A technical memorandum to

Mobility Measurement in Transportation (MMUT)

FHWA Pooled Fund Study

Submitted by the

TEXAS A&M TRANSPORTATION INSTITUTE

TECHNICAL MEMORANDUM

Quality of NPMRDS Data on Lower Volume Roadways

Task 1 – Evaluate Mobility Datasets

Authors:

Kartikeya Jha

Shawn Turner

Texas A&M Transportation Institute

August 2019

[This page intentionally left blank.]

EXECUTIVE SUMMARY

Speed data are used for a variety of transportation and traffic engineering analyses. This memorandum summarizes the findings of a study examining the nature and magnitude of differences between speed data captured by permanent monitoring stations installed by state departments of transportation (DOTs) and corresponding speed data in National Performance Measurement Research Data Set (NPMRDS) on lower volume roadways in three states – CO, TX and SC. The key findings of this technical memorandum are listed below, and more discussion is provided in the body of the memo.

Key Points

- Vehicle speeds at permanent monitoring stations were consistently higher than NPMRDS speeds by about 8 mph on average. These differences varied between sites and states. Colorado showed the highest differences in speeds (11 mph on average) while for Texas and South Carolina these differences were comparable and relatively lower (around 6 mph on average).
- The absolute percent difference for these speed data ranged between 15% for CO and around 10% for TX and SC.
- Overall, speed differences were higher on lower-volume arterial roads (9.2 mph on average) than on higher-volume Interstate highways (7.1 mph on average).
- The consistent differences in speeds recorded by these two speed data sources can be attributed to a few different factors. Terrain (topography along the roadway segment) is one of the contributing factors to speed bias, but not the only factor. Variable terrain often increases the differences between the two speed observations, but sometimes other contributing factors show a higher overall effect than terrain. The other influencing factors can include the difference in methodologies for capturing speed data and potential difference in vehicle fleet composition in captured sample leading to variable speed records in response to congested periods of travel.
- The effects of roadway segment length, amount of traffic volume (AADT) and proportion of missing observations were not found to be consistent or conclusive. It was expected that long segment lengths, low traffic volume and high percentage of missing observations could potentially increase the speed differences, but their effects on speed bias and mean absolute percent difference values were not observed to be as strong as anticipated.

The findings can help agencies get a deeper understanding of the quality of NPMRDS travel time data on their lower volume roadway network and the suitability of one database over the other depending on their intended uses. For instance, based on these results, permanent monitoring station data can be useful for safety analyses or to identify operating speeds for road design at specific identified locations on lower volume roadways vis-à-vis determining free-flow speeds for an extended segment of roadway, where NPMRDS speed database can be useful. However, agencies should be mindful of the underlying biases in NPMRDS data while using it in these situations. The study highlights the merits and limitations of both data sources, and underlines their applicability based on use case.

INTRODUCTION

The National Performance Measurement Research Data Set (NPMRDS) is a national database of recorded travel times on all road segments located on the National Highway System (NHS) network. The NPMRDS travel times are collected from Global Positioning System (GPS) enabled mobile devices (e.g., smart phones, personal navigation devices, truck fleet tracking and telematics systems, etc.) in the traffic stream. Some limitations of NPMRDS data, particularly on lower volume roadways, have been recognized by other studies *(1)*. This research examines the magnitude and nature of differences between traffic stream speeds measured at permanent monitoring stations and speed data for corresponding time periods in NPMRDS on lower volume road segments. Measured speeds are used for several activities including roadway system performance monitoring, roadway design, deriving free-flow speeds, identifying operating speeds for safety analyses, and funding allocation and infrastructure management strategies. Therefore, it will be helpful for state departments of transportation (DOTs) and agencies to get a deeper understanding of the quality of NPMRDS travel time data on their lower volume roadway network.

METHODOLOGY AND DATA DESCRIPTION

Lower volume sites (in terms of Annual Average Daily Traffic, AADT) were identified in three states – Texas (24 sites), Colorado (30 sites) and South Carolina (8 sites). The AADT for all analyzed locations ranged between 2,000 and 30,000 to cover a broad range of traffic volume levels. Speed records for hourly aggregated time intervals were collected from NPMRDS and permanent monitoring station databases separately in both directions of travel. Hourly aggregated speeds were utilized to allow for one-to-one comparisons between NPMRDS and DOT-collected speeds because the DOT-collected (permanent station) speeds were available hourly aggregated level *(2-4)*. Any data points where hourly averaged speed data pairs for both station-collected speed data and NPMRDS speeds were not available for any particular hour were excluded. This resulted in 32,918 one-hour aggregated speed comparisons.

The permanent monitoring stations are at fixed locations and therefore collect spot speeds of vehicles passing over installed detectors. These spot speeds are also referred to as time mean speeds (TMS) in traffic engineering parlance. NPMRDS speeds are collected on extended segments of roadways knows as Traffic Message Channel (TMC) and are based on point-paired data which capture vehicle travel times between detection locations. In this way, they represent space mean speeds (SMS).

Figure 1 shows an example of a data collection location illustrating the difference in data collection approaches at a permanent monitoring station and for NPMRDS.

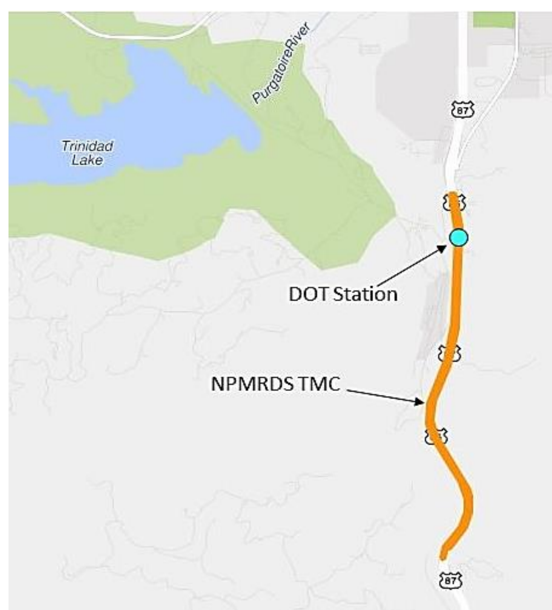


Figure 1 Permanent Station Located on a TMC

The lengths of TMCs for all analyzed locations ranged between 0.5 mile and 30 miles. Several study locations are on rural road segments (hence the low traffic volume) which are homogenous in terms of traffic volume and road configuration. Therefore, some of these TMC segments are defined to be reasonably long in the NPMRDS database. The median TMC lengths in Colorado, Texas and South Carolina were 9.3 miles, 3.6 miles, and 3.5 miles respectively, while the mean TMC lengths for the three states were 13.8 miles, 6.6 miles and 5.1 miles respectively. Given project resources, the DOT permanent monitoring station sites were considered the best readily available benchmarks for these comparisons, so the inherent difference in speed measurement methods (TMS vs SMS) could not be avoided.

The functional class of roadway sections analyzed include minor arterials (13% of total sites), principal arterials (58% of total sites) and Interstates (29% of total sites). Data for relatively higher traffic demand periods of the year (such as around end of summer, Thanksgiving, Christmas, etc.) were collected in order to make sample sizes for hourly speed aggregation as robust as possible for comparisons on these otherwise lower volume roadways. One- to two-week long periods of continuous data availability were identified for year 2017 around these durations of expected high traffic demand.

