DEPARTMENT OF TRANSPORTATION

Weigh-in-Motion Sensor and Controller Operation and Performance Comparison

Diwakar Gupta, Principal Investigator
Department of Industrial and Systems Engineering
University of Minnesota

January 2018

Research Project
Final Report 2018-03

To request this document in an alternative format, such as braille or large print, call 651-366-4718 or 1-800-657-3774 (Greater Minnesota) or email your request to ADArequest.dot@state.mn.us. Please request at least one week in advance.

Technical Report Documentation Page

1. Report No. MN/RC 2018-03	2.	3. Recipients Accession No.
4. Title and Subtitle Weigh-in-Motion Sensor and Controller Operation and Performance Comparison		5. Report Date January 2018
		6.
7. Author(s) Diwakar Gupta, Xiaoxu Tang, Lu Yuan		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Industrial and Systems Engineering University of Minnesota 111 Church Street S. E. Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS # 2016001
		11. Contract (C) or Grant (G) No. (c) 99008, (wo) 182
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Research Services & Library 395 John Ireland Boulevard, MS 330 St. Paul, Minnesota 55155-1899		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes http:// mndot.gov/research/reports/2018/201803.pdf		

16. Abstract (Limit: 250 words)

This research project utilized statistical inference and comparison techniques to compare the performance of different Weigh-in-Motion (WIM) sensors. First, we analyzed test-vehicle data to perform an accuracy check of the results reported by the sensor-vendor Intercomp. The results reported by Intercomp mostly matched with our own analysis, but the data were found to be insufficient to reach any conclusions about the accuracy of the sensor under different temperature and speed conditions. Second, based on the limited data from the Intercomp and IRD sensor systems, we performed tests of self-consistency and comparisons of measurements to inform the selection of a superior system. Intercomp sensor data were found to be not self-consistent but IRD data were. Given the different measurements provided by the two sensors, without additional data, we were not able to reach a conclusion regarding the relative accuracy or the duration of consistent observations before needing recalibration. Initial comparisons indicated potential problems with the Intercomp sensor. We then suggested alternate approaches that MNDOT could use to determine whether recalibration was required. Finally, we analyzed ten-month data from the IRD WIM system and four-month data from the Kistler WIM system to evaluate relative sensor accuracy. While both systems were found to be self-consistent within the data time frame, the Kistler system generated more errors than the IRD system. Conclusions regarding relative accuracy could not be reached without additional data. We identified the sorts of measurements that would need to be monitored for recalibration and the methodology needed for estimating future recalibration time.

17. Document Analysis/Descriptors Weigh in motion scales, Weigh in motion, Sensors, Weight measurement, Statistical inference, Performance measurement		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 127	22. Price

Weigh-in-Motion Sensor and Controller Operation and Performance Comparison

FINAL REPORT

Prepared by:

Diwakar Gupta
Xiaoxu Tang
Lu Yuan
Department of Industrial and Systems Engineering
University of Minnesota

January 2018

Published by:

Minnesota Department of Transportation
Research Services & Library
395 John Ireland Boulevard, MS 330
St. Paul, Minnesota 55155-1899

ACKNOWLEDGMENTS

We are grateful to Transportation Data and Analysis (TDA) staff for providing us with the data needed to perform this analysis. In particular, we acknowledge generous help received from Mr. Gregory Wentz, Mr. Joshua Kuhn, and Mr. Benjamin Timerson.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

There are currently three Weigh-In-Motion (WIM) systems installed at MnROAD. One of these WIM systems (including sensor and controller) is manufactured by Intercomp, a Minnesota company headquartered in Medina. The second WIM system has quartz piezo-electric sensors manufactured by Kistler and controllers manufactured by IRD. The third system has Kistler sensors and controller. The Kistler WIM sensors use a piezo-quartz technology to determine the weight of the vehicle as it passes over the sensors. The Intercomp WIM sensors use a strain-gauge-type technology for determining the weight of the vehicle. The Kistler sensors cost about $28,000 per lane in materials cost. For the purpose of this research study, the Intercomp sensors cost $21,340 per lane in materials cost. Installation cost is in the range of $15,000 to $18,000 per lane for both sensors. The objective of this research project is to utilize statistical inference and comparison techniques to compare the performance of the Kistler and Intercomp WIM sensors. To this end, we perform three sets of analyses, which are described in the three chapters following the introduction.

