Data-Driven Planning of Reliable Itineraries in Multi-Modal Transit Networks

Michael Redmond · Ann Melissa Campbell · Jan Fabian Ehmke

Abstract Multi-modal travel itineraries are based on traversing multiple legs using more than one mode of transportation. The more combinations of legs and modes, the more challenging it is for a traveler to identify a reliable itinerary. Transportation providers collect data that can increase transparency for reliable travel planning. However, this data has not been fully exploited yet, although it will likely form an important piece of future traveler information systems. Our paper takes an important step in this direction by analyzing and aggregating data from the operation of scheduled and unscheduled modes to create a reliability measure for multi-modal travel. We use a network search algorithm to evaluate itineraries that combine schedule-based long-distance travel with airlines with last-mile and first-mile drive times to efficiently identify the one with the highest reliability given a start time and travel time budget. Our network search considers multiple origin and destination airports which impacts the first and last mile as well as the flight options. We use extensive historical datasets to create reliable itineraries and compare these with deterministic shortest travel time itineraries. We investigate the amount of data that is required to create reliable multi-modal travel itineraries. Additionally, we highlight the benefits and costs of reliable travel itineraries and analyze their structure.

M. Redmond
University of Iowa
108 John Pappajohn Business Building
Iowa City, IA 52242 E-mail: michael-a-redmond@uiowa.edu

A. M. Campbell
University of Iowa
108 John Pappajohn Business Building
Iowa City, IA 52242 E-mail: ann-campbell@uiowa.edu

J. F. Ehmke
Universitatspl. 2
39106 Magdeburg, Germany E-mail: jan.ehmke@ovgu.de

1 Introduction

Multi-modal travel itineraries are based on traversing multiple legs using more
than one mode of transportation. The more combinations of legs and modes,
the more challenging it is for a traveler to choose an itinerary. Tools for multi-
modal travel planning have gained popularity in recent years due to travelers'
demand for integrated, door-to-door travel options. This demand is reflected
in travel websites like Rome2Rio and Coord, which offer routing interfaces
specifically designed for travelers to plan multi-modal itineraries (20, 22). Mu-
nicipalities are also seeing multi-modal travel planning as a way to combat
city congestion and improve access to public transit. Recently, the U.S. De-
partment of Transportation released a report that outlines the importance of
considering shared mobility services in the planning of multi-modal itineraries
(21) to not only reduce congestion and improve access to public transit, but
also for sustainability reasons.

Many transportation providers collect data that can increase transparency
for reliable travel planning. However, this data has not been fully exploited yet,
even though it will likely form an important piece of future traveler information
systems. Our paper takes an important step in this direction by analyzing and
aggregating data from the operation of scheduled and unscheduled modes to
create a reliability measure for multi-modal travel. As investments in multi-
modal travel planning expand, it is important to develop technology that helps
travelers choose between the many possible options.

Many empirical studies in recent years have shown that travelers value re-
liability of travel time about as much as short travel times (16, 19, 8, 12). Cur-
rently, travel websites do not consider reliability measures in the development
of multi-modal itineraries, but focus primarily on time and cost. Hence, there
exists a gap between the information available to travelers in their decision-
making process and what these travelers actually value in the choice of an
itinerary. So far, travelers can only consider reliability information on the
individual legs of a planned itinerary. For instance, the major airlines give his-
torical data for on-time statistics for individual flights in their itinerary details
and report this information using data only from the past month. However,
this data can often be incomplete and may not be the best predictor for the
planned travel. In addition, it does not allow travelers with multiple legs to
see the possibility of a missed connection or propagated delays. To plan for the
first and last mile, many travelers simply open their favorite navigation app to
look at the current predicted travel time from an airport to their final origin
or destination. However, this time can vary widely depending on the time of
day and day of week, which can affect the success of the entire multi-modal
itinerary.

This research seeks to develop a stochastic optimization model for finding the most reliable multi-modal itinerary given a start time and travel time budget. Since the number of combinations of different modes already makes the problem quite complex, the stochastic nature of multi-modal travel itineraries has been largely left out of itinerary planning. In our reliability model, we combine both schedule-based travel modes (airline network) and unscheduled travel modes (drive times reflecting ride-sharing or taxi transportation). We use a network search algorithm to evaluate itineraries. Our network search considers multiple origin and destination airports which impacts the first and last mile as well as the flight options. We consider the stochastic nature of travel in our model, with travel time distributions built using both discrete and continuous distributions. The main contributions of our research include:

- The development of a reliability measure for multi-modal itineraries,
- The use of historical data to form travel time distributions for different legs and modes of the itinerary,
- An efficient network search algorithm to identify the most reliable multi-modal itinerary,
- An analysis of the structure of reliable itineraries in different stages of the multi-modal itinerary compared to deterministic and other alternative approaches,
- An analysis of features that tend to impact the reliability of itineraries (e.g. times of day, month of the year, airports included).

In Section 2, we give an overview of related work on travel planning in multi-modal travel networks. Then, in Section 3, we present a formal model of our multi-modal travel network and explain how our reliability measure is defined and calculated across the stages of a multi-modal itinerary. Section 4 lays out the network search algorithm that finds the most reliability itinerary (MRI) and presents techniques that we use to reduce the computational run-time of the algorithm. Section 5 shows how we utilize extensive sets of real historical transportation data to build the probability distributions for the legs in the flight and driving networks. Section 6 analyzes how varying parameters across multiple origin-destination (OD) pairs affect the reliability throughout multi-modal travel itineraries. Section 7 summarizes the main takeaways of the paper and provides recommendations of how our work can be useful for both travelers and transportation providers.

2 Related Work

This section contains a summary of related work. First, we will address the relevance of travel time distributions for reliable, a priori travel planning. Then, related work on multi-modal network modeling is summarized, which has been discussed in the context of mobility as a service (MaaS) recently. One way to optimize for the most reliable multi-modal itinerary is the resource-constrained shortest path problem (RCSPP). Our work builds upon these concepts and

extends reliability across transfers through conditional probability and success of transfers to come up with the most reliable itineraries for travelers.

To consider reliability in a priori travel planning, we need not only expected travel times, but the ability to compute the probability that certain combinations of legs in an itinerary will yield successful connections. Several methods exist for determining delay distributions for different modes of transportation. (24) examine departure delay distributions through a statistical approach that takes into account seasonal trends, the daily propagation patterns and random residuals. They demonstrate good fit of the distribution and strong predictive performance. However, they only examine departure delays of single flights and not the entire itinerary. For city logistics, (11) model travel time variability along urban road networks using approximations of Burr XII Type 3 distributions based on realized standardized path durations. They use these time-dependent travel time distributions to simulate routes for arrivals at customers' locations. We expand on this approach by using Google's historical database of travel times to form travel time distributions for routes in urban road networks.

It is a particular challenge to find reliability across different modes given that there may exist multiple connections or a transfer from a scheduled mode of transportation to an unscheduled mode, such as a car trip to the airport. While not focused on reliability, several areas of research are looking at building these multi-modal itineraries as a service to travelers. For example, (28) develop a framework for MaaS and show the sustainability and need for different modes of travel across spatial and temporal efficiency. (26) demonstrate how they use their trip planner algorithm to offer alternatives to travelers based on different criteria or multi-criteria optimization. Their algorithm gives more varied solutions than existing trip-planning services from Bing or Google, and on average has better distribution of arrival times and less transfers for multi-modal trip planning. While our work can be expanded to fit into a multi-criteria version, the reliability measure we propose is focusing on the reliability of the planned multi-modal itinerary.

Many works look to establish multi-criteria optimization within the context of multi-modal travel planning. (3) provide exemplary work that surveys route-planning methods across various transportation networks. They propose building a graph for one transportation mode and then combine it with other transportation modes through "link arcs" to allow for multi-modal transfers. They then provide categories for solving this problem including cost combination, resource-constrained shortest paths and multi-criteria optimization. (2) look into modeling multi-modal travel as an optimization problem with a single linear utility function that considers not only shortest travel time, but other objectives that go into their TRANSIT optimization algorithm. Meanwhile, various techniques improve upon multi-modal network algorithms to make them tractable, including (9), who use techniques such as contraction hierarchies to reduce the size of the combined graphs significantly. Finally, (7) use multi-criteria optimization to create multi-modal travel itineraries and

develop a set of Pareto-optimal solutions, ultimately applying fuzzy logic to give a score to each of those itineraries in the set.

To create reliable itineraries in a travel network, the problem can be modeled as a RCSPP. (6) solve the Maximum Probability Shortest Path Problem by maximizing the probability that all resource constraints are satisfied given the cost of the path does not exceed a specified threshold. They assume that all arc resources are independent and normally distributed. (27) examine the RCSPP with stochastic link travel times by modeling it as a 0-1 integer programming model. They use Lagrangian relaxation methods to find the a priori optimal path with the least expected travel time. (1) solve a variation of the RCSPP called the Reliable h-Path Problem to find paths that minimize cost given a reliability threshold. The key differentiation between our problem and the RCSPP in the work above is that we have conditional probabilities on the edges of our graph, which significantly complicates the modeling and the network search, but makes the problem more realistic.

(4) present a dynamic programming version of time-dependent reliable itineraries in a public transit network. They use past timetables to check how many times the itinerary finished before a certain time to determine that itinerary's success rate. Our work uses this concept of utilizing historic timetables and a time budget for an itinerary and builds upon it with a network search algorithm dedicated to finding the most reliable itineraries. Our conditional model of reliability is inspired by the modeling of reliable train connections by (18). Similar to our problem, for each departure and arrival times for a connection, they calculate distributions of arrival and departure times and then update the delays in the network through delay propagation. We extend this problem to look at multiple connections while still maintaining the conditional probability in our reliability calculations through the network. We start from results presented in (23), our previous paper, where we developed a network search algorithm for a uni-modal travel network considering historical flight data and conditional probability along edges to find the most reliable uni-modal flight itinerary. The additional challenges tackled in the present paper are the combination of multiple modes, the creation of distributions for scheduled and unscheduled travel options as well as the sheer amount of operational data arising in multi-modal travel networks.

3 Model

In this paper, we extend the flight network model presented in (23) to include the first and last mile for door-to-door, multi-modal travel planning. This is done through the combination of continuous distributions derived from historical drive times with discretized empirical distributions created from historical flight data. We first describe the notation required for our multi-modal travel network, outline the foundation for computing our reliability metric, and then describe how the distributions for driving and flight itineraries are combined.

