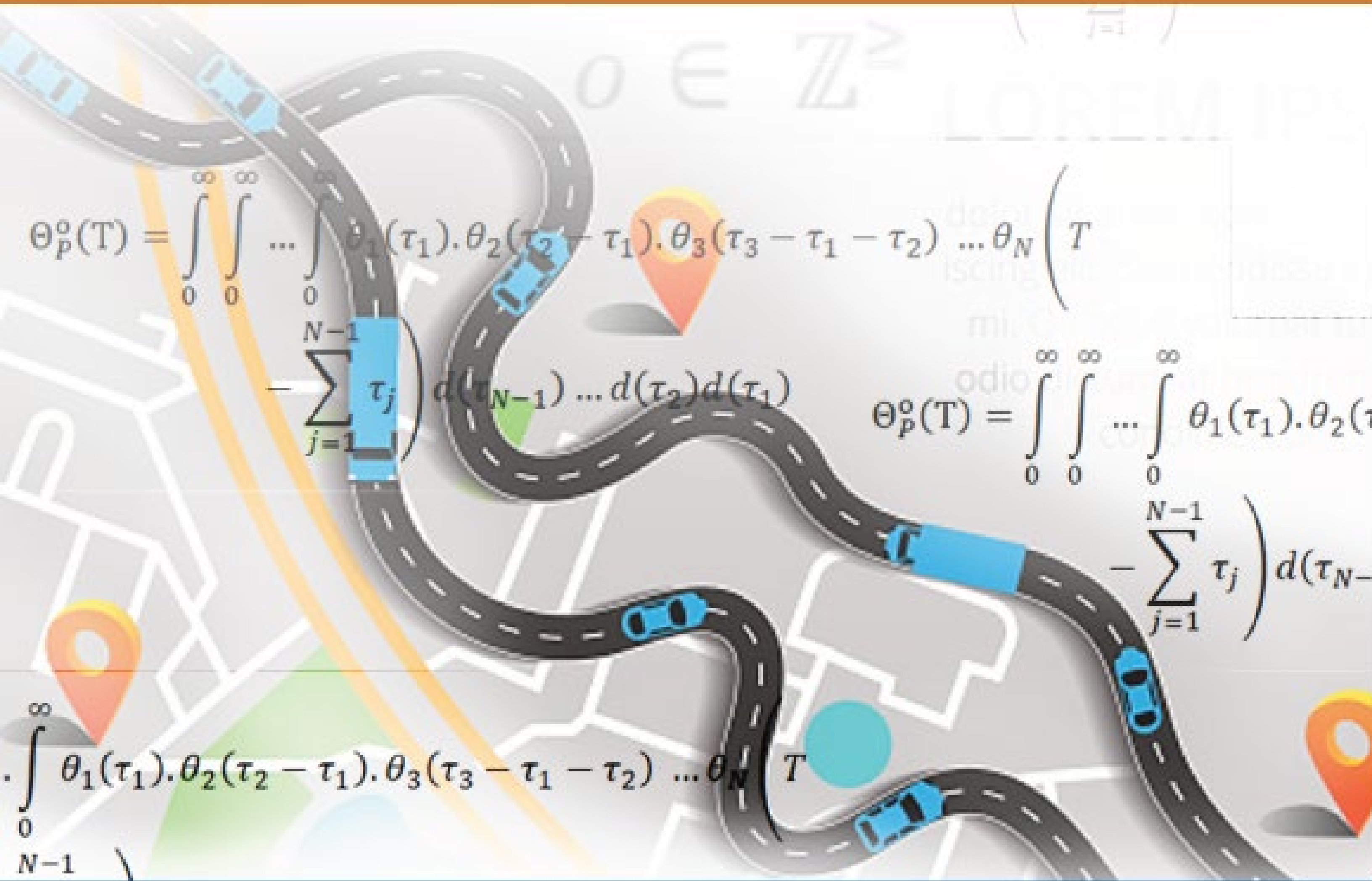
U.S. Department of Transportation
Federal Highway Administration

# ESTIMATION OF TRAVEL TIME DISTRIBUTIONS ALONG USER-DEFINED TRAVEL PATHS

## Application Reference



Cover Photo Source: Getty Images

# TECHNICAL REPORT DOCUMENTATION PAGE

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| FHWA-HOP-20-066 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Estimation of Travel Time Distributions Along User-Defined Travel Paths | May 2024 |
| | 6. Performing Organization Code |
| 7. Authors | 8. Performing Organization Report No. |
| Hani Mahmassani and Monika Filipovska | |
| 9. Performing Organization Name and Address | 10. Work Unit No. (TRAIS) |
| Leidos<br>11251 Roger Bacon Drive<br>Reston, VA 20190 | 11. Contract or Grant No.<br>DTFH61-16-D-00030/0021 |
| 12. Sponsoring Agency Name and Address | 13. Type of Report and Period Covered |
| U.S. Department of Transportation<br>Federal Highway Administration<br>1200 New Jersey Avenue, SE<br>Washington, DC 20590 | February 2019–February 2020 |
| | 14. Sponsoring Agency Code<br>HOTM |

**15. Supplementary Notes**

The government task manager was Douglas Laird.

**16. Abstract**

Travel time reliability has direct economic impacts on the decisions of logistics firms and delivery vehicles, and the service levels ultimately experienced by shippers, firms, and consumers. Underlying all measures of travel time reliability is the variability of travel times experienced along individual links, by various modes, at junctions and intermodal terminals, and through the individual trajectories of travelers and goods. Multiple challenges can arise in estimating travel time distributions for any path in a given network with variable travel times. This reference brings together modern advances in data availability and probabilistic modeling techniques to overcome these challenges, and synthesize travel time distributions at the path level in generalized networks. These may include different facility types and potentially modal services along a particular path. This reference provides an overview of the methodology and tools that can be applied to successfully synthesize and replicate traveler-based travel time distributions along continuous paths, regardless of facility type.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Travel time reliability, travel time distribution, origin-destination | No restrictions. |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 44 | N/A |

# SI* (MODERN METRIC) CONVERSION

| FACTORS APPROXIMATE CONVERSIONS TO SI UNITS | | | | |
|---|---|---|---|---|
| **SYMBOL** | **WHEN YOU KNOW** | **MULTIPLY BY** | **TO FIND** | **SYMBOL** |
| **LENGTH** | | | | |
| In. | inches | 25.4 | millimeters | mm |
| ft | feet | 0.305 | meters | m |
| yd | yards | 0.914 | meters | m |
| mi | miles | 1.61 | kilometers | km |
| **AREA** | | | | |
| $in.^2$ | square inches | 645.2 | square millimeters | $mm^2$ |
| $ft^2$ | square feet | 0.093 | square meters | $m^2$ |
| $yd^2$ | square yard | 0.836 | square meters | $m^2$ |
| ac | acres | 0.405 | hectares | ha |
| $mi^2$ | square miles | 2.59 | square kilometers | $km^2$ |
| **VOLUME** | | | | |
| fl oz | fluid ounces | 29.57 | milliliters | mL |
| gal | gallons | 3.785 | liters | L |
| $ft^3$ | cubic feet | 0.028 | cubic meters | $m^3$ |
| $yd^3$ | cubic yards | 0.765 | cubic meters | $m^3$ |
| NOTE: volumes greater than 1,000 L shall be shown in $m^3$ | | | | |
| **MASS** | | | | |
| oz | ounces | 28.35 | grams | g |
| lb | pounds | 0.454 | kilograms | kg |
| T | short tons (2000 lb) | 0.907 | megagrams (or "metric ton") | Mg (or "t") |
| **TEMPERATURE (exact degrees)** | | | | |
| °F | Fahrenheit | 5 (F-32)/9 or (F-32)/1.8 | Celsius | °C |
| **ILLUMINATION** | | | | |
| fc | foot-candles | 10.76 | lux | lx |
| fl | foot-Lamberts | 3.426 | candela/$m^2$ | cd/$m^2$ |
| **FORCE and PRESSURE or STRESS** | | | | |
| lbf | poundforce | 4.45 | newtons | N |
| lbf/$in.^2$ | poundforce per square inch | 6.89 | kilopascals | kPa |

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

## SI* (MODERN METRIC) CONVERSION (continued)

| APPROXIMATE CONVERSIONS TO SI UNITS | | | | |
|---|---|---|---|---|
| SYMBOL | WHEN YOU KNOW | MULTIPLY BY | TO FIND | SYMBOL |
| **LENGTH** | | | | |
| mm | millimeters | 0.039 | inches | in. |
| m | meters | 3.28 | feet | ft |
| m | meters | 1.09 | yards | yd |
| km | kilometers | 0.621 | miles | mi |
| **AREA** | | | | |
| $mm^2$ | square millimeters | 0.0016 | square inches | $in.^2$ |
| $m^2$ | square meters | 10.764 | square feet | $ft^2$ |
| $m^2$ | square meters | 1.195 | square yards | $yd^2$ |
| ha | hectares | 2.47 | acres | ac |
| $km^2$ | square kilometers | 0.386 | square miles | $mi^2$ |
| **VOLUME** | | | | |
| mL | milliliters | 0.034 | fluid ounces | fl oz |
| L | liters | 0.264 | gallons | gal |
| $m^3$ | cubic meters | 35.314 | cubic feet | $ft^3$ |
| $m^3$ | cubic meters | 1.307 | cubic yards | $yd^3$ |
| **MASS** | | | | |
| g | grams | 0.035 | ounces | oz |
| kg | kilograms | 2.202 | pounds | lb |
| Mg (or "t") | megagrams (or "metric ton") | 1.103 | short tons (2000 lb) | T |
| **TEMPERATURE (exact degrees)** | | | | |
| °C | Celsius | 1.8C+32 | Fahrenheit | °F |
| **ILLUMINATION** | | | | |
| lx | lux | 0.0929 | foot-candles | fc |
| $cd/m^2$ | candela/$m^2$ | 0.2919 | foot-lamberts | fl |
| **FORCE and PRESSURE or STRESS** | | | | |
| N | newtons | 0.225 | poundforce | lbf |
| kPa | kilopascals | 0.145 | poundforce per square inch | $lbf/in^2$ |

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

# TABLE OF CONTENTS


**CHAPTER 1. INTRODUCTION ........................................................................... 1**

**REFERENCE OBJECTIVE ........................................................................... 3**

**CHAPTER 2. TRAVEL TIME RELIABILITY APPLICATIONS OVERVIEW........................... 5**

**TRAVEL TIME RELIABILITY INDICES .............................................................. 5**

**APPLICATION TYPES ........................................................................... 7**

Performance Measurement Applications ........................................................................... 7

Performance Monitoring Applications ........................................................................... 8

Simulation Modeling Applications ........................................................................... 8