Collected data were compiled for all sites, and missing observations were removed before comparisons were made. The following metrics were calculated from compiled data:

- Mean speed difference bias (in mph) = $\frac{1}{n}\sum_{i=1}^{n} DOT\ Speed_i - NPMRDS\ Speed_i$
- Mean absolute speed difference bias (in mph) = $\frac{1}{n}\sum_{i=1}^{n} |DOT\ Speed_i - NPMRDS\ Speed_i|$
- Mean absolute percent difference (MAPE) (in %) = $\frac{1}{n}\sum_{i=1}^{n} \left|\frac{DOT\ Speed_i - NPMRDS\ Speed_i}{DOT\ Speed_i}\right| * 100$

Median values for all three metrics (median speed difference bias, median absolute speed difference bias, and median absolute percent difference) were also calculated.

RESULTS AND DISCUSSION

This section presents the overall results for all locations and discusses some key aspects behind the findings. Some interesting observations for speed patterns specific to a few locations are provided in Appendix A.

Table 1 summarizes the overall findings for speed comparisons in the three states. The second column "No. of observations" in Table 1 provides the number of hourly speed comparison data points for respective states. The median speed bias across all locations is found to be 8.5 mph, showing that speeds recorded by permanent monitoring stations are consistently higher than NPMRDS speeds. The median absolute percent difference in speeds is 13 percent. Overall, Interstate highway locations showed a slightly lower speed difference bias (7.1 mph on average) than non-Interstate highway locations (9.2 mph on average). In other words, speed differences were higher on lower-volume arterial roads than on higher-volume Interstate highways.

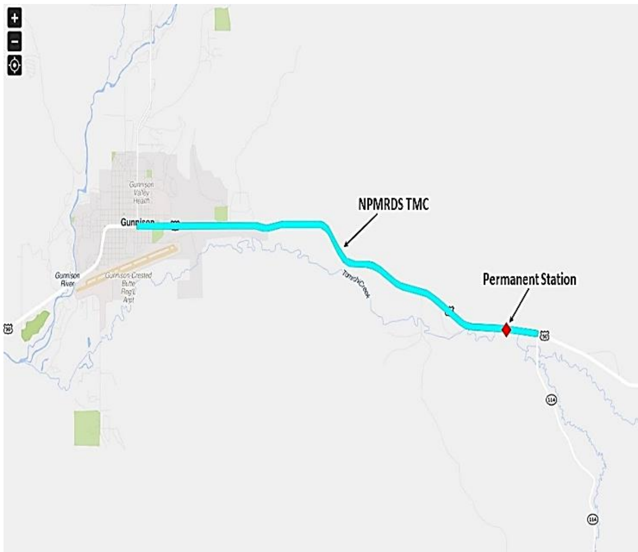
Some of the potential reasons for these observations can include:

i. Difference between TMS and SMS: As discussed in the previous section, time mean speed is associated with a point over time, whereas the space mean speed is associated with a section of roadway. "Space-mean speeds weight slower vehicles' speeds more heavily, as the slower vehicles are within the segment of interest for a longer period of time." *(5)* While TMS is the arithmetic mean of all speed observations, SMS is the harmonic mean of all speed observations. Time mean speed is always equal to or greater than space mean speed – the difference depends on the variability of individual speed observations. The length of TMCs can also influence these differences – longer TMCs can result in higher differences between TMS and SMS depending on geometric and geographical factors. Table 1 shows that speeds recorded at permanent monitoring stations (TMS) are greater than NPMRDS speeds (SMS) on average.
ii. Truck proportion bias in NPMRDS fleet: NPMRDS data can exhibit a disproportionate truck GPS sample. INRIX (NPMRDS data provider) collects location data from both commercial truck fleets and passenger cars. On rural highways (which typically have lower traffic volume, and are the focus of this analysis), the ratio of truck GPS samples to passenger car GPS samples may be much higher, which could cause slower truck speeds to have a disproportionate impact when calculating overall traffic stream speeds. This phenomenon has minimal to no effect on permanent monitoring station speeds.
iii. Time window of observations: In order to have a greater probability of robust sample sizes for hourly speed data on these lower volume road segments, periods of analysis were chosen around relatively higher traffic demand periods of the year (such as end of summer, Thanksgiving, Christmas, etc.). Sometimes, this can also result in higher levels of congestion, particularly during peak demand periods, exaggerating the effect of differences between TMS and SMS.

Table 1: Summary of Hourly Speed Comparisons

State	No. of observations (% of total)	Average (Mean) Values			Median Values		
		Average Absolute Speed Difference (mph)	Average Speed Difference Bias [DOT-NPMRDS] (mph)	Mean Absolute Percent Difference	Median Absolute Speed Difference (mph)	Median Speed Difference Bias [DOT-NPMRDS] (mph)	Median Absolute Percent Difference
CO	16,548 (50%)	12	11.2	18%	11	11.0	15%
TX	14,290 (44%)	8	6.8	12%	6	6.0	9%
SC	2,080 (6%)	8	6.1	14%	6	5.2	10%
Overall	32,918	10	8.9	15%	9	8.5	13%

The overall magnitude of differences is higher in Colorado than in Texas and South Carolina. A potential reason for this higher difference in speeds can be the difference in nature of terrain of roadway segments analyzed. Figure 2 shows one such example location where the median absolute percent difference between permanent monitoring station speeds and NPMRDS speeds is on the higher end (26% and 32% in Eastbound and Westbound directions respectively). As shown in the figure, the average slope along the full length of the TMC is 1.8% in Eastbound and -1.9% in Westbound directions respectively. However, there are several sections of relatively high slope in both directions (maximum of +17.0% and -22.4% slopes), potentially resulting in lower travel speeds through the entire segment of the road compared to the specific spot where the permanent monitoring station detector is installed, where the slope is 1.2% (located by the vertical marker in the elevation profile). This can result in higher speed observations at the location of the permanent monitoring station (hence higher DOT-recorded speeds) due to flatter terrain, but lower overall speeds on the entire TMC (relatively lower NPMRDS speeds) because of variable terrain.



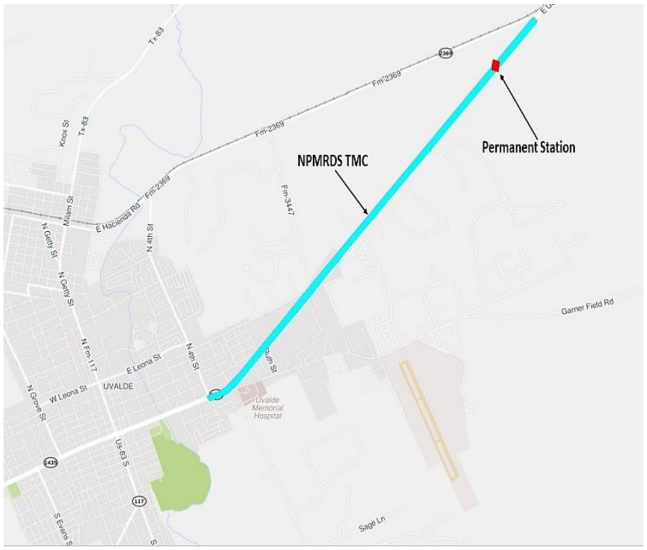
(a) TMC and Permanent Station Location

(b) Elevation Profile Along the TMC

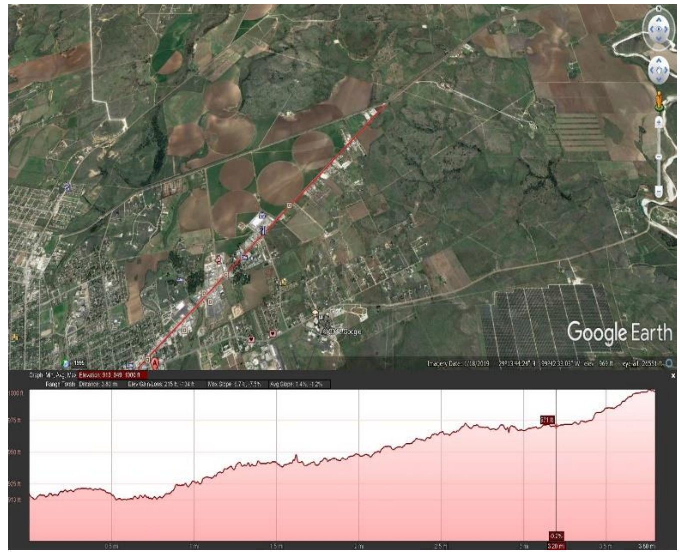
Figure 2 Example Location in Colorado Illustrating High Effect of Terrain on Speed Differences

