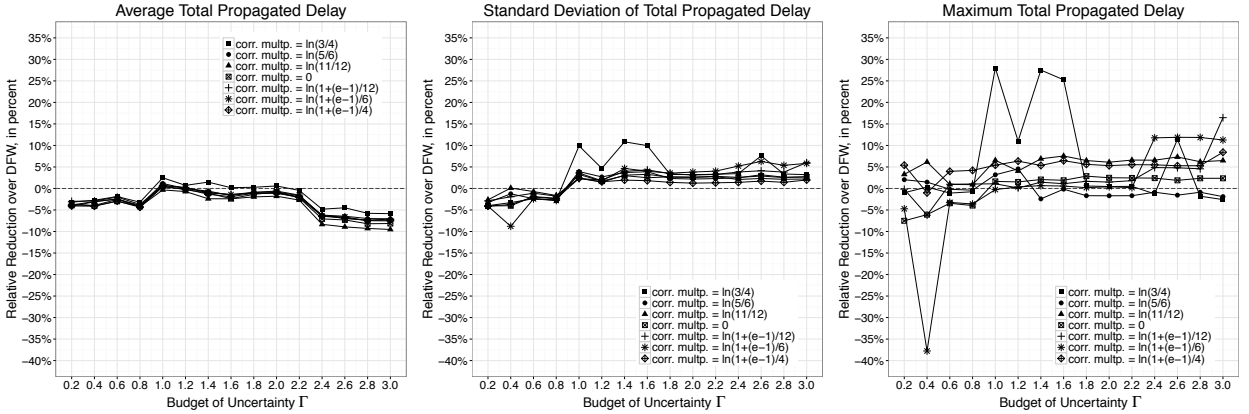


Figure 22 Impact of Deviation in Correlation (Training Delay Data: Truncated Normal Distribution / Testing Delay Data: Log-normal Distribution / Testing Flight Network: N_2)



Appendix B: Method to Perturb Spearman's Rank Correlation Coefficient Matrix

The method is inspired by Galeeva et al. (2007) where they provide a way to generate correlation matrices around a base correlation matrix by perturbing its eigenvalues. The detailed procedure we use is as follows,

1. Apply eigenvalue decomposition on the Spearman's rank correlation coefficient matrix ρ^{train} for the training data set,

$$\rho^{\text{train}}_{i,j} = \sum_{k,l=1}^{|\mathcal{F}|} V_{i,k}\Lambda_{k,l}V_{l,j},$$

where $\Lambda_{k,l} = \lambda_k \delta_{k,l}$. $\delta_{k,l} = 1$ if $k = l$; 0, otherwise. $\lambda_1 > \lambda_2 > \cdots > \lambda_{|\mathcal{F}|}$ are the $|\mathcal{F}|$ eigenvalues for ρ^{train}. V is the eigenvectors.

2. Create variables $\sigma_1, \sigma_2, \cdots, \sigma_{|\mathcal{F}|}$ for each eigenvalue, which satisfy the following set of equations

$$\lambda_1 e^{\sigma_1} = \lambda_2 e^{\sigma_2} = \cdots = \lambda_{|\mathcal{F}|} e^{\sigma_{|\mathcal{F}|}}$$

There are multiple solutions for the set of equations above. We fix $\sigma_1 = 1$, and the remainings can be calculated accordingly. Since λ_1 is the largest eigenvalue, it can be seen easily that $\sigma_2, \cdots, \sigma_{|\mathcal{F}|} > 1$.

3. Perturb these eigenvalues with a parameter α. The perturbed eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_{|\mathcal{F}|}$ will be calculated as $\hat{\lambda}_i = \lambda_i e^{\sigma_i \alpha},\ \forall i = 1, \cdots, |\mathcal{F}|$.

4. Normalize the perturbed eigenvalues to make $\sum_{i=1}^{|\mathcal{F}|} \hat{\lambda}_i = |\mathcal{F}|$.

5. Construct matrix ρ' with the perturbed eigenvalues and the eigenvectors for $\rho^{\text{train}}_{i,j}$,

$$\rho'_{i,j} = \sum_{k,l=1}^{|\mathcal{F}|} V_{i,k} \hat{\Lambda}_{k,l} V_{l,j},$$

where $\hat{\Lambda}_{k,l} = \hat{\lambda}_k \delta_{k,l}$.

6. Construct Spearman's rank correlation coefficient matrix by normalizing matrix ρ',

$$\rho^{\text{test}}_{i,j} = \frac{\rho'_{i,j}}{\sqrt{\rho'_{i,i}\rho'_{j,j}}}$$

The normalizing step makes sure that $-1 \le \rho^{\text{test}}_{i,j} \le 1$ and $\rho^{\text{test}}_{i,i} = 1,\ \forall i = 1, \cdots, |\mathcal{F}|$.