**CHAPTER 3. APPLICATION FRAMEWORK ........................................................ 11**

**SCOPE OF THE STUDY ........................................................................... 11**

**DATA REQUISITES ........................................................................... 11**

**DESIGN OF EXPERIMENTS ........................................................................... 12**

**ASSESSING THE RESULTS ........................................................................... 12**

**CHAPTER 4. CONCEPTUAL FRAMEWORK AND METHODOLOGY ........................... 13**

**CONCEPTUAL FRAMEWORK ........................................................................... 13**

Raw Data ........................................................................... 14

Trajectory Processor ........................................................................... 16

Link-Level Information Library ........................................................................... 17

Path Travel Time Distribution ........................................................................... 17

**SOLUTION METHODOLOGY ........................................................................... 18**

Numerical Solution Method ........................................................................... 18

Convolution Integrals on Subpaths ........................................................................... 20

**CHAPTER 5. GEOGRAPHIC INFORMATION SYSTEM PLATFORM ........................... 22**

**PURPOSE OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM ................ 22**

**FUNCTION OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM ............... 23**

**IMPLEMENTATION OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM ... 25**

**GENERALIZED USE OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM ........................................................................... 31**

**CHAPTER 6. NEXT STEPS FOR APPLICATION ................................................ 33**

**REFERENCES ........................................................................... 35**

## LIST OF FIGURES

**Figure 1. Diagram. Methodology to estimate path travel time distributions.......................14**
**Figure 2. Diagram. Combinatorial data-mining search approach for finding subpath-level information to use in estimation of path travel time distributions.............21**
**Figure 3. Diagram. Functionality overview of the geographic information system platform. ..........................................................................................................24**
**Figure 4. Diagram. Implementation details for a geographic information system (GIS) Application Platform consisting of a sequential implementation of three parts. .........................................................................................................................25**
**Figure 5. Screenshot. User interface for the geographic information system platform showing a prompt for users to input the network layer and provide information about it. ...........................................................................................27**
**Figure 6. Screenshot. Geographic information system platform user interface prompt for users to specify their input of travel time data................................................27**
**Figure 7. Screenshot. User interface of the geographic information system platform for the selection of the Origin-Destination (O-D) pair. ..........................................28**
**Figure 8. Screenshot. User interface of the geographic information system platform for path selection. ..............................................................................................29**
**Figure 9. Screenshot. User interface of the geographic information system platform prompting users to input departure information. .................................................30**

## LIST OF TABLES

**Table 1. Travel time reliability measures. ................................................................... 6**

# CHAPTER 1. INTRODUCTION

Travelers experience the transportation system along trajectories that comprise multiple facilities, and sometimes multiple modes of travel. User-centric system performance seeks to translate objective measurements of travel time and delay along links to overall time and delay measures of paths connecting two points in the network. Behavior studies have shown the importance of travel time reliability to users' satisfaction with a transportation network and their travel and activity choices (Fosgerau and Karlström 2010). Travel time reliability also has direct economic impacts on the decisions of logistics firms and delivery vehicles, and the service levels that shippers, firms, and consumers ultimately experience.

Underlying all measures of travel time reliability are the variable travel times along individual links, in various modes, at junctions and intermodal terminals, and through the individual trajectories of travelers and goods. Travel behavior literature has asserted that trajectories are the most complete representation of travel experience—particularly project L04 under the second Strategic Highway Research Program (SHRP2) (Mahmassani et al. 2014; Kim and Mahmassani 2015). The literature has similarly asserted pattern trajectories—spatial-temporal trajectories of travel activity, as representative of individual daily activity and travel patterns (Allahviranloo and Recker 2015). According to the literature, a main advantage of trajectories is their flexibility in computing performance measures at varying disaggregation and aggregation levels. Specifically, trajectories effectively provide a unifying concept that consistently ties together actual measurements, traffic model calibration, and network performance analysis (Hou 2015).

With the increasing proliferation of global positioning system (GPS)-enabled smartphones and GPS-connected vehicles, quasi-continuous trajectories have become available to characterize the travel time variability of individual travelers. A quasi-continuous trajectory records movement at a high sampling rate, thus delivering near-continuous representation of the traveler's experience. While full connectivity is not likely in the near future, agencies can still accomplish a great deal even with small samples of trajectories circulating in a multimodal network. Trajectory-derived measurements can be effectively augmented with other sources, especially those based on fixed sensors, which remain the workhorse of the traffic industry (Hou 2015). Such use can recognize that various sources of travel time measurements have distinct characteristics and can be based on sound physics to correctly define performance measures of interest. For example, distinctions between time-based and space-based definitions of average speeds are important when combining information from different sources. Unfortunately, many studies in practice have not used consistent and correct definitions of the basic quantities measured; the resulting travel time distributions end up incorrectly defined, or inconsistent with their intended purpose and interpretations.

A key question that arises from the perspective of a given user (traveler or goods shipping company) pertains to the variability of travel times along specific paths contemplated (E. D. Miller-Hooks and Mahmassani 2003). Related to that is the problem of finding paths that are in some way optimal, or meet certain reliability targets, when the network travel times are stochastic with nonstationary distributions (E. Miller-Hooks and Mahmassani 2003; Nie and Wu 2009; Zockaie, Nie, and Mahmassani 2014).

There are several ways to address the issue of modeling travel time variability along specific paths depending on the available data. If a very large sample of trajectories is available for a given network, the answer lies in constructing distributions of experienced travel times for all users who have traveled along the path of interest, assuming sufficient trajectory observations along the path. Typically, the analyst may be interested in time-dependent distributions (i.e., rather than characterize variability through a single distribution across the entire observation period, they may use conditional distributions for specific time periods). This would, however, entail a far larger number of observations. Furthermore, the intent of this study is to construct these distributions for any user-specified path—including paths over which very few observations may be available. For this purpose, link-level information can be combined to the path level. Normally, link-level distributions can be efficiently extracted from all the trajectories, as each trajectory contributes to the sample of link travel times.

Three challenges can arise in estimating travel time distributions for any path in a given network with variable travel times:

- First is the measurement and observability of the link travel time distributions; namely, whether there are a sufficient number of observations from available trajectories to characterize and estimate the distribution functions of these travel times.
- The second challenge is methodological, and results from the fact that travel times on different links are generally correlated over space and time. This significantly complicates computing path travel time distributions from those of their constituting links since there are no generalized methods for determining the distribution of the sum of correlated random variables.
- The third challenge is inferential: What does the travel time distribution represent, within-day variation or day-to-day variation of travel performance characteristics, and for what class of user is this distribution applicable to? Separation by component enables greater accuracy in prediction, and flexibility in customizing solutions to the wider user community for different purposes.

Most early work on path finding with stochastic travel times assumed that travel times are independent random variables, where variation on one link at a given time is treated independently from variation on adjacent links, or at different times. However, the main factors known to affect travel time variation, such as congestion and weather, tend to act on multiple links simultaneously, and their impact tends to linger long after the event itself may have cleared. This results in varying degrees of correlation in the travel times observed on different links in different time periods. While correlation tends to be stronger along contiguous or adjacent links, it is not limited to these situations (Zockaie, Mahmassani, and Kim 2016). Estimating general correlation patterns can be challenging, because it involves a much larger number of observations than estimating the first or second moments of these distributions. In addition, obtaining the path-level distributions entails convolving the link-level distributions, a process that typically does not have closed form analytical solutions, and for which numerical integration techniques may not always converge. For this reason, researchers have used statistical sampling methods, such as Monte Carlo techniques, for that purpose in a variety of areas (Srinivasan and Mahmassani 2005).

This reference brings together recent advances in data availability and probabilistic modeling techniques to overcome the above challenges, and synthesize travel time distributions at the path level in generalized networks, which may include different facility types, and potentially modal services, along a particular path. This would characterize the user experience of individual travelers or goods shippers as input when evaluating reliability performance of transportation systems and devising approaches to improve this performance.

## REFERENCE OBJECTIVE

This reference intends to provide information to transportation agencies, professionals and researchers on a methodology and tools that could potentially synthesize and replicate traveler-based distributions of time travel along continuous paths regardless of facility type. The methodology discussed in this reference is primarily based on research and testing and the tools have only been developed as prototypes. Through testing at different levels on simulated data, the research team has determined that the framework, processes, and tools discussed in this reference have practical applicability in estimating travel time distributions.

Specifically, the objectives of this reference are to:

- Describe the practical applications of the methods to performance measurement, performance monitoring, and simulation modeling.
- Systematically describe the various steps involved in applying the travel time reliability methodology, including an overview of the associated tools and how they function.
- Provide demonstrated evidence on how transportation agencies, professionals or researchers can apply the methodology.

# CHAPTER 2. TRAVEL TIME RELIABILITY APPLICATIONS OVERVIEW

The key question this project aims to help answer arises from the perspective of a given user and pertains to the reliability or variability of travel times along specific paths in the network contemplated by the user. The goal of the present study is to bring together modern advances in data availability and probabilistic modeling techniques to overcome the challenges associated with estimating travel time distributions for any path in a given network. Using these tools, the goal is to synthesize travel time distributions at the path level in generalized networks—including several facility types along a particular path. The estimated travel time distributions are intended to characterize the experience of individual travelers or goods shippers as input when evaluating the reliability performance of transportation systems and devising approaches to improve this performance.

In applying the estimation methods for reliability performance measurement and monitoring developed in this study, some travel time reliability measures are defined in order to facilitate use of the estimated travel time distributions obtained with the methodology. This chapter presents a synthesis of various available travel time reliability indices and provides general descriptions for three different application types.

## TRAVEL TIME RELIABILITY INDICES

Travel time literature presents a number of reliability performance measures based on travel time reliability; recommendations on their suitability for different purposes are also available. A report for project L04 under the second Strategic Highway Research Program (SHRP2) (Mahmassani et al. 2014) classified travel time measures based on their applicability to different levels of travel time distributions and reliability analysis. The levels of applicability—network level, origin-destination (O-D) level, and path/segment/link level analysis—resulted in categorizing reliability measures into three types: distance-normalized measures (type A), comparable travel time measures (type B), and same travel distance measures (type C). These reliability measures and categorizations are presented in table 1.

For network-level performance measures, travel times experienced by users are not directly comparable since distance traveled by vehicles may be significantly different, and as such measures in which trip travel times are normalized by the trip distance can be used. Each vehicle's travel time can be converted into a distance-normalized travel time, such as travel time per mile (TTPM), and the performance statistics can be extracted from the distribution of TTPMs, as shown in the type A measures. For O-D-level performance measures, travel times experienced by vehicles are comparable, despite potential differences in actual trip distances depending on the user-chosen path. O-D-level travel times and their distributions can be limited to trips between actual traffic analysis zones (TAZ), as is typical in O-D-level analysis or can be between any two points for this application. For such cases, reliability measures that can be used include conventional distribution characteristics, as shown in type B.