However, the following two examples in Figure 3 and Figure 4 illustrate that there can be other contributing factors to high differences in speeds between permanent monitoring station data and NPMRDS data. For instance, as shown in Figure 3 for an example location in Texas, the median absolute percent difference is about 30% even though the terrain is not that variable along the TMC (average slopes of 1.4% and -1.2% and maximum slopes of 6.7% and -7.5% in Eastbound and Westbound directions respectively). On the other hand, Figure 4 represents a location in Colorado, where the median absolute percent difference is relatively low (around 7% in both Northbound and Southbound directions) although the terrain is highly variable (average slopes of 3.3% and -3.6% and maximum slopes of 20.8% and -20.1% in Northbound and Southbound directions respectively).

These examples illustrate that the difference in permanent monitoring station speeds and NPMRDS speeds is influenced by a few different factors. Terrain is one of the contributing factors, but not the only factor. Highly variable terrain often increases the difference between the two speed observations, but sometimes other factors show a higher overall effect than terrain.

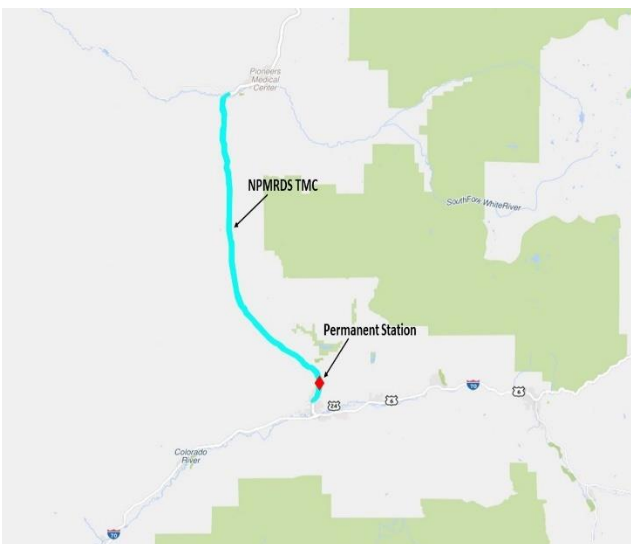


(a) TMC and Permanent Station Location

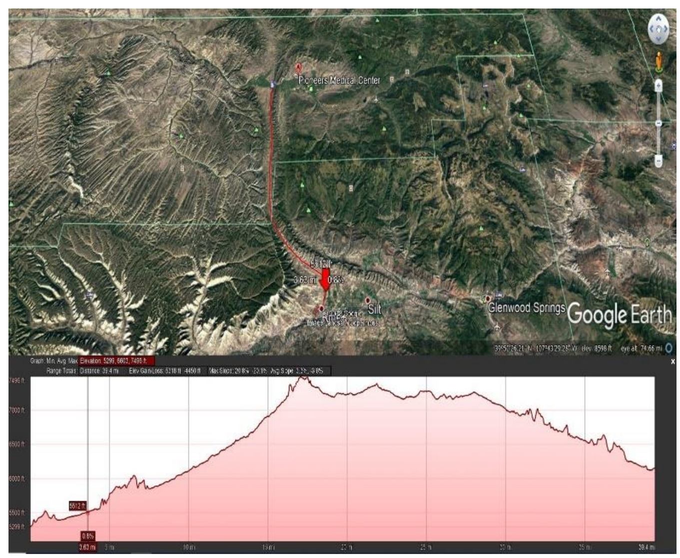


(b) Elevation Profile Along the TMC

Figure 3 Example Location in Texas Illustrating the Effect of Terrain on Speed Differences



(a) TMC and Permanent Station Location



(b) Elevation Profile Along the TMC

Figure 4 Example Location in Colorado Illustrating Low Effect of Terrain on Speed Differences

Figure 5 represents a typical pattern of differences in station and NPMRDS speeds at all locations. It shows a consistent bias of between 5-8 mph whereby permanent monitoring station speeds (blue data points) are higher than corresponding hourly NPMRDS speeds (maroon data points). Several other interesting patterns were observed for a few other locations, which are discussed in Appendix A.

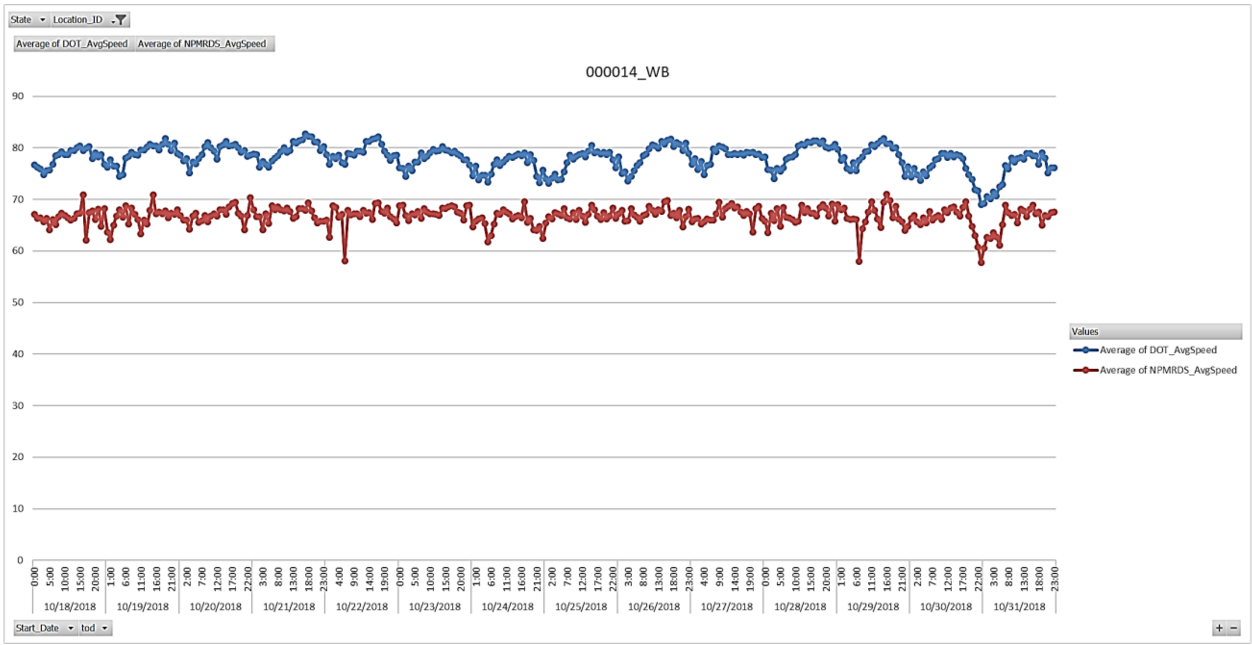


Figure 5 Typical Speed Pattern Difference Between Station Data and NPMRDS Speeds

The effects of roadway segment length, amount of traffic volume (AADT) or proportion of missing observations were not found to be consistent or conclusive. It was expected that long segment lengths, low traffic volume and high percentage of missing observations could potentially increase the speed difference, but their effects on speed difference bias and absolute percent difference values were not found to be as strong as anticipated. This is shown in Figure 6 and Figure 7. Although the bias appears to moderate with increasing AADT to an extent, it doesn't exhibit a consistent trend. This can be caused by a combination of other factors (such as different data collection methods (TMS vs SMS), inherent truck proportion bias in NPMRDS, etc.) which are not captured solely by changing AADT levels.

