Data Analysis for Ver-Mac

Juan Carlos Martinez Mori, Yanning Li, Daniel Work

1 Introduction

This report documents the data analysis for the project *Improving the Effectiveness of Smart Work Zone Technologies* (R27-155). The data collected by Ver-Mac traffic management systems deployed in two work zones was studied. Multiple findings regarding this data set are presented.

1.1 Project Description

This project aims to evaluate the effectiveness of smart work zone technologies by assessing the performance of different sensor network configurations in a microscopic simulation environment. In order to transfer the simulation results to the real world, the computer models need to be calibrated against field data. The following work zones were modeled and calibrated using the data collected by the Ver-Mac traffic management system:

1. I-57/I-64: IDOT Contract No. 78276, in Jefferson County, IL. In total, 25 sensors were deployed, including 22 radar sensors and 3 Remote Traffic Microwave Sensors (RTMS).
2. I-80: IDOT Contract No. 60Y64, in Will County, IL. In total, 30 sensors were deployed, including 18 radar sensors and 12 RTMS.

1.2 Data Set

The data set of each work zone can be accessed through the computer program JamLogic, developed by Ver-Mac. The following types of data were used for calibration:

- Vehicle speed
- Vehicle count

Ver-Mac provides data aggregated in different granularity, ranging from 30 seconds to 1 year. This analysis is based on the 5-minute granularity data set, which was found to provide a good balance between data resolution and sensor noise. A python code was developed to read the files outputted by JamLogic, and perform visualization, statistical analysis, and formatting (for the micro-simulation environment) on the data.

2 Methodology

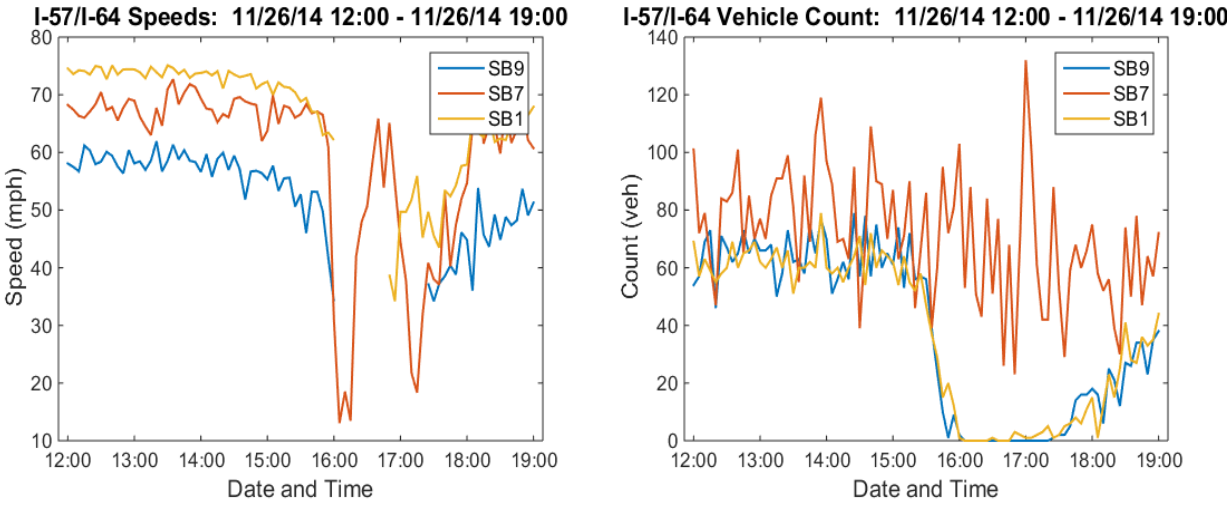
This section presents the procedure and metric used for analyzing the data set.

2.1 Preliminary Analysis

Field data set was known to have potential issues due to sensor malfunction, sensor failure, or communication timeout. A visual inspection is useful to identify and categorize any data abnormality.