3.1 Multi-modal Network Model

We will begin with defining flight and driving networks separately and then combine them as a multi-modal travel network. We represent the flight network by the graph $G^F = (\mathcal{A}, E^F)$. The set of nodes, $\mathcal{A}$, represents the airports in the network. The set E^F contains all the edges modeling the flights that connect the airports over a two-day span to allow for evening departures. Each edge $e^F \in E^F$ is characterized by a tuple such that $e^F = (l, a_1, a_2, \tau_1, \tau_2)$. In particular, a scheduled flight is associated with a departure airport $a_1 \in \mathcal{A}$, as well as a scheduled departure time τ_1, scheduled arrival airport $a_2 \in \mathcal{A}$, a scheduled arrival time τ_2, and an airline $l \in \mathcal{L}$.

The drive to and from the airport is represented by the driving graph $G^{\mathcal{D}} = (\mathcal{C}, E^{\mathcal{D}})$. The set of nodes, $\mathcal{C}$, represents the city centers considered as origins and/or destinations in the network. $\mathcal{A}^c$ is the set of the five closest airports to a city node $c \in \mathcal{C}$, and each airport is also in the above flight network, such that $\mathcal{A}^c \in \mathcal{A} \ \forall c \in \mathcal{C}$. The drive time between cities and airports (and vice-versa) is time-dependent. Similar to many stochastic time-dependent vehicle routing problems (STDVRP) such as (10), we use time blocks to distinguish the stochastic travel time distributions on an edge. The set of time blocks is $[1, ..., \Theta]$, where Θ is the number of time blocks. The edges of the driving graph can be defined by the tuple $e^{D1} = (c, a_1, \theta)$ (for the drive to the airport) and $e^{D2} = (a_2, c, \theta)$ (for the drive from the airport to the destination), where a_1 is the airport associated with the first-mile (and $a_1 \in \mathcal{A}^c$), a_2 is the destination airport for a last-mile journey (and $a_2 \in \mathcal{A}^c$), and $\theta \in [1, ..., \Theta]$ is the time block during which travel begins.

Thus, our multi-modal network model is made up of graph $\overline{G} = G^F \cup G^{\mathcal{D}}$ with nodes $\overline{\mathcal{A}} = \mathcal{A} \cup \mathcal{C}$ and edges $\overline{E} = E^F \cup E^{\mathcal{D}}$, where $E^{\mathcal{D}} = E^{\mathcal{D}1} \cup E^{\mathcal{D}2}$.

Hereafter, we will refer to a set of connecting driving and flying legs from the traveler's origin to final destination as a *complete itinerary* (CI). A *partial itinerary* (PI) will be defined as a set of driving and potentially flying legs from the origin city to an airport on the way to the destination. Both CIs and PIs can be comprised of several legs, and each leg will be represented by edges e as described above.

We want to identify the reliability of a CI from a specified starting location s to a specified destination location d where $s, d \in \mathcal{C}$. The CI is further restricted to begin at or after time τ_S with a travel time budget of B. With these assumptions, the latest allowed arrival time at the destination d will be $\tau_s + B$. We will use λ_i to represent the sequence of edges for an itinerary $i \in \mathcal{I}$, where $\mathcal{I}$ is the set of all possible PIs and CIs.

Each itinerary, i, whether complete or partial, is defined by a label that contains several pieces of information about that itinerary and is defined by the tuple $(\lambda_i, rel_i, \pi_i, Z_i)$:

- λ_i: Sequence of edges for itinerary i, where there are m flying edges in the itinerary embedded by two driving edges e_1^D and e_2^D
- rel_i: Probability of itinerary i reaching destination by $\tau_s + B$ (reliability)
- π_i: Scheduled travel time of itinerary i

– Z_i: Conditional distribution of tail edge of itinerary i

The scheduled travel time, π_i, reflects how much time is between τ_s and the scheduled arrival of the current tail flight e_m + median drive time from the final airport to the destination if it is a CI. The reliability rel_i of PI i reflects the probability of making all required connections in the PI through edge e_m. If i is a CI, the reliability rel_i represents the probability that this itinerary will lead to an arrival at the destination by $\tau_s + B$. More precisely, for a CI, rel_i is the probability that the traveler drives and arrives to the airport sufficiently before the departure time of the first flight, none of the flights of the itinerary are canceled, the traveler makes all of the connections and drives to the final destination d, arriving before $\tau_s + B$. The calculation of the reliability for itineraries will be explained in detail in Sections 3.2.1–3.2.3.

For our model, we take the following assumptions. We assume travelers on multi-flight itineraries only include flights within a partnering airline group. We do not assume flights wait for late drivers/travelers. Drive times are independent of the delays along the flight edges. Arrival/departure times of different flights are independent. In other words, because one flight is late, this does not tell us anything about the likelihood of other flights being late. Then, we can compute our reliability measure as described below. The required input data we need to create discrete probability distributions of departures and arrivals for all flights as well as for driving times is discussed in Sections 5.1 and 5.2, respectively.

In the label for the itinerary i, we designate the distribution for the tail edge of the itinerary as Z_i. We further define this distribution and differentiate between driving and flying edges. We will use Y for the random variables representing the travel time for the driving legs and X for the flight legs. For simplicity, we will drop the use of the e in defining the probability distributions:

– $P(Y^D \leq t)$: probability that departure of driving edge e^D leads to arrival at first flight or destination by time t,
– $P(X_{dep}^F = t)$: probability that departure of flight e^F occurs at time t,
– $P(X_{arr}^F = t)$: probability that arrival of flight e^F occurs at time t,
– $P(X_{arr}^F = t | X_{dep}^F = t')$: probability that arrival of flight e^F occurs at time t when it departs at t'.

For each scheduled flight leg e, we assume we also know the following:

– $min(X_{arr}^F)$: the earliest possible arrival for leg e^F,
– $max(X_{arr}^F)$: the latest possible arrival time for leg e^F,
– $min(X_{dep}^F)$: the earliest possible departure for leg e^F,
– $max(X_{dep}^F)$: the latest possible departure time for leg e^F.

We will now explain the modeling of our reliability metric following the edges of an itinerary step by step.

3.2 Reliability Calculation

3.2.1 Transfer from Driving to First Flight

To model the connection between the first driving edge ($D1$) and the first flight edge ($F1$), we evaluate the probability that the stochastic drive time Y^{D1} is less than the difference between t and the start time τ_S and the transfer time $\mathcal{T}_{D1}^{tr}$ in Equation (1). This is done for every t in the historical distribution of the first flight, and the summation gives the connection probability between the first driving edge and the first flight edge. The transfer time in this case is reflective of check-in time, ticketing and traversing terminals and security.

$$P(C_{D1,F1}) = \sum_{t=min(X_{dep}^{F1})}^{max(X_{dep}^{F1})} P(X_{dep}^{F1} = t) \cdot P(Y^{D1} \leq t - \tau_S - \mathcal{T}_{D1}^{tr}). \quad (1)$$

Thus, we are able to get the initial reliability of the connection between the driving edge and the initial flight.

3.2.2 Transfers Between Flights

Knowing how to compute this first connection between the driving and first flight edge, the reliability of subsequent flight edges can now be computed as follows. To arrive at this conditional probability, we must look at the summation of probabilities across all departure times of the first flight. Here, for each t' in the departure distribution of the first flight, we must compute its frequency, the probability that the driving edge reaches the airport with sufficient check-in time, and the probability associated with arrival time t given a traveler departs at t' for $F1$. This will give us the conditional probability of arrival of $F1$ at t. This is outlined in Equation (2):

$$P(X_{arr}^{F1} = t \mid C_{D1,F1})$$
$$= \frac{P(X_{arr}^{F1} = t) \cap P(C_{D1,F1})}{P(C_{D1,F1})}$$
$$= \frac{\sum_{t'=min(X_{dep}^{F1})}^{max(X_{dep}^{F1})} P(X_{dep}^{F1} = t')P(Y^{D1} \leq t' - \tau_S - \mathcal{T}_{D1}^{tr})P(X_{arr}^{F1} = t|X_{dep}^{F1} = t')}{P(C_{D1,F1})}.$$
$$(2)$$

The arrival time distribution in Equation (2) is then used to compute the probability of making the connection to the next flight in the itinerary and is shown in Equation (3):

$$P(C_{F1,F2}|C_{D1,F1}) = \sum_{t=min(X_{dep}^{F2})}^{max(X_{dep}^{F2})} P(X_{dep}^{F2} = t) \cdot P(X_{arr}^{F1} \leq t - \mathcal{T}_{F1}^{tr} \mid C_{D1,F1}),$$
$$(3)$$

where

$$P(X_{arr}^{F1} \leq t - \mathcal{T}^{tr} \mid C_{D1,F1}) = \sum_{t'=min(X_{arr}^{F1})}^{min[max(X_{arr}^{F1}),t-\mathcal{T}_{F1}^{tr}]} P(X_{arr}^{F1} = t'|C_{D1,F1}). \quad (4)$$

We repeat this process for each flight in the itinerary. We can generalize Equation (3) to Equation (5) to compute the conditional probability of making a connection between $F(k)$ and $F(k+1)$ given all the previous connections have been made. Equation (5) holds for $k = 2 \ldots m - 1$.