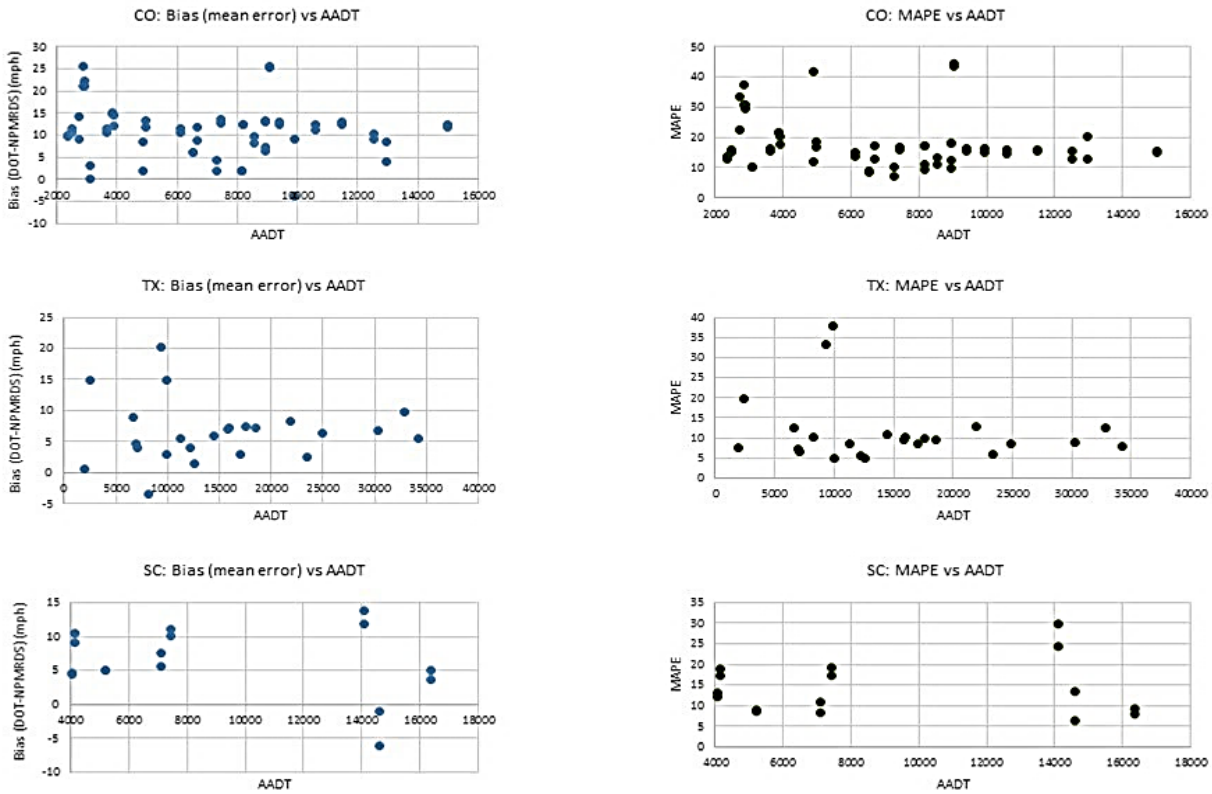


Figure 6 Effect of AADT on Speed Bias

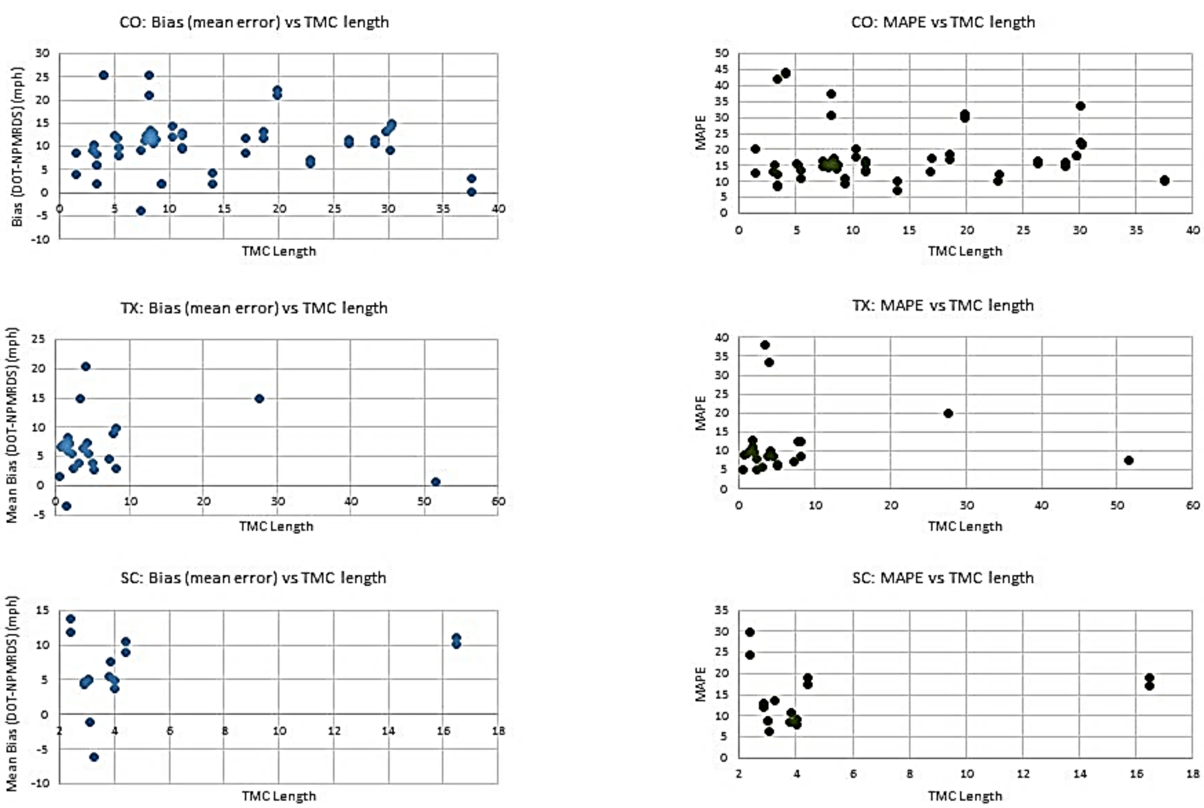


Figure 7 Effect of TMC Length on Speed Bias

CONCLUSIONS

This research examined the nature and magnitude of differences between speed data captured by permanent monitoring station installed by state DOTs and corresponding speed data in NPMRDS on lower volume roadways in three states – CO, TX and SC. It was found that speeds recorded at permanent station locations were consistently higher than NPMRDS speeds by about 8 mph on average. These differences varied between sites and states. Colorado showed the highest differences in speeds (11 mph on average) while for Texas and South Carolina these differences were comparable and relatively lower (around 6 mph on average). The absolute percent difference for these speed data ranged between 15% for CO and around 10% for TX and SC.

