INRIX Signal Analytics

Data Validation & Metrics FAQ

CONTENTS

INRIX SIGNAL ANALYTICS FAQ

What is INRIX Signal Analytics?

INRIX Signal Analytics is a cloud-based analytics application that uses crowd sourced vehicle waypoint data to help transportation professionals assess and improve their signalized intersection performance. The application summarizes individual vehicle statistics without the need for permanent vehicle detection infrastructure.

The platform is scalable, easy-to-use, and cost-effective; it is intuitive and there is nothing to install. By transforming trillions of data points into visualizations, daily reports, and interactive dashboards, Signal Analytics allows transportation professionals to identify, rank, and prioritize a variety of traffic signal projects. These tools empower transportation professionals to take a proactive approach for the management and monitoring of signalized intersections, while reducing data collection and reporting costs.

This document outlines how we know this process works today. In this document, we will discuss the data collection process, how the data are used to identify individual vehicle metrics, how those metrics are aggregated, and how we initially validated the aggregate metrics. This document concludes with insights related to data quantity and current use cases for the signal analytics platform.

What data is being used in INRIX Signal Analytics?

The data being used are probe trajectory data. These data are collected from connected vehicles and include individual waypoint information every few seconds.

Key Details:

- The waypoint data allow a vehicle to be traced through an intersection, where valuable insights can be extracted and aggregated to understand and improve the signal performance at an intersection (Figure 1).
- INRIX Signal Analytics sources data purely from high-quality, low ping frequency (< 5 second) data providers to produce a series of signal performance measures.
- The metrics collected at the vehicle level are approach speed, travel time, stops, and entering and exiting heading. The data are processed at the movement level, then aggregated at the intersection level to provide scalable metrics at every intersection.

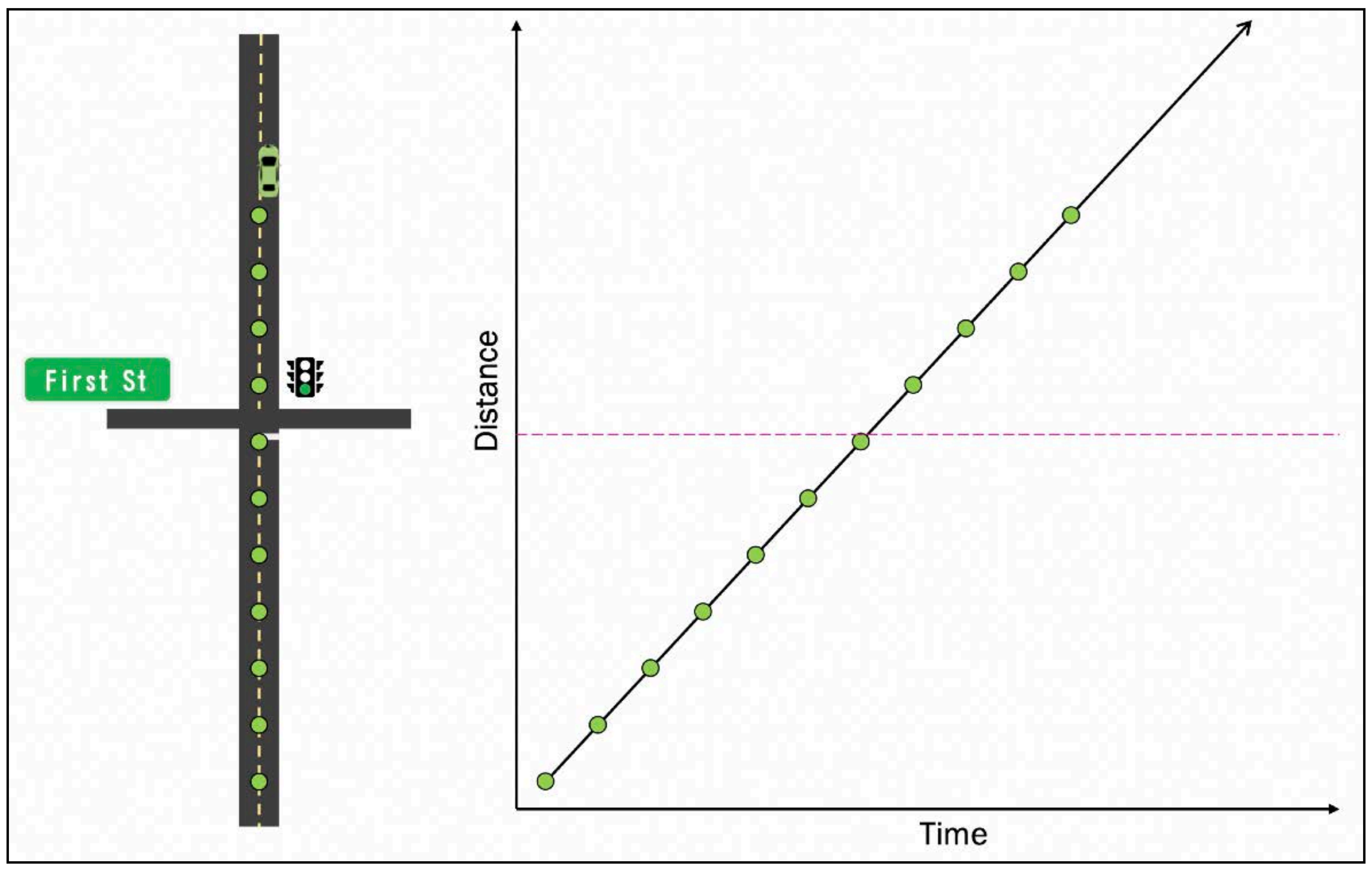


Figure 1. Trajectory Data and Time Space Diagram Example

What metrics are being collected in INRIX Signal Analytics?

Leveraging the high frequency waypoint data, a vehicle's journey through the intersection can be characterized. Figure 2 shows examples of three different trips through the intersection.

Key Details:

- The green vehicle traveled through the intersection in 12 seconds with a constant speed and no stops. It can logically be assumed that this vehicle arrived at an intersection when the signal was green and experienced little to no delay.
- The yellow vehicle traveled through the intersection in 32 seconds, and slows to a stop prior to the signalized intersection. This vehicle is assumed to have arrived on a red signal and experienced a minor amount of delay (~20 s) as it made the journey through the intersection.
- The red vehicle traveled through the intersection in 100 seconds, with two observed stops. The vehicle experienced significant delay (~88 s), and because the vehicle had to stop two distinct times, it likely experienced a split failure, or had to sit through more than one cycle at the intersection. Logically, these visuals are intuitive, but there are certainly assumptions that need to be made to automate these insights.