$$P(C_{F(k),F(k+1)}|C_{D1,F1} \cap \cdots \cap C_{F(k-1),F(k)})$$

$$= \sum_{t=min(X_{dep}^{F(k+1)})}^{max(X_{dep}^{F(k+1)})} P(X_{dep}^{F(k+1)} = t) \cdot P(X_{arr}^{F(k)} \leq t - \mathcal{T}_{F(k)}^{tr} \mid C_{D1,F1} \cap \cdots \cap C_{F(k-1),F(k)})$$

$$(5)$$

In order to compute the probability of an arrival time t for the kth flight on the itinerary, a similar conditional probability argument as in Equation (3) is used. Equation (6) evaluates this conditional probability for flight $F(k)$ in the itinerary. We can now define conditional arrival times for $F(k)$ as

$$P(X_{arr}^{F(k)} = t \mid C_{D1,F1} \cap \cdots \cap C_{F(k-1),F(k)}) =$$

$$\frac{\sum_{t'=min(X_{dep}^{F(k)})}^{max(X_{dep}^{F(k)})} P(X_{dep}^{F(k)} = t') P(X_{arr}^{F(k-1)} \leq t' - \mathcal{T}_{F(k-1)}^{tr}|C_{D1,F1} \cap \cdots \cap C_{F(k-2),F(k-1)}) P(X_{arr}^{F(k)} = t|X_{dep}^{F(k)} = t')}{P(C_{F(k-1),F(k)} \mid C_{D1,F1} \cap \cdots \cap C_{F(k-2),F(k-1)})}$$

$$(6)$$

where

$$P(X_{arr}^{F(k-1)} \leq t' - \mathcal{T}_{F(k-1)}^{tr}|C_{D1,F1} \cap \cdots \cap C_{F(k-2),F(k-1)}) =$$

$$\sum_{t''=min(X_{arr}^{F(k-1)})}^{min[max(X_{arr}^{F(k-1)}),t'-\mathcal{T}_{F(k-1)}^{tr}]} P(X_{arr}^{F(k-1)} = t''|C_{D1,F1} \cap \cdots \cap C_{F(k-2),F(k-1)}).$$

$$(7)$$

If an itinerary i is a PI, the conditional probability for the distribution of the tail flight $F(m) - P(X_{arr}^{F(m)} = t \mid C_{D1,F1} \cap \ldots C_{F(m-1),F(m)})$ – will become the values for X_{arr}^i in the label for the itinerary. This allows us to use the calculations for PIs to quickly compute the values for itineraries built from adding a leg to a particular PI. For a PI, the value of rel_i will be computed based on Equation (8):

$$rel_i = \prod_{k=1,..m-1} P(C_{F(k),F(k+1)}|C_{D1,F1} \cap \cdots \cap C_{F(k-1),F(k)}). \quad (8)$$

3.2.3 Final Driving Edge for Complete Itinerary

For a CI, the probability of arriving at the final destination is computed based on the probability values for the prior connections and the likelihood the destination is reached within the travel time budget B for the final driving edge $D2$. Here, $D2$ represents the continuous distribution of drive times that depart at the conditional arrival times of the last flight on the itinerary $F(m)$. We sum the values for $P(X_{arr}^{F(m)} = t \mid C_{D1,F1} \cap \cdots \cap C_{F(m-1),F(m)})$ from $t = min(X_{arr}^F(m))$ to $\tau_S + B$ to get $P(X_{arr}^m \le \tau_S + B \mid C_{1,2} \cap \ldots C_{F(m-1),F(m)})$. The transfer time $\mathcal{T}_{D2}^{tr}$ is determined by the amount of time required to unload and find unscheduled transportation options. Then, our final calculation of reliability for the CI is computed as follows:

$$
rel_i = \sum_{t=min(X_{arr}^{F(m)})}^{\tau_S+B} P(X_{arr}^{F(m)} = t | C_{D1,F1} \cap \cdots \cap C_{F(m-1),F(m)}) \cdot P(Y^{D2} \le \tau_S + B - \mathcal{T}_{D2}^{tr} - t)
$$

$$
\cdot \prod_{k=1,..m-1} P(C_{F(k),F(k+1)} | C_{D1,F1} \cap \cdots \cap C_{F(k-1),F(k)}). \tag{9}
$$

In Equation 9, we can see that the final leg of the itinerary $D2$ is a time-dependent travel time from the conditional arrival times of $F(m)$ to the final destination. How to model the distribution of the random variables Y^{D1} and Y^{D2} is outlined below.

3.3 Data-Driven Modeling of Drive Times

This paper seeks to combine historical data of drive times with historical information of flight times to create reliable itineraries in a multi-modal travel network. In doing so, we provide a more realistic model of the travel time's reliability for a CI from origin to destination between distant city centers. Instead of modeling driving to and from the airport based on a detailed road network model as mentioned in (9), we derive travel time distributions of the drive to and from the airport and combine these with edges of our flight network model. This will be sufficient to determine the most reliable door-to-door itinerary.

In Sections 3.2.1 and 3.2.3, we modeled the random variable for driving edges as Y^{D1} and Y^{D2}. As shown in Section 3.1, these driving edges can be defined by a tuple of (i, j, θ), where i and j are either city nodes $c \in \mathcal{C}$ or airport nodes $a_1, a_2 \in \mathcal{A}^c$. Thus, we can also write these edges as Y_{ij}^θ denoting the nodes they connect and their travel time block θ. Furthermore, Y_{ij}^θ can be defined as:

$$
Y_{ij}^\theta = R_{ij}^\theta + \delta_{ij}^\theta, \tag{10}
$$

where R_{ij}^θ is the baseline (lower bound) travel time for time block θ. Here, δ_{ij}^θ is a random variable from $[0, d_{ij}^\theta]$, where d_{ij}^θ is also a fixed value for that time block and edge. Therefore, Y_{ij}^θ can fall anywhere in the interval of $[R_{ij}^\theta, R_{ij}^\theta + d_{ij}^\theta]$.

We assume that there is a data source that gives us the required information to derive these travel time distributions. Further explanation on how we use this data source to build distributions for the driving edges $D1$ and $D2$ can be found in Section 5.2.

4 Solution Approach

In the following, we present a stochastic network search algorithm that can determine the MRI including stochastic information (1) on drive times to and from the airport and (2) on the performance of flights. We extend a stochastic network search for the most reliable combination of flights presented by (23). In Section 4.1, we present the pseudo-code of the network search algorithm. As the proposed algorithm would require a complete exploration of the investigated travel network and hence lead to exhaustive runtimes, we introduce multiple techniques that help reduce the runtime of the algorithm in Section 4.2.

4.1 Network Search Algorithm

We present the pseudo-code of the network search algorithm in Algorithm 1. Initially, the reliability of the MRI is set to 0 (line 1). Alternatively, we can initialize the reliability using stronger lower bounds. This is discussed in Section 4.2.1. Next, in lines 5-10, initial PIs are created and added to the priority queue ($LIST$) using the initial time-dependent driving edges from the origin to the closest airports. While the priority queue is non-empty, the PI with the highest reliability measure is chosen (line 13), and new itineraries are created for all edges emanating from the tail node of the PI, i_{max}. If the reliability of a new PI is greater than the current best value for the MRI, then it is added to the priority queue (line 23). If the tail airport of the added edge is in $\mathcal{A}^d$, the set of airports closest to the destination, then a time-dependent driving edge is added to the itinerary to form a CI, as shown in lines 28-32. If this CI has a higher reliability than the MRI^*, then it serves as the incumbent MRI (line 34). This process continues until the priority queue is empty, and then the incumbent MRI^* becomes the most reliable itinerary, MRI.

4.2 Runtime Reduction Techniques

Extending the ideas of the stochastic network search presented in (23), there are three options to reduce the runtime required for the creation of reliable itineraries: (1) the improvement of lower bounds, (2) network reduction techniques, and (3) smart queue management. We will summarize the different

---

Algorithm 1 Network Search: Finding the MRI

1: $rel_{MRI^*} = 0, LIST = \emptyset$
2: $k = 1$
3: $\theta =$ time block at τ_S
4: **for all** $a \in \mathcal{A}^s$ **do** // Create labels for first driving edges
5:     $\lambda_k \leftarrow (s, a, \theta)$
6:     $\pi_k \leftarrow \mathbf{E}(Y_{s,a}^\theta)$
7:     $rel_k \leftarrow 1$
8:     $Z_k \leftarrow Y_{s,a}^\theta$
9:     $LIST \leftarrow LIST \cup (\lambda_k, \pi_k, rel_k, Z_k)$
10:     $k = k + 1$
11: **end for**
12: **while** $LIST \neq \emptyset$ **do**
13:     $i_{max} = PI$ with highest rel_i of $LIST$ labels
14:     $LIST \leftarrow LIST \setminus \{i_{max}\}$
15:     $e_m =$ tail edge of PI i_{max}
16:     $\mathcal{E}_m =$ all edges adjacent to tail edge
17:     **for all** $e_j \in \mathcal{E}_m$ **do** // Create label for new PI k
18:         $\lambda_k \leftarrow \lambda_{i_{max}} \cup e_j$
19:         Compute rel_k as indicated in Equation (8)
20:         $\pi_k \leftarrow \tau_1^j - \tau_2^m + \pi_{i_{max}} + \pi_j$
21:         Compute X_{arr}^k as indicated in Section 3
22:         **if** $A_2^j \notin \mathcal{A}^d$ and $rel_k > rel_{MRI^*}$ **then**
23:             $LIST \leftarrow LIST \cup (\lambda_k, rel_k, \pi_k, Z_k)$
24:             $k = k + 1$
25:         **else if** $rel_k < rel_{MRI^*}$ **then**
26:             Move to next $e_j \in \mathcal{E}_m$
27:         **else**
28:             $\theta =$ time block at $\tau_2^k + \mathcal{T}_{D2}^{tr}$
29:             $\lambda_k \leftarrow (a_2^k, d, \theta)$
30:             Compute rel_k as in Equation (9)
31:             $\pi_k \leftarrow \pi_{k-1} + \mathbf{E}(Y_{a_2,d}^\theta)$
32:             $Z_k \leftarrow Y_{a_2,d}^\theta$
33:             **if** $rel_k > rel_{MRI^*}$ **then**
34:                 $MRI^* \leftarrow (\lambda_k, rel_k, \pi_k, Z_k)$
35:             **end if**
36:             $k = k + 1$
37:         **end if**
38:     **end for**
39: **end while**
40: $MRI = MRI^*$

---

ideas that we use from the previous paper here and detail how we extended them below.

4.2.1 Improvement of Lower Bounds

B1 Shortest Path Lower Bound: We can use the shortest travel time itinerary, SP, to determine whether there is a feasible itinerary at all and to find a lower bound on the reliability of the MRI. It is determined with Dijkstra's algorithm.

Adaptations:

– **Shortest Path Lower Bound (B1)**

The initial shortest path reliability lower bound considers the shortest path from the origin airport to the destination airport. To create the shortest scheduled multi-modal travel time itinerary, we adapt this by using the median travel time of the drive time distribution and the scheduled departure and arrival times of the flights. Then, the reliability of this path can be calculated using Equation (9), and it replaces rel_{MRI^*} in the first line of Algorithm 1. This then serves as a stronger lower bound for rel_{MRI^*} throughout the algorithm.

4.2.2 Network Reduction Techniques

N3 Removing Unneccessary Flights: We remove all flights that cannot be part of a feasible solution if the scheduled start and arrival times would not be within the limits of the budget.