The consistent differences in speeds recorded by these two speed data sources can be attributed to a few different factors. Terrain (topography along the roadway segment) is one of the contributing factors to speed bias, but not the only factor. Variable terrain often increases the differences between the two speed observations, but sometimes other contributing factors show a higher overall effect than terrain. The other influencing factors can include the difference in methodologies for capturing speed data and potential difference in vehicle fleet composition in captured sample leading to variable speed records in response to congested periods of travel.

Overall, speed differences were higher on lower-volume arterial roads (9.2 mph on average) than on higher-volume Interstate highways (7.1 mph on average). Interestingly, no consistent or conclusive effect of TMC segment length, value of AADT or proportion of missing observations was observed on speed difference bias and absolute percent difference values within the constraint of scope of locations studied. It was expected that long segment lengths, low traffic volume and high percentage of missing observations could potentially increase the speed differences, but their effects on speed difference bias and mean absolute percent difference values were not observed to be as strong as anticipated.

The findings can help agencies get a deeper understanding of the quality of NPMRDS travel time data on their lower volume roadway network because measured speeds are used for several transportation and traffic engineering related activities. Agencies can decide to use one database or the other depending on their intended use of recorded vehicle speeds. For instance, based on these results, permanent monitoring station data can be useful for safety analyses or to identify operating speeds for road design at specific identified locations on lower volume roadways vis-à-vis determining free-flow speeds for an extended segment of roadway, where NPMRDS speed database can be useful.

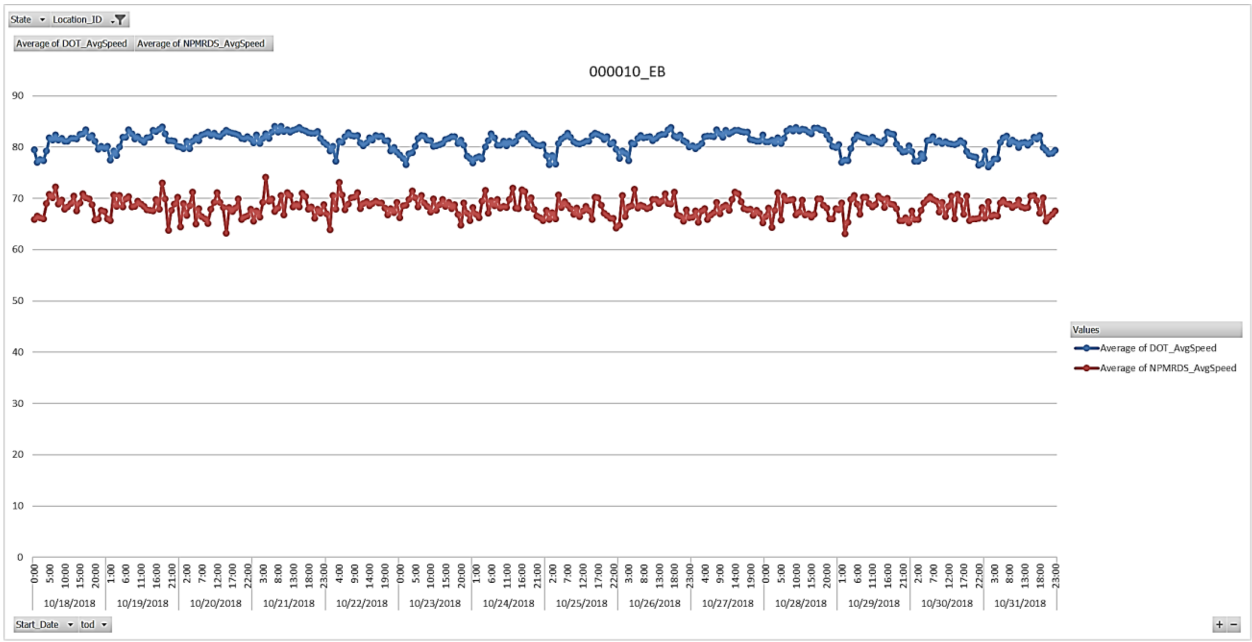
REFERENCES

1. Turner, S. M., and P., Koeneman. *Validating the National Performance Management Research Data Set (NPMRDS) for South Dakota*. No. SD2013-08-F. South Dakota Department of Transportation Office of Research, March 2018.
2. Transportation Data Management System, Colorado Department of Transportation. Accessed via https://cdot.ms2soft.com/tcds/tsearch.asp?loc=Cdot&mod=
3. Traffic Count Database System (TCDS), Texas Department of Transportation. Accessed via https://txdot.ms2soft.com/tcds/tsearch.asp?loc=Txdot&mod=

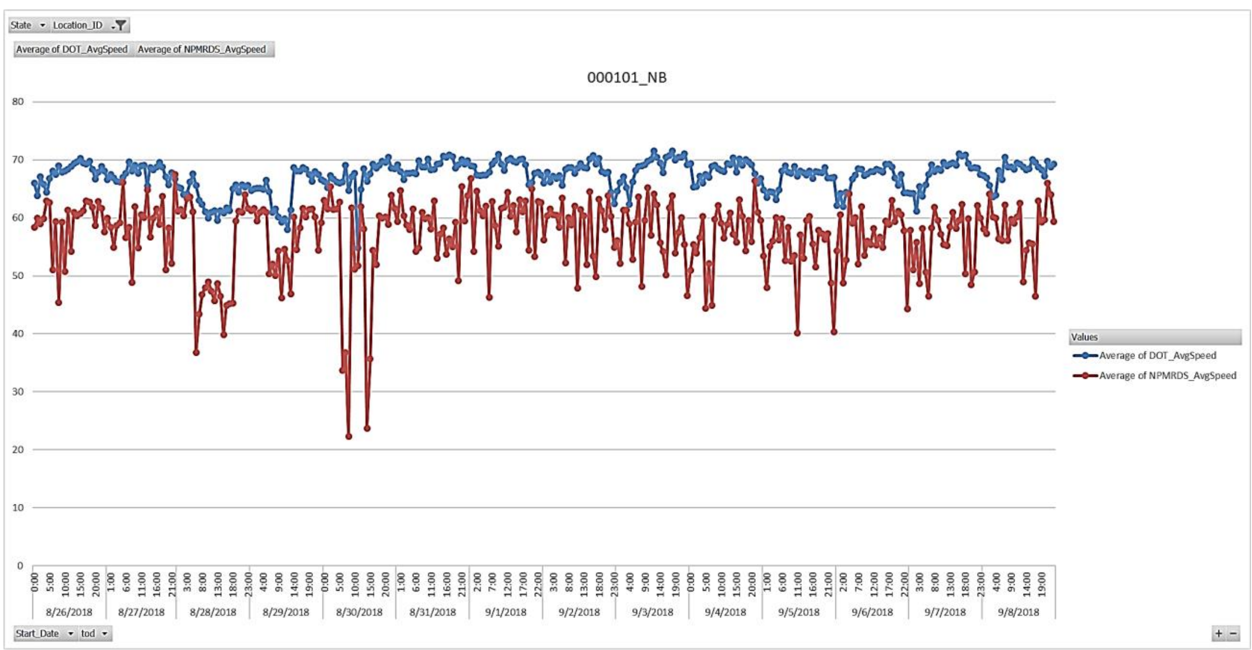
4. Transportation Data Management System, South Carolina Department of Transportation. Accessed via https://scdot.ms2soft.com/tcds/tsearch.asp?loc=Scdot&mod=
5. Travel Time Data Collection Handbook, Chapter 1 – Introduction. Federal Highway Administration. Accessed via https://www.fhwa.dot.gov/ohim/handbook/chap1.pdf

