**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. **Setting**

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 needed to be calibrated against field data. The following work zones were modeled and calibrated using the data of their respective Ver-Mac traffic management system:

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 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-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.

1. **Methodology**

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

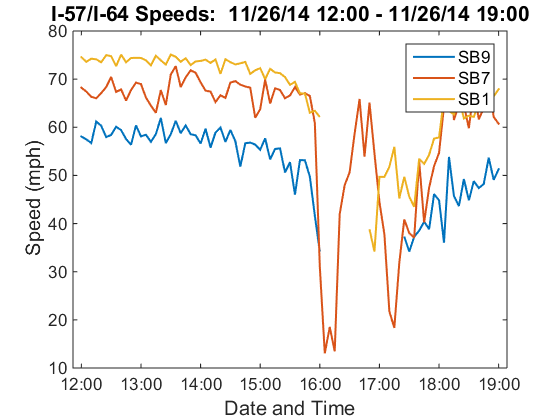
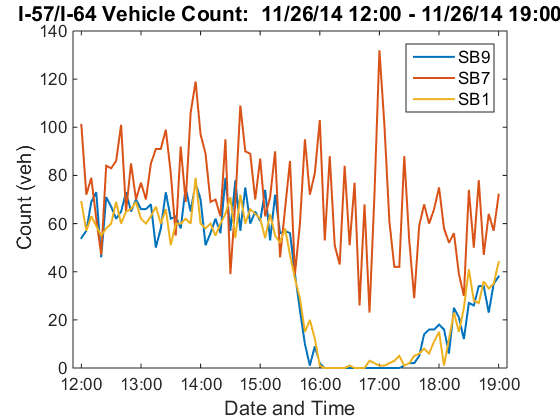
* 1. **Preliminary Analysis**

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). The issue was observed at other times in which 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 present abnormal readings as traffic congestion starts

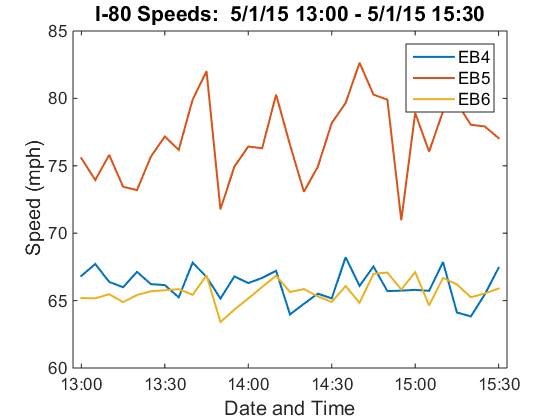
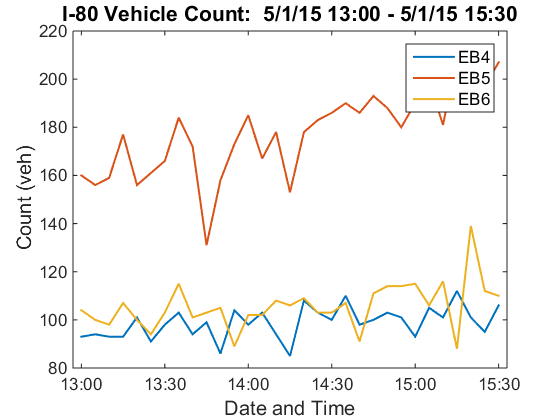


As shown above, sensors SB9 and SB1 (radar) presented the described issue, while SB7 (RTMS) did not. This complication was present in the remaining radar sensors of the work zone. Calibrating against such unrealistic readings would produce poor results. However, disregarding the radar sensors and relying solely on the RTMS sensor would not capture the entirety of the traffic dynamics.

* + 1. **Data Consistency**

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

In the 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, at such close locations and 5-minute aggregation intervals, their count and speed measurements should be approximately the same in free flow. Fig. 2 shows their readings during a free flow interval on May 1, 2015.



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

As seen above, 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.

* 1. **Full Scale Analysis**

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

* + 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 user specified sensors and time intervals. The next equation was followed:

This was applied on the data of sets of sensors under the following time interval categories:

* Peak hours: Daily peak hours were assumed to be from 16:30 to 17:30. The average amount of missing peak hour data among the sensors in a specified section of the network (e.g. south bound) was computed for a specified month and presented as a single value.
* Congested hours: Days with particularly severe congestion within a specified month were chosen (by visual inspection). The missing data of a set of sensors was computed for the relevant time interval (with a few hours before and after congestion). This was presented per sensor and as a set monthly average.
  + 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 time intervals and pairs of sensors. The next equation was followed:

This procedure was applied on pairs of sensors in segments of road with no ramps. The analysis was conducted on a typical free flow interval of time and on one month of data. The results were presented as averages in both cases. The sensor pair configurations were:

* RTMS and radar sensors
* Pairs of radar sensors

The data consistency between pairs of RTMS sensors could not be assessed since a case in which two RTMS sensors were located in a segment of road with the described characteristics could not be found.

1. **Findings**

is analysis as -granularity Note that in the following tables, the symbol ‘\*’ denotes a RTMS sensor.

|  |  |  |
| --- | --- | --- |
| **Congested Hours Missing Data**  **November 26, 2014 12:00 – 19:00** | | |
| **Sensor** | **Missing Speeds** | **Missing Counts** |
| **SB1** | 9.41% | 9.41% |
| **SB2** | 22.35% | 22.35% |
| **SB3** | 23.53% | 23.53% |
| **SB4** | 14.12% | 14.12% |
| **SB5** | 14.12% | 14.12% |
| **SB6** | 34.12% | 34.12% |
| **SB7\*** | 0.00% | 0.00% |
| **SB8** | 27.06% | 27.06% |
| **SB9** | 18.82% | 18.82% |
| **Average** | 18.17% | 18.17% |

|  |  |  |
| --- | --- | --- |
| **Peak Hours Missing Data**  **November 2014** | | |
| **Direction** | **Missing Speeds** | **Missing Counts** |
| **South Bound** | 4.73% | 4.73% |
| **East Bound** | 1.17% | 1.31% |
| **North Bound** | 1.03% | 1.60% |
| **Average** | 2.31% | 2.55% |

Note: The missing counts section of this table does not consider sensors NB1, NB3, NB4 and EB2, since they did not report any count readings.







|  |  |  |
| --- | --- | --- |
| **Sensor Pairs Percent Change**  **May 2015** | | |
| **FirstSecond** | **% Difference Speed** | **% Difference Count** |
| **EB4EB5\*** | 16.57% | 73.04% |
| **EB4EB6** | -0.29% | 19.13% |
| **EB8EB9\*** | -7.98% | 79.66% |
| **EB8EB10** | 13.68% | 23.47% |

|  |  |  |
| --- | --- | --- |
| **Sensor Pairs Percent Change**  **May 1, 2015 13:00 – 15:30** | | |
| **FirstSecond** | **% Difference Speed** | **% Difference Count** |
| **EB4EB5\*** | 16.31% | 79.85% |
| **EB4EB6** | -0.75% | 7.27% |
| **EB8EB9\*** | 10.81% | 93.73% |
| **EB8EB10** | -6.93% | 13.68% |

1. **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.