**Table 1. Travel time reliability measures.**

| | Network Analysis | Origin-Destination Analysis | Path/Segment/Link Analysis |
|---|---|---|---|
| **Travel times for vehicles** | Not comparable | Comparable | Comparable |
| **Travel distances for vehicles** | Different | Different | Identical |
| **Distance- normalized measures** (type A) | (a) | (a) | (a) |
| **Measures for comparable travel times** (type B) | Not applicable | (b) | (b) |
| **Measures for the same travel distance** (type C) | Not applicable | Not applicable | (c) |

(a) Mean of TTPM (travel time per mile);
Standard deviation (std.dev.) of TTPM;
95th/90th/80th percentile TTPM.

(b) Average travel time
Std.Dev. of travel times
Coefficient of variation *(std.dev. of travel times ÷ mean travel time)*
95th/90th/80th percentile travel time
Buffer index *((95th percentile travel time − mean travel time) ÷ mean travel time)*
Skew index *((90th percentile travel time − median travel time) ÷ (median travel time − 10th percentile travel time))*
Percent on-time arrival *(percent of travel times < 1.1 × median travel time)*

(c) Travel time index *(mean travel time ÷ free- flow travel time)*
Planning time index *(95th percentile travel time ÷ free-flow travel time)*
Misery index *(mean of the highest 5% of travel times ÷ free-flow travel time)*
Frequency of congestion *(percent of travel times > 2 × free-flow travel time)*

At the path/segment/link-level, the travel times for different vehicles are comparable, and their travel distances are identical, which allows for the calculation of unique free-flow travel time for a given path and the use of additional measures that utilize this information. In this category are measures such as the Travel Time Index, Planning Time Index, Misery Index and Frequency of Congestion, shown in Type C. In addition to the application levels ranging from general to specific, the measures types from A to C also range from general to specific, where measures of Type A can be applied at all levels of analysis, measures of Type B can be applied at the two lower levels of analysis (OD-level and Path-level), and Type C measures are most specific and can only apply to the Path/Segment/Link-level analysis.

For the purposes of this study, performance measures are envisioned as primarily applying to path-level travel time distributions. In such cases, considering a single path of choice, the travel distances are identical and travel times comparable, and thus type C measures are applicable and offer the most detailed information, but type A and type B measures can also be applied. In a more general case, if the measures are used for comparing travel time distributions of two or more separate paths for the same given O-D points (i.e., their travel times are comparable despite having different travel distances), the performance relative to one another can be measured using type B measures, but type A measures can also be applied. Finally, for comparison of travel time performance along multiple distinct paths whose travel times are not necessarily comparable, type A measures should be used.

## APPLICATION TYPES

Methods for estimating travel time distributions along continuous user-defined paths have many applications in traffic operations and transportation planning where travel time reliability is an important performance factor, and vehicle trajectory data are available or "particle-based" traffic simulation modeling tools are employed—thus producing individual traveler or vehicle trajectories. Potential applications are generally classified into three categories for the purposes of this project: performance measurement, performance monitoring, and simulation modeling. A description of the possible applications under these categories follows.

### Performance Measurement Applications

Analysts can use the developed methods for estimating path travel time distributions in a range of performance measurement applications.

#### *Policy-Level Applications*

At the policy level, metropolitan planning organizations or agencies responsible for planning the roads in a network may wish to understand the status of the road network at the present time or in the future. Travel time reliability, specifically the descriptive statistics derived from travel time distributions, indicates one aspect of the operating conditions on a road network. Compared to other performance metrics, such as network travel times or average route travel times that do not convey the variability and stochastic aspects of those characteristics, travel time reliability measures specifically focus on the expected variability rather than the simple summary statistics. The methods developed as part of this project enable transportation agencies to estimate travel time distribution along any path in the network, including less frequently traveled paths. Travel time reliability metrics can be applied to better describe the overall status of the road network.

Analysts can conduct tests of the network on various sources of variability, such as recurring congestion due to inadequate base capacity or exogenous factors, such as weather, incidents, and other disruptive events.

#### *Online and Real-Time Applications*

Online and real-time applications for estimating travel time distributions can provide more detailed traveler information, specifically when comparing the relative reliability of alternative routes in a network. Unknown variability in travel times may be a factor that a large portion of users are unable to readily incorporate into their trip planning. Knowing the variability of travel times on the available routes could be desirable, and may even be crucial, in users' decision making. Thus, travel time reliability information in near real time would increase the value of the information supplied to travelers.

## Performance Monitoring Applications

Analysts can use travel time variability information on specific paths, between O-D pairs or on a network level, for performance measurement. There are also various performance monitoring applications.

### *Policy-Level Applications*

Transportation agencies and practitioners may be interested in measuring the effects of a proposed plan or improvement option relative to a base case, or comparing multiple options as part of an evaluation process to identify a preferred solution. Such metrics may include levels of service, travel speeds, travel time, or operating cost savings. The majority of these measures are based on the network's operation under optimal conditions where travel reliability is not affected. For a more comprehensive evaluation, analysts may introduce travel time variability as an additional aspect of measuring performance.

### *Project-Level Applications*

Another performance monitoring application is testing networks under planned-event conditions, such as construction zones, festivals, and sporting events. Planned events often lead to changes in the base daily travel demand and/or variations in travel patterns, and may be accompanied by interventions such as lane or approach closures. Once analysts have collected data for use during planned events, they can apply the developed methods for estimating travel time distributions; then, they can use appropriate travel time reliability measures and indices to assess the performance of the network under the planned-event conditions. This type of performance monitoring requires a characterization and understanding of the base network operating conditions. Therefore, performance monitoring under planned events should be done relative to the network performance in the base case. After analysts specify the framework for such performance monitoring, and select corresponding travel time reliability performance measures based on characteristics of the specific application, they can apply the developed methods as solutions to the intermediate step of estimating path travel time distributions along any path in the network.

## Simulation Modeling Applications

Simulation modeling applications are related and can further extend performance measurement or performance monitoring applications. This section outlines five simulation modeling applications, based on the performance measurement and monitoring applications identified above.

### *Model Calibration and Validation*

The calibration and validation of simulation models is one of the most important applications for this study. Specifically, analysts can use user-centric network reliability measures to model user responses based on network conditions. The travel times actually experienced by users on a given trip may be different from estimates based on physical measures of average travel time and reliability. The user perspective is inherently an entire-trip perspective, meaning that the reliability measures for travel models and network simulation tools have to be synthesized at the O-D route level (Mahmassani et al. 2014). Therefore, the calibration and validation of such models are a direct application of the estimation models developed in this project for synthesizing O-D route-level travel time distributions. Analysts can then directly determine travel-time reliability measures from these distributions.

Mahmassani et al. (2014) incorporated reliability into demand and network simulation models both implicitly (by treating travel time as a random variable with its distribution described analytically or based on historical data), and explicitly (by directly simulating travel time distribution through multiple model runs). They incorporated reliability measures for travel time distribution estimation into the demand side (for example, an activity-based model) with respect to trip mode choice, destination choice, and departure time choice. Additionally, accessibility measures that account for reliability can impact upper-level choice models of car ownership and activity-travel patterns. On the other hand, an operational network simulation model can incorporate travel reliability, where route choice incorporates the reliability measure of interest, and the demand model imports this measure at the O-D or path level.

Model developers can implement this by replacing a generic travel time variable in route choice with a path-level travel time reliability measure computed from the path travel time distribution.

### *Policy Measures Performance*

Analysts can test policy levers, such as demand management, road pricing strategies, and improved traffic management, in a simulation modeling environment which allows for producing scenarios that show variation in traffic conditions. The resulting data enable application of the travel time distribution estimating approaches to testing network performance, in terms of travel time reliability, under the effects of different policy measures.

### *Advanced Traffic Management Systems*

Current advanced traffic management systems (ATMS) incorporate near-real-time predictive modeling to assess various traffic management strategies prior to implementing them. Through predictive modeling that employs “particle-based” simulation models, analysts could add a travel time reliability measure to the modeling process. In such an application, the operations models recognize the dynamic and probabilistic nature of traffic flow, compute travel time reliability using the methods developed for estimating travel time distributions, and disseminate this information to users through the traveler information system. Thus, users have access to travel time information, reliability of travel time, optimal paths in terms of reliability, and least cost paths. Dong (2009) demonstrated such a framework.

### *Traffic Control Measures Performance Monitoring*

Analysts can also perform performance monitoring of travel time reliability for a variety of traffic control measures as part of a simulation modeling application. They can design tests to examine travel time reliability over a particular area of the road network or within the corridor being modified. From multiple simulation runs, an appropriate travel time reliability index could illustrate the level of travel time reliability for each traffic control scheme, which analysts can then use to compare, evaluate, and select options.

### *Planned-Event Performance Monitoring*

Analysts can assess performance monitoring in a specific network or portion of the network in terms of travel time reliability in a simulation setting. They can use simulation models that allow changes from the supply side, in terms of restricting capacity or eliminating certain links or approaches, to simulate a planned event with lane or link closure measures. Similarly, analysts can use simulation models that allow changes from the demand side, in terms of increasing demand for specific O-D pairs or traffic flows along specific paths or links, to simulate a planned event expected to produce significant demand shifts. Using the developed estimation methods, they can estimate travel time distribution for specific paths or O-D pairs from the simulation data and results, and then apply it to some of the desired indices to assess changes in travel time reliability compared to a base case data set or simulation.

# CHAPTER 3. APPLICATION FRAMEWORK

The methods developed in this project for estimating travel time distributions for user-defined paths provide a range of approaches for obtaining travel time distribution estimates along specific paths, even when the paths are not frequently traveled. This chapter outlines the overall steps for implementing a framework to use these methods and provides a brief discussion of the general approaches to performing each step. This chapter addresses the following steps: scope of the study, data requirements, design of experiments, and assessing the results.

## SCOPE OF THE STUDY

Scoping the study involves defining the study problem or objective, the spatial and temporal scope

of the analysis, and the factors for which travel time reliability are to be tested.

The spatial scope of the analysis may include selecting a specific network, but more specifically may include selecting a specific set of paths for which travel time distributions are studied.

Analysts can define the temporal scope such that analysis focuses on a weekday, weekend, or possibly a peak period special event.

Additionally, the scope of the study is directly related to the definition of analysis scenarios, i.e., the factors that will be considered in the analysis. These may include specific exogenous factors, demand levels, or agency-implemented measures and controls. At this stage, the analyst will define the base case, and identify specific scenarios to assess travel time distributions relative to the base case.

## DATA REQUISITES

Acquisition of relevant data is a fundamental step that defines the ability to properly assess the reliability impacts associated with the network, corridor, segment, etc. Depending upon the problem at hand, specific data should be collected to create the various testing scenarios. Analysts could acquire data related to the various exogenous factors affecting travel time reliability to populate the scenario manager, again depending upon the problem to be analyzed. This additional data could include road closure information, collision data, weather data, special event data, etc. Analysts could collect the data corresponding to the spatial and temporal limits defined earlier.

