Addendum to Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (2019 Update)

Reliability Analysis Guidance Addendum

September 2023

U.S. Department of Transportation
Federal Highway Administration

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HOP-20-T013	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Addendum to Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (2019 Update)		5. Report Date September 2023
		6. Performing Organization Code:
7. Author(s) David Hale, Paul Morris, Nick Semeja		8. Performing Organization Report No.
9. Performing Organization Name and Address Leidos 11251 Roger Bacon Drive Reston, VA 20190		10. Work Unit No.
		11. Contract or Grant No. DTFH61-16-D-00053
12. Sponsoring Agency Name and Address Office of Operations Federal Highway Administration 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered
		14. Sponsoring Agency Code HOP
15. Supplementary Notes The government task managers were Neil Spiller and John Halkias.		
16. Abstract This addendum provides up-to-date guidance on how to incorporate travel time reliability principles into the microsimulation analysis process. Microsimulation is the modeling of individual vehicle movements on a second or subsecond basis for the purpose of assessing the traffic performance of highway and street systems, transit, and pedestrians. Microsimulation analyses are increasingly visible and important—fostered both by the continued evolution of microsimulation software capability and increasing application within transportation engineering and planning practices. These guidelines provide practitioners with guidance on the appropriate application of microsimulation models to traffic analysis problems, with an overarching focus on existing and future alternatives analysis.		
17. Key Words Traffic analysis tools, traffic simulation, highway capacity	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov	

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 23	22. Price N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

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 (2,000 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

APPROXIMATE CONVERSIONS FROM 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 (2,000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for 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. PURPOSE

This document is an addendum to *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (2019 Update)* (Federal Highway Administration Report No. FHWA-HOP-18-036)[1] and reflects up-to-date guidance on incorporating travel time reliability (TTR) in the Traffic Analysis Toolbox (TAT). The addendum consists of:

- Updates to the existing Toolbox volume text
- Additional content to be appended to the Toolbox volume

[1]Wunderlich, K., M. Vasudevan, P. Wang, et. al. 2019. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (2019 Update).* Report No. FHWA-HOP-18-036. Washington, DC: FHWA.

CHAPTER 2. UPDATES TO EXISTING TOOLBOX VOLUME TEXT

INTRODUCTION

Page 1:

- With reference to the first “Guiding Principle of Microsimulation: Ensure the Analysis Has a Clear Objective and Well-Defined Performance Measures”:
    - A modern traffic analysis can be considered robust if it produces multiple performance measures related to TTR (e.g., buffer time, travel time index, planning time index, probability of on-time arrival) and accompanying visualizations (e.g., scatterplots, histograms, probability density functions). By contrast, peak-hour analyses based on simple averages may produce optimistic outcomes rarely observed in the field. A traffic analyst should seek to observe and understand the full range of traffic conditions that occur throughout the day, week, and year.

CHAPTER 5. “MODEL CALIBRATION”

Page 58:

- With reference to section 5.2, “Identify Representative Days”:
    - For travel conditions involving inclement weather, multiple methods and options are available for calibrating a simulation model. Weather can affect car-following headways, lane-changing aggressiveness, free-flow speeds, driver reaction times, vehicle deceleration capabilities, and so on. Some of these options are discussed in the report FHWA-HRT-04-131, *Identifying and Assessing Key Weather-Related Parameters and Their Impacts on Traffic Operations Using Simulation* (Zhang et al. 2004).

GLOSSARY

Page 107:

- Travel Time Reliability (TTR) is defined in 23 CFR 490.101 as the “consistency or dependability of travel times, as measured from day to day, across different times of the day, or both.” Dependability of travel times is often further explained by academics as consistency of proximity to free-flow travel times, as opposed to consistently excessive travel times.

CHAPTER 3. ADDITIONAL CONTENT TO BE APPENDED TO THE TOOLBOX VOLUME

RELIABILITY ANALYSIS OPTIONS FOR MICROSIMULATION

At the time of this writing, the scenario generator and trajectory processor are the primary tools available to support TTR analysis via microscopic traffic simulation. However, not all products are compatible with these tools, and not all analysts have the access, expertise, or necessary input data to use these tools. In these cases, analysts may consider the spatiotemporal traffic matrix (STM) approach, the manually generated scenario approach, or both. The STM is a key concept for congestion identification and probe data analysis. Each cell of this matrix simultaneously represents a specific roadway segment and time period. In the previous century, traffic engineers were taught to perform peak-hour analysis, which only represents a few cells within this matrix. By carefully analyzing traffic demand variability that occurs throughout the year, it becomes possible to account for the effects of weather, work zones, incidents, and even seasonal effects. In a microsimulation, the study section would be subdivided by links, which would be specified in an upstream-to-downstream manner. Two-dimensional grids would be analyzed one arterial at a time. The study period might be subdivided into time periods of 5, 10, or 15 minutes. Time periods longer than 15 minutes may fail to properly capture the effects of demand variability. To achieve a robust reliability analysis, the analyst would need sufficient operating condition data (i.e., reflecting varying demands, weather, incidents, and events) to model many days of the year.

NUMERIC PERFORMANCE MEASURES

Daily Measures

The analyst may define one or more cutoff speed thresholds to differentiate between congested and uncongested conditions. Cutoff speeds may differ between links if, for example, the posted speed limits differ among links. Duration represents the longest continuous time period within the analysis box, during which the segment is congested (i.e., colored red). Extent represents the longest continuous spatial length within the analysis box, during which the time interval is congested. Intensity is a two-dimensional performance measure, covering both space and time, which represents the percentage of the analysis box that is congested. Speed drop is the average percentage difference between actual speeds and cutoff speeds and is only averaged over the congested red area of the analysis box.

Vehicle delay is computed according to the following equations, which are documented within the *Most Congested Freeways in 2013 Report and Methodology* (Iteris 2014). However, instead of using a fixed value of 35 miles per hour (mph), cutoff speeds are obtained from the analyst, as described earlier. Bottleneck volumes should be measured immediately downstream of the downstream end of congestion. Vehicle delay is an important performance measure for comparing and ranking bottlenecks because it captures the effects of both speed drops and volumes.

$$Delay\ on\ each\ TMC\ for\ one\ vehicle = \frac{Length\ of\ TMC}{Average\ TMC\ Speed} - \frac{Length\ of\ TMC}{35}$$

Source: Federal Highway Administration[2]
TMC = traffic message channel.