2.1.1 Data Completeness

The vehicle speed and count data from SB1, SB7, and SB9 in I-57/I-64 work zone on Nov 26, 2014 was plotted in Fig. 1. It was observed that the radar speed data from 16:00 to 17:30 (when we believe a severe traffic congestion occurred) was missing, while the radar count data was unrealistically low. The issue was observed at other times when severe traffic congestion seemed to have occurred.



a) Vehicle speed starts missing in radar sensor. b) Vehicle count drops to zero in radar sensors.

Figure 1: Radar sensors (SB1, SB9) present abnormal readings as traffic congestion starts

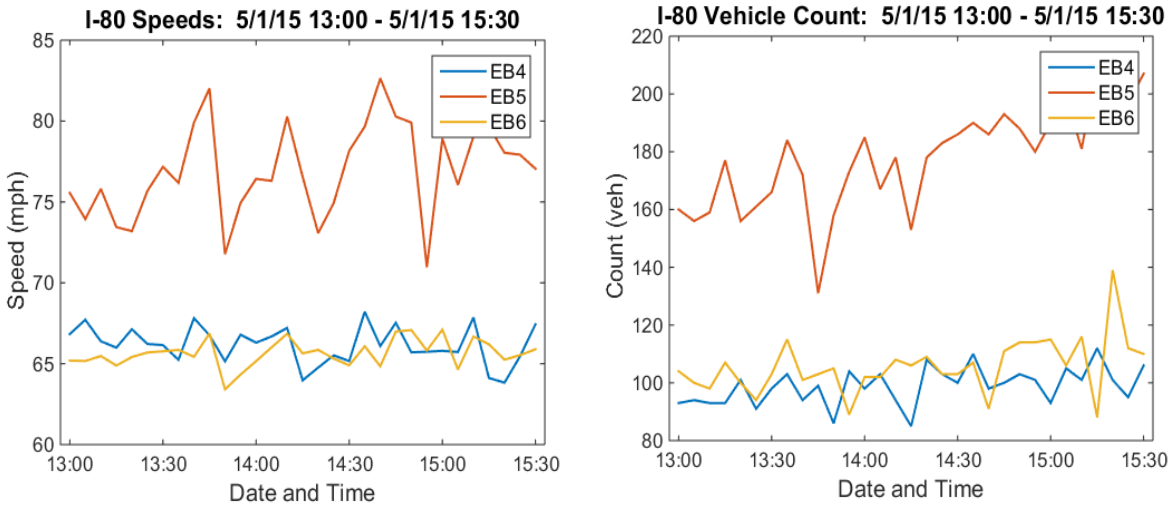
Overall, radar sensors seem to be malfunctioning when severe congestion occurs. Calibrating against such unrealistic readings would produce poor results. However, disregarding the radar sensors and relying solely on the RTMS sensor would not provide sufficient data to capture the traffic dynamics (e.g. only one RTMS was deployed at each direction in I-57/I-64 work zone).

2.1.2 Data Consistency

Though sensors have different levels of accuracy, sensors measuring the same traffic quantities should have relatively similar outputs within the noise error bound.

In I-80 work zone, three sensors EB4 (radar), EB5 (RTMS), and EB6 (radar) were deployed at 0.5 miles apart with no ramps in between and no significant structural changes. Theoretically, in free flow, at such close locations and 5-minute aggregation intervals, their speed and count measurements should be approximately the same.

Fig. 2 shows their readings during 13:00 to 15:30 on May 1, 2015. Significantly different readings were observed between two types of sensors. This behavior violates the principle of mass conservation, meaning that at least one of the sensor types was consistently presenting incorrect values. A possible explanation is that, considering that the radar sensors are mounted at a low height on the roadside, they may not capture the traffic at the further lane. This situation was repeated on other pairs of radar and RTMS sensors in this work zone.



a) Inconsistent vehicle speed. b) Inconsistent vehicle count.

Figure 2: Inconsistency between sensors EB4 and EB6 (radar) and EB5 (RTMS) during free flow

2.2 Full Scale Analysis

Based on the preliminary analysis, both incompleteness and inconsistency were observed in the data set of the two work zones. This statistical analysis aims to quantitatively evaluate the level of the incompleteness and inconsistency at a larger extent than the preliminary analysis.

2.2.1 Metric for Data Incompleteness

The following metric was used in the analysis.