Analysts could consider the following types of data sources towards satisfying the data requisites: 1) fixed sensor, 2) floating car data, and 3) GPS information from equipped vehicles. Fixed sensors are limited in the coverage and the information they can provide. In general, they provide occupancy and speed data, which would not satisfy the data requisites for applying these methods. Floating car data (FCD), provide information such as speed, location, direction of travel from active mobile phones in the vehicles. This provides a better coverage and more information as compared to fixed sensors and provides direct information on travel times—which are of main concern in this study. To further enhance the quality of the information, analysts can use data from the vehicles equipped with GPS. Therefore, the data for this study would not only depend on the spatial and temporal scope defined for the analysis and the analysis scenarios to be considered, but should also include vehicle trajectory data, thus providing complete information on travel times for a given vehicle's trips.

At this stage, analysts clean and process the data. As described in the conceptual framework, cleaning of trajectory data helps to prevent contamination with noise and erroneous measurements. Analysts can process the clean, reformatted trajectory data to create a library containing travel time distributions for individual links. They can then map vehicle trajectories, typically given in latitude and longitude, onto the corresponding network at the link level. This ensures that during analysis, one has access to travel times for the entire path in the trajectory data, but is also free to use the data for any subpath that can be extracted from the given paths, down to the link level.

## DESIGN OF EXPERIMENTS

Based on the project's goals and objectives, the defined temporal and spatial constraints, and the analysis scenarios, the analyst designs a set of experiments. This involves selecting paths to be tested, the number of analysis scenarios available for each path, and the desired information for each defined case. At this stage, the analyst tests the objective or problem of the study on a diverse set of paths and defines the analysis scenarios to avoid dependencies. They can then use the available data, along with the developed methods for estimating travel time distributions, for all paths and scenarios in the designed experiments.

Once estimated travel time distributions have been obtained, they can be used to obtain any relevant travel time reliability indices based on the objective of the study, and according to the type of problem, as shown previously in table 1.

## ASSESSING THE RESULTS

Assessing the results will depend on the objectives of the specific study. Analysts can assess results from a system or traveler standpoint.

From a system standpoint, transportation management agencies should measure reliability performance levels of given transportation systems at a network level, in a subarea, or for a specific corridor. These agencies can use the results (i.e., the overall travel time distribution and associated reliability measures) to answer the following questions: How disperse are travel times on this system? What proportion of travelers experience serious congestion along this road? How variable and uncertain is travel time on a given road compared to another road?

From a traveler standpoint, information on the reliability or variability of travel times that would be experienced, based on selected path and departure time period, could be crucial. Agencies can deliver this information using traveler information systems or variable message signs. The agency can obtain the travel time distribution for a particular departure time interval to assess the probability that a particular traveler departing at this interval experiences a specified level of congestion. Another important user-level measure is the experienced schedule delay, which is generally the difference between actual and desired arrival times for that individual.

# CHAPTER 4. CONCEPTUAL FRAMEWORK AND METHODOLOGY

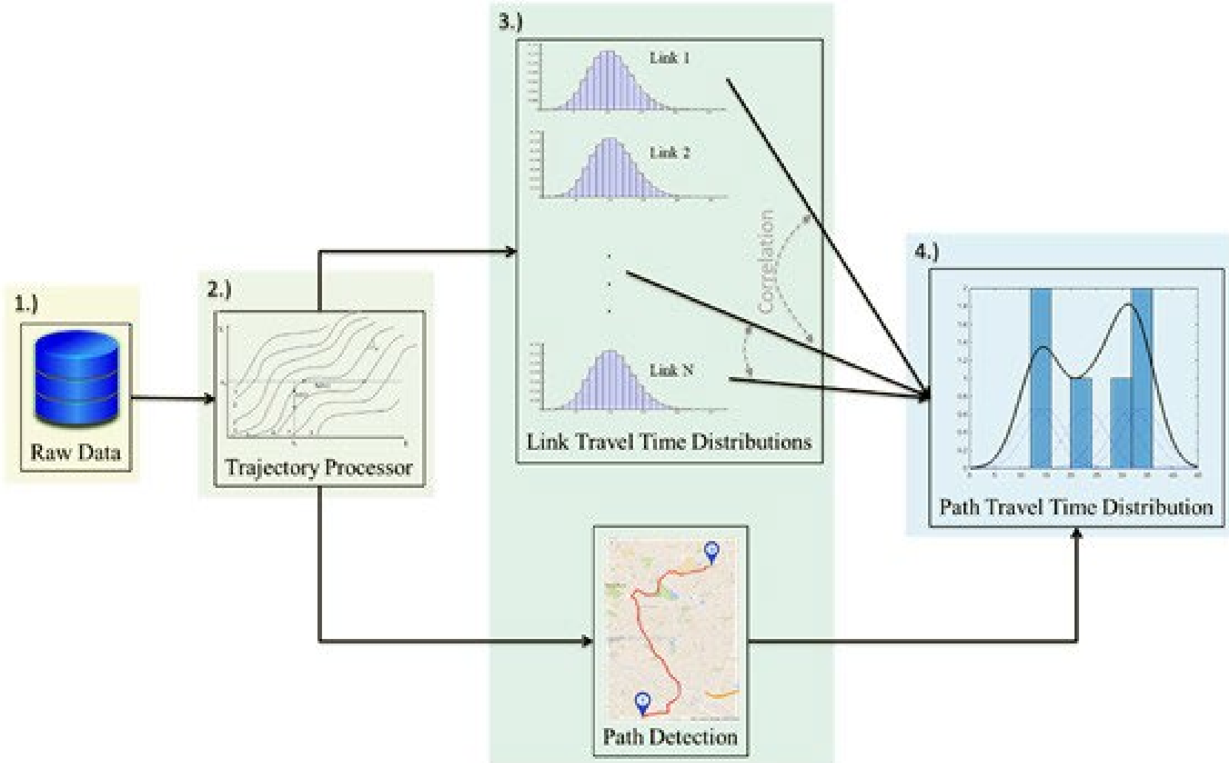
In the interest of addressing the aforementioned subjects, the study aims to: 1) use vehicle trajectory data to identify the paths traversed by drivers, recognizing that these paths are continuous but may comprise of different facility types, and 2) use the same data set to develop methods to successfully synthesize and replicate traveler-based distributions of travel time along the identified paths. The methods formulate and develop the path-level distributions as convolutions of link-level travel times that follow general distributions (i.e., recognizing spatial-temporal correlations).

## CONCEPTUAL FRAMEWORK

Figure 1 presents the steps involved in such a study. The sequential framework generally consists of four modules: raw data, trajectory processor, link-level information library, and path travel time distribution.

a) Raw data: This module retrieves real data from field observations that are frequently contaminated with noise, and erroneous measurements. In data-driven studies, a preliminary step is to clean and scrub the data to make them available for investigation. After cleaning, data are reformatted per the analyst's desire.
b) Trajectory processor: This module processes cleaned and reformatted trajectories that are processed to create a library containing travel time distributions of individual links. In this step, the analyst identifies the path (i.e., set of consecutive links) between a given origin-destination pair.
c) Link-level information library: For the identified paths and with the library, this module subsets links comprising the paths from the library, and identifies correlation relationships among the paths.
d) Path travel time distribution: Finally, with the information retrieved in the previous steps, this module solves the convolving integral with correlated random variables to determine the underlying travel time distribution.

This chapter describes each step in the conceptual framework, focusing on step D for estimating path travel time distributions.




Source: Federal Highway Administration

**Figure 1. Diagram. Methodology to estimate path travel time distributions.**

## Raw Data

### *Data Types and Sources*

Popular data sources found for this study include 1) fixed sensor, 2) floating car data, and 3) GPS information from equipped vehicles. Fixed sensors are generally limited in the coverage and the information they can provide. In general, they provide occupancy and speed. This leaves the analyst to estimate the flow and density. Floating car data (FCD), provide information such as speed, location, and direction of travel from active mobile phones in the vehicles. FCD generally provide a better coverage and more information compared to fixed sensors, while eliminating additional hardware installation or any approximations. Finally, data from vehicles equipped with GPS provide similar information as FCD, though typically with better quality of the collected information.

The proposed conceptual framework for travel-time distribution estimation along user-specified paths uses high-resolution vehicle trajectory data. This type of travel time data had not been available until recent technological developments and may almost exclusively be provided by floating car data (FCD) from probe vehicles or GPS information from equipped vehicles.

Because the proposed solution methodology is based on travel times reported (and/or estimated) per vehicle trajectory, the travel time data to support application of this methodology should satisfy the following:

- Report travel times by vehicle trip on a trajectory basis, which provides *x-y*
- coordinates and time stamp at each reported location.
- Report vehicle information (i.e., its position, at least) with an observation every 20–60 seconds.
- Represent sufficient sampling and time-series to allow statistically meaningful analysis.

The spatial and temporal span of the data should be application specific. Spatially, the application may specify that the data covers the network of one or multiple cities, in one or more countries. Temporally, applications should typically use data over multiple days, but sometimes they may cover all weather seasons, or a precise period of the year. The specifics of these requisites are based on the application and the question that it aims to answer or address.

Trajectory data should also ideally possess the following general characteristics for travel time reliability analysis:

- Capture both types of congestion (recurring and non-recurring).
- Cover the range of road facilities that may be included in the subject area analysis, from freeways to arterial roads and (possibly) managed lanes.
- Allow statistically meaningful analysis of data through availability for a relatively long period of time (e.g., long enough to cover seasonal variation).
- Provide travel time at disaggregated levels (e.g., vehicle travel time) and at fine time intervals in addition to average travel times to capture time-of-day variation and vehicle-to-vehicle variation.

For certain applications, the data should provide sufficient information on components, causes, and other characteristics of congestion so that appropriate parameterization can be established for testing.

The emergence of probe data over the past few years has opened the opportunity to capture all information for this type of analysis, since such data can be available all the time for all major roads in the network, including major arterials. Probe-based trajectory data represent a significant increase in quality and quantity of relevant information. The detail in such data makes it possible to analyze travel time data according to network and route components (e.g., on link and path bases) and geographic aggregations (e.g., on an O-D zone basis).

### *Data Cleaning and Pre-processing*

The preliminary step is to clean and pre-process data to ensure they are readily available for use. In this case, given that the analyst would be working with vehicle trajectory data, the steps for cleaning and pre-processing data would, in part, depend on the data's format and information. In general, cleaning data focuses on finding and resolving missing and/or incorrect information, while pre-processing focuses on transforming data from the current format to a format appropriate for analysis.