Figure 1. Equation. Calculation of delay per vehicle from the space-time matrix.

$$\begin{aligned} &Delay\ on\ each\ TMC\ (vehicle - hours) \\ &\qquad = Delay\ on\ each\ TMC\ for\ one\ vehicle \times Bottleneck\ Volume \end{aligned}$$

Source: Federal Highway Administration
TMC = traffic message channel.

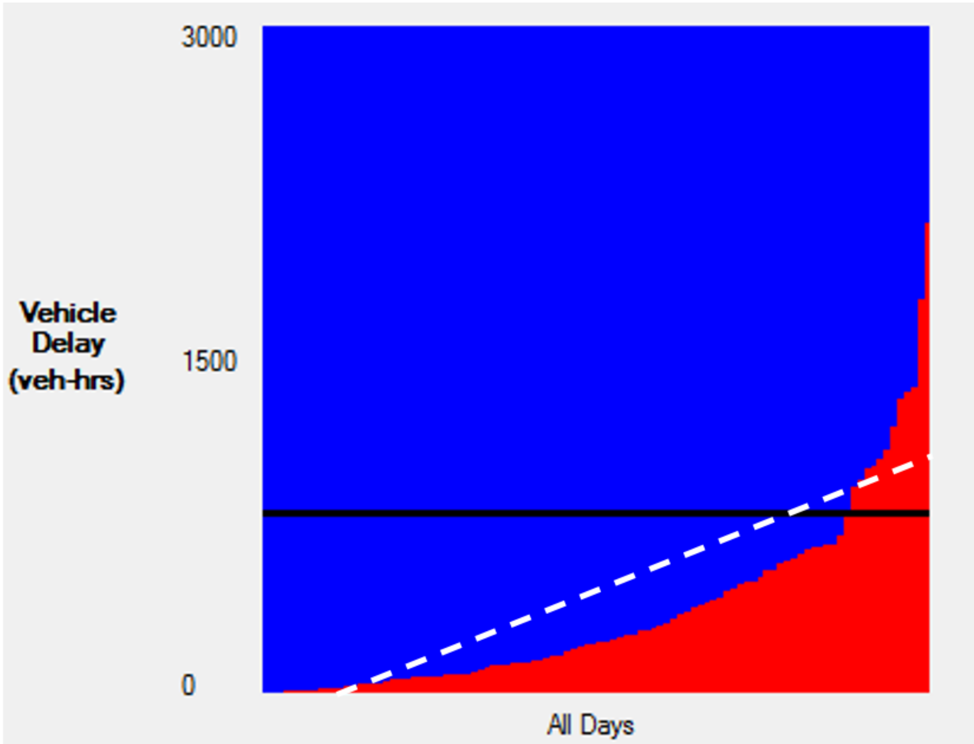
Figure 2. Equation. Calculation of traffic message channel delay.

Annual Measures

Once the analyst constructs an STM based on simulation outputs, they may report on a rich set of percentile performance measures. Percentile measures will affect which day is identified during the year. For example, the 85th percentile intensity will correspond with whatever day of the year exhibits the 85th percentile worst intensity. The 85th percentile worst intensity means 85 percent of days in the simulation analysis will have a lower intensity than that day. The analyst may omit certain months of the year to perform a seasonal analysis.

The annual reliability matrix (ARM) concept can be used to obtain additional measures. In Figure 3, the x-axis now contains all days of the year and reflects the percentile worst days of the year. The y-axis now denotes vehicle hours of delay. The ARM displays all daily delays of the year in ascending order, with the lowest delay day on the far left, and the highest delay day on the far right. The total red area is the total amount of delay that occurred throughout the year. However, to better assess the reliability of the facility, the slope and shape of the ARM must be taken into consideration.

[2]Hale, D., R. Jagannathan, M. Xyntarakis, P. Su, X. Jiang, J. Ma, J. Hu, and C. Krause. *Traffic Bottlenecks: Identification and Solutions*, Report No. FHWA-HRT-16-064. Washington, DC: FHWA. http://www.fhwa.dot.gov/publications/research/operations/16064/16064.pdf, last accessed March 22, 2023.



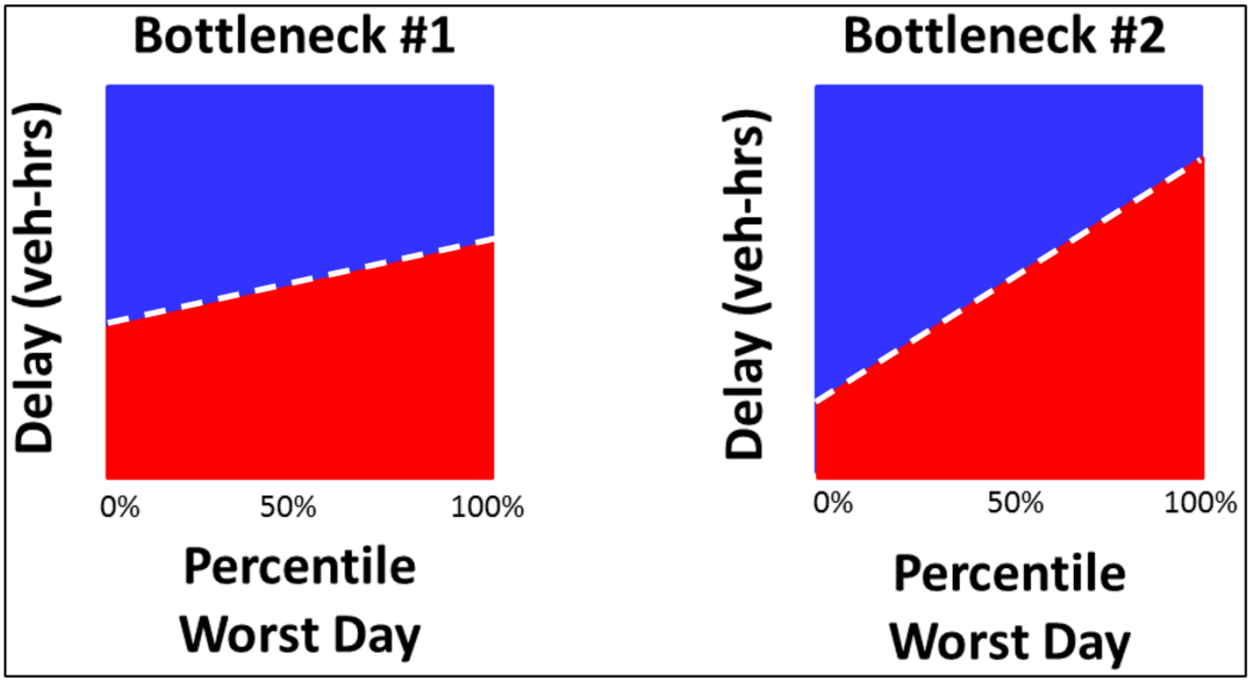
Source: Federal Highway Administration Congestion and Bottleneck Identification Tool (CBI Tool)[3]