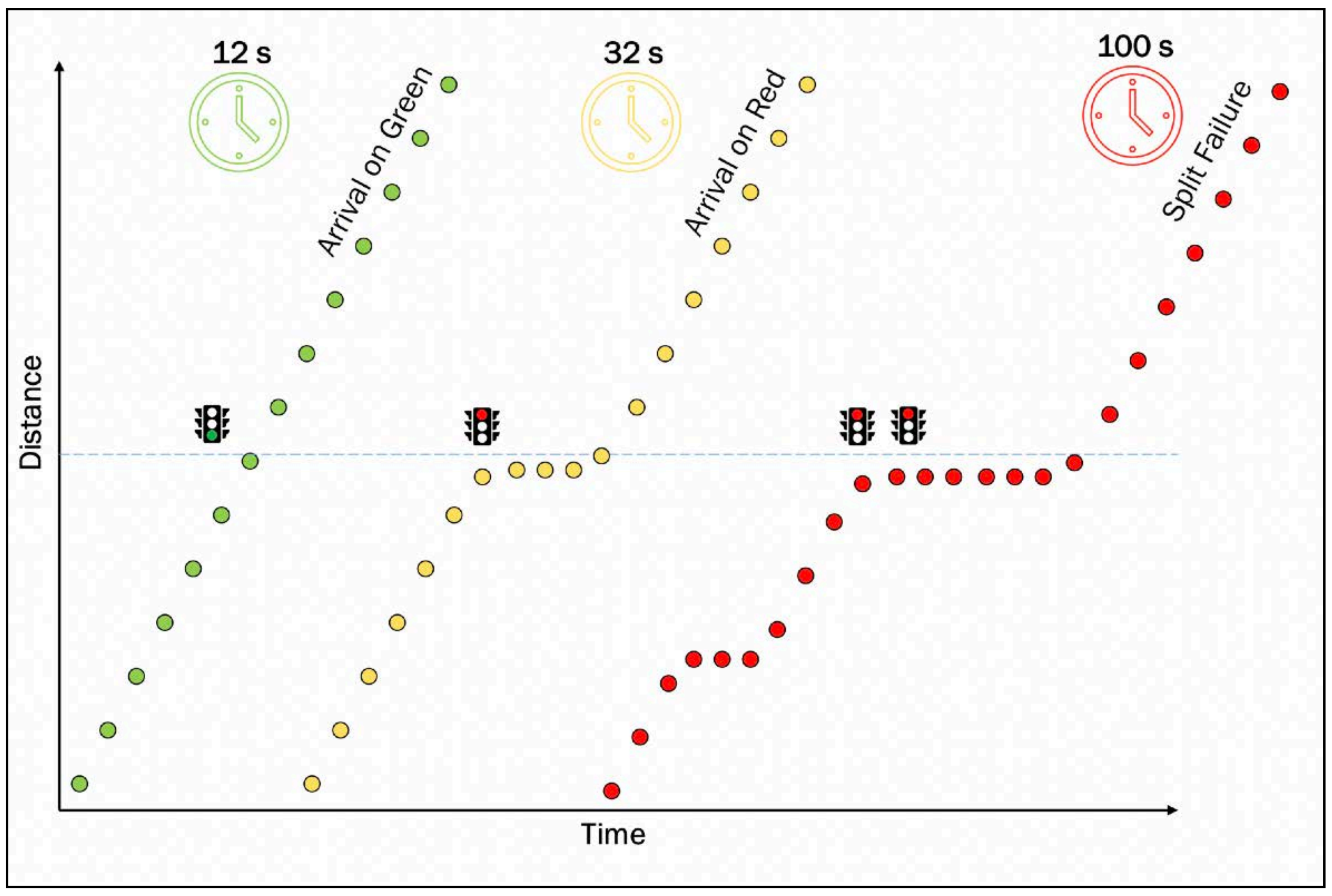


Figure 2. Vehicle Experiences Captured

What assumptions were made and how do you know those assumptions were correct?

A series of assumptions were necessary to create performance metrics at the intersection.

For each vehicle traveling through the intersection the following assumptions were made:

- The intersection metrics consider an inbound length of 150 meters (~492 ft) prior to the stop bar and an outbound length of 80 meters (~262 ft) past the stop bar. The inbound length is used to determine if a vehicle stopped prior to the intersection. Both the inbound and outbound lengths are used to determine the travel time of the vehicle through the intersection.

- A vehicle is considered to have stopped at the intersection if the speed dropped below 10 kph (6.2 mph) for 2 seconds (or one vehicle waypoint) in the inbound length.

- The reference travel time, used to determine a typical travel time through the intersection, is considered the 5th percentile travel time of all vehicles that did not stop while making the same movement during the selected time period.

Figure 4 illustrates the assumptions made above.

These assumptions allow us to consider every vehicle we observe traveling through an intersection and define the characteristics of that vehicle including:

- **Arrival on Green (AOG)** – Arrivals on green represent a vehicle that did not have to stop at a signalized intersection.

- **Travel Time** – The time a vehicle takes to travel the inbound and outbound length of the movement.

- **Approach Speed** - Maximum speed of a vehicle using waypoint pairs on the inbound length of an intersection.

- **Control Delay** - The difference between the actual travel time for a vehicle to move through the intersection versus the reference travel time.

- **Split Failure** – A split failure is defined when a vehicle is forced to stop more than once at a traffic signal.

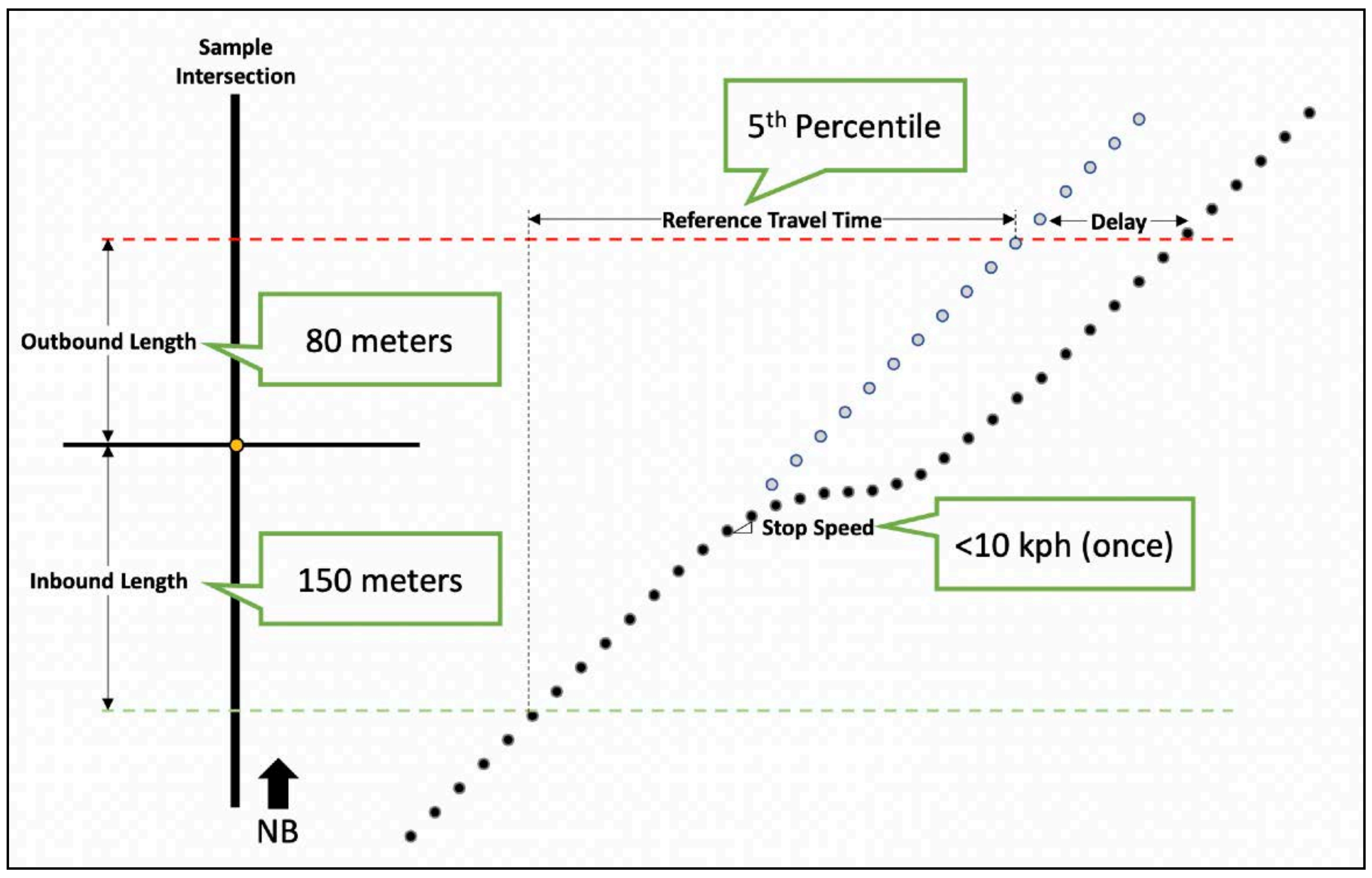


Figure 4. Assumptions for Signal Analytics

We can then take those individual characterizations and aggregate them at the movement (left, thru, right), approach (NB, SB, EB, WB), and intersection (6th and Main) levels.

A preliminary study was performed using a detector based automated traffic signal performance measure (ATSPM) system to confirm these assumptions were providing reasonable metrics for percent arrivals on green (POG), control delay, and split failures. The Utah DOT Automated Traffic Signal Performance dashboard was used to make these initial comparisons.

Examples of Validation for INRIX Signal Analytics Data

Percentage Arrival on Green (POG): Example 1

A series of intersections were investigated and compared using two weeks of trajectory data and a single day of Utah ATSPMs from January 2020. The intersection of 4100 S @ 2700 W was used to try a variety of assumptions including: shifting the inbound and outbound lengths, shifting the stop speed, and shifting the stop duration. It is important to note that the POG measured using the trajectory data and the ATSPM information are not identical methodologies.