N4 Travel Time Bound: This improvement technique eliminates potential flights that would make the itinerary fall outside the bounds of travel $[\tau_S, \tau_S + B]$. In particular, we compute a minimum travel time between airports in the network. Then, we prune flights from the network that would not be possible to take due to the minimum flight time getting to the destination or arriving from the origin outside of the time constraints.

Adaptations:

– **Travel Time Bound (N4)** The travel time bound uses the median travel time for the driving edges and a lower bound on the travel time of the flight edges. As a lower bound for flight travel time, we calculate the haversine distance between two airports and assume a constant speed of 500 MPH to compute the bound for travel time. Each flight edge has a tuple that contains information for the scheduled arrival airport and time and scheduled destination and time $(a_1, a_2, \tau_1, \tau_2)$. If, beginning at τ_S, the minimum combined flight travel time and driving time from s to a_1 is later than τ_1, or if the minimum travel time from a_2 to d is after $\tau_S + B$, then the flight is removed from the network at the beginning of the algorithm.

– **Only Qualifying Flights** We limit flights to only those with at least 15 observations. Note this number occurs with around 97% of the flights in the experiments that were tested. We require 15 observations because we observed having few observations can skew reliability results. We also limit multi-flight itineraries to connections between partnering airlines.

4.2.3 Managing the Priority Queue

M7 Dominant Tail Flight Rule: We analyze partial flight itineraries, $PFIs$, with regard to whether they have the same tail flight. If so, only the PFI with the highest reliability should remain in the priority queue. This is due to the fact that the PFI with an identical tail flight and lower reliability will not be the MRI.

Adaptations:

– **Dominant Tail Edge Rule (M7)** For each itinerary that is a candidate
   to be added to the priority queue, it must go through three checks to
   determine if it should be added. These are done to prevent itineraries with
   identical or sub-optimal tail edges from being added to the priority queue.
   – If there exists a PI from $LIST$ whose reliability as calculated in Equa-
      tion (8) is greater than the candidate itinerary's reliability and has the
      same tail airport, airline and an earlier scheduled arrival time than the
      candidate, then the candidate PI is not added to the priority queue.
   – If there exists a PI that has been previously selected from $LIST$ that
      shares and contains the tail flight of the candidate PI and has a greater
      reliability than the candidate PI, then the candidate PI is not added
      to $LIST$.
   – If neither of the above items are true, the candidate PI will be added
      to $LIST$. However, there may exist other itineraries on the priority
      queue that have a later arrival time at the tail airport and a lower
      reliability than the candidate PI. If such partial itineraries exist, they
      are removed from the priority queue as they would not be viable for
      the MRI.

5 Building Edge Distributions

Since we follow a data-driven approach using large amounts of publicly avail-
able data, in Section 5.1, we will analyze what amount and time period of data
is beneficial to create the empirical distributions for our flight edges. Section 5.2
examines how we use Google data and time blocks to create distributions for
the driving edges in our network.

5.1 Creating Most Accurate Flight Distributions

The distribution of travel times for our flight edges are derived from historical
flight data, which can be found in the United States Bureau of Transportation
Statistics' (USBTS) historical database at *http://www.transtats.bts.gov*. The
database provides commercial domestic flight data dating back to 1987. For US
certified air carriers that account for at least one percent of domestic scheduled
traveler revenue, this database provides information on scheduled and actual
departure and arrival times. In this paper, we use data from the years 2016 and
2017. This represents data from over 16 million recorded flights. We seek to
understand what amount of data and what time period relative to the flight
will give us travel time distributions that improve the predictive ability of
the reliability metric. Once the appropriate amount of prior data to use for
the flight distributions has been determined, we can use the corresponding
probability mass functions (PMF) in our calculations of reliability.

   With most airline websites, the reported data for on-time percentage of a specific flight is based on its performance over the previous month. However, data from the previous month is not necessarily representative of the actual delays that will be realized for that flight in the current month. Thus, we will experiment to see if other time periods can improve the prediction of flight travel times and thus itinerary performance. In order to minimize the error between *predicted* and *actual* reliability of an itinerary, we examine an itinerary's observed reliability.

   The predicted reliability will be built on empirical distributions from historical data from a varying number of previous months. Since scheduled flight times and numbers can change from week to week, we consider the flight to be the same if it has a scheduled departure within an hour of the scheduled departure of the chosen flight. For example, if we are analyzing an itinerary from June, we include in a flight's empirical travel time distribution any flight from May (or another combination of prior months) that departs within 60 minutes of the scheduled departure time with the same origin, destination, and airline. Then, we use these empirical distributions in Equations (1-9) to calculate the predicted reliability. The actual reliability will be the outcomes of an itinerary over all instances in one month. The itinerary is considered successful if it reaches its destination within the budget and makes all connections. The (average) actual reliability of the realized itineraries over the current month is then compared to the predicted reliability measure based on data from flights of previous months. For example, if one day in April a flight connection was missed and another day the tail flight arrived late outside the budget, the itinerary would be successful in 28/30 instances for an actual reliability of 93.3%. This is then compared to the predicted reliability of that itinerary (90.1%, for example) and the error measure is 3.2%. Our goal is to find which set of months to include in the empirical distributions of the predicted reliability to reduce this error between predicted and actual reliabilities.

Table 1 Airports Used for Distribution Testing

| OD Pairs	
Long (>1800 miles apart)	
Newark, NJ (EWR)	Los Angeles, CA (LAX)
San Jose, CA (SJC)	Detroit, MI (DTW)
Panama City, FL (ECP)	Bozeman, MT (BZN)
Norfolk, VA (ORF)	Colorado Springs, CO (COS)
Medium (800-1400 miles apart)	
Omaha, NE (OMA)	Savannah, GA (SAV)
Bentonville, AR (XNA)	Boston, MA (BOS)
Philadelphia, PA (PHL)	New Orleans, LA (MSY)
Wichita, KS (ICT)	Fort Wayne, IN (FWA)
Short (<400 miles apart)	
Phoenix, AZ (PHX)	San Diego, CA (SAN)
Houston, TX (HOU)	Oklahoma City, OK (OKC)
Greensboro, NC (GSO)	Washington, DC (DCA)
Cedar Rapids, IA (CID)	Chicago, IL (MDW)

In the following, we want to analyze systematically what historical data creates the most accurate predicted reliability of itineraries across a range of budgets, months and OD pairs. To do this, we look at a number of different itineraries from short to medium to long distances between OD pairs. The 12 OD pairs we use for these experiments are listed in Table 1, and both directions of the trip are considered. In addition, days for different months (April, July, October, December) are considered to evaluate the impact of seasonality on reliability as well as two different starting times (6:00 and 13:00) and three different budget multipliers (1.1, 1.25, 1.50). The budget multiplier is a multiple of the travel time of the SP. Thus, if SP takes 200 minutes to get from origin to destination, then with a budget multiplier of 1.25, B would be 250 minutes. The itineraries that are analyzed are based on the particular SP for a given day in the month and start time, and the MRI for that day, start time and budget. In order to be considered a valid itinerary, there need to be at least 15 recorded instances for all legs on the specified itinerary for predicted and actual distributions.

The reliability measure as outlined in Equation (9) is then calculated using each different predicted distribution. We consider a number of different combinations of historical distributions that are used to determine which minimizes the error measure between the predicted and actual reliability. These combinations include the previous month, the previous 3 months, the same month from 1 year ago, the previous month plus the month from 1 year ago, and the previous month plus 3 months from one year ago. These combinations are chosen to test an effect of recency as well as seasonality on the predictive power of the reliability measure. More data was added to distributions, such as the previous 6 months or 4 months from one year ago, to test if adding data would improve predictability. However, adding even more data did not seem to create improvements, as measured by the average error or the variance of the error. The Root Mean Squared Error ($RMSE$) is chosen as the error measure due its sensitivity to large outliers (5), a key component that we look to avoid with the reliability metric. As shown in Table 2, the previous month of data has an $RMSE$ for the SP and MRI of 1.26 and the $RMSE$ is minimized for both the SP and MRI itineraries when the previous month of data is used along with the 3 months from 1 year ago.

Table 2 Comparison of the Predictive Power of Different Flight Distributions

Distribution Type	$RMSE\ SP$	$RMSE\ MRI$
(1) Past month	1.26	1.26
(2) Past 3 months	1.39	1.24
(3) Same month − 1 year ago	1.48	1.29
(4) Previous month + month from 1 year ago	1.24	1.18
(5) Previous month + 3 months from 1 year ago	**1.20**	**1.07**

This is further demonstrated in Figure 1, which shows a boxplot of the absolute differences between predicted and actual reliability values for the dif-

Fig. 1 Comparison of the Difference in Reliability Across Distributions
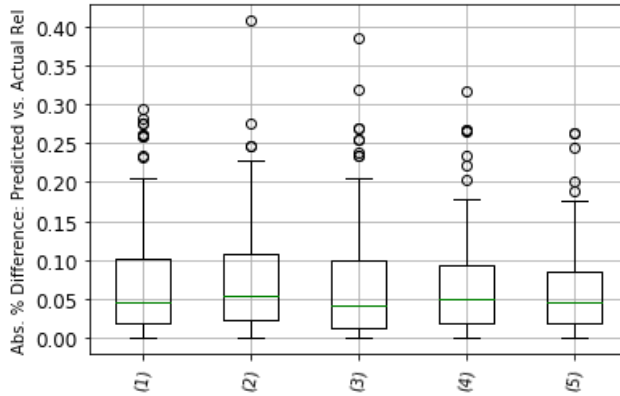
ferent itineraries. The 75th percentile of the instances from the distribution for the prior month plus 3 months from 1 year ago (5) have a predictive error of less than 8% for their reliability values. This is 2% better than all other distributions tested. There are also less extreme outliers for that distribution than for other distributions. In some months, distribution (5) performed significantly better than others, including December and October, when the seasonal component is more important for predicting reliability. In April and July, recency was slightly more important. However, overall, the 3 months from one year ago plus the previous month were best at keeping the error for most of the instances low while managing against many large outliers like distributions (1) and (3). For these reasons, we chose to create predictive distributions consisting of 3 months from 1 year ago plus the prior month as the flight distribution in our experiments.