To understand the meaning of parameter $\alpha \in (-\infty, 1]$:

- if $\alpha = 1$, $\hat{\lambda}_1 = \hat{\lambda}_2 = \cdots = \hat{\lambda}_{|\mathcal{F}|} = 1$, thus $\hat{\Lambda} = I$ and $\rho' = V\hat{\Lambda}V^T = VV^T = VV^{-1} = I$. This leads to $\rho^{\text{test}} = I$, which means the testing data will have independent primary delays.

- if $\alpha = 0$, then $\hat{\lambda}_i = \lambda_i,\ \forall i = 1, \cdots, |\mathcal{F}|$. This means the testing data will have the same Spearman's rank correlation coefficient matrix as the training data.

- for the case $\alpha \to -\infty$, we have

$$\rho^{\text{test}}_{i,j} = \frac{\rho'_{i,j}}{\sqrt{\rho'_{i,i}\rho'_{j,j}}} = \frac{\sum_{k,l=1}^{|\mathcal{F}|} V_{i,k}\hat{\Lambda}_{k,l}V_{l,j}}{\sqrt{\sum_{k,l=1}^{|\mathcal{F}|} V_{i,k}\hat{\Lambda}_{k,l}V_{l,i}}\sqrt{\sum_{k,l=1}^{|\mathcal{F}|} V_{j,k}\hat{\Lambda}_{k,l}V_{l,j}}} = \frac{\sum_{k=1}^{|\mathcal{F}|} V_{i,k}\hat{\lambda}_k e^{\sigma_k\alpha}V_{k,j}}{\sqrt{\sum_{k=1}^{|\mathcal{F}|} V_{i,k}\hat{\lambda}_k e^{\sigma_k\alpha}V_{k,i}}\sqrt{\sum_{k=1}^{|\mathcal{F}|} V_{j,k}\hat{\lambda}_k e^{\sigma_k\alpha}V_{k,j}}}$$

Divide both the numerator and denominator by e^{α}, we have

$$\rho^{\text{test}}_{i,j} = \frac{V_{i,1}\hat{\lambda}_1 V_{1,j} + \sum_{k=2}^{|\mathcal{F}|} V_{i,k}\hat{\lambda}_k e^{(\sigma_k-1)\alpha}V_{k,j}}{\sqrt{V_{i,1}\hat{\lambda}_1 V_{1,i} + \sum_{k=2}^{|\mathcal{F}|} V_{i,k}\hat{\lambda}_k e^{(\sigma_k-1)\alpha}V_{k,i}}\sqrt{V_{j,1}\hat{\lambda}_1 V_{1,j} + \sum_{k=2}^{|\mathcal{F}|} V_{j,k}\hat{\lambda}_k e^{(\sigma_k-1)\alpha}V_{k,j}}}$$

Since $\sigma_2, \cdots, \sigma_{|\mathcal{F}|} > \sigma_1 = 1$, as $\alpha \to -\infty$ we have

$$\lim_{\alpha\to-\infty} \rho^{\text{test}}_{i,j} = \frac{V_{i,1}\hat{\lambda}_1 V_{1,j}}{|V_{i,1}|\hat{\lambda}_1|V_{1,j}|} = \frac{V_{i,1}V_{1,j}}{|V_{i,1}V_{1,j}|} = \begin{cases} 1, & V_{i,1}V_{1,j} > 0 \\ -1, & V_{i,1}V_{1,j} < 0 \end{cases}$$

This means the testing data will be perfectly correlated. Whether it is positively or negatively correlated depends on the eigenvectors of ρ^{train}.

References

Ahmadbeygi, Shervin, Amy Cohn, Marcial Lapp. 2010. Decreasing airline delay propagation by re-allocating scheduled slack. *IIE Transactions* **42**(7) 478–489.

Bai, Yuqiong. 2006. Analysis of aircraft arrival delay and airport on-time performance. Ph.D. thesis, University of Central Florida Orlando, Florida.

Barnhart, Cynthia, Dimitris Bertsimas, Constantine Caramanis, Douglas Fearing. 2012. Equitable and efficient coordination in traffic flow management. *Transportation Science* **46**(2) 262–280.

Barnhart, Cynthia, Natashia L Boland, Lloyd W Clarke, Ellis L Johnson, George L Nemhauser, Rajesh G Shenoi. 1998a. Flight string models for aircraft fleeting and routing. *Transportation Science* **32**(3) 208–220.