The POG using the ATSPM suite relies on advanced detection approximately 400 to 500 feet upstream of the stop bar, an assumed speed of the vehicle, and signal timing information. The POG logic using trajectory data relies on knowing if the vehicle slowed to a stop prior to traveling through the intersection. Initial logic shown in Figure 5b shows the comparison between POG for ATSPM and trajectory metrics when the inbound length was 150m, and the stop threshold was <1 kph, meaning a vehicle had to be observed at 0 kph to be characterized as a stop. A later version of the logic, shown in Figure 5b, loosened the stop threshold to 15 kph, which resulted in a reduction in POG for the trajectory data.

Figure 5. POG by approach and time of day with logic statements

Figure 5a. 4100 S @ 2700 W

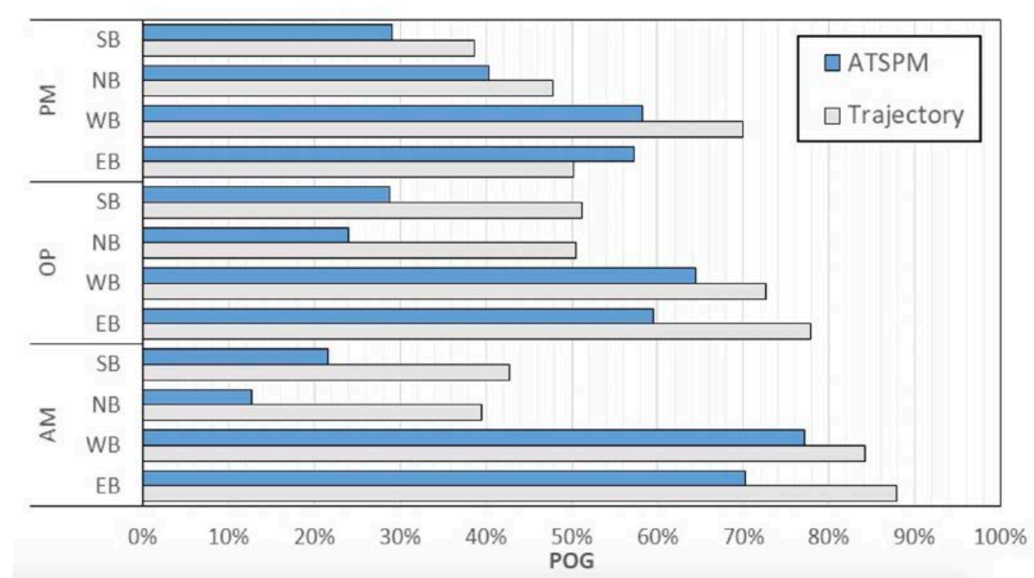


Figure 5b. Early logic

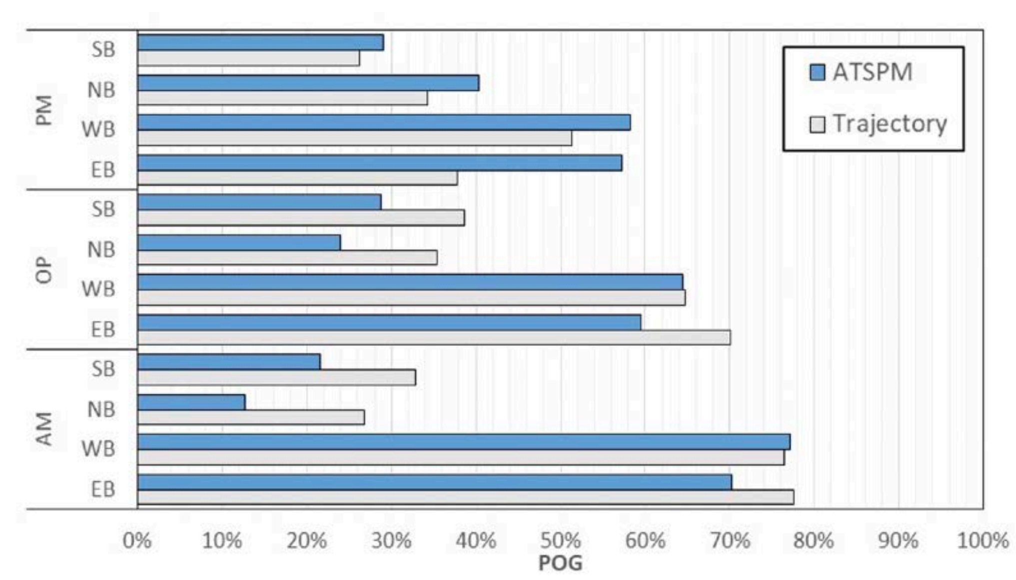
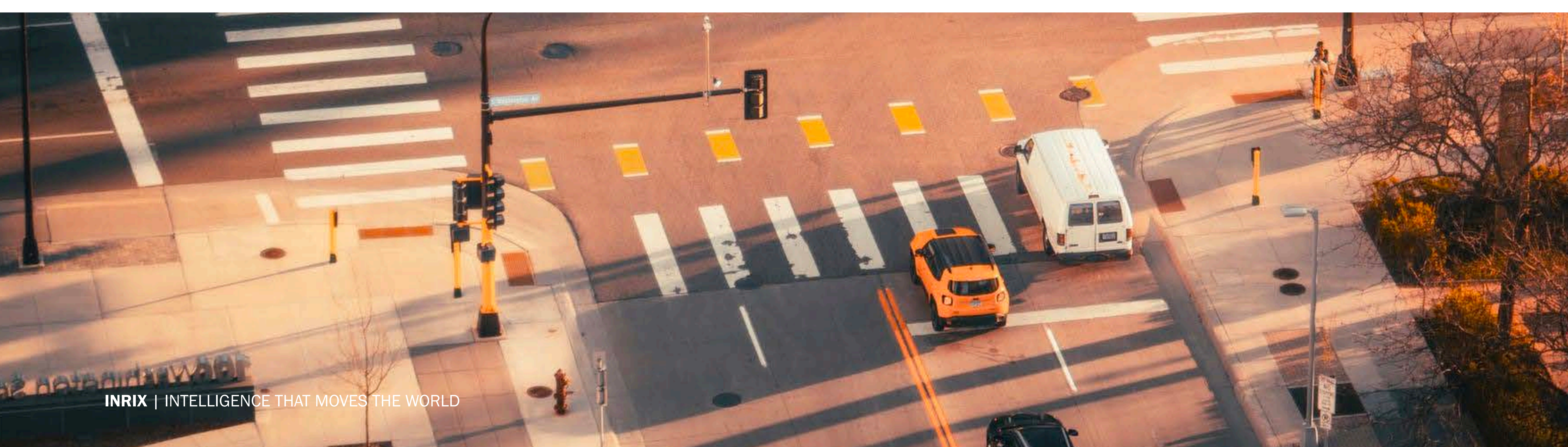


Figure 5c. Later logic



Examples of Validation for INRIX Signal Analytics Data

Percentage Arrival on Green (POG): Example 2

Another example of a comparison between trajectory data and ATSPM data used the intersection of 3500 S @ 3200 West, which is located southwest of Salt Lake City, to perform an analysis of the percent arrivals on green. The EB approach and two weeks of trajectory data were compared with a single day of ATSPM data (in the form of a Purdue Coordination Diagram). The percent differences for each of the timing periods are shown in the callouts (+3%, -2%, -3%, -5%). These values were obtained by taking the POG from the single day of ATSPM and subtracting the POG from the aggregated two weeks of trajectory data.