In this study, raw data should be cleaned from observations where the position information for a vehicle may be missing or incorrect for some observations. Missing observations are easy to detect because they are often reported as a blank field, none value, or fixed number, such as 0 or 100, for all missing observations. Cleaning can be done by simply removing those observations, or applying an appropriate methodology for interpolation. Detecting incorrect or erroneous observations in the data is harder and involves checking that all-time steps and positions are consecutive and physically possible in terms of a vehicle's maximum possible speed of travel.

In this study, the research team pre-processed data by mapping the position (*x*-*y* coordinate) information provided in the trajectory data onto the appropriate links in the network. Two main things to consider are that position observations have a margin of error that should be accounted for when mapping on the network, and the network structure restricts vehicle movement information, or at least information useful for the study.

## Trajectory Processor

The vehicle trajectory processor further processes the cleaned and reformatted trajectory data. At this step, the processor maps the trajectory for each vehicle onto the network to find the specific path (i.e., set of consecutive links) the vehicle traversed to travel between its given O-D pair. The team processed the data to determine the a) link- and b) route-level travel time distributions.

a) Link-level travel time distribution:

$$w_{\phi_i}^{l} = \frac{overlap\ of\ \phi_i\ with\ l}{length\ of\ l}$$

The team retrieved link-level information from the trajectories. They counted all trajectories that significantly overlapped with the link in consideration toward the travel time distribution of the link. To begin the analysis at this level, the team divided the whole urban network studied into pre-defined links, and then mapped the trajectories onto the links. Not all the trajectories crossed the entire link in consideration. In the case when only a part of the trajectory overlapped the link, they assigned a weight (<1). The weight $\phi_i$, of trajectory, toward the travel time distribution of link *l* is calculated as:

Hence, the team determined a weighted travel time rate distribution for each link. The trajectories with weight greater than 0.5 were taken into consideration.

The developed travel time distribution forms the library of the of link-level time distribution as presented in figure 1. The library is referred to retrieve link-level travel times to solve the convoluting integral of correlated random variables.

b) Route-level travel time distribution:

A route is generally understood as a set of consecutive links. Information at this level of analysis forms the ground truth. The team determined travel time distribution of a route under consideration through the trajectories overlapping with the entire route.

## Link-Level Information Library

For the identified paths and with the library, the conceptual framework subset links comprise the paths from the library and identify correlation relationships among the link travel times. A link-level information library allows us to have the data mapped onto a network, while remaining at the highest level of discretization possible in the network. This is important to be able to apply and test a variety of different solution methods and to make the most of the available data.

Specifically, the link-level information library allows for easy retrieval of the entire trip information for a given vehicle and thus its total path travel time for its origin-destination pair, while also allowing use of the travel-time information for any subset of that vehicle's path down to the link level. The data comprising the link-level information library can include:

- Trajectory ID.
  - A unique identifier for each new trip or trajectory. This identifier can also be referred to as a probe or vehicle identifier and can be used as such provided that it is unique for each trip even if performed by the same probe or vehicle.
- Link ID.
  - A unique link identifier, for each link traversed in a given trip (trajectory ID). The link identifier should correspond to an available network dataset that should contain at least the forward or backward star information for the network.
- Travel time (for the specified trajectory, on the specified link).

Though not necessary, the framework can include the following additional information in the link-level information library to allow for using other features of the estimation approaches:

- Start time.
  - A specific start time, for each trajectory ID and at each link traversed in its trajectory.
- Date or day of the week.
- Weather information.
  - Weather information consisting of one or more of the following: precipitation level (inches), snow level (inches) and visibility (mi) each trajectory ID and each link traversed in its trajectory.

## Path Travel Time Distribution

Finally, with the information retrieved in the previous steps, this step focuses on using the solution methodology to solve the convolving integral with correlated random variables to determine the underlying travel time distribution for any user-specified path in the network. The next section presents specific details of this solution methodology.

# SOLUTION METHODOLOGY

The key question this project aims to answer arises from a given user's perspective (traveler or goods shipping company) about the reliability or variability of travel times along specific paths in the network contemplated by the user. There are several ways to answer this question, depending on available information for the time period of interest. If a very large sample of trajectories is available for a given network, the analyst can obtain an answer by constructing distributions of experienced travel times for all users who have travelled along the path of interest, assuming a sufficient number of trajectories traverse the particular path. Typically, one would be interested in time-dependent distributions, i.e., conditional distributions for given time period, which would in turn entail a far larger number of observations. Furthermore, since the intent is to be able to construct distributions for any user-specified path—including paths over which very few observations may be available, information should be used from smaller segments of the path of interest that may be frequently traveled even when the complete path is not.

The fundamental problem is to formulate and solve a convoluting integral where random variables are correlated. Moreover, the random variables may have distribution types that are different from each other. In this section, an integral that includes these aspects is formulated first.

A path, $P$, consists of $N$ consecutive sections $\in \{l_1, l_2, l_3, \dots, l_N\}$, where sections of the path $P$ can be subpaths of varying lengths, and a special case being where the path is broken down to each individual link. Let $\theta_i$ denote the cumulative distribution function of the travel time on link $l_i$, where consequently $\theta_i(\tau)$ is the cumulative probability of a travel time τ occurring on link $l_i$. The methodology obtains cumulative distribution of travel time P, $\theta_P(T)$, from solving the integrals presented in the equations below. In the simplest case, the methodology assumes link travel times to be independent, then introduces time-dependence, and then further refines the model by introducing nonstationary distributions and operational conditions.

## Numerical Solution Method

### *Convoluting Integral With Independent Random Variables*

The methodology solves convolution integral with an independent random variable using the method of induction. That is, the methodology solves the integral with two random variables at the time where the result from the previous step forms a new random variable for the next step, as shown in equation 1.

$$\Theta_P(\mathrm{T}) = \int_0^\infty \dots \left( \int_0^\infty \left( \int_0^\infty \theta_1(\tau_1) \,.\, \theta_2(\tau_2 - \tau_1) d(\tau_1) \right) . \, \theta_2(\tau_3 - \tau_1 - \tau_2)\, d(\tau_2) \right) \dots \qquad (1)$$

$$\theta_N \left( T - \sum_{j=1}^{N-1} \tau_j \right) d(\tau_{N-1})$$

At this step, the methodology assumes link travel times to be independent, i.e., uncorrelated. This step is based on the following iterative procedure:

- Pick one realized travel time randomly for each of the links. Construct a set of link travel times.
- Sum picked link travel times together
- Reiterate this process multiple times to obtain a set of path travel time estimates.

***Convoluting Integral With Dependent Random Variables***

When the correlation is assumed, then the methodology obtains distribution of travel time on the path $P$, $\Theta_P^o(\mathrm{T})$, from solving the following integral, as shown in equation 2:

$$\Theta_P^o(\mathrm{T}) = \int_0^\infty \int_0^\infty \dots \int_0^\infty \theta_1(\tau_1).\,\theta_2(\tau_2-\tau_1).\,\theta_3(\tau_3-\tau_1-\tau_2)\,\dots\theta_N\left(T - \sum_{j=1}^{N-1}\tau_j\right) d(\tau_{N-1})\dots d(\tau_2)d(\tau_1) \quad (2)$$

In the formulation, the order $o \in \mathbb{Z}^{\geq}$ is the maximum number of upstream sections that are considered to be correlated. Accordingly, $\theta_i$ the methodology assumes to be correlated with the previous $n$ = min$\{i, o\}$ sections. In equation 2, probability distribution type of $\theta_i$'s is not restricted to any specific form. Furthermore, sections within a path can have multiple distribution types of travel time.

There are vehicles that do not travel the entire path, but only a portion. Before they turn away from the path, these vehicles may slow down and affect the continuing vehicles. Accounting for these vehicles in the same way as the continuing vehicles will give an incorrect estimate of path travel time. To address this issue, the methodology gives more weight to vehicles continuing to the next section on the path from the current section to compute section travel time distribution while solving the integral in equation 2.

The outcome of the integral is a distribution of travel time that will then be fitted into an analytical form. The first attempt is to fit a standard distribution type, like log-normal distribution. Another approach is to estimate a Fourier series. A decision criterion other than root mean square error (RMSE) should be used because such estimation techniques are inaccurate around the mode values of the distributions.

The integral presented in equation 2 is stationary; to add a temporal dimension to it the methodology modifies the integral to take the following form, as shown in equation 3:

$$\Theta_P^{o,t'}(\mathrm{T}) = \int_0^\infty \int_0^\infty \dots \int_0^\infty \theta_1^{t'}(\tau_1).\,\theta_2^{t'}(\tau_2 - \tau_1).\,\theta_3^{t'}(\tau_3 - \tau_1 - \tau_2) \dots \theta_{N-1}^{t'}\left(\tau_{N-1} - \sum_{j=1}^{N-2} \tau_j\right).\,\theta_N^{t'}\left(T - \sum_{j=1}^{N-1} \tau_j\right) d(\tau_{N-1}) \dots d(\tau_2)\, d(\tau_1) \quad (3)$$

In equation 3, $t'$ is prevailing time interval, which simply can be four time periods: 1) early morning, 2) midday, 3) evening, and 4) late evening to midnight. $t'$ can also be finer time interval bins of 1 hour. Former time intervals save on computational efforts and the later gives a precise estimation. However, for the later time intervals, there might not be enough data for estimation. Furthermore, there is an effect from operational conditions (OC). Accordingly, the methodology also adds weather or seasonal attributes to equation 3, to represent different OCs.