In Chapter 2, we analyzed test-vehicle data to perform a check of accuracy of results reported by the sensor-vendor Intercomp. We performed this analysis in two iterations – one using the actual weights data that Intercomp used, and the other using truck-ticket weights as actual weights. Results from our first analysis largely confirmed the Intercomp-report results. In particular, measurement errors were not associated with observation number. Our second analysis using truck-ticket weights, however, indicated some differences. Details are presented in Chapter 2. In both sets of analyses, we determined that the data were insufficient to support any conclusion regarding the accuracy of the sensor under different temperature and speed conditions. Additionally, we found that axle count, vehicle class, and axle type were associated with measurement accuracy.

In Chapter 3, we analyzed limited data from the two sensor systems (Intercomp and IRD) to lay the foundation for analysis that leads to the selection of a superior system in terms of better accuracy and lower life-cycle cost in the future. Our analysis included tests of self-consistency within a single system and comparisons of measurements provided by the two sensor systems. By examining vehicle counts, vehicle weights (especially Class-9 vehicles), and vehicle speeds, we concluded that Intercomp sensor data were not self-consistent, whereas the IRD sensor data were. We also concluded that the measurements provided by the two sensors were different. Without additional data, we could not conclude which sensor was more accurate, or which sensor provided a longer period of consistent observations before needing recalibration. However, initial comparisons identified a potential issue with the Intercomp sensor. Our analysis also provided alternate approaches that MNDOT could use to determine the need of recalibration.

During the course of this project, Intercomp sensors failed in late February 2016 and we had to rely on alternative data from Kistler to perform the analysis reported in Chapter 4. Specifically, we analyzed ten-month data from the IRD WIM system and four-month data from the Kistler WIM system to evaluate sensor accuracy. We examined the self-consistency of the two systems and compared them in terms of daily observations and error counts, gross vehicle weights (especially for Class-9 vehicles) and speed. Our analyses indicated that both systems were self-consistent within respective timeframes. However, the measurements provided by the two systems were different. The Kistler system generated more errors than the IRD system. Without additional data, we were unable to conclude which system was more accurate. Our analysis also identified the sorts of measurements that need to be monitored for recalibration and the methodology for estimating future recalibration time.

CHAPTER 1: INTRODUCTION

Intercomp, a Minnesota company headquartered in Medina, manufactures Weigh-In-Motion (WIM) sensors and controllers. They have installed a WIM system at MnROAD. MnDOT has also installed a WIM at MnROAD that has quartz piezo-electric sensors manufactured by Kistler and a controller manufactured by IRD. The Kistler WIM sensors use a piezo-quartz technology to determine the weight of the vehicle as it passes over the sensors. They cost about $28,000 per lane in materials cost. The Intercomp WIM sensors use a strain gauge type of technology for determining the weight of the vehicle. For the purpose of this research study, the Intercomp sensors cost $21,340 per lane in materials cost. Installation cost is in the range of $15,000 to $18,000 per lane for both sensors.

The objective of this research project is to utilize statistical inference and comparison techniques to compare the performance of the Kistler and Intercomp WIM sensors. Depending on the results of the initial test, MnDOT would test a couple of different options, including the development of an interface of Intercomp sensors with the IRD controller, and an evaluation of modifications needed to make to its existing WIM polling software to be able to communicate and download data from the Intercomp system.

This research project consists of three tasks. The analyses and findings for each of these tasks are documented in separate chapters. Chapter 2 evaluates the baseline performance of Intercomp's sensors by utilizing the available test-vehicle data. We first verify the results documented in Intercomp's report (Kroll, Young, & Kroll, 2015) by performing the same analyses and tests as described in that report, including regression models using error percentage as the response variable and observation number as the explanatory variable. The observation number, in particular, is taken as a proxy for time and ambient conditions (temperature, humidity, and barometric pressure). The authors of that report also examine the relationship between dynamic weights and static weights. Thus, we perform the analyses relying on two sets of weights, the actual weights as used by the Intercom's report, and the truck-ticket weights (i.e., the true vehicle weights). We also discuss the appropriateness of the models documented in Intercomp's report and perform additional analysis to evaluate the baseline performance using methods that were not considered in Intercomp's report. In particular, we use truck-ticket weights and consider the effects of speed, axle count, vehicle number, temperature, and axle weights (steer, drive, and trailer) on error percentage.

With the baseline performance evaluated in Chapter 2, Chapter 3 proceeds to compare and contrast the reported performance of the Intercomp and the IRD WIM systems. The goal is to lay the foundation for analysis that leads to the selection of a superior system in terms of better accuracy and lower life cycle cost in the future. We analyze the CSV-formatted data provided by MnDOT (OTSM), which covers the time period of January 22 to March 28 of 2016 for Intercomp and October 1, 2015 to March 28, 2016 for IRD. We then develop a methodology for comparing and contrasting the two sensor systems' performances based on a variety of metrics.