Figure 6. POG comparison on a single approach using a PCD

Figure 6a. 3500 S @ 3200 W

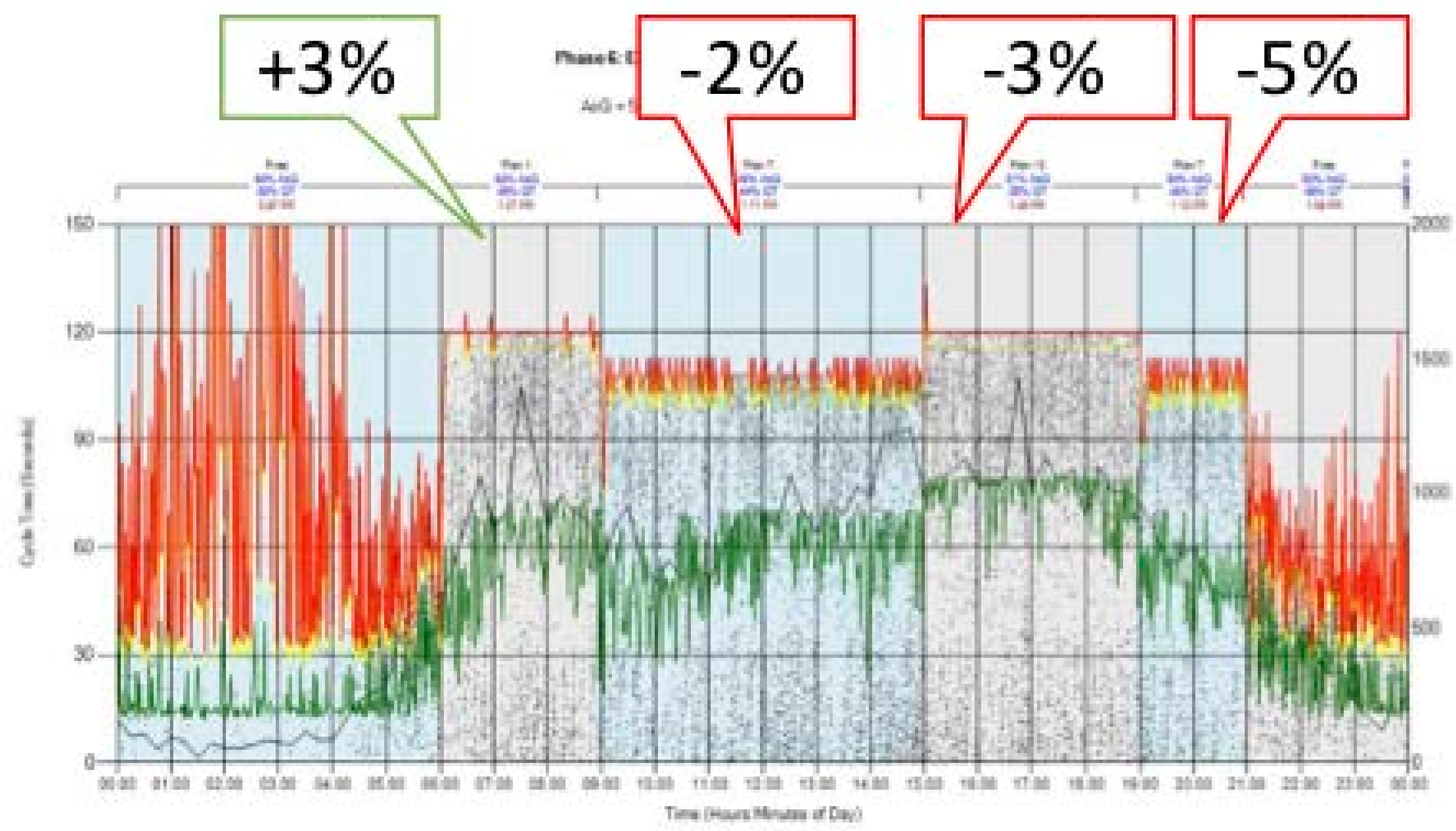


Figure 6b. Purdue Coordination Diagram

Percentage Arrival on Green (POG): Example 3

Another example of a comparison between trajectory data and ATSPM data used the intersection of 3500 S @ 3200 West, which is located southwest of Salt Lake City, to perform an analysis of the percent arrivals on green. The EB approach and two weeks of trajectory data were compared with a single day of ATSPM data (in the form of a Purdue Coordination Diagram). The percent differences for each of the timing periods are shown in the callouts (+3%, -2%, -3%, -5%). These values were obtained by taking the POG from the single day of ATSPM and subtracting the POG from the aggregated two weeks of trajectory data.

Figure 7. POG comparison on a single approach using a PCD

Figure 7a. 5400 S @ 2700 W

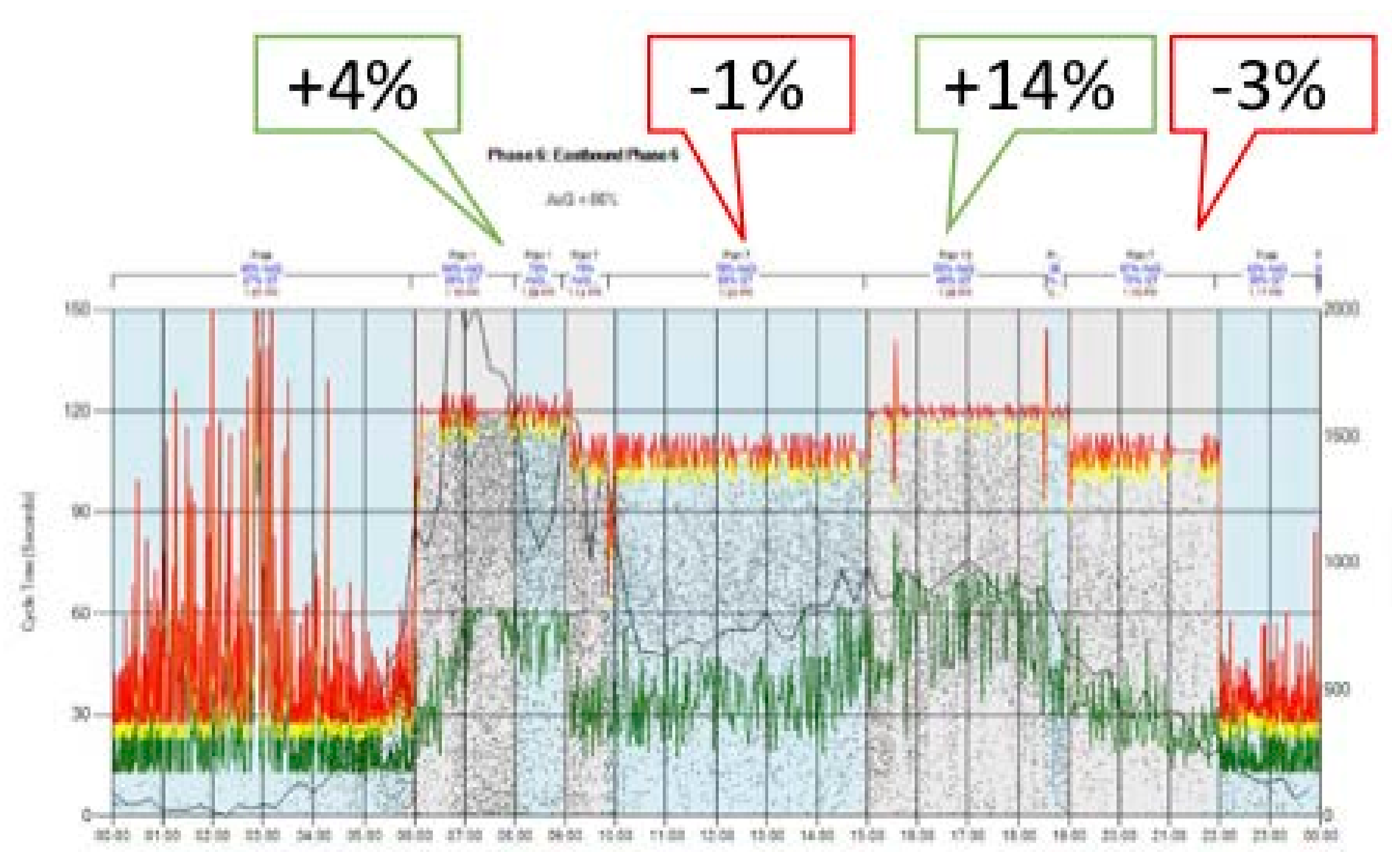


Figure 7b. Purdue Coordination Diagram

Examples of Validation for INRIX Signal Analytics Data

POG Analysis Shows Little or No Bias In Trajectory Data Based Metrics

The entire POG analysis used data available from 10 intersections in the Salt Lake City area to gain a comfort level with the metrics that were being produced using the trajectory data. The final comparison leveraged 100 comparisons of different intersections, times of day, and direction. The average difference was 0% and the average percent difference was also 0%. This highlights that there was little to no bias in the trajectory data metrics. The absolute average difference was 5.6%, showing there was some variability between the ATSPM measurements and the trajectory-based metrics. This difference can be attributed to numerous factors including comparing one day of ATSPM data to two weeks of trajectory data, as well as uncertainty in the locations of the detection at the intersection and what movements were being captured.

Figure 8. Possible irregular lane usage on the EB approach.

Control Delay

A similar comparison of control delay was also performed. The control delay values from the trajectory data were almost always higher than those of the ATSPM system. This makes sense as the ATSPM system uses a simplified delay approach where the 'delay' metric is only capturing the time between the detector activation and when the signal turns green, while the trajectory-based metrics include the deceleration and acceleration time. When comparing thru and right movements with the delay produced by the ATSPMs, the delay produced by the trajectory data was 78% higher on average. Although this appears to be a large number, the ATSPM delay calculation is crude compared to the measured trajectory values.