5.2 Creating Driving Distributions

Since we do not have extensive historical data for the driving edges, we use predictive markers to form continuous probability density functions (PDF) and cumulative distribution functions (CDF) for evaluation of reliability in Equations (1) and (9). To create travel time distributions on driving edges, we access the Google Maps API (14) to retrieve the driving times between a given origin and a destination. For an OD pair in a road network, Google Maps provides an "optimistic", "best guess" and "pessimistic" drive time between the two specified locations. We transform these three markers into a log-normal distribution of continuous drive times following the ideas of (15). Those authors reference numerous papers that suggest that the log-normal distribution is a viable way to model traffic congestion and the uncertainty of drive times. This

Table 3 Drive Time Blocks

| Driving Time Block	Time Window
Morning Peak Driving	07:00 − 09:00
Midday Driving	09:00 − 16:00
Afternoon Peak Driving	16:00 − 18:00
Free Flow Driving	18:00 − 07:00

distribution has non-negative properties and a long tail that can model large traffic delays.

To model the time dependency of the drive times in the course of the day, we divide the day into time blocks as seen, for example, with Uber movement data (25). We use four separate time blocks defined by a range of times as listed in Table 3. Each block is defined by three numbers based on a representative time within that block. For example, for the Morning Peak Driving, the pessimistic, best guess and optimistic travel times from the Google Maps API for 08:00 are used. This is then fit to a log-normal distribution, where the mean of the log to the distribution is the log of the best guess travel time. Meanwhile, the standard deviation of the distribution is the difference between the log of the best guess drive time and the log of the pessimistic drive time divided by 1.645. This represents that the pessimistic driving time is at roughly the 95th percentile of the distribution (or two standard deviations from the mean). Once we have the mean and standard deviation to define this distribution, we are able to include these stochastic driving times into our flight network model.

An example showing why we make the travel time along driving edges time-dependent is given in Figure 2. The fitted distributions represent an example driving edge from New York City center to LaGuardia Airport (LGA) at different time blocks. They vary widely by both median and variance when started in different time blocks. The off-peak block has a smaller travel time median as well as little variance, while the afternoon peak block has more variance and a longer tail to reflect rush hour delays. Note that, depending on data availability, these can be further discretized by day of week and finer time intervals, but for simplicity of this model, we have used the four daily time blocks. We also enforce that these driving edges follow FIFO rules as described in (13) ensuring that leaving the origin at a later time will not get to the destination sooner.

Finally, we will demonstrate how we combine first and final driving edge distributions with flight edge distributions. For the first driving edge, we use the fitted log-normal distributions from the time block associated with the start time. As shown in Figure 3, the continuous drive time distribution is used to determine the missed connection with the historical distribution of the first flight. Figure 3 demonstrates that at the most frequent departure (18:10) of the flight from Newark airport (EWR), around 89% of the driving times would make it to the airport on time. Thus, we are able to use Equation (1) to calculate the conditional flight distribution after this first driving leg. Figure 3 shows, on average, driving from the city center to Newark will make it in time

Fig. 2 Time Block Distributions Throughout the Day – New York City Center → LGA
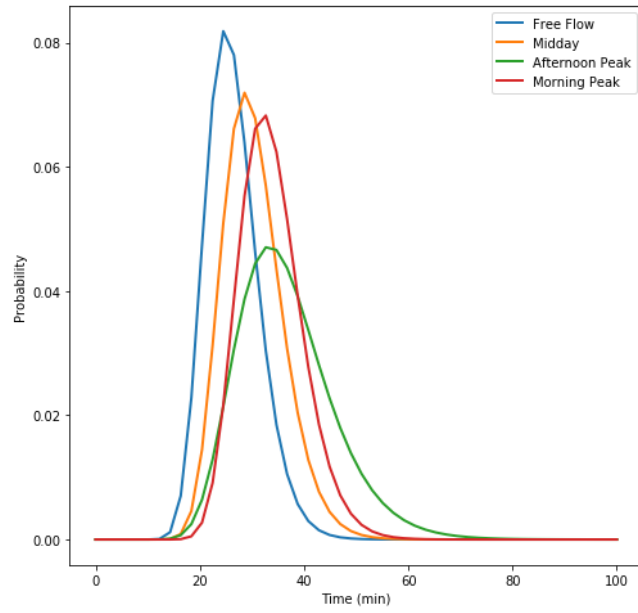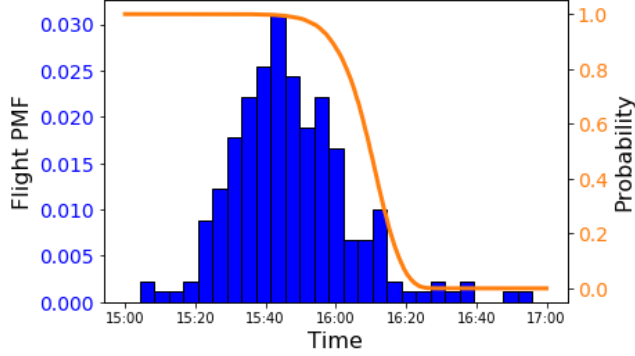

Fig. 3 First Driving Edge – New York City Center ($\tau_S = 17{:}10$) → EWR ($\tau_{dep} = 18{:}10$)
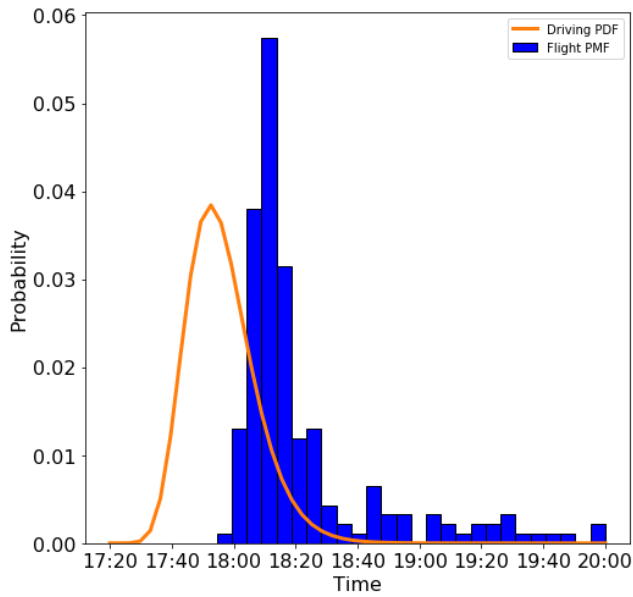
for all instances of the 18:10 flight. However, the PDF representing the driving edge contains a long tail due to the pessimistic possibility of the driving edge, which could cause a missed connection with this initial flight.

Fig. 4 Final Driving Edge – EWR ($\tau_{arr} = 16:05$) → NY City Center ($\tau_S + B = 16:45$)

For the final driving edge, we use the CDFs for the drive time distribution from each airport to the destination city center and time block. To compute the reliability of making it to the destination within the budget, we combine the distribution of discretized flight arrival times that we have from the flight network search and add the distribution of driving times to get a new CDF that we can compare to the budget, as shown in Equation (9). An example of the interaction between a flight arrival PMF and the CDF that represents reaching the destination within the travel budget is shown in Figure 4. This particular cross-country flight from Los Angeles to Newark (EWR) gets in earlier than scheduled more than 80% of the time. Thus, at the median arrival time of the flight (15:35), the traveler is able to reach the destination within the budget near 100% of the time. However, late arrival times after 16:05 PM indicate a near zero percent chance of reaching the destination within the travel time budget. This modeling of drive time edges can help us gain more insight into the interaction between these long-distance and first-mile modes in further experiments.

6 Results

Using the distributions built in Section 5, we perform a set of experiments to understand the characteristics and the value of reliable multi-modal itineraries. We will describe the design of experiments in Section 6.1. We will break down and analyze the results for different itineraries by budget, distance, city size, and start time in the subsequent subsections.

6.1 Design of Experiments

We use 30 OD pairs of city centers selected from the 1,000 most populated cities in the United States (17). These OD pairs are displayed in Figure 5. As noted in Table 4, we choose the 30 OD pairs such that there are 10 long-distance, 10 medium-distance, and 10 short-distance pairs. For each ten, we select two city center OD pairs by population, such that we have two Large-Large, Medium-Medium, Medium-Large, Small-Large and Small-Small combinations.

Table 4 Parameters Tested Across Itineraries

Distance between cities	City Size
Long-distance (>1800 miles)	Large – Top 100 populated cities >200,000
Medium distance (800-1400 miles)	Medium – Between 75-100,000
Short Distance (<400 miles)	Small – 900-1000 largest cities <50,000
Start Times of Itineraries	Budget Multipliers
06:00	1.1 – 110% of Shortest Travel Time
10:00	1.25 – 125% of Shortest Travel Time
16:40	1.5 – 150% of Shortest Travel Time
Itineraries Tested	Days of Travel
MRI	April 17th, 2017
SP	July 17th, 2017
$Closest$	October 17th, 2017
$Biggest$	December 17th, 2017

We test the reliability of itineraries across a variety of categories as shown in Table 4. These tests aim to show how the distance between cities, the population of the cities, the start time during the day, the various budget sizes and the different days of travel affect the reliability for multi-modal itineraries created with different objectives. These itineraries include:

- MRI – most reliable itinerary,
- SP – shortest travel time itinerary based on median drive times and scheduled flight times,
- $Closest$ – most reliable itinerary using the closest airport by median drive time to the origin and destination,
- $Biggest$ – most reliable itinerary using the largest airport based on number of daily outgoing (origin) or incoming (destination) flights.

These different itineraries were tested with all of the parameters listed in Table 4. In our experiments, the transfer time for the first driving edge ($\mathcal{T}_{D1}^{tr}$) is 30 minutes to represent check-in time at the airport, and the transfer time for the final driving edge ($\mathcal{T}_{D2}^{tr}$) is 15 minutes to represent de-planing and getting an Uber, taxi, or ride-share. Note that with access to more information, these transfer times could be made stochastic and time or airport-specific. All algorithms were implemented in Python and run on an Intel Core i5 2.3 GHz processor with 16 GB of RAM. Over 99.5% of MRI instances were solved within 30 seconds with more than 77% of them solved in less than 1 second.

Across the 2,160 different experiments, there were a number of interesting trends that were revealed by the data.

Fig. 5 Our 30 City Center OD Pairs

6.2 Comparison Across Budgets

First, we want to understand how the different modes impact the reliability of the itineraries. We want to see how this changes for different itineraries as well as different budgets.