veh = vehicle; hrs = hours.

Figure 3. Graph. Example of the annual reliability matrix.

Researchers designed the ARM and associated numeric measures for the purpose of comparing and ranking traffic bottlenecks (Hale et al. 2016). These performance measures convey both the annual intensity and variability of traffic congestion. Some performance measures in the industry convey annual intensity while ignoring annual variability and reliability. For example, Figure 4 demonstrates a comparison of two hypothetical ARMs. The amount of red area is essentially equal for both bottlenecks. By some industry standards, these two bottlenecks would be considered equivalent priorities. However, bottleneck No. 2 (on the right-hand side) should be ranked as the higher priority because more time would be needed to ensure an on-time arrival.

[3]FHWA. n.d. "Congestion and Bottleneck Identification (CBI) Tool Software Download" (webpage). https://highways.dot.gov/research/resources/software/congestion-bottleneck-identification-cbi-tool-software-download, last accessed March 22, 2023.



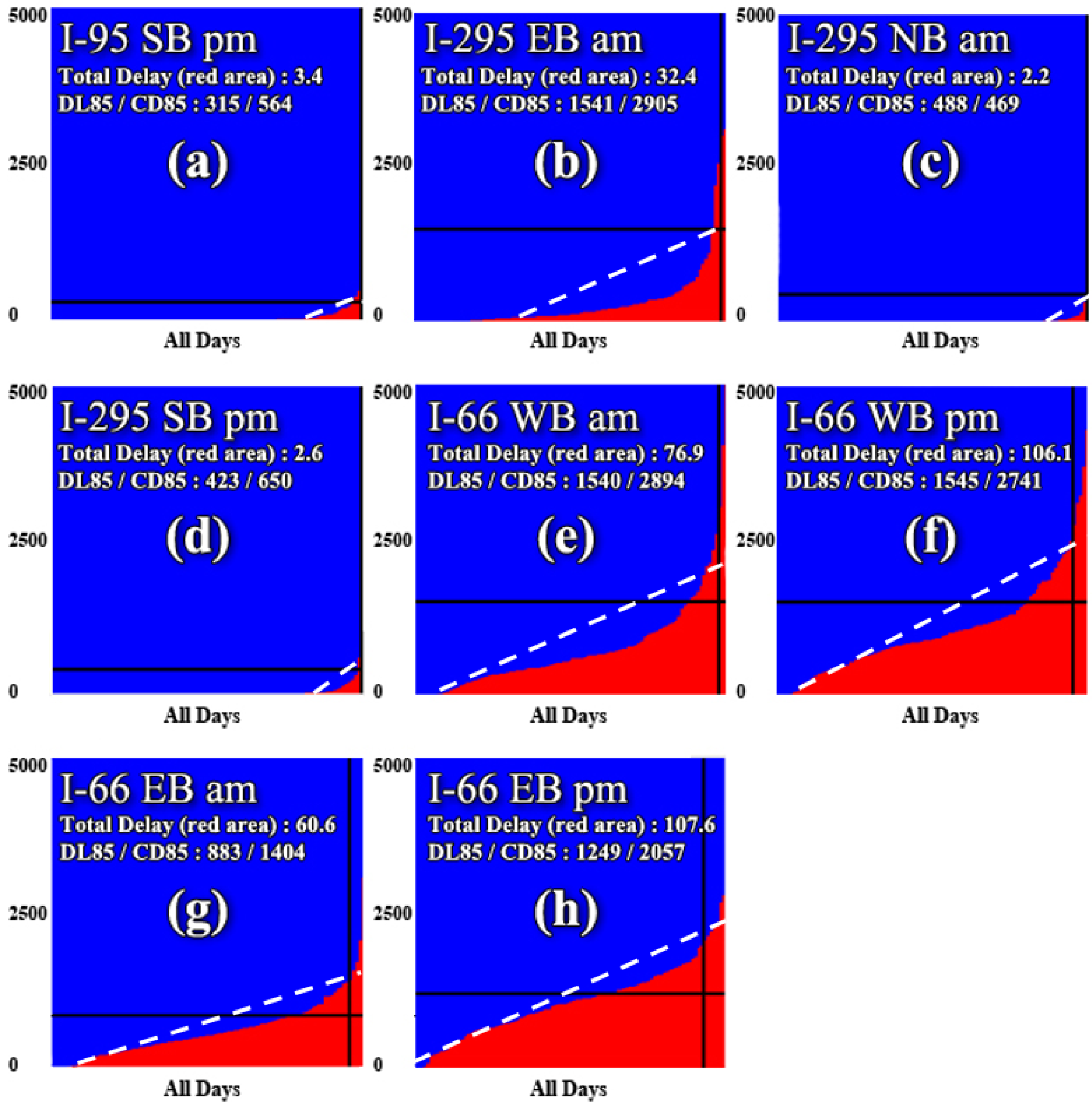
Source: Federal Highway Administration

veh = vehicle; hrs = hours.