Examples of Validation for INRIX Signal Analytics Data

Split Failures

The final metric we processed for some initial insights was the split failure measurement. The southbound left turn movement of the intersection at 300W and North Temple in Salt Lake City was used to demonstrate our split failure approach.

We define a split failure as a vehicle that must stop twice prior to the intersection. The ATSPM approach to identify split failures relies on the red occupancy ratio (ROR) and the green occupancy ration (GOR) of the stop bar detection at the intersection. If the detector is occupied for over 85% of time for the entire green phase and over 85% of the first five seconds of the red phase, a split failure is assumed to have occurred. Using the southbound left turn shown in Figure 9, the Purdue split failure graph was taken for a weekday in January 2020 (pre-covid – Figure 10a) and a weekday in April 2020 (during covid – Figure 10b).

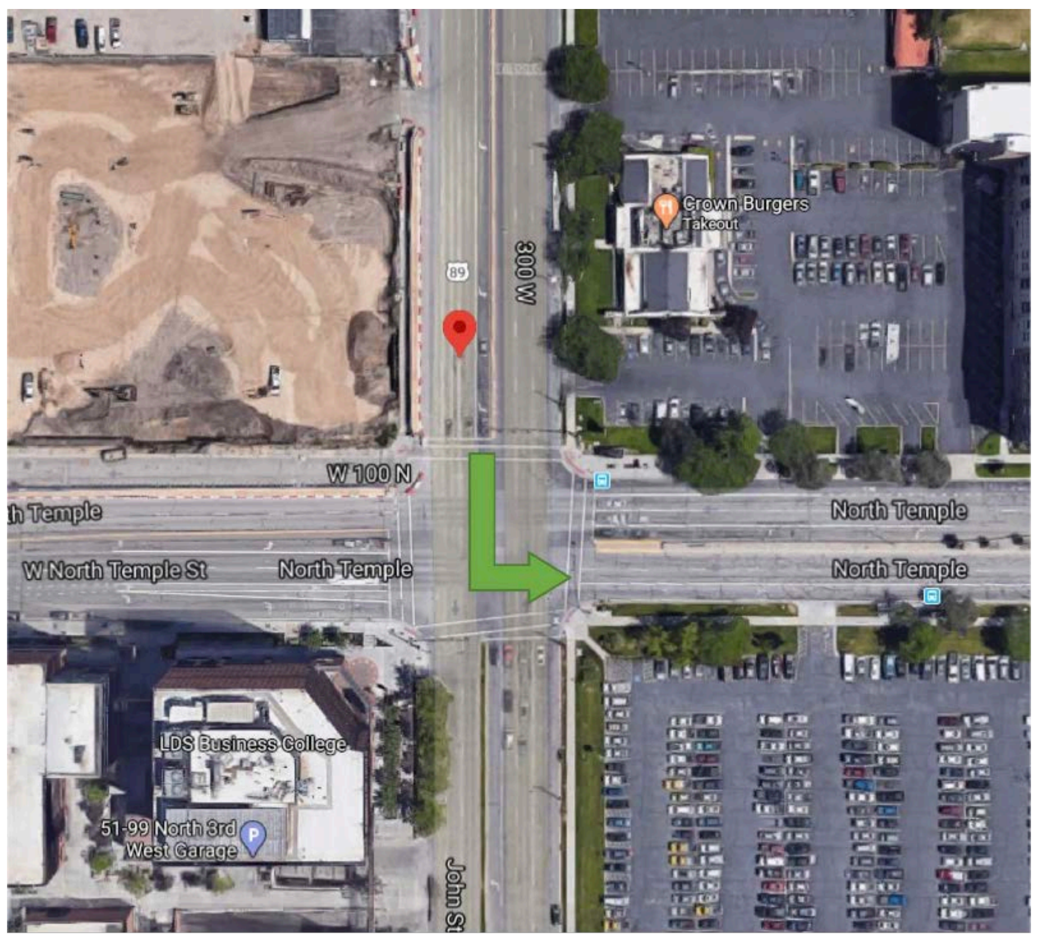


Figure 9. 300 W and North Temple Intersection.

Figure 10. Split Failure Approach

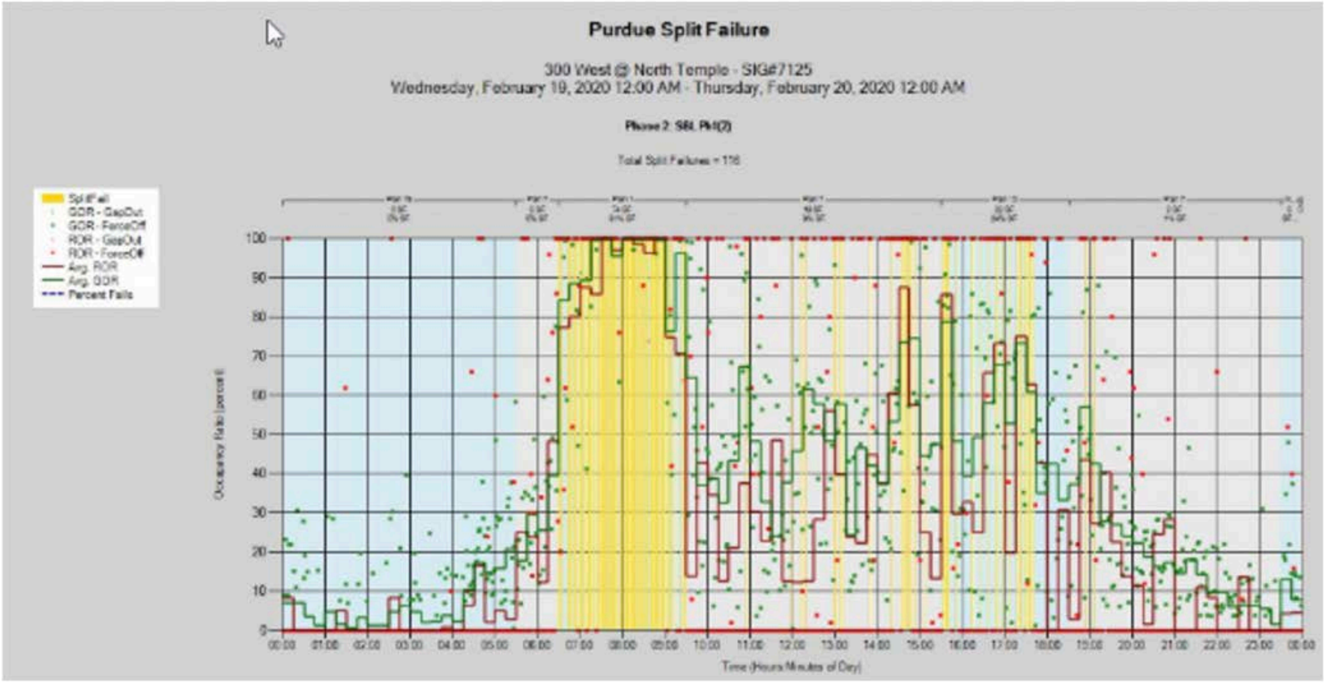


Figure 10a. Purdue split failure diagram (pre-covid)

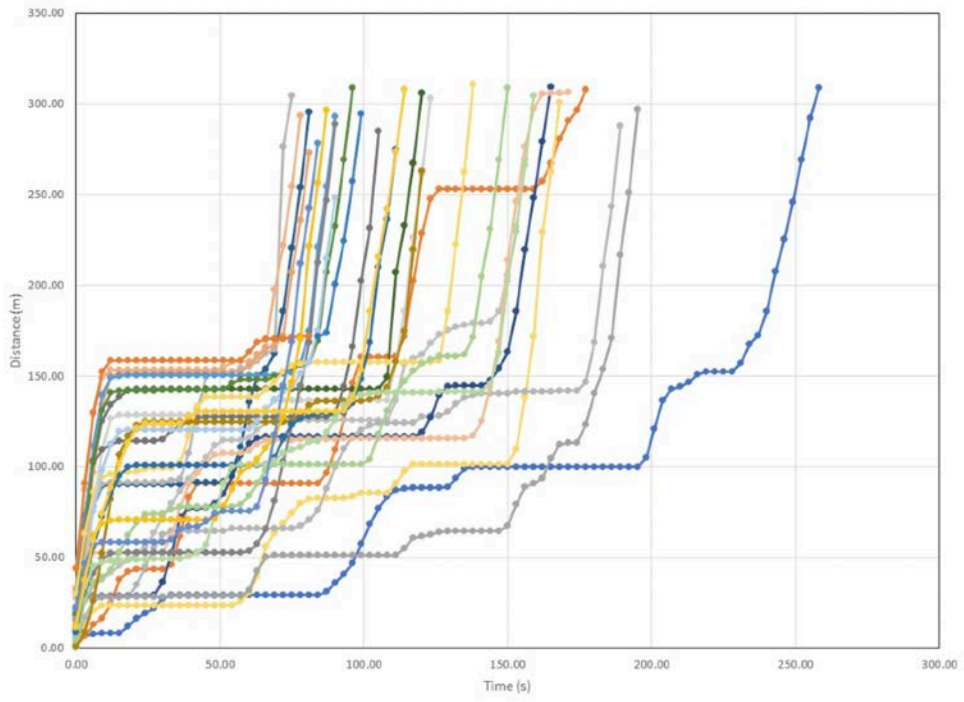


Figure 10b. Corresponding trajectory data (>1 stop)

