**Data Analysis for Ver-Mac**

**Juan Carlos Martinez Mori, Yanning Li, Daniel Work**

1. **Introduction**

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

* 1. **Setting**

This project aims to evaluate the effectiveness of smart work zone technologies by assessing the performance of different smart work zone sensor networks in a microscopic simulation environment. In order to transfer the result in a simulated environment to the real world work zones, a critical part of the project implementation is to calibrate the traffic micro-simulation model. Two work zones were modeled in the micro-simulation environment and Ver-Mac field traffic data was used to calibrate these two models:

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

The data set of each work zone can be accessed and downloaded through the computer program JamLogic developed by Ver-Mac. The following types of data were used, along with their respective timestamp:

* 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-min granularity data set, which we found to be a good balance between data resolution and sensor noise.

A python code was developed to perform visualization, analysis, and formatting of the data.

1. **Methodology**

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

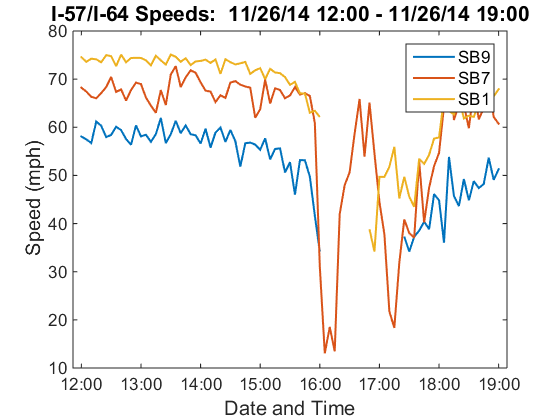
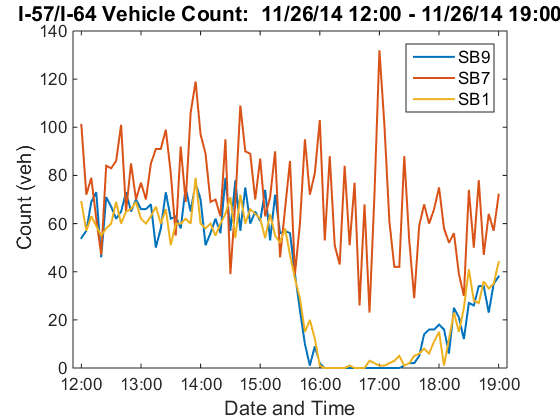
* 1. **Preliminary Analysis**

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 any data abnormality.

* + 1. **Data Completeness**

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

a) Vehicle speed starts missing in radar sensor. b) Vehicle count drops to zero in radar sensors.   
Figure 1: Radar sensors present abnormal readings as traffic congestion starts



As shown above, sensors SB9 and SB1 (radar) presented the described issue, but SB7 (RTMS) did not. This situation was recurrent in all of the remaining sensors in the work zone; radar sensors had the problem, but not the RTMS. This inconsistency made unclear which sensors were correct.

* + 1. **Data Consistency**

Sensors have different levels of accuracy. However, sensors measuring the same traffic quantities should have similar outputs.

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

a) Inconsistent vehicle counts. b) Inconsistent vehicle speeds.  
Figure 2: Inconsistency between sensors EB4 and EB6 (radar) and EB5 (RTMS) during free flow

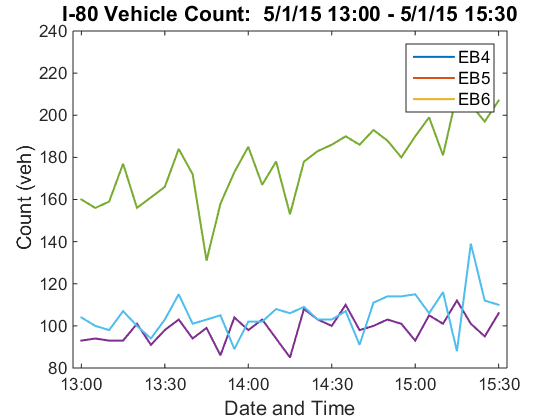
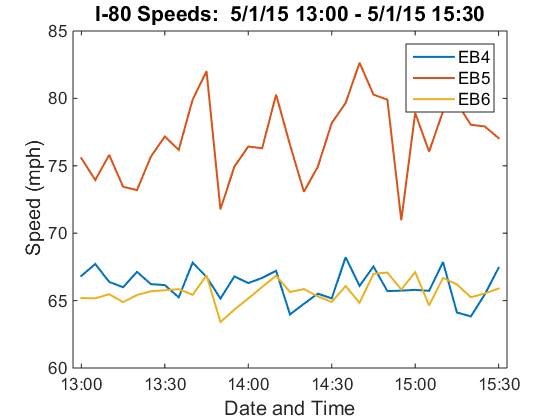


Fig. 2 shows the count and speed plot of EB4, EB5 and EB6, deployed in I-80 work zone, on May 1st, 2014. Significantly different readings were observed between two types of sensors.

In general, RTMS sensors have greater readings for both speed and count. This violates mass conservation, meaning that at least one of the sensor types was consistently presenting incorrect values. Considering that the radar sensors are mounted at a low height on the roadside, it may not capture the traffic at the further lane.

* 1. **Full Scale Analysis**

Based on the preliminary analysis, both incompleteness and inconsistency was observed in the data set in two work zones. This statistical analysis aims to quantitatively evaluate the level of the incompleteness and inconsistency of the data.

* + 1. **Metric for Data Incompleteness**

The following metric was used in the analysis.

**Arithmetic Mean**

The arithmetic mean of data (speed or count) was determined for continuous user specified time intervals. The next equation was followed:

**Percent Missing**

The percent of missing data (speed or count) was determined for continuous user specified time intervals. The next equation as followed:

Percent missing in peak hours

Percent missing in congested hours

Your other numbers

* + 1. **Metric for Data Inconsistency**

**Percent Change**

The percent of change between the readings (speed or count) of two sensors was determined for user specified sensors and time intervals. The next equation was followed:

Inconsistency between RTMS and Radar

Inconsistency between Radars

Inconsistency between RTMS

1. **Findings**

Overall, this report documents the data analysis procedure and findings for Ver-Mac data collected in two work zones. Our conclusions are:

1. Data incompleteness and inconsistency issues are common in field data, especially if collected in noisy environment such as work zones []. These issues were also observed in Ver-Mac data set in the two work zones.
2. Data issues add difficulty in the model calibration for the project since the real traffic condition is not available.
3. The analysis procedure is documented in this report and the python code is also 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.

The systematic issues present in each of the work zones made the determination of the sensors’ accuracy problematic. In the case of I-57/I-64, there is no way to interpolate with sufficient certainty the large gaps of missing data. This means that the exact traffic behaviors at those times intervals remains unknown. In the case of I-80, since both of the sensors types had relatively precise readings, it is troublesome to discern which one had correct measurements (or at least readings that were close to reality). In fact, if not for the obvious violation of the principle of mass conservation, it would have been hard to identify the presence of the systematic errors in first place. Arriving to these conclusions would have been more arduous if all of the sensors had been placed next to entrance or exit ramps, for example.

The findings presented in this report are of importance for the three parties involved. IDOT may want to ensure that the sensors produce accurate readings. This would result in more reliable traffic estimations, which translates into an increased user safety. Ver-Mac may want to identify the sources of the odd sensor behavior, so as to improve the quality of their traffic management systems and collaborate with IDOT’s objectives. Our research group found it challenging to deal with the issues presented, and had to implement creative solutions to move our project forward.