Figure 4. Graphs. Example comparison of annual intensity and reliability.

Case Study Comparing and Ranking Bottlenecks

Figure 5 illustrates a case study comparison of eight real-world bottleneck locations. The total red area is shown on the lower right of each ARM diagram. The first delay value shown is the horizontal delay level (shown as a horizontal black bar), below which 85 percent of the red area resides. The second delay value shown is the highest delay value that occurred throughout the year (shown as a vertical black bar).



Source: Federal Highway Administration

CD85 = 85th percentile cumulative delay; DL85 = 85th percentile delay level; EB = eastbound; NB = northbound; SB = southbound; WB = westbound.

Figure 5. Graphs. Example comparison and ranking of bottlenecks (by vehicle hours of delay).

Process of Comparison

The ARM diagram and three associated numeric measures could now be used for direct, apples-to-apples comparisons and rankings. A multivariate approach (i.e., based on multiple measures) provides a more robust process of comparison than a univariate approach (i.e., based on a single measure). The following case study narrative illustrates why.

Overall annual delay is probably the most appropriate performance measure to begin comparison for this reason: If one bottleneck has significantly more overall annual delay than a second bottleneck, the first bottleneck almost certainly offers less TTR than the second bottleneck. However, if two bottlenecks have similar values of overall annual delay, the 85th percentile delay level should be an effective tiebreaker for indicating which bottleneck exhibits superior reliability. The 100th percentile delay level should clarify the results in some situations.

In the particular case study shown in Figure 5, (h) and (f) would be ranked as the worst bottlenecks according to their total red areas of 107.6 and 106.1, respectively. However, it seems that the difference between 107.6 and 106.1 is not enough to confidently assert that (h) is the worst bottleneck in terms of reliability. In fact, bottleneck (f) has a much higher 85th percentile delay level (1,545 vehicle hours of delay) than (h) (1,249 vehicle hours of delay). Indeed, the bottleneck (f) ARM in Figure 5 appears to have a steeper annual slope than the bottleneck (h) ARM. Therefore, bottleneck (f) appears to be the worst bottleneck in terms of annual reliability.

Next, based on the total red area, (e) and (g) are the next worst bottlenecks (after (h) and (f)). Although the total red area of (e) appears to be significantly larger than that of (g) by a ratio of 76.9 to 60.6, the 85th percentile red area of (e) (1,540 vehicle hours of delay) is also much greater than that of (g) (883 vehicle hours of delay). Therefore, (e) is worse than (g).

Interestingly, if 85th percentile delay level were used as a univariate measure for ranking bottlenecks, (e) (1,540 vehicle hours of delay) and (f) (1,545 vehicle hours of delay) would have graded out as equal bottlenecks. However, because (f) has much more total annual delay than (e) by a ratio of 106.1 to 76.9, (f) is demonstrably less reliable. Similarly, if 85th percentile delay level were used as a univariate measure, (e) (1,540 vehicle hours of delay) and (b) (1,541 vehicle hours of delay) also would have graded out equally; however, because (e) has much more total annual delay by a ratio of 76.9 to 32.4, (e) is demonstrably less reliable. The question becomes, how could (e) and (b) have such unequal total annual delays while having similar 85th percentile delay levels? The data implies that (e) must have had one or two extremely bad days to skew the results. Indeed, the worst day at (e) (9,826 vehicle hours of delay) was the worst delay day among all eight bottlenecks. Thus, the final rankings would be (f) (worst), (h), (e), (g), (b), (a), (d), and (c) (best).

Manually Generated Scenario Approach

Without special preprocessor and postprocessor tools to support TTR analysis, a microscopic traffic simulation user may consider manually developing a relatively small number (e.g., a dozen or so) of core scenario datasets. The first consideration is defining the reliability reporting period (RRP). Terms such as RRP, STM, analysis box, and analysis cube are sometimes used interchangeably, but they all involve the prerequisite choice of temporal and spatial analysis boundaries.

Once RRP is defined, traffic volume demands associated with that particular physical network and time horizon could be grouped into a small number of scenarios (e.g., low, medium, high).

Similarly, there could be additional scenarios to reflect the impacts of nonrecurring events (i.e., weather, incidents, work zones, special events). Many options are available for modeling nonrecurring events in microsimulation including adjustments to car following, lane changing, free-flow speed, route choice, and traffic demand. In the manually generated scenario approach described here, the microscopic simulation user would apply engineering judgment to implement proper adjustments within the scenario datasets.

The number of scenarios used in a manually generated approach would presumably be smaller than the number used with a tool like the scenario generator. Table 1 illustrates what would happen if the demand, weather, and incident variations throughout RRP were grouped into a very small number of scenarios (i.e., 3×2×2=12). In this manner, expanding from two weather groupings (ideal, poor) to three (ideal, light precipitation, heavy precipitation) could inflate the overall number of scenarios from 12 to 18; adding a fourth grouping with two options (e.g., work zone inactive, work zone active) would multiply the number from 18 to 36, and so on.

Table 1. Obtaining a travel time distribution via manually generated scenarios.

Core Scenario	1	2	3	4	5	6	7	8	9	10	11	12
Demand	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Demand Frequency	20%	40%	40%	20%	40%	40%	20%	40%	40%	20%	40%	40%
Weather	Ideal	Ideal	Ideal	Poor	Poor	Poor	Ideal	Ideal	Ideal	Poor	Poor	Poor
Weather Frequency	70%	70%	70%	30%	30%	30%	70%	70%	70%	30%	30%	30%
Incidents	None	None	None	None	None	None	Yes	Yes	Yes	Yes	Yes	Yes
Incident Frequency	70%	70%	70%	70%	70%	70%	30%	30%	30%	30%	30%	30%
Relative Frequency	9.8%	19.6%	19.6%	4.2%	8.4%	8.4%	4.2%	8.4%	8.4%	1.8%	3.6%	3.6%
Random # Seed Realization	98	196	196	42	84	84	42	84	84	18	36	36

Source: Federal Highway Administration

Table 1 further illustrates what would happen if the analyst applied frequency-weighting factors to each scenario. In this example, high demand occurs 40 percent of the time, clear weather 70 percent of the time, and so on. Relative frequency (second-to-last row) is the product of demand, weather, and incident frequencies for each core scenario. Combined with a number of random seed realizations weighted by frequency of occurrence, the Table 1 exercise would initially produce 1,000 travel time outcomes. Typical reliability performance measures (e.g., travel time index, 85th percentile day, buffer index) and accompanying visualizations (e.g., scatterplots, histograms, probability density functions, cumulative density functions) could be obtained from this set of 1,000 outcomes. If the analyst did not have time to process 1,000 runs, the number of runs could be reduced proportionally (e.g., divided by 10).