Percent Missing

The percent of missing data (speed or count) was determined for each sensor in a time interval:

$$Percent\ missing = \frac{No.\ of\ missed\ readings}{No.\ of\ time\ intervals} * 100$$

Besides the percent missing in a large time period, the percentages of missing data in specific scenarios were also important. Particularly, the following two scenarios were considered:

- Peak hours percent missing: The average amount of missing peak hour data during a time period among the sensors in a specified section of the network (e.g. south bound). Daily peak hours were assumed to be from 16:30 to 17:30.
- Congested hours percent missing: The average percent of missing data in a set of sensor during a time period. Particularly, severe congestions within the time period were chosen by visual inspection. Congested hours are defined as when the decrease of speed is observed.

2.2.2 Metric for Data Inconsistency

Percent Change

The percent of change between the readings (speed or count) of pairs of sensors was determined in specified time intervals:

$$Percent\ Change\ (x_f, x_o) = \frac{x_f - x_o}{x_o} * 100$$

x_o = *Reading from first sensor*
x_f = *Reading from second sensor*

This procedure was applied on pairs of sensors in segments of road with no ramps. The analysis was conducted on a typical free flow condition.

3 Findings

This section presents the numerical results of the quality assessment procedures performed on the data sets in each of the work zones. Note the data set used for this analysis is in 5-minute granularity.

3.1 I-57/I-64

The period from Nov 1st, 2014 to Nov 30th, 2014 in south bound traffic in I-57/I-64 work zone was investigated. The congested hours are identified as follows: xx to xx on day xx; xx to xx on day xx. Table 1 summarizes the percent missing for each sensor and Table 2 provides the percent change between representative sensor pairs that should have produced similar outputs.

Sensors	Entire period Percent Missing		Peak Hours Percent Missing		Congested Hours Percent Missing	
	Speed	Count	Speed	Count	Speed	Count
SB1						
SB2						
SB3						
SB4						
SB5						
SB6						
SB7*						
SB8						
SB9						
Average			4.73%	4.73%		

Table 1: The percent missing for I-57/I-64 South Bound. '' denotes the RTMS sensor.*

Sensor Pair	Percent Change	
	Speed	Count
SB 5 -> SB6		
SB 5 -> SB7*		

Table 2: The average percent change of speed and count in free flow traffic for sensor pairs on I-57/I-64 in November 2014. '' denotes RTMS sensors.*

3.2 I-80

The period from May 1st, 2015 to May 31st, 2015 in east bound traffic in I-80 work zone was investigated. The congested hours in this period are identified as follows: xx to xx on day xx; xx to xx on day xx. The percent missing and percent change statistics are provided in Table 3 and Table 4 respectively.

Sensors	Entire period Percent Missing		Peak Hours Percent Missing		Congested Hours Percent Missing	
	Speed	Count	Speed	Count	Speed	Count
EB3						
EB4						
EB5*						
EB6						
EB7*						
EB8						
EB9*						
EB10						
EB11						
EB12*						
EB14						
EB15						
EB16						
Average						

Table 3: The percent missing for I-80 East Bound. '' denotes the RTMS sensor.*

	Percent Change	
Sensor Pair	**Speed**	**Count**
EB4→EB5*	16.57%	73.04%
EB4→EB6	-0.29%	19.13%
EB8→EB9*	-7.98%	79.66%
EB8→EB10	13.68%	23.47%

Table 2: The average percent change of speed and count in free flow traffic for sensor pairs on I-80 in May 2015. RTMS sensors are denoted by ''.*

4 Conclusions

This report has documented the data analysis procedures and findings for the Ver-Mac data collected in two work zones. These are our conclusions:

1. Data incompleteness and inconsistency issues are common in field data, especially if collected in noisy environments such as work zones. These issues were found in the data of both work zones.
2. The systematic issues present in each of the work zones made the determination of the sensors' accuracy unfeasible.
3. Data issues add difficulty in the model calibration for the project since the real traffic conditions remain unknown wherever the sensors malfunctioned.
4. The analysis procedure is documented in this report and the python code has been made available. We would like to share the findings with Ver-Mac in case they would like to investigate and identify the cause of the data issues.