Table 5 Largo, FL → Santa Barbara, CA $\tau_S = 10{:}00$; $B = 14$ hours

Itinerary Type	Rel Lost 1st Drive	Rel Lost Flights	Rel Lost Last Drive	Total Rel	Airports	Travel Time
SP	8.4%	3.0%	<0.1%	88.6%	TPA → IAH → SFO	11 h 14 min
Closest	<0.1%	15.7%	9.7%	74.5%	TPA → IAH → SFO → SBA	12 h 52 min
Biggest	1.2%	6.5%	0.7%	91.6%	MCO → ATL → LAX	12 h 38 min
MRI	0.3%	3.4%	<0.1%	96.2%	TPA → DFW → LAX	11 h 34 min

As an example, we illustrate the differences between the itinerary types from the city centers of Largo, FL to Santa Barbara, CA on December 13, 2017. Table 5 highlights the different itineraries chosen based on different itinerary types. It includes the reliability lost on the first driving edge (*Rel Lost 1st Drive*), the reliability lost on flight connections/cancellations (*Rel Lost Flights*) and the reliability lost on the final driving edge (*Rel Lost Last Drive*). Note that the origin and destination airports can change according to the itinerary type. The duration of the itineraries varies from 11 hours 14 minutes to 12 hours 52 minutes. *SP* picks the closest airport to combine the first driving edge with the first flight edge (TPA) and loses most of its reliability due to a tight connection driving and checking into the first flight. Going from the two closest airports (TPA,SBA) to the destination, we see the reliability is lower (74.5%) due mostly to flight connections and delays. The *Biggest* itinerary chooses to fly from Orlando (MCO) to Los Angeles (LAX) and drive almost 3.5 hours to give a reliability of 91.6%. This route does have 1 hour and 25 minutes additional travel time over the *SP*. Finally, the *MRI* has a slightly longer scheduled travel time at 20 minutes more than the *SP*, but it offers an itinerary with a 96.2% reliability, an increase of 7.6% over the reliability of *SP*.

Fig. 6 Comparing Reliability Across Budgets and Itinerary Types
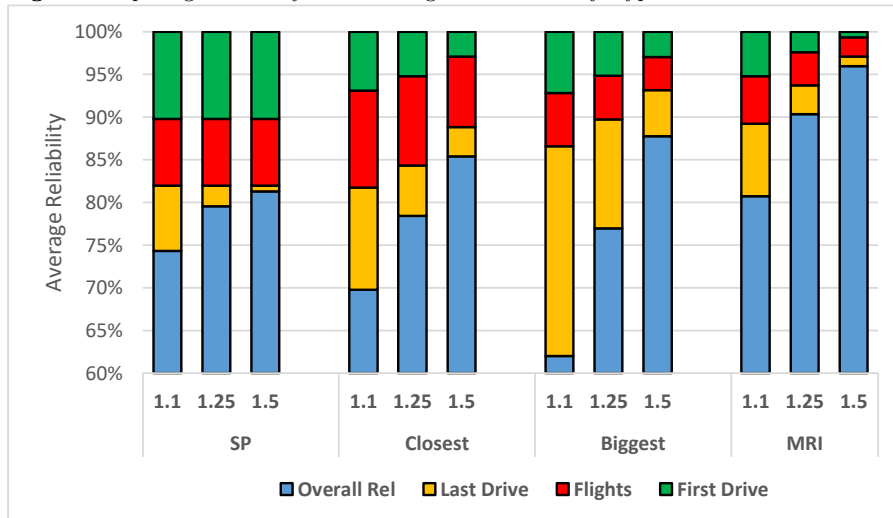

We next aggregate these reliability results from the experiments for all OD pairs, itinerary dates and start times in Figure 6. We use different colors to represent the average amount of reliability loss from a particular mode. Specifically, green represents loss from the first driving edge, red represents reliability loss from flight connections/cancellations, and yellow shows reliability loss from the final driving edge. The overall reliability is then presented

in blue. We analyze across the four different itinerary types and three different travel time budgets.

Since SP always chooses the same itinerary regardless of budget, the 10% reliability loss on the first driving edge and 8% on flight connections will be the same across all budgets. The final driving edge varies in reliability loss as the budget increases, giving more of a buffer for late-arriving flights. For $Closest$ itineraries with a budget multiplier of 1.1, the reliability after the first drive to the airport is 93.1%, 81.8% after flight connections/cancellations, and 70.0% overall after the final driving edge. $Closest$ itineraries have relatively small reliability losses on driving edges, which makes sense due to the shorter drives. However, across budgets, the flight reliability loss is between 8-12%, which is larger than with other itinerary types. This could be due to flying out of nearby, smaller airports with less numerous and reliable flight selections. The lack of a buffer on the final driving edge is highlighted in $Biggest$ itineraries. While first driving edge reliability and flight reliability are both better than for the SP and $Closest$ itineraries, these larger airports can be further away making the final drive into the destination difficult to reach within the budget. This causes a loss of nearly 25% in reliability from the final driving edge. Finally, the reliability of MRIs is considerably higher with an average of 80.7%, 90.3% and 96.0% across budget multipliers of 1.1, 1.25 and 1.5. MRIs are able to find routes that can go to airports with more reliable flights while still creating a driving time buffer that is within the travel time budget.

Fig. 7 Comparing Travel Time Across Budgets and Itinerary Types
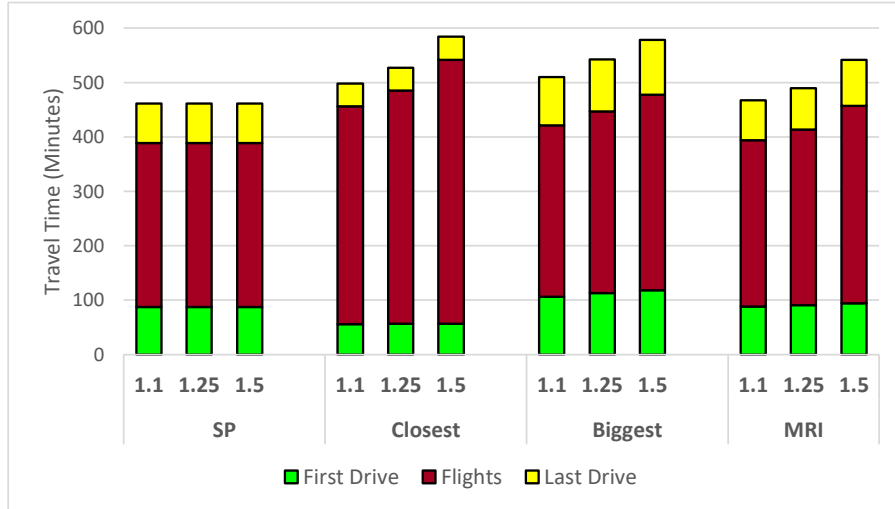
We also want to understand the change in travel time associated with these changes in reliability. Figure 7 outlines the average travel time used across the different stages of the itinerary. SP uses an average of 461 total

Table 6 Comparing Buffer Time for Each Multi-modal Stage

Itinerary: Budget Multiplier 1.1	First Driving Edge Buffer	Avg Flight Edge Buffer	Final Driving Edge Buffer
SP	1 hr 18 min	1 hr 22 min	1 hr 6 min
Closest	1 hr 39 min	1 hr 24 min	47 min
Biggest	1 hr 47 min	1 hr 12 min	35 min
MRI	1 hr 46 min	1 hr 30 min	57 min

minutes of travel time, with 302 minutes spent flying or at airports and 87 and 72 minutes for first and last mile travel (including transfer time). The travel itineraries to *Closest* airports do not have much drive time, but they are accompanied by long waiting times at airports, while the opposite is true for *Biggest* itineraries. Both closest and biggest itineraries have around 40, 60 and 120 minutes more travel time than the *SP* for budget multipliers of 1.1, 1.25 and 1.5, respectively. The *MRI* only has 6, 28 and 81 minutes of additional travel time for reliability gains of 6%, 10% and 14% over *SP* as shown in Table 6. Thus, depending on the situation, it may be beneficial to give up some travel time to gain reliability.

In order to understand where this additional travel time is used and how it affects reliability, it is important to look at how each itinerary utilizes buffer times between legs. Table 6 outlines the buffer time for the respective stages of multi-modal travel with a tight budget of 1.1. The *First Driving Edge Buffer* is the difference between the scheduled time of the initial flight and the median drive time to the airport. The average flight buffer is between the scheduled arrival of one flight and the scheduled departure of the subsequent flight. The final driving edge buffer is the difference between the budget and the median drive time arrival at the destination.

The *SP* itinerary attempts to get to the destination as quickly as possible, and the lack of buffer time is apparent, especially in the first driving edge. The large amount of reliability loss (25%) for the *Biggest* itineraries can be traced to a lack of buffer time on the final driving edge at only 35 minutes on average. This can be problematic over long drives into city centers that have large airports. When comparing *SP* and *MRI*, the reliability lost in each stage is comparable to the buffer time given. The first driving edge for the *SP* loses 10% reliability compared to 5% lost for *MRI* driving edges due to the additional 28 minutes of buffer for the *MRI*. However, the *MRI* leaves 57 minutes of buffer for the final driving edge compared to 1 hour and 6 minutes for the *SP*, and this results in higher reliability loss for the *MRI* (8.5% vs. 7.6%). The *MRI* is able to identify reliable legs and allot buffer time while not sacrificing a large amount of travel time. This shows the necessity of an algorithm to not only offer the shortest travel time or cheaper travel from large airports, but also find the best combinations of flying and driving legs to increase the reliability of an itinerary.

6.3 Comparison Across Trip Distance

It is also important to know how the reliability of multi-modal itineraries changes over trips of different lengths. Figure 8 compares results for long ($>$1800 miles apart), intermediate (800-1400 miles apart) and short ($<$400 miles apart) distance for both the SP and MRI itineraries. The reliability loss for the first driving edge on SP itineraries stays consistent regardless of distance at around 10%. However, as the distance decreases, the reliability lost on flight legs decreases significantly due to more direct flights. Meanwhile, the final driving edge reliability loss increases from 2.2% to 5.0% as the budgets are almost 400 minutes shorter for short-distance flights, making reaching the destination in time with a stochastic driving edge more difficult.