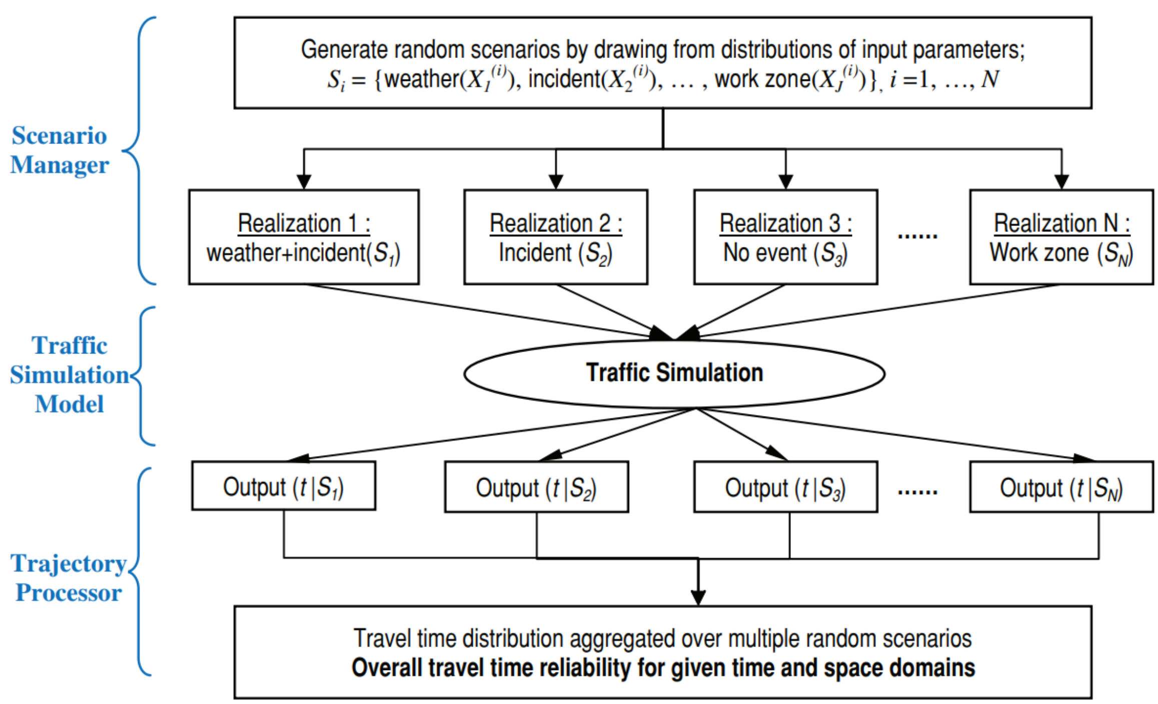
Scenario Development Methods

Approaches to Assessing Reliability

TTR is a relative concept, in that it depends on the temporal and spatial boundaries for which travel times are observed. For example, TTR of weekdays is different from that of weekends on the same road network. Therefore, defining time and space domains can precede assessing reliability. In general, the time domain is specified by a date range of the overall time period (e.g., June 1, 2012 to August 31, 2012), day of week (Monday to Friday), and time of day (6 a.m. to 10 a.m.). The time domain could also be a specific season or day of each year (e.g., Thanksgiving Day). The space domain defines at which level travel times are collected, and the reliability measures are calculated (e.g., network level, origin destination [O-D] level, path level, and link level). Two different approaches can assess TTR for given time and space domains: (1) Monte Carlo and (2) mix and match (Mahmassani et al. 2014). The former tries to generate all possible scenarios that could occur during the given temporal and spatial boundaries to introduce realistic variations in the resulting travel time distribution; the latter constructs scenarios by manually choosing various combinations of scenario components. These approaches are discussed in more detail in the following sections.

Monte Carlo Approach

Many of the travel time unreliability factors fall into the area in which the randomness can be parameterized, and probabilities can be assigned based on the known parameters of the demand, supply, or both (Mahmassani et al. 2014). This approach uses Monte Carlo simulation to prepare input scenarios aimed at propagating uncertainties in selected scenario components *(X)* into uncertainties in the generated scenarios (S_i ($i = 1, \ldots, N$)), which can be translated into the resulting travel time distribution. As depicted in Figure 6, the scenario manager performs Monte Carlo simulation to generate hundreds or thousands of input scenarios by sampling from the joint probability distribution of scenario components. Each scenario from the sampling process is equally likely, thus allowing the trajectory processor to simply aggregate travel time distributions from many simulation runs to obtain the most likely (probable) outcome of a set of reliability performance indicators for the given time and space domains.

