Appendix F | Technical Details for Commercial Speed Data Processing

The commercial speed data processing, which ultimately converted the raw Traffic Message Channel (TMC) link data into average peak period speeds on every CMP segment, consisted of four steps described below.

Figure F-1: Data Analysis Procedures for Commercial Speed Data

Further explanation of each step is provided below.

F.1 | Step 1. Mapping TMC links onto CMP Segments

Commercial speed data collected by INRIX was reported against lengths of roadway called TMC links. TMC links are typically short links of roadway averaging 0.4 miles in length (range: 19 feet to 4.2 miles)[38].

For this project, it was required that the average speed be reported against an Alameda CTC CMP segment. CMP segments are typically longer segments of roadway averaging approximately 1.2 mile in length (range: 0.2 to 5.0 miles).

Therefore, TMC links needed to be aligned against or mapped onto the CMP segments. This mapping was created as a part of the 2013 validation project and updated for the 2014 and 2016 TMCs.

It should be noted that for some CMP segments, the ends of the CMP did not align with the ends of the TMCs. **Figure F-2** shows a schematic example to explain this concept. It shows one CMP segment that is made up of four TMC links. However the end of the last TMC link does not align with the end of the CMP segment. In these instances, only the overlapping portion of the TMC length was used to calculate the average speed.

Figure F-2: End points of CMP and TMC do not align

[38] TMC length statistics are based on TMCs used in this monitoring project.

F.2 | Step 2. Filter Raw Data

The raw INRIX data was filtered to remove:

- Times outside the morning and afternoon peak periods;
- Days other than Tuesdays to Thursdays;
- Data points impacted by special events i.e. spring break, incidents, construction, major sporting events; and
- Data points with lower data quality scores.

INRIX includes a data quality score that accompanies every INRIX data point. The score value is defined as:

- Score of 30: Data are exclusively generated from real-time sources.
- Score of 20: A mix of historical and real-time sources are used.
- Score of 10: Data are exclusively generated from historical data.

Only raw speeds that were directly measured were used for computing LOS in the CMP network. As such, data points with scores of 10 and 20 were removed, and only data with a score of 30 were used.

The quantity of remaining data points was tracked so the sample size of score 30 was known. The sample sizes are presented in conjunction with all associated commercial speed data results.

Note that Steps 2 and 3 were undertaken using the open source software R. This software is widely used in data analytics and statistics for managing medium size quantities of data (as was the case in this project). Datasets of this size would be difficult to manipulate in a spreadsheet program. Iteris wrote R scripts that performed these processes.

F.3 | Step 3. Spatial and Temporal Data Aggregation - Average Speed Computations

This section discusses the methodology of aggregating the data both spatially and temporally. The input to this step was 11 million data points of INRIX speed data. **Table F-1** displays two such sample data points. The output from this were the average speed and sample size of each CMP segment. A sample of the output is included in **Table F-2**.

Table F-1: Sample INRIX Input Data

TMC Code	Time Stamp	Speed (mph)	Travel time (min)	Score
105+04359	2016-03-01 07:00:39	69	1.17	30
105N04358	2016-03-01 07:00:39	65	0.59	30

Table F-2: Sample Output from Step 3 – Average Speed on CMP Link

ID	CMP Route	Jurisdiction	Length (mi)	Sample Size	Speed (mph)
F1	I-80 – EB: SF County Line to Toll Plaza	Oak	2.01	2795	62.2
F2	I-80 – EB: Toll Plaza to I-580 SB Merge	Oak	1.3	1889	63.2

The following steps describe how the dataset was restructured to obtain the results in **Table F-2**. This involved spatial and temporal aggregation.