Appendix A
Interesting Speed Pattern Differences at Specific Locations

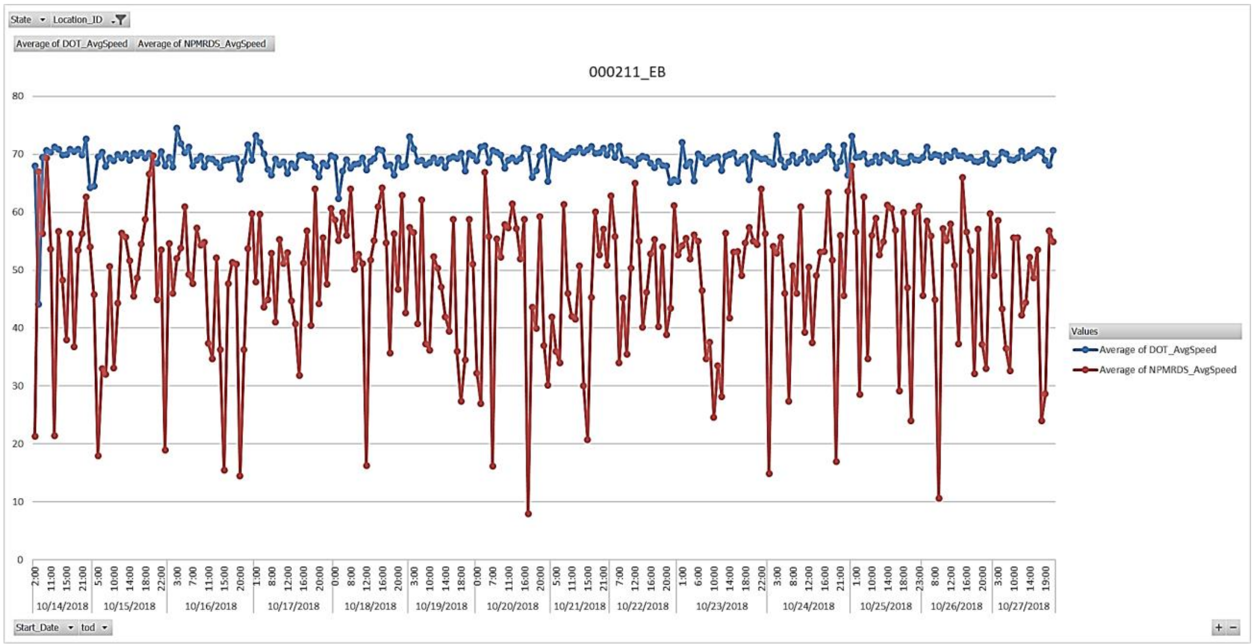
Colorado – consistent pattern



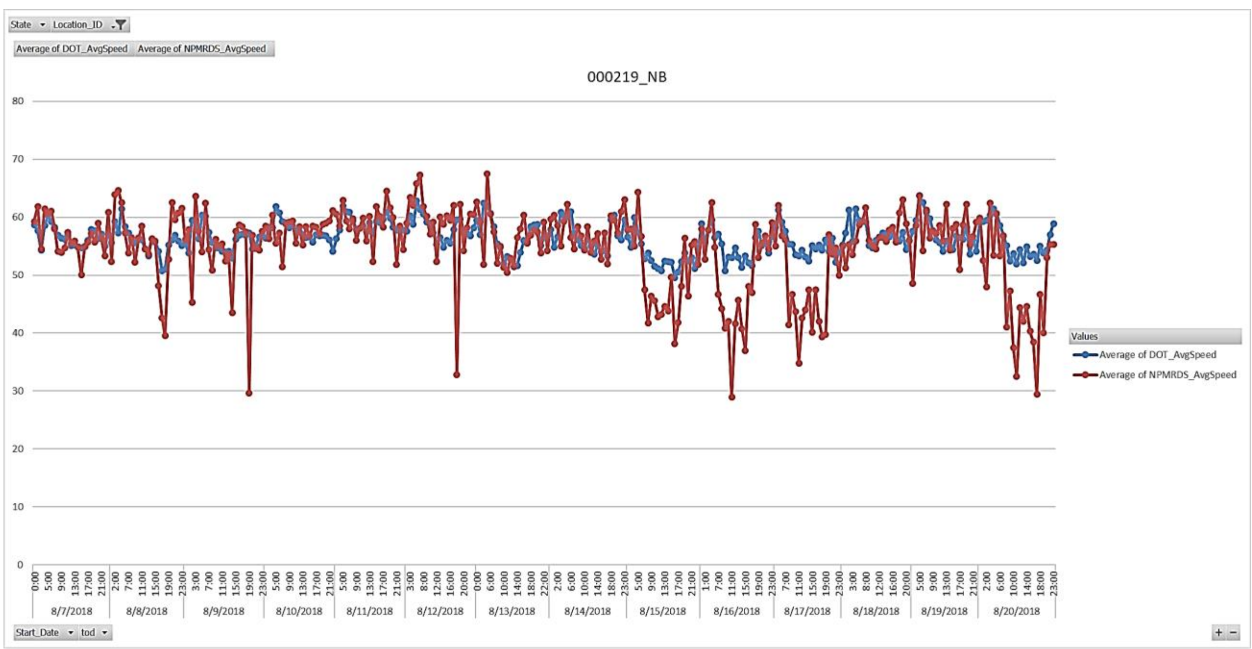
Colorado – variable NPMRDS readings but consistent trends



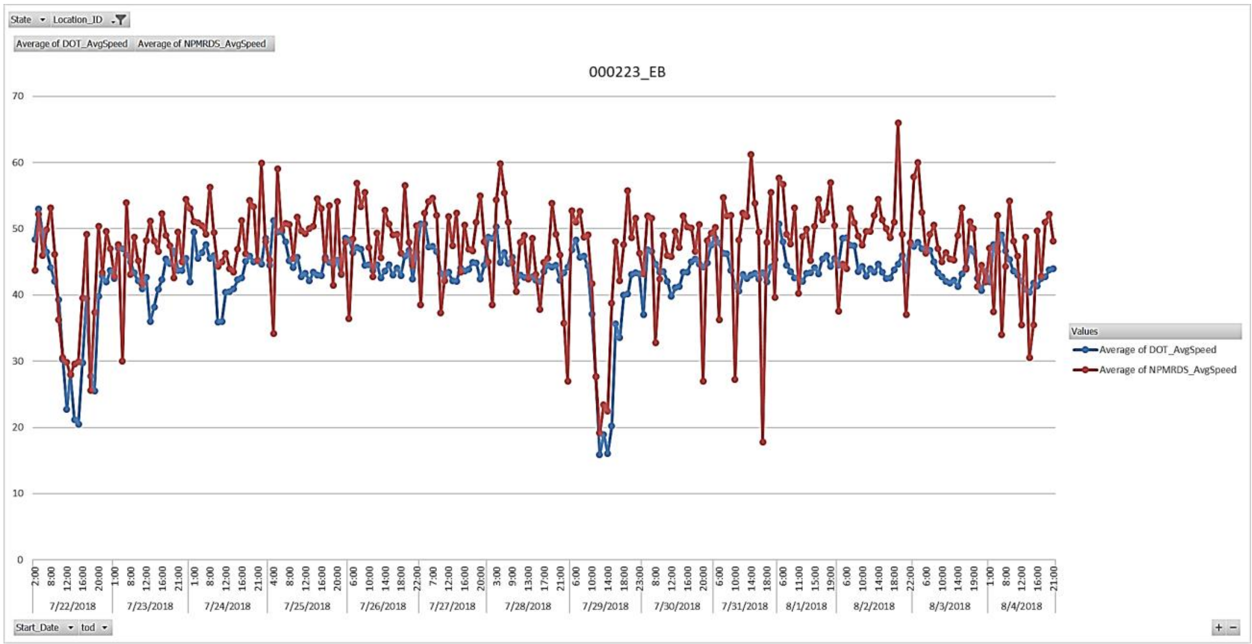
Colorado – variable NPMRDS readings and unexplained trends