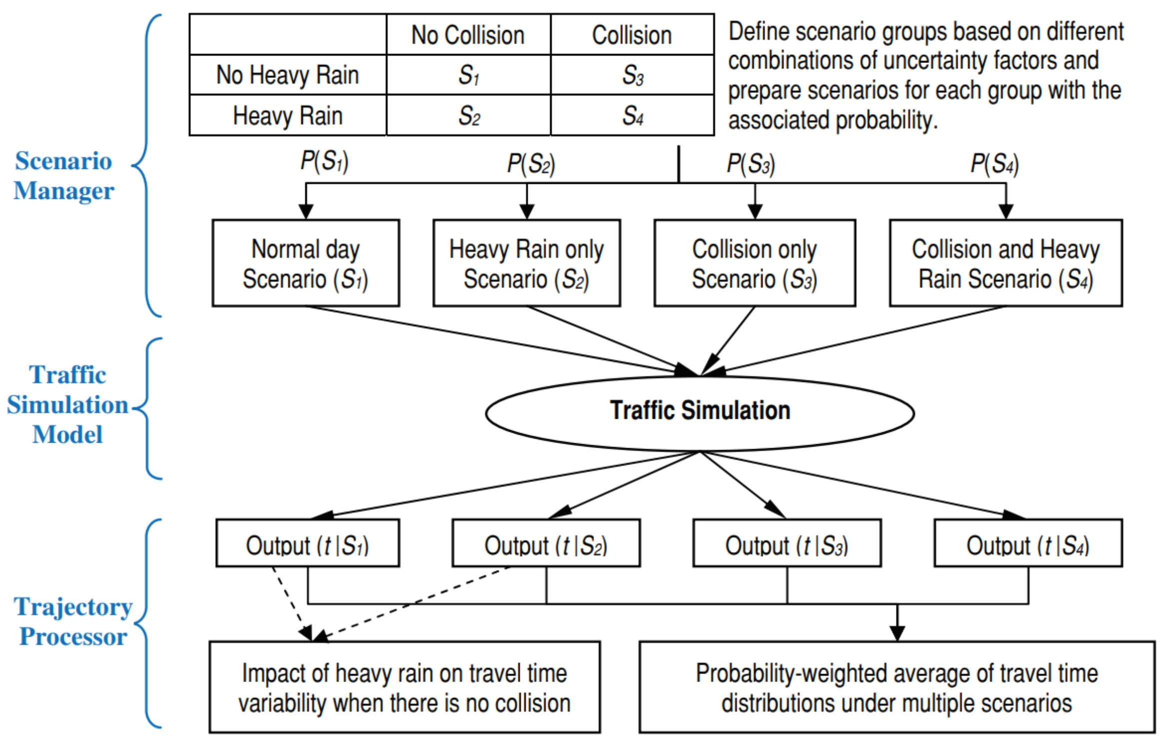


Source: Mahmassani et.al. 2014.

Figure 6. Illustration. Monte Carlo reliability analysis approach.

Mix-and-Match Approach

Instead of randomly generating scenarios, given the underlying stochastic processes, one could explicitly specify scenarios with historical significance or policy interest (Mahmassani et al. 2014). The mix-and-match approach aims to construct input scenarios in a more directed manner, either by mixing and matching possible combinations of specific input factors or by directly using known historical events or specific instances (e.g., holiday, ball game). Figure 7 shows a schematic diagram illustrating this approach with a simple example. Consider two scenario components—collision and heavy rain—where each component has two discrete states: occur and not occur. The Cartesian product of the two components' states defines four possible scenario groups. Suppose there is a representative scenario for each group, and the scenario probability is assigned based on the joint probability of collision and heavy rain events. In that case, a probability-weighted average of travel time distributions under all four scenarios can be used as the expected travel time distribution to approximate overall reliability measures. A more informative use of this approach is to understand the impact of a particular scenario component on travel time variability by investigating gaps between different combinations of output results.



Source: Mahmassani et.al. 2014.

Figure 7. Illustration. Mix-and-match reliability analysis approach.

Combined Approach

Unlike the simple example in Figure 7, it is often necessary to allow randomness in scenarios within each group, especially when there is no predefined representative scenario. It is also possible to have no probability value for each scenario group known to users. In both cases, the

Monte Carlo approach can be used in conjunction with the mix-and-match approach—that is, sampling random scenarios from their conditional distributions given each group (for the former) and generating many scenarios for the entire scenario space and categorizing them into the associated groups to obtain the group probabilities (for the latter).

Cluster Selection

Phoenix Pilot Cluster Methodology

The following information summarizes a case study demonstrating how to effectively select sufficient types and the number of clusters for reliability analysis. The Second Strategic Highway Research Program (SHRP 2) Project L04 pilot project team developed this documentation.

Introduction to Cluster Analysis

The full-year analysis process pilot tested in this project relied upon a large amount of observed baseline data and individual simulations. This reliance was necessary to systematically and accurately represent the wide range of traffic and environmental conditions that occurred over the course of the year for which actual conditions were observed and simulated. A well-tested mathematical procedure known as a hierarchical cluster analysis was used to minimize the number of simulation runs needed to develop a good representation of the full-year travel time distribution profile. The term cluster analysis encompasses many different algorithms developed for grouping large numbers of objects with similar characteristics into much smaller discrete sets or taxonomies that can be more efficiently analyzed.

The cluster analysis algorithm employed in this pilot test is embedded in an open-source statistical software package called R. The R software package was applied according to the following two-step procedure:

1. Select an appropriate measure for quantifying distance between clusters. This project employed the commonly used Euclidean distance as the means for calculating composite distance between observed data points and for calculating distance between centroids of the respective clusters. The equation shown in Figure 8 was used for these purposes. All variables used in this project were normalized to values between 0 and 1, so the result of applying the equation in Figure 8 is a relative distance measurement that has no dimensional units associated with it.

$$D_{ij} = \sqrt{\sum_{k=1}^{n} \left(x_{ki} - x_{kj} \right)^2}$$

Figure 8. Equation. Calculation of composite distances between observed data points.

Where:
D_{ij} = distance between cases i and j.
x_{kj} = value of variable x_k for case j.
for each of data points i and j.