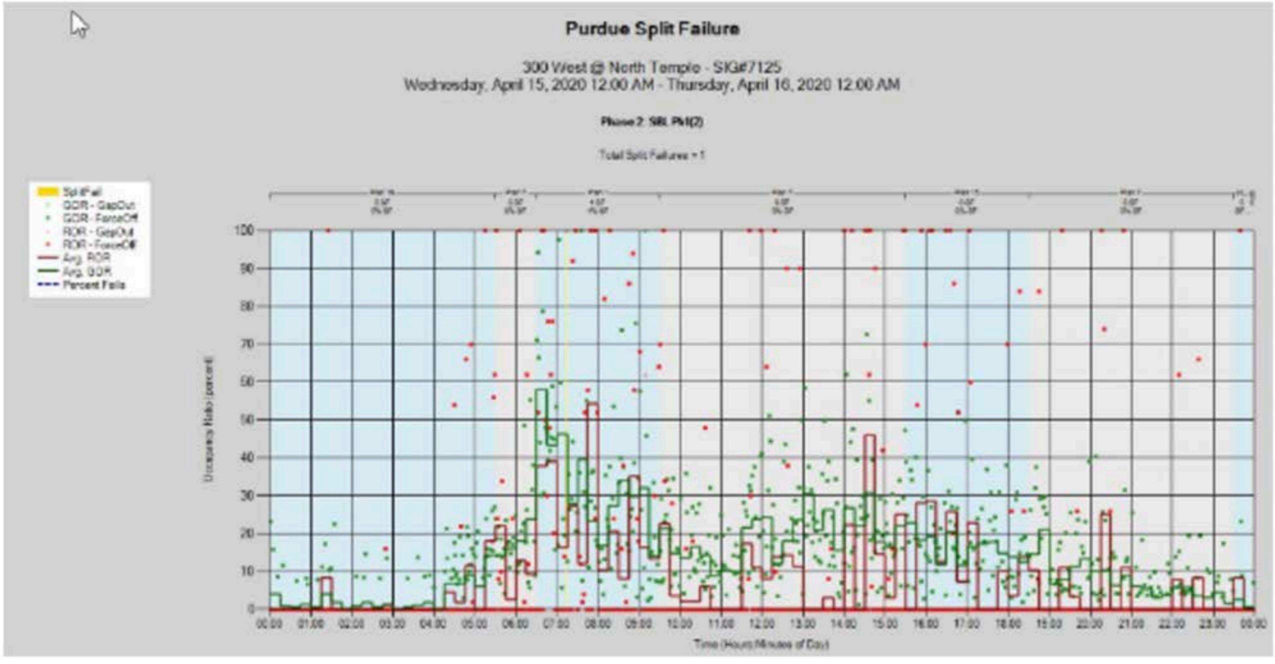


Figure 10c. Purdue split failure diagram (during-covid)

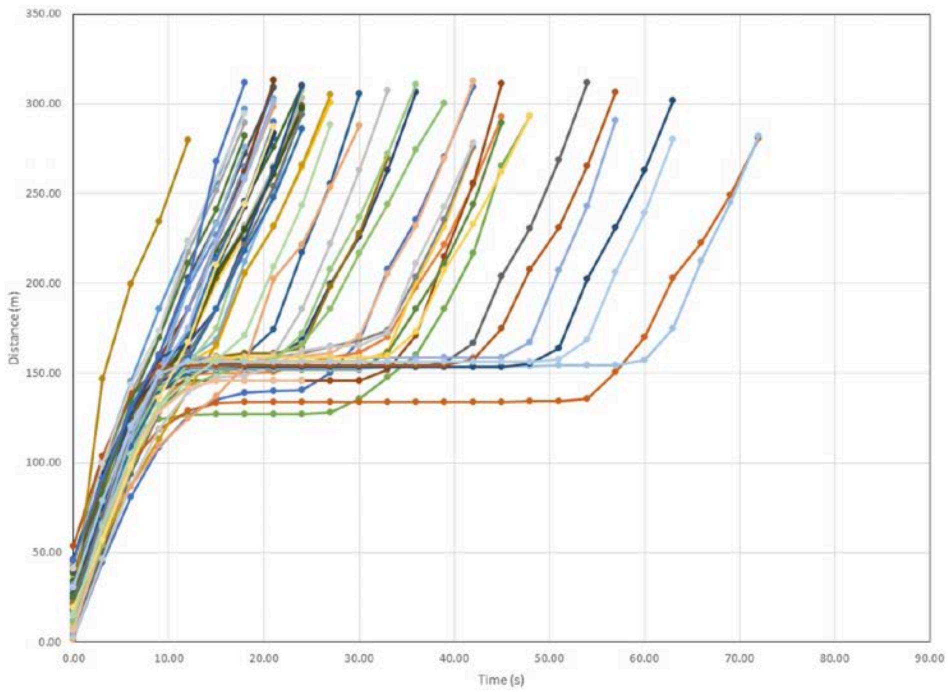


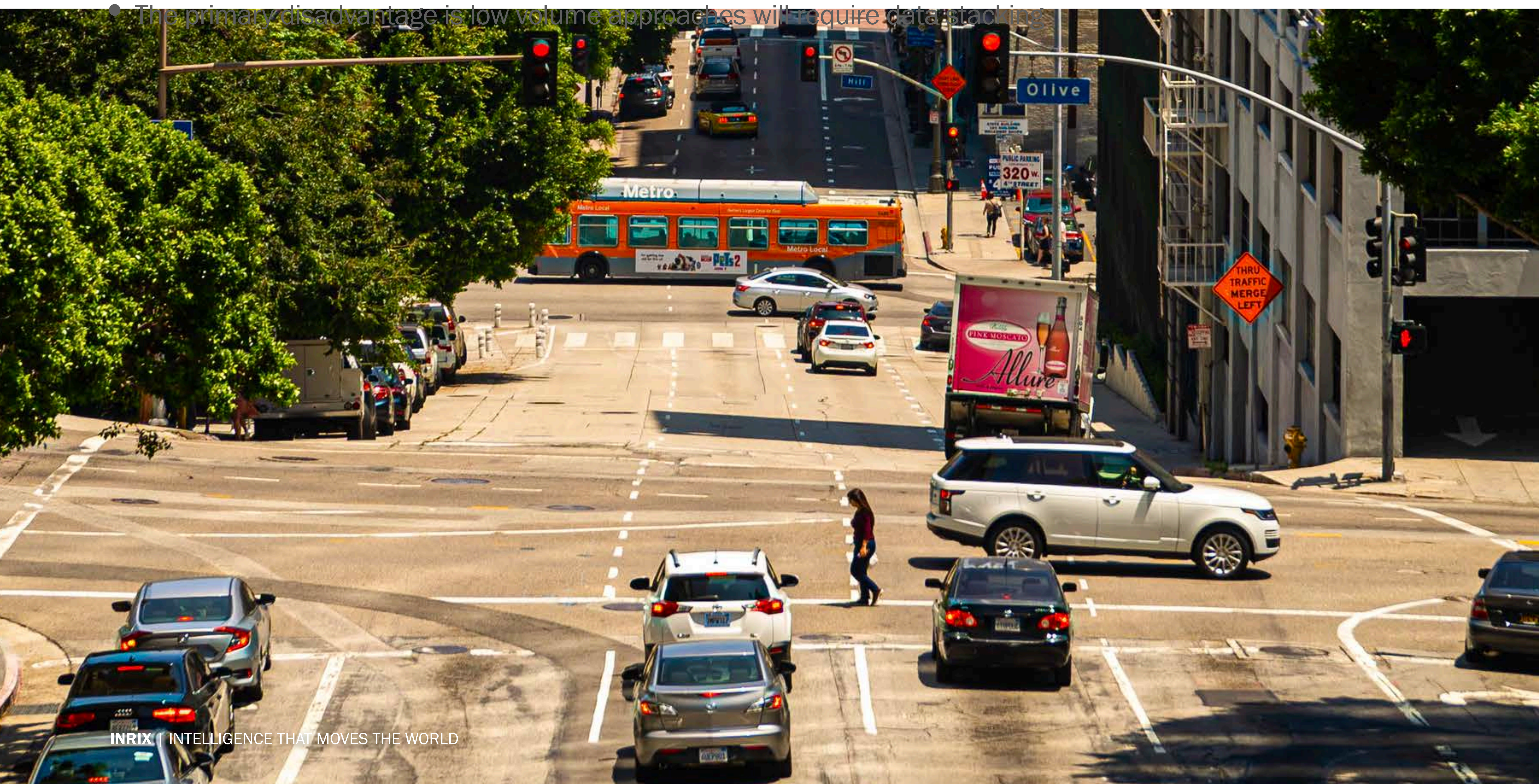
Figure 10d. Corresponding trajectory data (>0 stops)