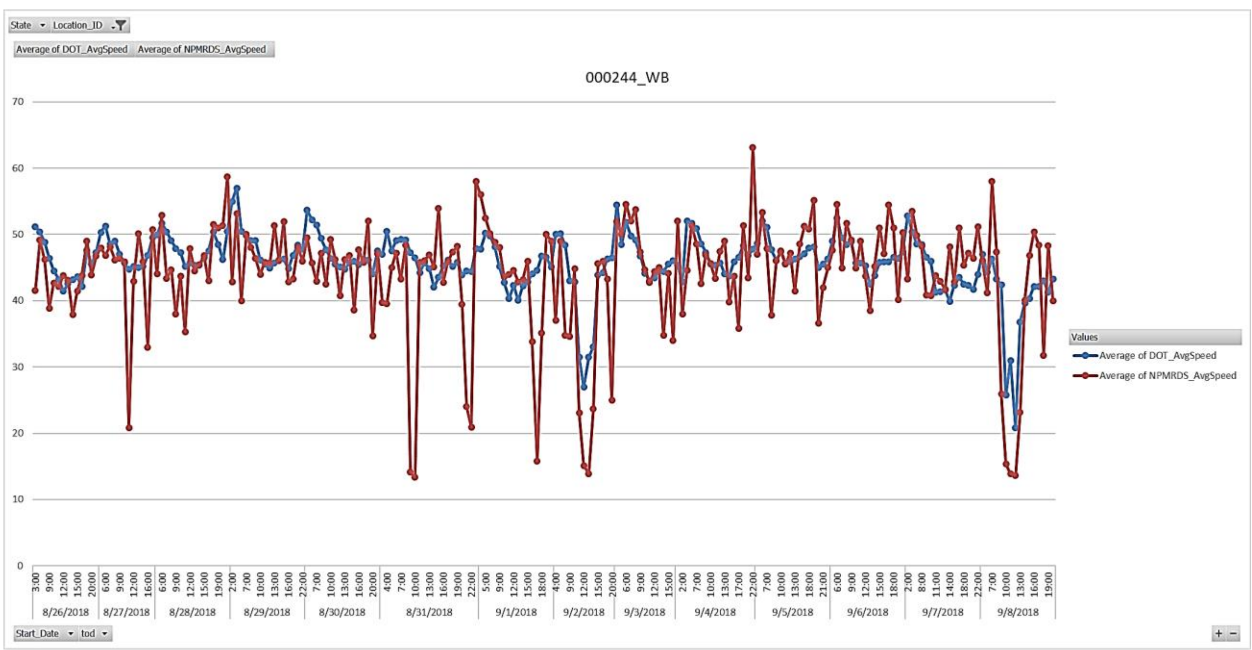
Colorado – variable NPMRDS readings and exaggerated congestion



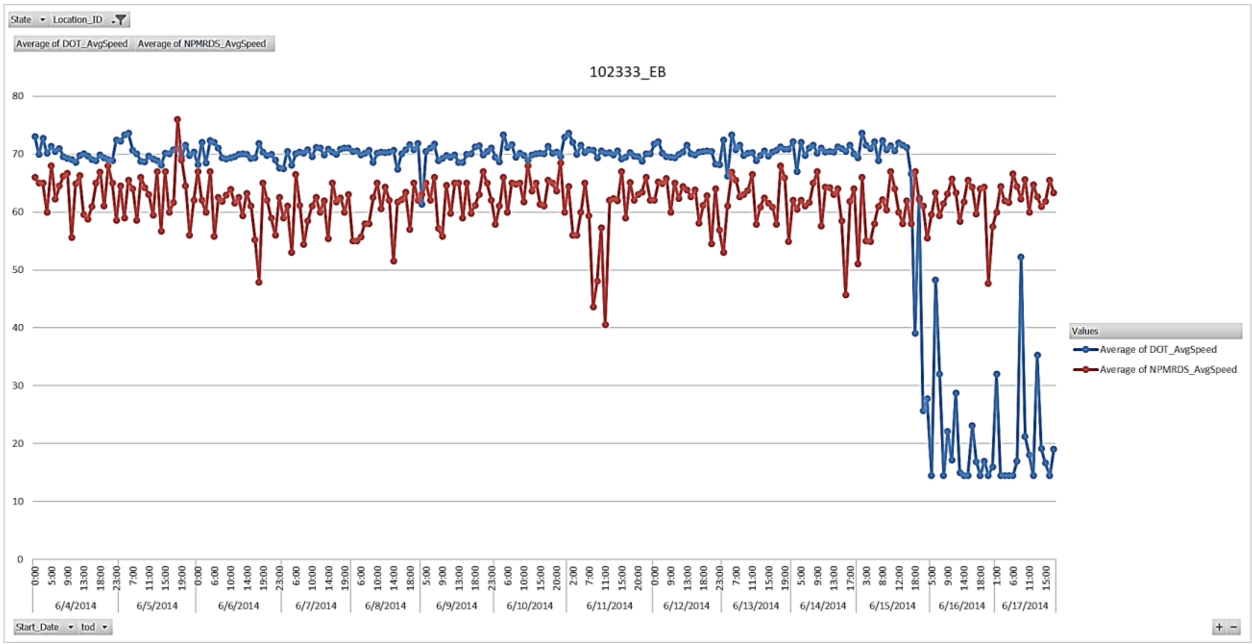
Colorado – NPMRDS readings > station readings (only in EB direction)