In equation 4, $\zeta$ is an OC indicator. The analyst can divide into many categories, but this will increase computational efforts. If the analyst uses too few categories, this can reduce the efficiency of the method. Hence, there is a tradeoff. The methodology can find an optimal balance through conducting a sensitivity analysis.

$$\Theta_P^{o,\zeta,t'}(\mathrm{T}) = \int_0^\infty \int_0^\infty \dots \int_0^\infty \theta_1^{\zeta,t'}(\tau_1).\,\theta_2^{\zeta,t'}(\tau_2 - \tau_1).\,\theta_3^{\zeta,t'}(\tau_3 - \tau_1 - \tau_2) \dots \theta_{N-1}^{\zeta,t'}\left(\tau_{N-1} - \sum_{j=1}^{N-2} \tau_j\right).\,\theta_N^{\zeta,t'}\left(T - \sum_{j=1}^{N-1} \tau_j\right) d(\tau_{N-1}) \dots d(\tau_2)\, d(\tau_1) \quad \mathbf{(4)}$$

Here, the methodology assumes link travel times to be time dependent. The steps involved in this process are as follows:

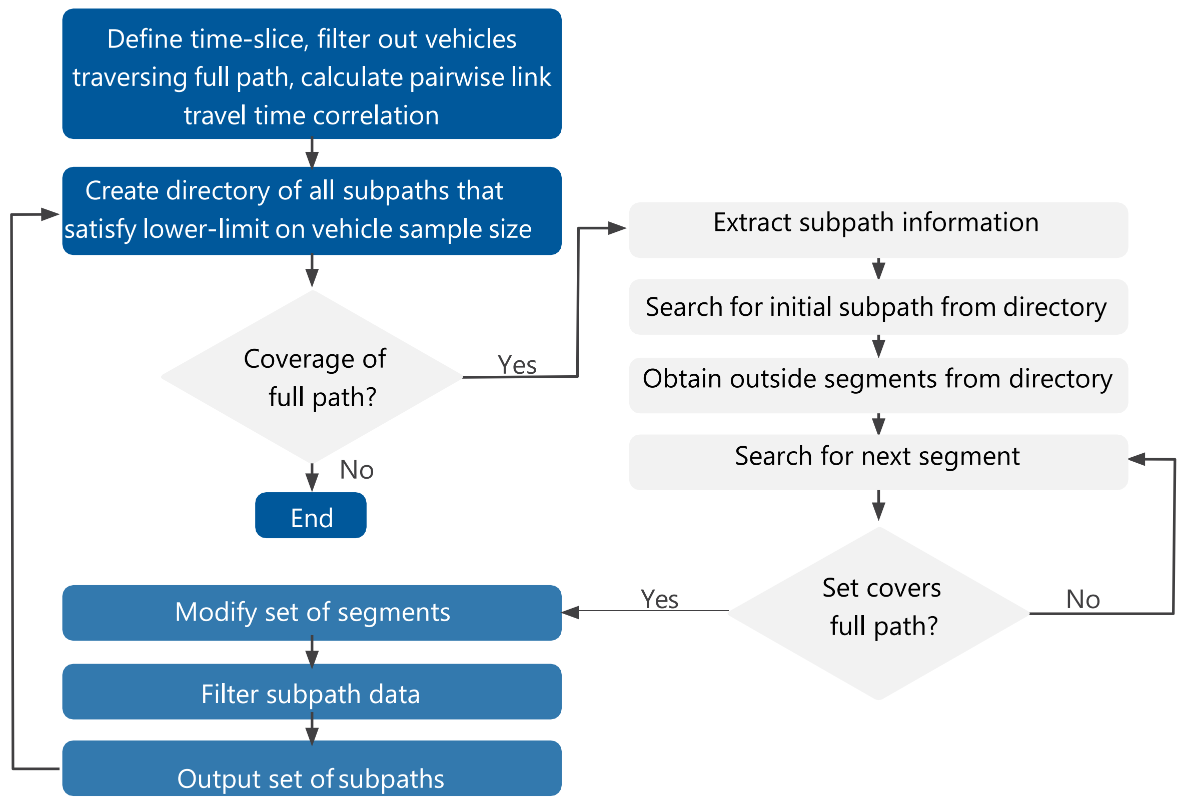
- Bin whole time duration in uniform t-minute intervals.
- Randomly choose a time bin.
- Pick a realized link travel time for each of the constituting links from the chosen time bin until the time adds up to the size of the bin, then pick travel times from the next time bin, until the end of the path, to form a vector.
- Sum elements of the vector to get a path travel time estimate for the time bin.
- Repeat the process multiple times.

## Convolution Integrals on Subpaths

As mentioned above, the analyst can apply each of the approaches described when using travel time distributions on known subsections of the path to form the path travel time distribution. The simplest and most general case is when these subsections are the links of which the path is composed. However, in some cases it may be useful to consider using subpaths of the path in question, instead of links. This simplifies the convolution process and can capture some of the correlations between link travel time within known subpath travel time distributions.

In order to apply these methods to the convolution of subpath travel time distributions, the analyst should apply an approach to obtain an appropriate set of nonoverlapping subpaths that covers the full path. The criteria for forming these subpaths should focus on:

- Sufficient sample size (number of vehicle traversals) for each subpath.
- High correlation between link travel times within the subpath.
- Low correlation of the subpath with outside links and subpaths.
- Consistency between the time periods covered in the sample for each subpath.




Source: Federal Highway Administration

**Figure 2. Diagram. Combinatorial data-mining search approach for finding subpath-level information to use in estimation of path travel time distributions.**

The approach for finding the subpaths to use for the convolution is shown in figure 2. This recursive approach produces multiple groups of subpaths, where each of the groups are nonoverlapping subpaths that cover the entire path. Therefore, the methodology produces an estimation of the path travel time distribution from each group or covering of the path and combines these to give the complete path travel time distribution.

# CHAPTER 5. GEOGRAPHIC INFORMATION SYSTEM PLATFORM

The final product of this project is an ArcGIS-based platform plug-in that provides a graphical user interface (GUI) through which the analyst can use the developed solution methodologies for the intended application in this project, namely the estimation of travel time distributions along user-defined travel paths.

This platform is intended to perform the following general functions:

- Use available historical data and direct input from the user via the GUI and mapping capabilities of the geographic information system (GIS) software.
- Apply the solution methodology to solve a convoluting integral with spatial-temporal characteristics of correlated random variables.
- Report the estimated travel time distribution and associated reliability measures back to the user via the GUI and map representation of the GIS software.

The remainder of this chapter provides a more in-depth description of the motivation for developing a GIS platform, and discusses its purpose, the intended functionality, an overview of the platform implementation, and generalized use of the platform.

## PURPOSE OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM

This report demonstrates a unified approach with broad applicability to a variety of planning and operations analysis problems that enables agencies, transportation professionals, and individual users to incorporate travel time variability and reliability in decisionmaking and evaluation. The conceptual framework and solution methodology detailed in chapter 4 are completely independent of specific software or analysis tool requirements, enabling wide adoption.

A GIS-based platform allows for easy prototyping of key concepts in this project, and enables easy delivery, application, testing, and evaluation of developed methods. The platform provides a GUI and GIS-based mapping functions to allow users to specify paths along which reliability measures are desired, as well as support path finding algorithms that go beyond conventional skimming to consider reliability and risk consideration. The tool is entirely based on the methodology developed for this project, outlined in chapter 4, and enables user-friendly application by removing from users the difficult task of developing the methodology. As described in chapter 4, the tool is intended to support conditional distributions in the path-level numerical integration methods, enabling estimation of path travel time distributions by time of day and by origin-destination (O-D) pairs.

The purpose of the GIS platform is twofold:

- Provide a prototype of the developed solution methodology for application, testing, and evaluation within the scope of this project.
- Allow for delivery, application, testing, and evaluation of the solution methodology, using

a GUI and map representation, by agencies, professionals, and other users for application purposes outside the scope of this project.

Having discussed the intended purpose of the GIS platform, the next section describes the platform functionality to satisfy the intended purpose, outlines the platform implementation to achieve that functionality, and describes the generalized use of the platform for purposes of concern in this reference.

## FUNCTION OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM

To satisfy the purposes previously described, the function of the GIS platform is to provide a GUI through which users can input their data; input their desired trip and path information using a map; and receive travel time distribution, variability, and reliability information. The ability for users to input their desired network and travel time data enables them to use the platform for their specific applications; inputting their desired trip and path information enables them to obtain more specific information or perform a more in-depth analysis, where appropriate.

The main functionality of the tool is broken down into three steps:

**Step 1:** O-D

- Users input their desired network of analysis and the O-D pair.
- The platform performs checks of the network data and desired O-D pair.
- The platform maps O-D pair points to the closest nodes on the network and provides back to users as output.

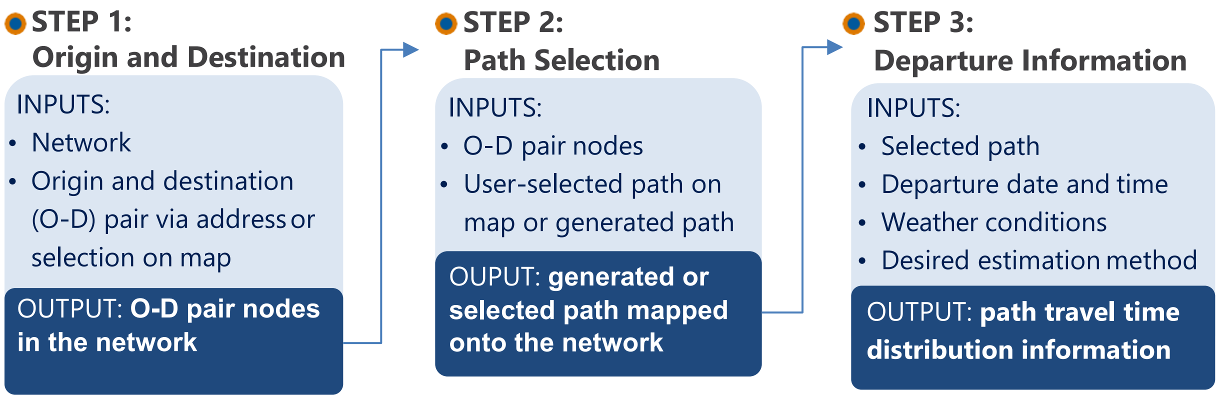
**Step 2:** Path selection

- The output of step 1 is taken as input.
- Users select a route on the map or ask for a path to be generated by the platform.
- The platform performs checks of the provided path.
- The platform maps the user-input path onto the links of the network
- and provides back to the user as output.

**Step 3:** Departure information

- The outputs of step 1 and step 2 are taken as input.
- Users input departure date and time, provide weather conditions (or generate them by the platform), and select the desired estimation method and output.
- The platform performs checks of the inputs.
- The platform calls estimation functions, estimates path travel time distribution, computes desired travel time reliability indices or variability measures, and reports these as output to the user.

Figure 3 shows the three steps that comprise the GIS platform's functionalities and their relationship to one another.