F.3.1 | Spatial Aggregation

Using the mapping created in Step 1 and the filtered INRIX data from Step 2, the TMC data was spatially aggregated on the CMP segments. In cases where multiple TMC links span a single CMP segment, the travel time was summed for all TMCs.

$$CMP\ Travel\ Time = TMC_1 + TMC_2 + \cdots + TMC_n$$

F.3.2 | Temporal Aggregation

Temporal aggregation involved the translation of the CMP travel time metric for each minute of data into one average speed value corresponding to each CMP segment for the entire monitoring period. The following formula was used for this:

$$Average\ CMP\ Speed = \frac{\sum CMP\ Length}{\sum CMP\ Travel\ time}$$

Sample size information was retained to assess the confidence level in the computed statistics.

F.3.3 | Sample Size

The sample size is the number of data points that contributed to the final calculation of average speed. The sample size varied on each TMC through removal of data points during the filtering process and through the processes discussed below.

Removal of TMC data points with scores of 20 and 10 (Step 2 above) eliminated data for particular one-minute time periods from one or more of the TMCs that comprise certain CMP segments. The example shows a longer CMP segment which is comprised of four TMCs. The table shows the data scores for each TMC for each one minute time period. In time periods 1, 2, and 7, one of the TMCs had a data score of 20 and therefore the record from that TMC was excluded for those

minutes. In time period 6, two of the TMCs had data scores of 20 and similarly, these TMC records were also excluded for time period 6 (**Figure F-3**).

Time period	TMC1	TMC2	TMC3	TMC4	Length
1	20	30	30	30	¾
2	20	30	30	30	¾
3	30	30	30	30	1
4	30	30	30	30	1
5	30	30	30	30	1
6	20	30	20	30	½
7	30	30	20	30	¾

Figure F-3: Example of Filtering Process

Iteris performed a check to ensure that any time periods that had too many TMCs removed were not included in the analysis. Where TMC data were available for less than 99% of the TMCs that were chosen for mapping, that one-minute time period was removed. To extend the above example further, if TMC1 was less than 1% of the CMP segment length, then it would still be possible to use the data in Time periods 1 and 2 (in addition to time periods 3, 4 and 5). This can be justified, because TMC1 does not contribute significantly to the distance-based average speed calculation.

In a small minority of cases, using the 99% threshold resulted in removal of too many time periods and an inadequate sample size. In these cases, the threshold was lowered to 70% to ensure that the sample size was adequate. A minimum sample size of 50 was used.

The remainder of this section gives information about the sample sizes observed on all CMPs. Note that there are 327 CMP segments measured with commercial speed data each having an AM and PM measurement of average speed. This totals 654 measurements.[39] **Figure F-4** shows a frequency plot of the sample sizes obtained for each CMP (AM and PM recorded separately). For example, there were 78 CMP measurements that had a sample size between 1000 and 2000 data points. The data points with lower sample sizes were typically located on the arterial network (Tier 2).

[39] Segments measured using floating car surveys were excluded from this analysis of sample size.

The assumptions made by Iteris in this section have been confirmed with Alameda CTC for their reasonableness.

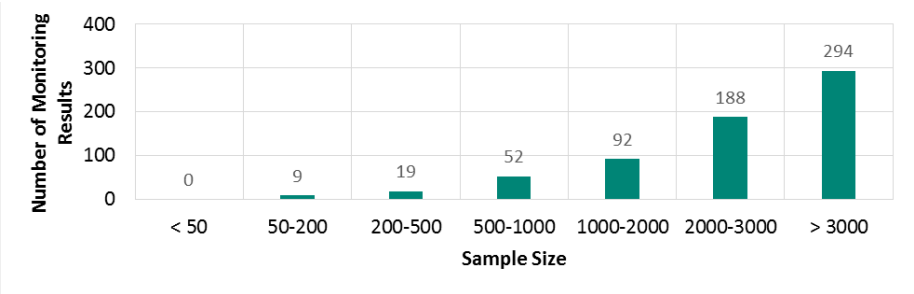

Figure F-4: Histogram of Sample Sizes for CMP Segments Monitored using Commercial Speed Data (Tuesday to Thursday time period)