**Fig. 8** Comparing Reliability Across Distance Between OD Cities
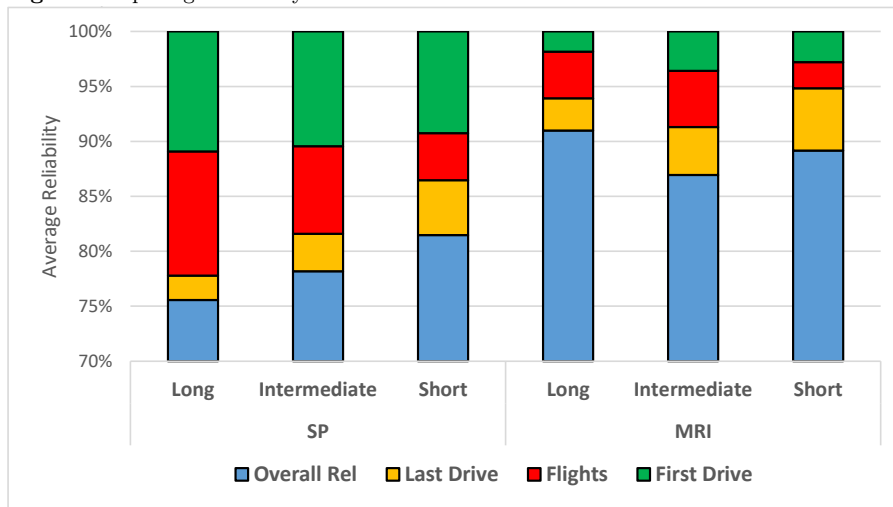
The MRI reliability loss does not follow this same pattern across distances. Long distance flights actually have a higher overall reliability than intermediate and shorter flights. This is due to drive times for all three distance types being close to two hours in total for the MRI. However, long, intermediate and short itineraries differ drastically in budget with 797, 576, and 402 minutes, respectively. Since there exists more of a buffer on driving edges but around the same median drive times, this allows long MRI itineraries to be the most reliable. Short MRI itineraries overcome their smaller travel budgets to be more reliable than intermediate itineraries. This is due to having greater than 80% direct flights compared to 34% direct flights for intermediate itineraries.

Overall, the MRI improves the reliability of short, intermediate, and long itineraries by 7.7%, 8.7% and 15.4% over SP itineraries. Thus, the MRI can improve reliability across all trip types and within a multi-modal context, particularly with longer distance itineraries.

6.4 Comparison Across Different Times of Day

Fig. 9 Comparing MRI Reliability Throughout the Day
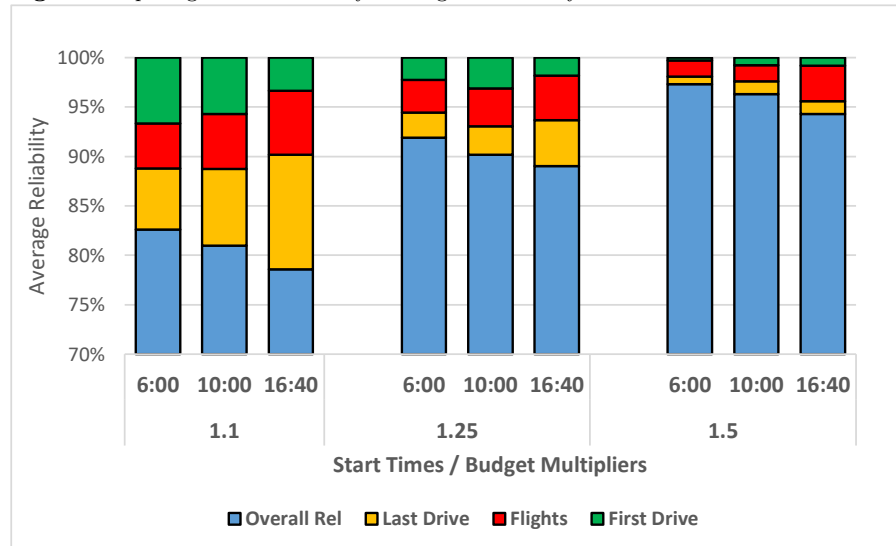
Like with itinerary distance, reliability also changes with different start times throughout the day. We examine early morning, mid-morning, and late afternoon start times across all OD pairs, and the results are reported in Figure 9. The overall average reliability is lower as the day progresses, presumably because of delay propagation and less flights being available.

There is, on average, a 10% increase in reliability for every start time if the budget multiplier is increased from 1.1 to 1.25 and then another 5% increase as the budget multiplier goes from 1.25 to 1.5. This shows the marginal gain available from increasing the travel time budget to allow for more buffer between edges. We also see the MRI itinerary with larger budgets being able to mitigate the reliability loss on driving edges to less than 1%, but with itineraries started later in the day, reliability is still lost on flight connections and cancellations. Overall, these results indicate that leaving the origin at an earlier time in the day will generally result in higher reliability.

6.5 Comparison Across City Size

We have also analyzed OD pairs that have different populations to see if any differences exist based on city size. We found that the airport selected in the MRI for medium size cities (75,000-100,000 people) tends to favor those airports that are far away and large in size. This is shown in Table 7, which

Table 7 Comparing Reliability Across City Size

City Itinerary	Daily Avg. OD Flights (MRI)	Daily Avg. OD Flights ($Closest$)	MRI Rel	$Closest$ Rel	MRI Drive Time
Big-Big	822	394	89.0%	79.9%	1 hr 26 min
Medium-Medium	993	173	88.0%	70.3%	2 hr 45 min
Small-Small	584	355	88.3%	82.1%	2 hr 25 min

gives a number of possible explanations why this trend occurs. The sum of the daily average flights out of the origin and the daily average flights into the destination is given by *Daily Avg. OD Flights*. The MRI reliability is relatively the same across city size, but is much lower for the *Closest* itineraries for medium cities. The reason medium cities may be traveling further and to larger airports could be explained due to this reduced reliability of closest airports.

Table 7 shows the reliability choosing closest airports in Medium-Medium itineraries is 70.3% while this same *Closest* itinerary for Big-Big and Small-Small cities are 79.9% and 82.1%, respectively. This could be due to medium cities being closest to smaller airports, as demonstrated by the 173 average daily flights out of closest airports opposed to 394 or 355. This leads to driving further for the MRI (2 hours 45 minutes vs. 1 hour 26 minutes). This could all be due to bigger cities having an airport with more frequent and reliable flights, whereas medium-sized city travelers must travel farther to find such airports, and when they do travel far, they select hub airports.

6.6 Comparison Across Months

Travelers also need to know which time of the year can lead to changes in reliability in case they are planning for a future vacation or business trip. As shown in Table 8, there exists a difference among months in average reliability for $MRIs$. This trend holds with SP as well, and it can most likely be attributed to the higher airport traffic during the month of July as shown by the 1000-2000 more flights per day compared to other months. This heavy traffic in travelers and flights can lead to delay propagation throughout the network and cause a drastic decrease in average reliability. Unique discrepancies in reliability such as this one can help travelers form trip itineraries that avoid situations or time periods that could result in a riskier journey with higher probability of delays or missed connections.

Table 8 Differences in Reliability Across Months

Day of Travel	Number of Flights	MRI Rel
December 13th, 2017	15230	90.9%
October 16th, 2017	16382	90.9%
July 17th, 2017	**17260**	**85.0%**
April 17th, 2017	16343	89.2%

6.7 Comparison of Airports

We examine which airports lost the most reliability per flight to understand the impact that individual airports have on itinerary reliability. Across the 2,160 instances tested, we find the most reliable itinerary. For these itineraries, we evaluate the average reliability lost for flights that go through each airport. The chart below shows the 5 airports (of those with more than 100 flights) with the most reliability lost per flight on average. As Figure 10 shows, O'Hare Airport in Chicago (ORD) loses nearly 6% reliability on average from cancellations or delays that caused missed connections. Additionally, Texas airports, Dallas-Fort Worth (DFW), Houston Hobby (HOU) and Dallas Love (DAL), lost over 4% reliability along with Minneapolis (MSP). Based on these findings, it would be wise for travelers to consider the possible delays or cancellations they may face when using these airports.

Fig. 10 Most Reliability Lost Per Flight at Major Airports
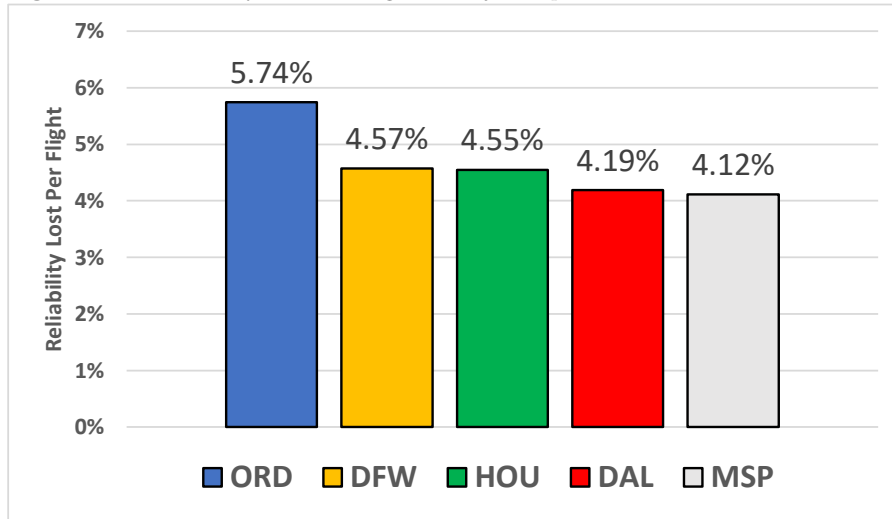
7 Conclusion

State-of-the-art traveler information systems can create multi-modal travel itineraries, but they usually do not consider the reliability of the chosen transportation services when creating these itineraries. In this paper, we have introduced techniques for identifying reliable itineraries that could be integrated in future traveler information systems. In particular, we have modeled scheduled and unscheduled transportation services in a stochastic multi-modal network.

Our experiments show that reliable multi-modal itineraries differ significantly from itineraries created with deterministic approaches. It is not obvious

which combination of services can bring the traveler to his or her destination in the most reliable way. Our approach can help improve decision making through updated traveler information systems as follows:

− Consider traveler value of time versus value of reliability through the use of a travel time budget,
− Use historical and predictive information to create reliable itineraries with appropriate buffers across different modes,
− Increase transparency of systematic expected delays at particular times of the day, seasons of the year, etc.,
− Consider not only local alternatives, but also include driving to larger airports that are further away to increase reliability.