The yellow lines in these figures suggest a split failure has occurred. Figure 10b and 10d show the corresponding trajectory data for vehicles who have experienced one or more stop during that week. Figure 10b shows the corresponding trajectories that experienced greater than one stop, which can visibly be interpreted as numerous split failures occurring. Figure 10d shows the same movements where one or more stops have occurred after demand patterns changed as a result of the pandemic.

Summary of Validation for INRIX Signal Analytics Data

In summary our initial assessment of the trajectory-based Signal Analytics resulted in the following conclusions:

- Comparing ATSPM and trajectory-based metrics is not an apples-to-apples comparison.
- Throughout the validation effort trends have generally aligned, outliers do exist.
- Levers exist to move the data, but consistent parameters will provide a systematic comparison
  - Increasing the outbound and inbound lengths increases the control delay values
  - Decreasing the stopping speed decreases the number of “stops”
- Trajectory based metrics have numerous advantages:
  - Better delay approximation (includes deceleration and acceleration delay)
  - Better “turning movement” understanding – the user has the ability to understand the metrics at each movement level
  - Avoids having to make certain assumptions about the vehicles trip before or after it drives over the detection
- The primary disadvantage is low volume approaches will require data stacking



How much data are available?

The Signal Analytics platform uses data that currently covers all 50 states. Nationwide, a snapshot for the month of January 2020 has 1.05 BILLION trips and 9.75 BILLION Miles Traveled, which is approximately 3.8% of the total national VMT from FHWA. Figure 3 below shows a map of the coverage by state for January 2020.

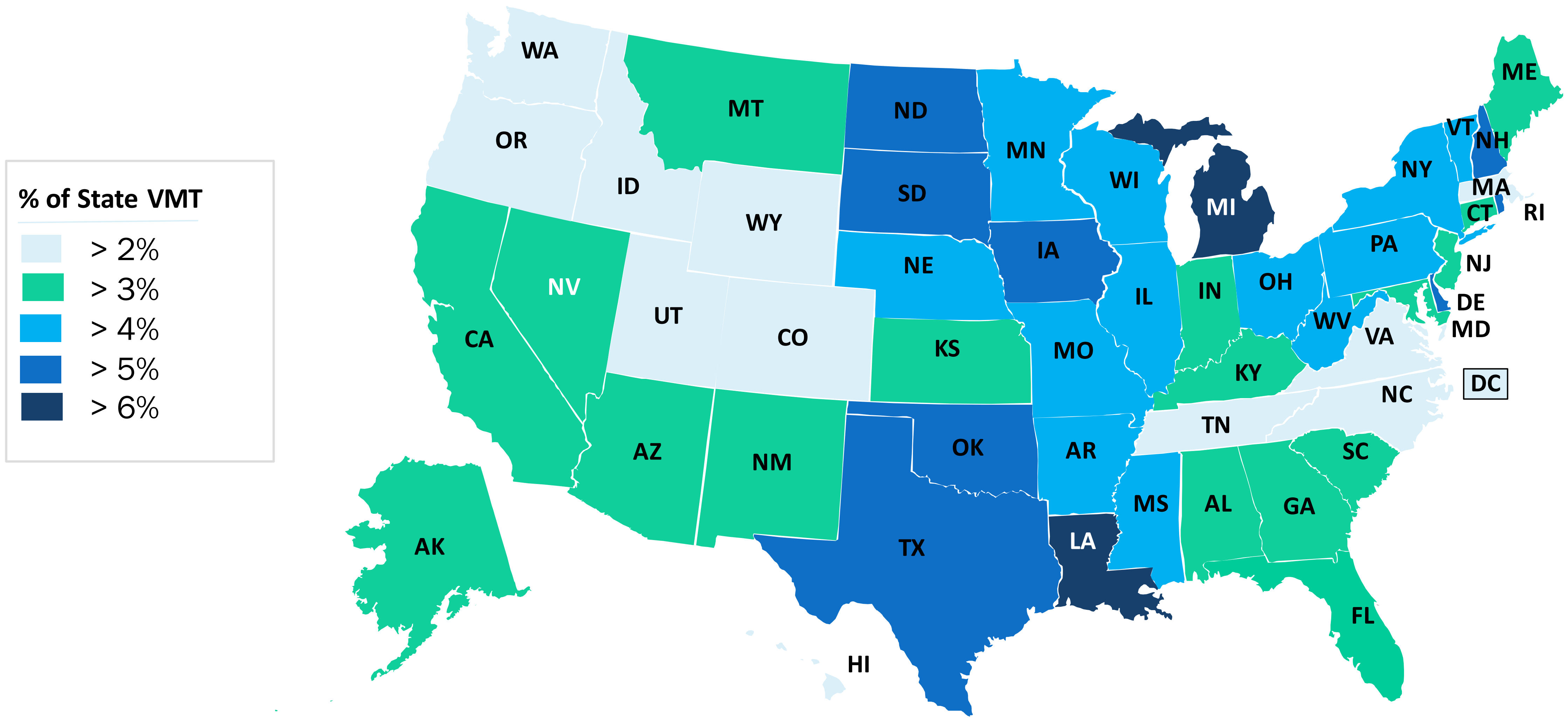


Figure 11. Data Coverage as % of State VMT (~3.8% of Total National VMT)

How much data do you need?

There has often been discussion around how many samples are necessary versus how many can be affordably collected when discussing traffic metrics such as travel time and delay. We will not dive too deep into statistics, but sample size is always an interesting question. The variables that need to be understood to determine an adequate sample size are:

- Margin of Error (E) – The acceptable margin of error of a particular metric is a delay value plus or minus 10 seconds acceptable, or does it need to be closer to plus or minus 5 seconds for a given purpose)
- Confidence interval or Z-score (Z)– How confident you want to be that the mean falls within the margin of error (common intervals are 90, 95 and 99%)
- Standard Deviation(σ) – the standard deviation of the population. (a measure of the variation of the population)

Assuming we have a somewhat large population (n > 30 vehicles), we can assume the sample standard deviation is the same as the population standard deviations. We can then leverage the confidence interval to create a formula that would estimate the total sample required (N):

$$N = \left(\frac{Z\sigma}{E}\right)^2$$

Data Validation Example: Austin, Texas

While investigating a movement during the PM peak period (4 PM – 7 PM), how many vehicles would need to be observed to obtain a delay value +/- 5 seconds with a 95% confidence level. The number of samples necessary would depend on the standard deviation of the population. If the standard deviation were 10 seconds, 15 samples would be required. If the standard deviation were five seconds, four samples would be required. If the margin of error were expanded to +/- 10 seconds, four samples and one sample would be needed, respectively.

Using this basis, the question may change from "do we have enough data to draw a conclusion?" to "how long do we have to collect data to draw a meaningful conclusion?".

Using the above logic 15 samples would be required to determine average delay +/- 5 seconds at a 95% confidence interval or four samples at +/- 10 seconds in the PM peak hour. Using 11 intersections on Congress Avenue (Figure 12) in Austin, Texas as an example, counts were pulled for one day, one week, and one month (Table 1, Table 2, and Table 3).