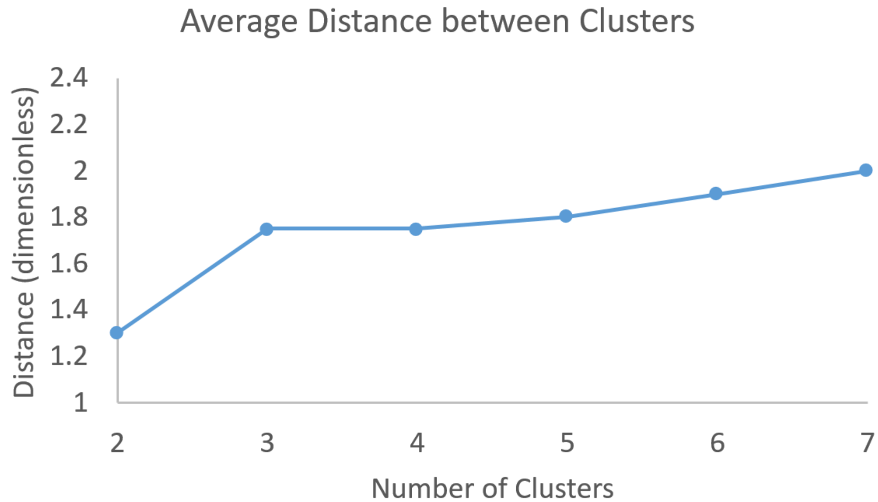
2. Determine the appropriate number of clusters using the K-mean cluster analysis technique. The K-mean cluster analysis technique is a well-documented method for partitioning a set of observed data points into clusters, wherein each observed data point is assigned to the particular cluster within a pre-established group of clusters that possesses the nearest mean. At the outset of the analysis, the user establishes the number of clusters to be created, with two clusters being the minimum. With this input, the K-mean cluster analysis technique then assigns each data point to one of the pre-established clusters to maximize the Euclidean distance between each cluster. The mean for each cluster then becomes reflective of the total of all data points assigned to that cluster. This mean will thereafter serve as the prototype for all observed data points assigned to that particular cluster. Determining an appropriate number of clusters is an iterative process. It usually begins with the minimum two clusters and then increases the number of clusters by one with each iteration until the point of diminishing returns is identified.

Application of Cluster Analysis in the Phoenix Pilot Test

The observed dataset for base year conditions on the Phoenix-area freeway system consisted of observations across 253 separate weekdays (weekday holidays were excluded). The dataset covers 19 two-way corridors in the Phoenix area, for a total of 38 one-way corridors. The temporal coverage of the data is from 3 p.m. to 7 p.m. in 2014.

In the pilot test, two separate cluster analyses were conducted. The first focused on identifying significant seasonal differences in the observed data. The second focused on identifying significantly different data clusters within each of the previously identified seasons. The seasonal analysis was conducted by including date information as one of the variables in the cluster analysis process. It was expected and found that including a date variable at this analysis stage resulted in a high likelihood of data observed during the same week or month being assigned to the same seasonal cluster.

The results of the iterative cluster analysis procedure conducted at the seasonal analysis stage are presented in Figure 9. Based on the results, it was concluded that three is the appropriate number of seasons to use for the 2014 observed data in the Phoenix pilot test site.



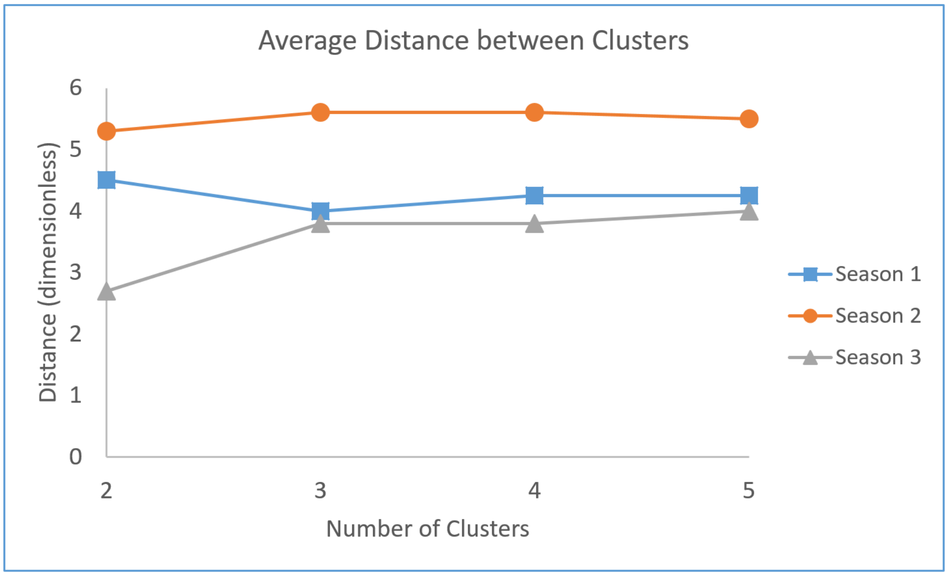
Source: Federal Highway Administration

Figure 9. Graph. Seasonal cluster analysis results for Phoenix pilot test site.

For each of the three seasonal data clusters that resulted from the previous analysis, an additional cluster analysis was conducted to evaluate the need for separate data clusters within each season. The only difference between the seasonal cluster analysis conducted earlier and these cluster analyses is the exclusion of the date variable for each within-season cluster analysis. The result of the iterative process of cluster investigations within seasons one, two, and three, respectively, is shown in Figure 10. Figure 11 summarizes the process used for the cluster analysis and its role in relation to the scenario development stage of the reliability analysis.

The analysis results presented in these figures indicate that no significant benefit will be gained from analyzing more than two clusters in any of the three seasons. This finding also highlights a key characteristic of the cluster analysis methodology that can have an important effect on the analyst's ultimate workload: To calculate the maximum distance between clusters, one must begin the analysis with at least two clusters. Thus, the remaining unanswered question is whether even two clusters are necessary.

To answer this question, the project team determined the centroid of the single cluster for each season and then applied the cluster analysis methodology manually to calculate the Euclidean distance between the single-cluster centroid for each season and the centroids of the two initial clusters developed for each season. The results of this analysis are presented in Table 2 for season 1 and show that, in this case, two clusters were found to be better than one. For this project, two within-season clusters were found to be the most appropriate number of clusters for each of the three seasons.