There are also advantages for transportation providers implementing our techniques for identifying reliable itineraries:

− Foster the implementation of standards for airports and airlines that not only look at on-time percentage, but also focus on number of missed connections and cancellations for the entire itinerary,
− Adjust capacity and resources for airports to address loss of reliability at peak times of year and times of day,
− Publish easy-to-understand reliability metrics embedded in multi-modal travel itineraries on travel websites to increase awareness and encourage return business for transportation services.

In future work, we plan to expand our reliability model to scheduled transit modes, such as buses or trains, as well as first-mile and last-mile travel with delayed or non-guaranteed supply, such as bike or ride sharing services. We plan to partner with local governments and companies to obtain this data and turn our focus to regional itineraries. It will be interesting to see how our models will perform given the characteristics of regional transportation networks.

8 Appendix

Table 9 examines key statistics from each of the OD pairs considered in the experiments from Section 6. $SP \#\_Flights$ and $MRI \#\_Flights$ is the average number of flights on itineraries between these cities. For example, New York to Los Angeles always chooses direct flights while Bozeman, MT to Panama City, FL averages more than 3 flights. The reliability for each stage of the SP and MRI itineraries are also displayed. This OD summary table can help trace which cities have reliability loss at different stages and how to use this to plan more reliable itineraries.

References

1. Andreas, A., Smith, J.C., Küçükyavuz, S.: Branch-and-price-and-cut algorithms for solving the reliable h-paths problem. Journal of Global Optimization **42**, 443–466 (2007)

Table 9 30 OD Pairs Summary Table

OD Pair		SP #_Flights	MRI #_Flights	SP_1st_Drive_Rel	SP_Flights_Rel	SP_Tot_Rel	MRI_1st_Drive_Rel	MRI_Flights_Rel	MRI_Tot_Rel
New York	Los Angeles	1.0	1.0	0.833	0.824	0.794	0.976	0.976	0.929
Virginia Beach	Colorado Springs	2.2	1.9	0.814	0.703	0.680	0.937	0.887	0.857
Philadelphia	New Orleans	1.5	1.3	0.827	0.798	0.768	0.929	0.915	0.871
Wichita	Fort Wayne	2.1	2.0	0.948	0.835	0.795	0.957	0.881	0.844
Winston-Salem	Washington	1.1	1.0	0.842	0.828	0.801	0.964	0.961	0.917
Columbus	Greensboro	1.9	1.8	0.896	0.836	0.803	0.991	0.958	0.924
San Jose	Warren	1.8	1.7	0.848	0.771	0.753	0.988	0.968	0.947
Wausau	Sacramento	2.3	2.2	0.932	0.751	0.726	0.966	0.886	0.850
Baltimore	Greenacres	1.0	1.0	0.883	0.879	0.828	0.944	0.944	0.888
Annapolis	Raleigh	1.0	1.0	0.964	0.964	0.920	0.973	0.973	0.929
Prescott Valley	Chula Vista	1.0	1.0	0.907	0.907	0.862	0.947	0.947	0.898
Omaha	Hilton Head Island	2.0	2.0	0.877	0.820	0.800	0.947	0.909	0.879
Racine	Santa Ana	1.1	1.1	0.861	0.839	0.805	0.969	0.964	0.928
Corpus Christi	Bellingham	2.5	1.9	0.848	0.717	0.703	0.962	0.923	0.904
Boca Raton	Forth Worth	1.0	1.0	0.933	0.933	0.882	0.968	0.968	0.912
Bryan	Tucson	1.3	1.2	0.891	0.852	0.818	0.973	0.952	0.907
Chino	Stockton	1.0	1.0	0.827	0.827	0.770	0.950	0.950	0.890
Lawton	Houston	1.2	1.1	0.856	0.819	0.766	0.922	0.915	0.854
Boca Raton	Mountain View	1.6	1.4	0.922	0.867	0.836	0.982	0.970	0.940
Santa Barbara	Largo	1.8	1.5	0.834	0.740	0.718	0.972	0.946	0.916
Sioux City	Scranton	2.0	2.0	0.831	0.713	0.678	0.927	0.869	0.822
Albany	Duluth	1.8	1.6	0.890	0.803	0.771	0.962	0.917	0.870
Nashua	Trenton	1.0	1.0	0.866	0.866	0.781	0.967	0.967	0.878
Medford	Yakima	1.7	1.5	0.921	0.802	0.767	0.972	0.912	0.855
Clovis	Atlantic City	2.6	2.3	0.938	0.782	0.767	0.984	0.926	0.901
Bozeman	Panama City	3.5	3.1	0.950	0.783	0.773	0.995	0.945	0.927
Holyoke	Bentonville	2.3	2.1	0.900	0.800	0.778	0.959	0.905	0.878
Cape Girardeau	Moorhead	2.1	2.0	0.816	0.729	0.702	0.933	0.870	0.822
Cedar Falls	Calumet City	1.1	1.0	0.949	0.921	0.867	0.980	0.980	0.925
Sherman	Friendswood	1.0	1.0	0.878	0.878	0.811	0.920	0.920	0.846

2. Antsfeld, L., Walsh, T.: Finding optimal paths in multi-modal public transportation networks using hub nodes and transit algorithm. Artificial Intelligence and Logistics (2012)

3. Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., Werneck, R.: Route planning in transportation networks. Algorithm Engineering pp. 19–80 (2016)

4. Böhmová, K., Mihalák, M., Pröger, Tobias Srámek, R., Widmayer, P.: Robust routing in urban public transportation: How to find reliable journeys based on past observations. ATMOS '13 pp. 27–41 (2013)

5. Chai, T., Draxler, R.R.: Root mean square error (rmse) or mean absolute error (mae)? - arguments against avoiding rmse in literature. Geoscientific Model Development **7**(3), 1247–1250 (2014)

6. Chen, J., Lisser, A.: Maximum probability shortest path problem. Discrete Applied Mathematics **192**, 40–48 (2015)

7. Delling, D., Dibbelt, J., Pajor, T., Wagner, D., Werneck, R.: Computing multimodal journeys in practice. International Symposium on Experimental Algorithms pp. 260–271 (2013)

8. Devarasetty, P., Burris, M., Shaw, W.: Do travelers pay for managed-lane travel as they claimed they would? before-and-after study of travelers on katy freeway, houston, texas. Transportation Research Board: Journal of the Transportation Research Board **2297**, 58–65 (2012)

9. Dibbelt, J., Pajor, T., Wagner, D.: User-constrained multimodal route planning. Journal of Experimental Algorithmics (JEA) **19** (2015)

10. Duan, Z., Sun, S., Sun, S., Li, W.: Stochastic time-dependent vehicle routing problem: Mathematical models and ant colony algorithm. Advances in Mechanical Engineering **7**(11) (2015)

11. Ehmke, J., Campbell, A.: Customer acceptance mechanisms for home deliveries in metropolitan areas. European Journal of Operational Research **233**, 193–207 (2014)

12. Ehreke, I., Hess, S., Weis, C., Axhausen, K.: Reliability in the german value of time study. Transportation Research Record: Journal of the Transportation Research Board **2495**, 14–22 (2015)

13. Fleischmann, B., Gietz, M., Gnutzmann, S.: Time-varying travel times in vehicle routing. Transportation science **38**(2), 160–173 (2004)

14. GoogleMaps: ”`https://developers.google.com/maps/documentation/`” (2018). Accessed 13 December 2018

15. Guessous, Y., Aron, M., Bhouri, N., Cohen, S.: Estimating travel time distribution under different traffic conditions. Transportation Research Procedia **3**, 339–348 (2014)

16. Hossan, M.S., Asgari, H., Jin, X.: Investigating prefernce heterogenieity in value of time (vot) and value of reliability (vor) estiamtion for managed lanes. Transportation Research Part A: Policy and Practice **94**, 638–649 (2016)

17. Jones, R.: 1000 largest u.s. cities by population with geographic coordinates. ”`https://gist.github.com/Miserlou/c5cd8364bf9b2420bb29`” (2013). Accessed 13 December 2018

18. Keyhani, M.H., Schnee, M., Weihe, K., Zorn, H.P.: Reliability and delay distributions of train connections. In: OASIcs-OpenAccess Series in Informatics, vol. 25. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2012)

19. Kou, W., Chen, X., Yu, L., Qi, Y., Wang, Y.: Urban commuters' valuation of travel time reliability based on stated preference survey: A case study of beijing. Transportation Research Part A: Policy and Practice **95**, 372–380 (2017)

20. Maan, M.: Rome2rio launches all-new app for ios. ”`https://www.rome2rio.com/blog/2018/06/12/new-rome2rio-app/`” (2018). Accessed 13 December 2018

21. McCoy, K., Andrew, J., Glynn, R., Lyons, W.: Integrating shared mobility into mutlimodal transportation planning: Improving regional performance to meet public goals. Tech. Rep. DOT-VNTSC-FHWA-18-13; FHWA-HEP-18-033, Federal Highway Administration: Office of Planning, Environment and Realty (2018)

22. Musulin, K.: Coord unveils routing api to support multimodal trip planning. ”`https://www.smartcitiesdive.com/news/coord-routing-api-multimodal-trip-planning/524525/`” (2018). Accessed 13 December 2018

23. Redmond, M., Campbell, A.M., Ehmke, J.F.: The most reliable flight itinerary problem. Networks pp. 1–19 (2018)

24. Tu, Y., Ball, M., Jank, W.: Estimating flight departure delay distributions - a statistical approach with long-term trend and short-term pattern. Journal of the American Statistical Association **103**(481), 112–125 (2008)

25. UberTechnologies: ”`https://movement.uber.com`” (2018). Accessed 13 December 2018

26. Ulloa, L., Lehoux-Lebacque, V., Roulland, F.: Trip planning within a multimodal urban mobility. IET Intelligent Transport Systems **12**(2), 87–92 (2017)

27. Wang, L., Yang, L., Gao, Z.: The constrained shortest path problem with stochastic correlated link travel times. European Journal of Operational Research **255**, 43–57 (2016)

28. Wong, Y., Hensher, D., Mulley, C.: Emerging transport technologies and the modal efficiency framework: A case for mobility as a service (maas) (2017)