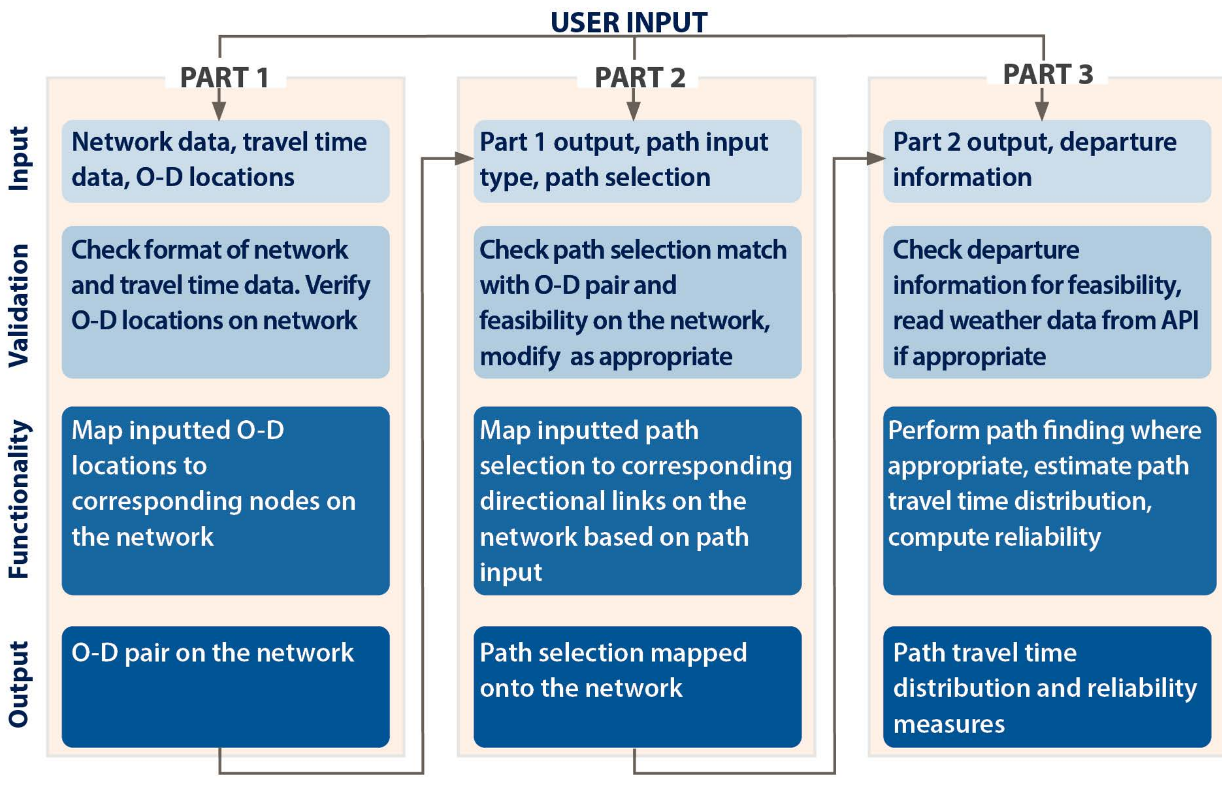



Source: Federal Highway Administration

**Figure 3. Diagram. Functionality overview of the geographic information system platform.**

## IMPLEMENTATION OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM

The three major steps outlined in the previous section provide a high-level overview of the GIS Application Platform functionality. This section describes the implementation of this functionality via a more detailed description of the specific steps. Figure 4 graphically represents the specific details behind these implementation steps.



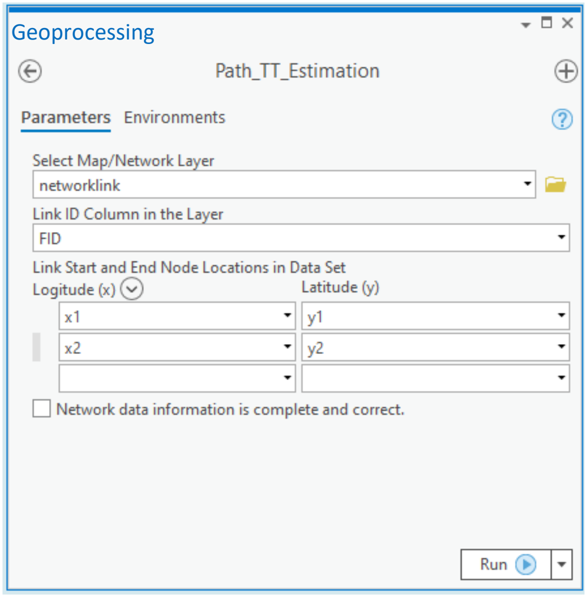

Source: Federal Highway Administration

**Figure 4. Diagram. Implementation details for a geographic information system (GIS) Application Platform consisting of a sequential implementation of three parts.**

The following section describes the tool's functionality in more detail. Figure 5 to figure 9 show representations of the user interface.

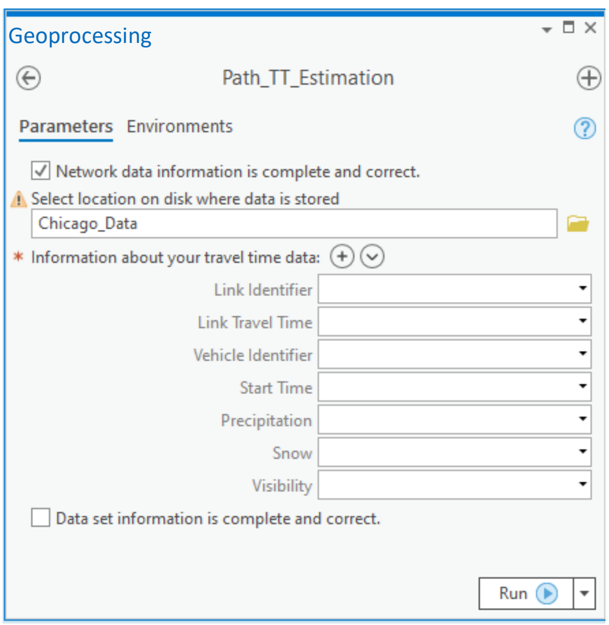
**Step 1**: O-D

- User defines the network and analysis area.
- Display the network and map of the analysis area.
- Perform checks to ensure provided network data can be used for the desired analysis, at minimum providing:
  - Link identification.
  - Start and end nodes for each link.
  - Node longitude and latitude (i.e., *x-y* location).
- Take user input for travel time data (ideally from vehicle trajectory data).
- Perform checks to ensure provided travel time data can be used for the desired analysis.
  - Check that data are in the form of a link-level travel time information library.
- Take user input for a desired trip O-D pair through address input or interactive selection on the map.
- Perform checks to ensure the O-D pair input corresponds to the user-provided network.
  - Check that O-D pair locations are within the network area.
- Map user input of O-D pair to the corresponding O-D nodes on the provided network, and output back to the user.



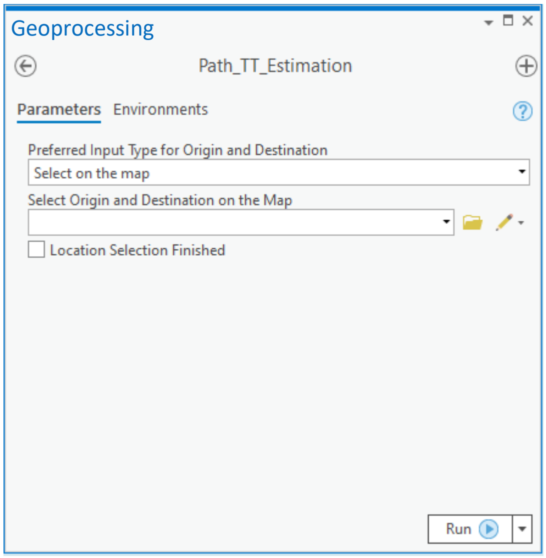

Source: Federal Highway Administration

**Figure 5. Screenshot. User interface for the geographic information system platform showing a prompt for users to input the network layer and provide information about it.**

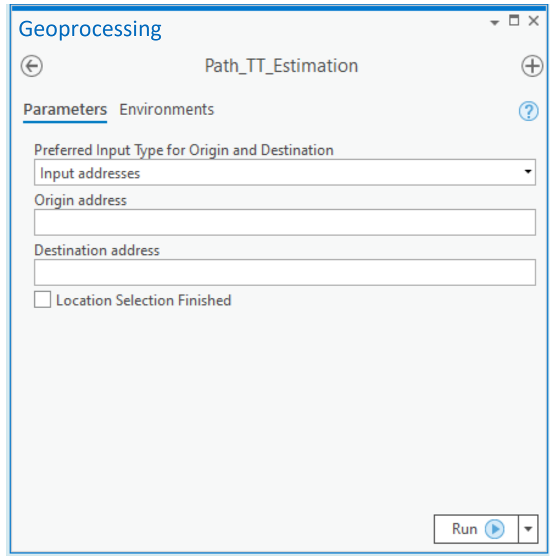



Source: Federal Highway Administration

**Figure 6. Screenshot. Geographic information system platform user interface prompt for users to specify their input of travel time data.**




a) User interface for selecting the O-D pair on the map.




b) User interface for selecting the O-D pair via address inputs.

Source: Federal Highway Administration

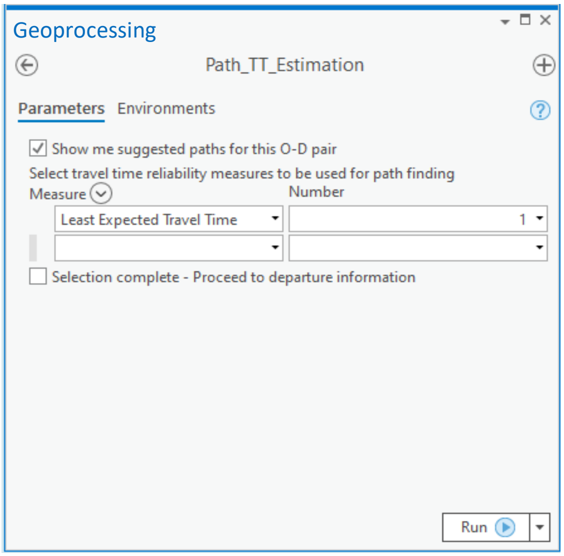
**Figure 7. Screenshot. User interface of the geographic information system platform for the selection of the Origin-Destination (O-D) pair.**

**Step 2:** Path selection

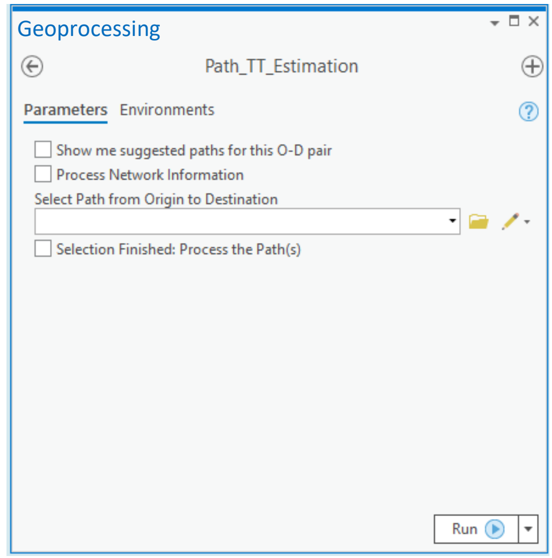
- Take the input and output from step 1.
- Take user input for the choice of a user-selected travel route or a platform-suggested travel route.

If the user chooses to input a travel route:

- Take user input for the desired travel route for the given O-D pair, through interactive tracing on the map.
- Perform feasibility checks on users' travel route input.
  - » Check that the route corresponds to the O-D pair nodes.
  - » Check that the route follows the directionality of links in the network.
  - » Check that the route input is a single connected path of consecutive links on the network.
  - » Modify the travel route input to meet these criteria if appropriate.
  - » Map the travel path onto the network and output it as a feature layer on the network (i.e., a subset selection of the network selected by the user).




a) User interface for selecting a user-defined input path.




b) User interface for preferences on generating suggested paths.