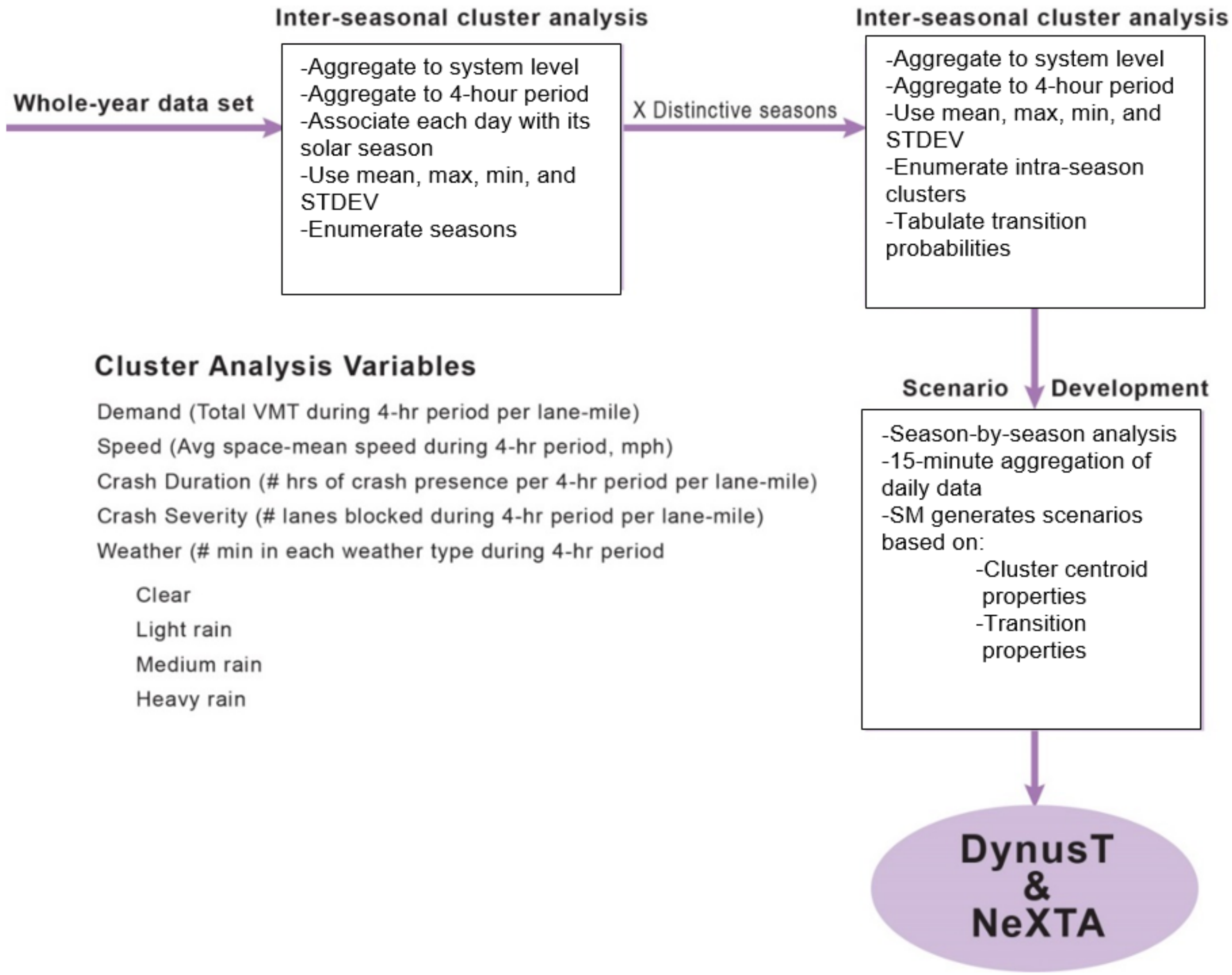


Source: Federal Highway Administration

Figure 10. Graphs. Within-season cluster analysis for seasons one, two, and three.

Table 2. Comparison of single cluster versus two-cluster analysis results.

Scenario	Scaled distance
Comparison of single-cluster centroid with centroid of cluster 1	3.184
Comparison of single-cluster centroid with centroid of cluster 2	1.045
Comparison of cluster 1 centroid with cluster 2 centroid	4.434



Source: Federal Highway Administration

Avg = average; SM = scenario manager; STDEV = standard deviation; VMT = vehicle miles traveled

Figure 11. Illustration. Role of cluster analysis in relation to scenario development.

REFERENCES

Culotta, K., V. Fang, F. Habtemichael, and D. Pape. 2019. *Does Travel Time Reliability Matter?* Report No. FHWA-HOP-19-062. Washington, DC: Federal Highway Administration.

Hale, D., R. Jagannathan, M. Xyntarakis, P. Su, X. Jiang, J. Ma, J. Hu, and C. Krause. 2016. *Traffic Bottlenecks: Identification and Solutions*. Report No. FHWA-HRT-16-064. Washington, DC: Federal Highway Administration.

Iteris. 2014. *2013 Most Congested Freeways Report and Methodology*. San Francisco, CA: Metropolitan Transportation Commission.

Mahmassani, H. S., J. Kim, Y. Chen, Y. Stogios, A. Brijmohan, and P. Vovsha. 2014. *Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools*. Report No. S2-L04-RR-1. Washington, DC: National Academy of Sciences, Engineering, and Medicine (NASEM).

NASEM. 2010. *Highway Capacity Manual*. Washington, DC: NASEM

Wunderlich, K., Vasudevan, M., Wang, P., et. al. 2019. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (2019 Update).* Federal Highway Administration Report No. FHWA-HOP-18-036.

Zhang, L., P. Holm, and J. Colyar. 2004. *Identifying and Assessing Key Weather-Related Parameters and Their Impacts on Traffic Operations Using Simulation*. Report No. FHWA-HRT-04-131. Washington, DC: Federal Highway Administration.

U.S. Department of Transportation
Federal Highway Administration

U.S. Department of Transportation
Federal Highway Administration
Office of Operations
1200 New Jersey Avenue, SE
Washington, DC 20590

Office of Operations website
https://ops.fhwa.dot.gov

September 2023