Green values are where counts exist that could provide a delay value +/- 5 seconds of the expected value. Yellow values are where counts exist that could provide a delay value +/- 10 seconds of the expected values. Red values are where there are not sufficient counts to make any of the claims above. As expected, the main north/south movement along the corridor can be sufficiently assessed using one single day of data, while side street movements may need to be stacked over time to develop statistically significant results.

Figure 12. 11 intersections on Congress Avenue in Downtown Austin, TX

	Westbound			Northbound			Eastbound			Southbound		
Intersection (1 day)	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right
Congress Avenue & West Cesar Chavez Street	24	44	10	10	34	10	8	68		0	48	9
Congress Avenue & West 2nd Street	0	0	0	5	40	5	0	0	0	0	56	0
Congress Avenue & East 3rd Street	0	0	0	0	37	0	0	0	0	0	54	0
Congress Avenue & East 4th Street	0	0	0	0	32	0	0	5	0	0	44	0
West 5th Street & Congress Avenue					32	6	5	58	7	0	42	
Congress Avenue & West 6th Street	13	44	5	10	30						31	0
Congress Avenue & East 7th Street					25	10	0	32	0	6	32	
Congress Avenue & West 8th Street	9	13	0	0	22						29	0
Congress Avenue & East 9th Street					22	0	0	0	0	0	25	
Congress Avenue & West 10th Street	11	8	0	0	19						14	0
West 11th Street & Congress Avenue	6	20		10		10		21	8			

Table 1. PM Peak Observed Counts on Congress Avenue (1 day – April 7, 2021)

	Westbound			Northbound			Eastbound			Southbound		
Intersection (1 week)	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right
Congress Avenue & West Cesar Chavez Street	108	242	23	48	197	77	39	308		24	216	36
Congress Avenue & West 2nd Street	12	11	9	17	208	27	9	14	7	10	240	8
Congress Avenue & East 3rd Street	5	8	0	7	202	22	5	16	11	5	235	14
Congress Avenue & East 4th Street	13	0	8	0	189	17	10	17	14	5	210	5
West 5th Street & Congress Avenue					179	38	41	233	26	12	195	
Congress Avenue & West 6th Street	55	232	23	60	163						148	20
Congress Avenue & East 7th Street					119	64	10	153	15	34	154	
Congress Avenue & West 8th Street	54	59	8	19	105						137	6
Congress Avenue & East 9th Street					94	14	9	15	28	8	107	
Congress Avenue & West 10th Street	33	36	0	15	90						67	0
West 11th Street & Congress Avenue	30	81		35		59		78	31			

Table 2. PM Peak Observed Counts on Congress Avenue (1 week – April 5 - 9, 2021)

	Westbound			Northbound			Eastbound			Southbound		
Intersection (1 month)	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right	Left	Thru	Right
Congress Avenue & West Cesar Chavez Street	457	1123	115	236	837	358	144	1392		113	840	176
Congress Avenue & West 2nd Street	57	53	56	59	910	99	38	100	49	46	929	48
Congress Avenue & East 3rd Street	14	18	10	31	885	94	33	80	43	34	934	44
Congress Avenue & East 4th Street	51	20	27	42	833	70	34	69	61	17	821	58
West 5th Street & Congress Avenue					788	146	157	1102	170	66	734	
Congress Avenue & West 6th Street	226	1079	123	287	680						574	84
Congress Avenue & East 7th Street					470	313	37	624	79	119	560	
Congress Avenue & West 8th Street	207	251	34	50	441						479	14
Congress Avenue & East 9th Street					407	57	32	79	100	40	383	
Congress Avenue & West 10th Street	142	146	9	48	394						228	7
West 11th Street & Congress Avenue	130	319		169		250		337	90			

Table 3. PM Peak Observed Counts on Congress Avenue (1 month – April 2021)

Conclusion: Trajectory Data Provides Reliable Insights

The message here is that it is possible to draw meaningful conclusions using trajectory data aggregated at the movement level. The main factor is understanding the variability, or standard deviation, of the metric you are trying to measure. Figure 13 shows some examples of required sample sizes based on the expected standard deviation of the metric. The assumed standard deviation value of 10 may be true for reliable movements, while higher values may be necessary for less reliable intersections. Stacking data over time can grow the sample to be statistically significant at some level for a movement of interest.

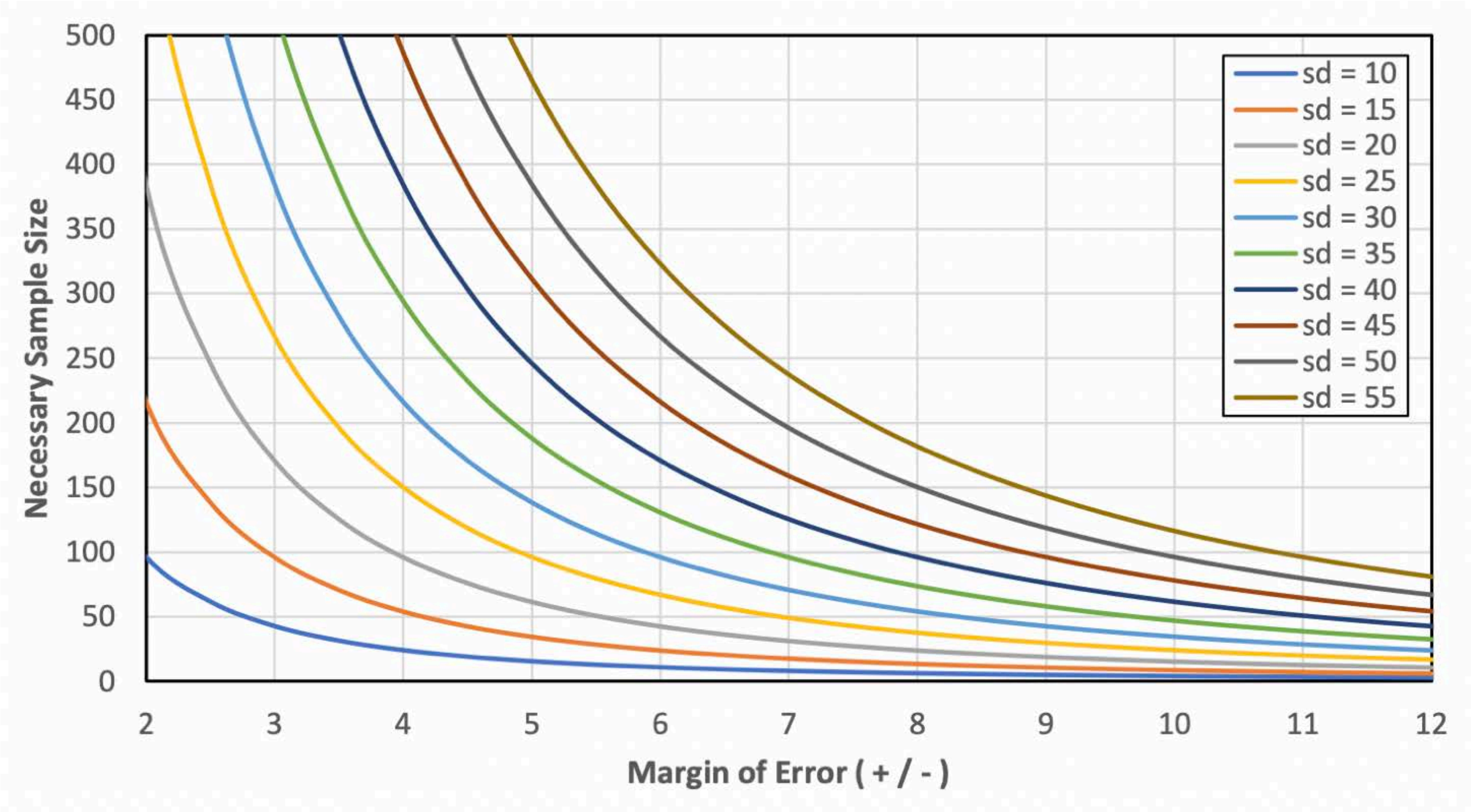


Figure 12. Sample size necessary for certain margins of error given the standard deviation

INRIX IQ Signal Analytics empowers everyone from traffic engineers to maintenance technicians to identify and solve performance issues faster, smarter and safer – all without ever leaving their desks. By utilizing anonymous data from connected vehicles, Signal Analytics eliminates the traditional cost and time expenditures for traffic signal improvement.

Go to INRIX.com/signals to sign-up for more details and to get a free demo or claim your free 14-day trial at iq.inrix.com.

INRIX

For more information contact us at sales@inrix.com