Source: Federal Highway Administration.

**Figure 8. Screenshot. User interface of the geographic information system platform for path selection.**

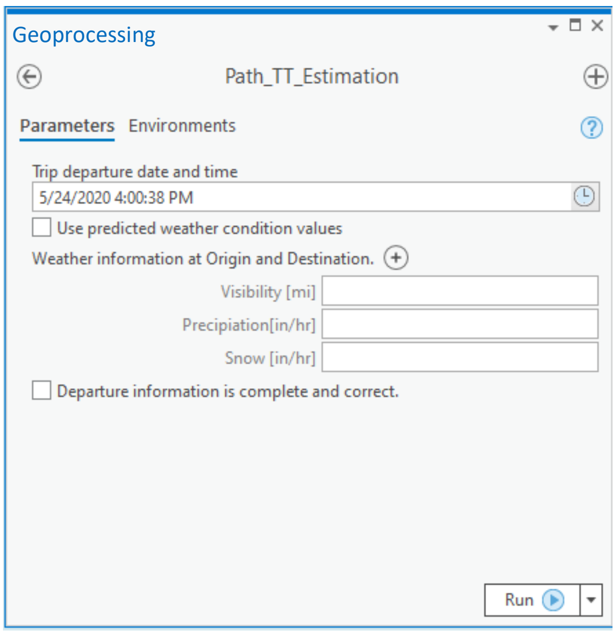
If the user chooses for platform-suggested paths to be generated for them:

- Take user input for the travel time reliability measures to be used for path finding and the number of best paths (up to three) to be generated for each of the selected measures.

**Step 3:** Departure information

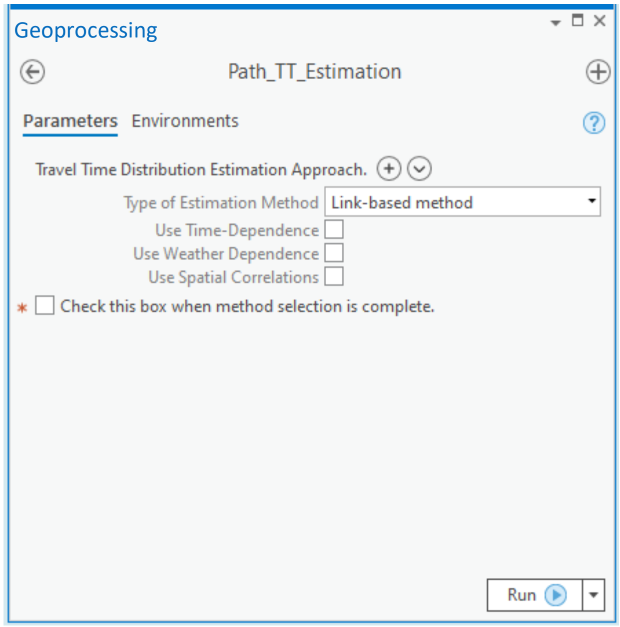
- Take the input and output from step 2
- Take user input for desired departure time and date through an interactive calendar and clock input.
  - » Restrict inputs to times and dates for which travel time and weather data are available.
- Take user input for the choice of manual input of weather or a reading and an output of weather conditions from an application programming interface (API).
  - » If user chooses manual input of weather conditions: Take user input of weather conditions for visibility, rain, and snow precipitation levels for the desired network area and departure date and time.
  - » If user chooses weather conditions from an API: Connect to a weather data API to read the current or predicted weather conditions for the user's desired network area and departure date and time.

- Take user input for the desired path travel time estimation approach and characteristics, based on the methods developed in this study.
- Read the network and travel time data provided by the user.
- Connect to and call the appropriate path travel time distribution estimation functions based on the user input.
- Run path travel time distribution estimation algorithms specified by the user.
- Compute desired travel time reliability indices and variability measures selected by the user and provide them as outputs to the user.
- Display a plot of the cumulative distribution functions and probability density functions for each of the selected (or generated) paths, along with summary statistics of their distributions including the minimum, maximum, expected value and variance of the path travel time distributions.




Source: Federal Highway Administration

**Figure 9. Screenshot. User interface of the geographic information system platform prompting users to input departure information.**




Source: Federal Highway Administration

**Figure 10. Screenshot. User interface of the geographic information system platform prompting users to select the specifications of the distribution estimation method.**

## GENERALIZED USE OF THE GEOGRAPHIC INFORMATION SYSTEM PLATFORM

As previously discussed, one of the main purposes of the GIS platform is to enable agencies, professionals, and other users to implement and apply the solution methodology developed in this study. For this purpose, it is important the GIS platform is suitable for use on any network and data that users may have access to or find an application for. As such, it is designed on a widely used GIS mapping and analytics platform, which allows for a map representation and visualization of the user's data. The GIS platform is also designed to take the user's input of travel time data and a network to adequately read and use the supplied data.

To input network data, users can specify the features in their data that contain the following information:

- Link identification number.
- Their corresponding starting and ending nodes identification, if available.
- Longitude and latitude (x-y location) of those nodes or the start and end points of the links.

Once these specifications are taken as user inputs, the platform performs the analysis on their provided network data.

To input travel time data, users can specify, at minimum, the features that contain information on:

- The link identification that corresponds to their inputted network.
- Link travel time.

To perform the highest level of analysis, the following additional information should be provided:

- Vehicle identification numbers.
- Time of day.
- Weather conditions.

Designing the platform for this level of generalization means the information or parameters are not hard coded or fixed within the platform; rather, the platform can read and use any data format, as long as users provide the data structure information. Because of the added step of inputting information, users should have some knowledge of the data they are using, but users should not have to read and process data themselves, since the platform provides that functionality.

# CHAPTER 6. NEXT STEPS FOR APPLICATION

This study focuses on estimating travel time distributions along user-defined paths, thus providing user-centric and reliability-based measures for network performance, its assessment, and modeling. Central to the user perspective is that it is inherently an entire-trip perspective, where travelers experience the transportation system through trajectories that comprise multiple facilities and sometimes multiple modes of travel. Thus, user-centric measures of system performance seek to translate objective measurements of attributes such as travel time and delay along links to overall time and delay measures at the level of a path connecting two points in the network.

This study brings together recent advances in data availability and probabilistic modeling techniques to synthesize travel time distributions at the path level in generalized networks. This provides a characterization of the user experience of individual travelers or goods shippers, as input to evaluating the reliability performance of transportation systems and devising approaches to improve this performance.

This reference provides an overview of where and how analysts can apply the methodologies, processes, and tools developed in this project. Specifically, this reference first summarizes a range of time-reliability indices and measures, with a focus on how they apply for the purposes of this study. Then, it gives a description of the practical applications of the methods to performance measurement, performance monitoring, and simulation modeling. Additionally, it provides a systematic description of the various steps involved in applying the travel time reliability methodology including an overview of the associated tools and how they function to apply the developed methodology. To allow for increased usage of the developed methods in application and minimize the effort on the part of the user for implementation, this project also provides a GIS Application Platform. Since the developed methods for the estimation of travel time distributions for any path in the network are highly technical and computationally heavy, a computer program performs the estimation. One important step is the development of a user interface that, coupled with network visualization, enables the user to have control over each portion of their study and application-specific implementation without implementing a computer program for the estimation approaches.

The GIS application platform is developed to provide a prototype of the developed solution methodology and enable its delivery, application, testing, and evaluation, using a graphical user interface (GUI) and map representation. The function of the platform, as described in this report, is as follows: to use available historical data and direct input from the user via the GUI and mapping capabilities of a GIS software, to apply the solution methodology to estimate travel-time distributions along user-specified paths, and to report the distribution and associated reliability measures back to the user via the GUI and the map representation in a GIS software.

# REFERENCES

Allahviranloo, Mahdieh, and Will Recker. 2015. "Mining Activity Pattern Trajectories and Allocating Activities in the Network." *Transportation* 42 (4): 561–79.

Dong, Jing, and Hani S. Mahmassani. 2009. "Flow Breakdown and Travel Time Reliability." *Transportation Research Record: Journal of the Transportation Research Board*. https://doi.org/10.3141/2124-20.

Fosgerau, Mogens, and Anders Karlström. 2010. "The Value of Reliability." *Transportation Research Part B: Methodological*. https://doi.org/10.1016/j.trb.2009.05.002.

Hou, Tian. 2015. "Network Traffic Modeling with Vehicle Trajectory Data" (doctoral dissertation, Northwestern University). ProQuest Dissertations and Theses.

Kim, Jiwon, and Hani S. Mahmassani. 2015. "Spatial and Temporal Characterization of Travel Patterns in a Traffic Network Using Vehicle Trajectories." *Transportation Research Procedia*. https://doi.org/10.1016/j.trpro.2015.07.010.

Mahmassani, Hani S., Jiwon Kim, Ying Chen, Yannis Stogios, Andy Brijmohan, and Peter Vovsha. 2014. *Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools*. Washington, DC: Transportation Research Board.

Miller-Hooks, Elise D., and Hani S. Mahmassani. 2003a. "Path Comparisons for a Priori and Time-Adaptive Decisions in Stochastic, Time-Varying Networks." *European Journal of Operational Research*. https://doi.org/10.1016/S0377-2217(02)00231-X.

Miller-Hooks, Elise D., and Hani S. Mahmassani. 2003b. "Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks." *Transportation Science* 34 (2): 198–215. https://doi.org/10.1287/trsc.34.2.198.12304.

Nie, Yu (Marco), and Xing Wu. 2009. "Shortest Path Problem Considering On-Time Arrival Probability." *Transportation Research Part B: Methodological*. https://doi.org/10.1016/j.trb.2009.01.008.

Srinivasan, Karthik K., and Hani S. Mahmassani. 2005. "A Dynamic Kernel Logit Model for the Analysis of Longitudinal Discrete Choice Data: Properties and Computational Assessment." *Transportation Science* 39 (2): 160–81.

Zockaie, Ali, Hani S. Mahmassani, and Jiwon Kim. 2016. "Network-Wide Time-Dependent Link Travel Time Distributions with Temporal and Spatial Correlations." Proceedings of the 95th Annual Meeting of the Transportation Research Board.

Zockaie, Ali, Yu M Nie, and Hani S. Mahmassani. 2014. "Simulation-Based Method for Finding Minimum Travel Time Budget Paths in Stochastic Networks with Correlated Link Times." *Transportation Research Record* 2467 (1): 140–48.

U.S. Department of Transportation
Federal Highway Administration

U.S. Department of Transportation
Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC 20590
https://ops.fhwa.dot.gov

May 2024

FHWA-HOP-20-066