Barnhart, Cynthia, Ellis L Johnson, George L Nemhauser, Martin WP Savelsbergh, Pamela H Vance. 1998b. Branch-and-price: Column generation for solving huge integer programs. *Operations Research* **46**(3) 316–329.

Ben-Tal, Aharon, Arkadi Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations Research Letters* **25**(1) 1–13.

Bertsimas, Dimitris, David B Brown, Constantine Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, Dimitris, Iain Dunning, Miles Lubin. 2014. Reformulations versus cutting planes for robust optimization. Available at Optimization Online.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2013a. Data-driven robust optimization. *arXiv preprint arXiv:1401.0212* .

Bertsimas, Dimitris, Eugene Litvinov, Xu Andy Sun, Jinye Zhao, Tongxin Zheng. 2013b. Adaptive robust optimization for the security constrained unit commitment problem. *Power Systems, IEEE Transactions on* **28**(1) 52–63.

Bertsimas, Dimitris, Dessislava Pachamanova, Melvyn Sim. 2004. Robust linear optimization under general norms. *Operations Research Letters* **32**(6) 510–516.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Bertsimas, Dimitris, Aurélie Thiele. 2006. A robust optimization approach to inventory theory. *Operations Research* **54**(1) 150–168.

Borndörfer, Ralf, Ivan Dovica, Ivo Nowak, Thomas Schickinger. 2010. Robust tail assignment. *Proceedings of the fiftieth annual symposium of AGIFORS*.

Bureau of Transportation Statistics. 2013. Research and innovative technology. URL `http://www.transtats.bts.gov/ot_delay/OT_DelayCause1.asp?pn=1`.

Chiraphadhanakul, Virot, Cynthia Barnhart. 2013. Robust flight schedules through slack re-allocation. *EURO Journal on Transportation and Logistics* **2**(4) 277–306.

Dunbar, Michelle, Gary Froyland, Cheng-Lung Wu. 2012. Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. *Transportation Science* **46**(2) 204–216.

Dunbar, Michelle, Gary Froyland, Cheng-Lung Wu. 2014. An integrated scenario-based approach for robust aircraft routing, crew pairing and re-timing. *Computers & Operations Research* **45** 68–86.

Friedman, Jerome, Trevor Hastie, Robert Tibshirani. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3) 432–441.

Froyland, Gary, Stephen J Maher, Cheng-Lung Wu. 2013. The recoverable robust tail assignment problem. *Transportation Science* .

Galeeva, Roza, Jiri Hoogland, Alexander Eydeland, Morgan Stanley. 2007. Measuring correlation risk. Tech. rep.

Hsieh, Cho-Jui, Inderjit S Dhillon, Pradeep K Ravikumar, Mátyás A Sustik. 2011. Sparse inverse covariance matrix estimation using quadratic approximation. *Advances in Neural Information Processing Systems*. 2330–2338.

Klabjan, Diego, Andrew J Schaefer, Ellis L Johnson, Anton J Kleywegt, George L Nemhauser. 2001. Robust airline crew scheduling. *Proceedings of TRISTAN IV*. Azores, Portugal.

Lan, Shan, John-Paul Clarke, Cynthia Barnhart. 2006. Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. *Transportation Science* **40**(1) 15–28.

Marla, Lavanya, Cynthia Barnhart. 2010. Robust optimization: Lessons learned from aircraft routing. Working Paper, Massachusetts Institute of Technology, Cambridge, MA.

MathWorks. 2014. Simulating dependent random variables using copulas. `http://www.mathworks.com/help/stats/examples/simulating-dependent-random-variables-using-copulas.html#zmw57dd0e179`.

Mueller, Eric R, Gano B Chatterji. 2002. Analysis of aircraft arrival and departure delay characteristics. *AIAA's Aircraft Technology, Integration and Operations (ATIO) Technical Forum*. Los Angeles, CA.

Rosenberger, Jay M, Ellis L Johnson, George L Nemhauser. 2004. A robust fleet-assignment model with hub isolation and short cycles. *Transportation Science* **38**(3) 357–368.

Shebalov, Sergey, Diego Klabjan. 2006. Robust airline crew pairing: Move-up crews. *Transportation Science* **40**(3) 300–312.

Smith, Barry C, Ellis L Johnson. 2006. Robust airline fleet assignment: Imposing station purity using station decomposition. *Transportation Science* **40**(4) 497–516.

Tu, Yufeng, Michael O Ball, Wolfgang S Jank. 2008. Estimating flight departure delay distributionsa statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association* **103**(481) 112–125.

Yen, Joyce W, John R Birge. 2006. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science* **40**(1) 3–14.