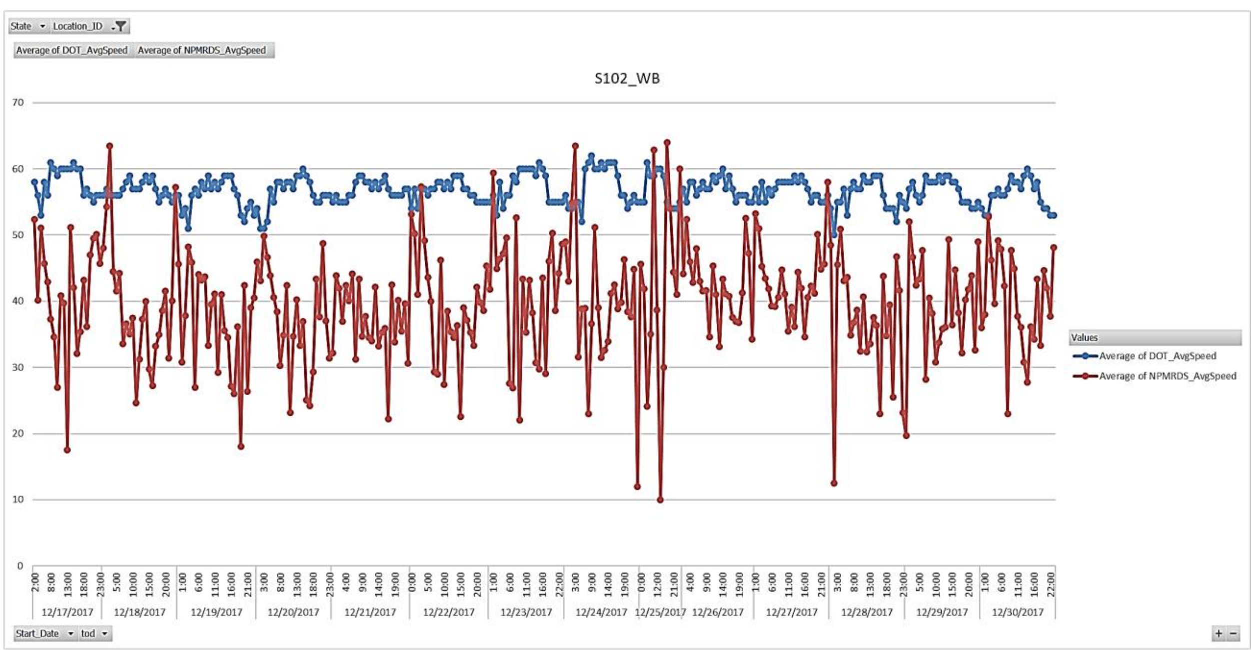
Colorado – variable NPMRDS readings but trend consistent with station readings



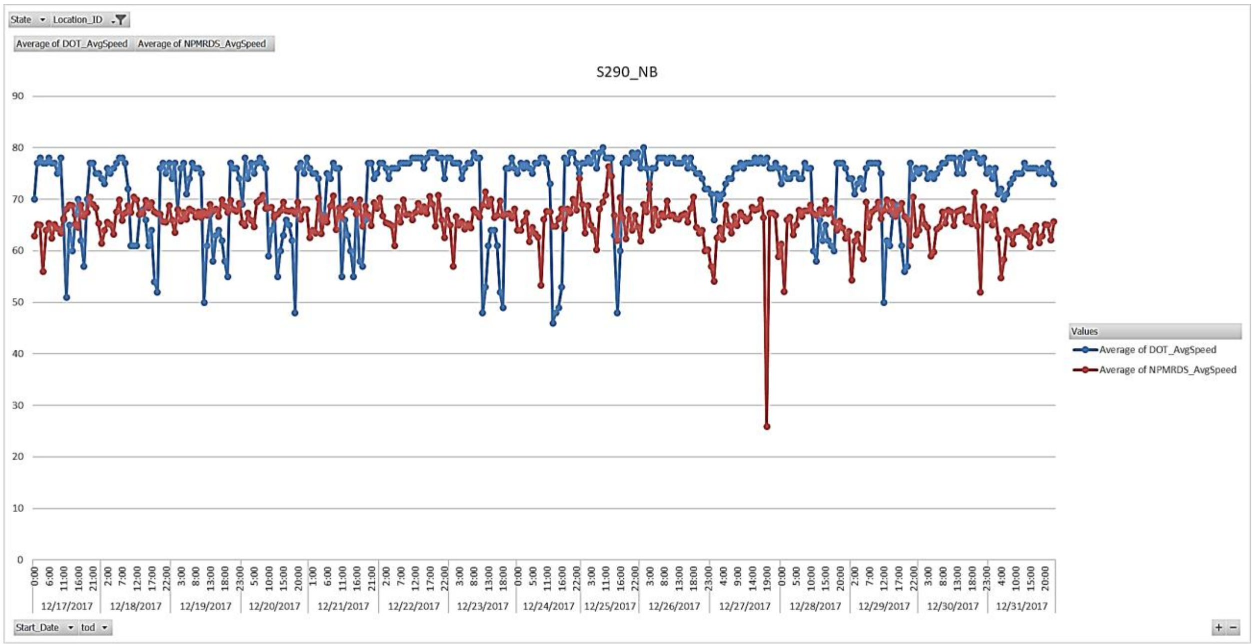
Colorado – possibly a station reader malfunction



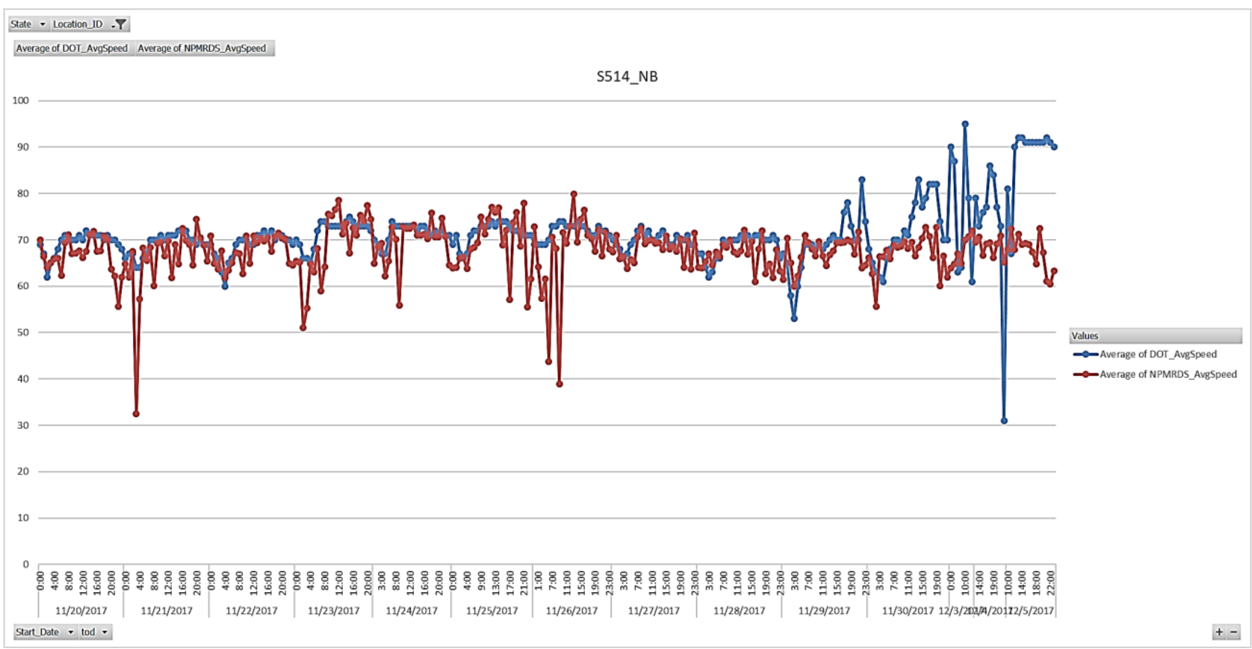
Texas – highly variable NPMRDS readings



Texas – station readings show repeated exaggerated mid-day congestion



Texas – possibly station reader malfunction (both directions)



Texas – station and NPMRDS readings closer than average