An important criterion of sensor performance is self-consistency. A sensor's measurements may change gradually over time, or exhibit a sudden and sharp turning point. Whereas the former may be caused by a variety of reasons, it is typically the case the latter occurs when the sensor fails and requires recalibration. Thus, in Chapter 3, we first perform a self-consistency analysis of each system for a selection of performance metrics: daily observation count, gross vehicle weight, and speed.

Daily observation counts are the frequencies of the data records from each day. The counts equal to the volumes of daily traffic captured by the sensor. Average daily traffic counts are important for MNDOT because these counts are reported to the FHWA. While the daily counts may vary by the day of week, season, and holidays, a sudden and significant change in the pattern of average daily traffic count is a signal of sensor-system failure and possible need for recalibration. Thus, daily observation count is selected as a key performance metric. Similarly, gross vehicle weight (GVW) is also selected for use in our analysis. Speed is another important metric for traffic monitoring purposes, as the speed affects the accuracy of vehicle-class detection, axle spacing detection and error code generation. Because different vehicle classes have different gross vehicle weight distributions and our primary focus is on Class-9 vehicles (which is important to MnDOT), all the analyses of gross vehicle weights and speed are carried out for Class-9 vehicle data only within this chapter. Analyses of other classes can be performed in a similar fashion. We examine time consistency for each of the metrics in various aspects, such as calculating and comparing statistical measures (e.g., mean, variance, and distributions) by data type (with vs. without error), lane (lane 1 vs. lane 2), and/or month.

Intercomp and IRD sensors are installed near each other and there is no entry to or exit from the highway between the locations of the two sensor systems, which means they measure the same vehicles during the same time interval. Therefore, we may compare their sensor observations for the same time interval to assess differences. The Intercomp data starts from January 22, 2016 and ends March 28, 2016, which is a subset of dates over which the IRD data spans. Therefore, we use only the January 22nd to March 28th data for both sensors in these comparisons. Our analysis does not indicate which sensor is more accurate, but informs whether their observations are similar or not. Accuracy comparisons would be possible if we had true vehicle weights.

We then compare the two sensors for each of the three aforementioned measures by reporting and evaluating the time consistency of statistical mean, variance and distribution. Specifically, the two-sample T-test is used to test the equality of the means, F-test is used to check whether the two sensors have same variability, and Kolmogorov–Smirnov test is used to test distributional differences. Utilizing statistical techniques, we are able to identify instances of self-inconsistency and further recommend actions that MNDOT can take both to improve performance evaluation and assess the need for recalibration.

During the course of this project, Intercomp sensors failed in late February 2016. As a result, we had to rely on alternative data from Kistler to perform the analysis reported in Chapter 4. The goal of this chapter is to compare and contrast the reported performance of the Kistler and the IRD WIM systems. Our analysis in this chapter follows the same approach as used in Chapter 3. Ten months of the IRD-system data (i.e., October 1, 2015 to July 28, 2016) and four months of the Kistler-system data (i.e., April 1, 2016 to July 28, 2016) in CSV format were provided by MNDOT for the purpose of this analysis. The data was produced by the installed sensors and controllers. Since the data for the Kistler system only contains Lane-1 (driving lane) measurements, the comparison of the two systems is based on lane 1 data only. As with Chapter 3, we examine the self-consistency for each of the two systems and then compare the two systems in terms of daily observations counts, gross vehicle weights (especially for Class-9 vehicles), and speed. Our analyses indicate that both systems were self-consistent within respective timeframes. However, the measurements provided by the two systems were different and the Kistler system generated more errors than the IRD system. Without additional data, we are unable to conclude which system was more accurate. Our analysis identifies the sorts of measurements that need to be monitored for recalibration and the methodology for estimating future recalibration time.

CHAPTER 2: EVALUATION OF INTERCOMP'S TEST VEHICLE DATA

2.1 INTRODUCTION

2.1.1 Research Goal

The goal of this chapter is to evaluate the baseline performance of Intercomp's sensors by utilizing the available test vehicle data. We perform the analysis in two parts. First, we verify the results documented in Intercomp's report (Kroll, Young, & Kroll, 2015). Second, we perform additional analysis to evaluate the baseline performance using methods that are not used in Intercomp's report.

2.1.2 Data Description

Two data sets are utilized for the analysis performed in this chapter. One is the raw data produced by the WIMLOGIX controller. It contains seven Excel files, with records of all vehicles passing through the sensor on August 8, October 1, October 2, October 18, October 22, October 23, and December 12 of 2013. This data set is referred to as Data Set 1. The other data set is an Excel file that contains records of the test vehicles only (identified by OtherID=x). This data set is referred to as Data Set 2. The report that Intercomp published uses data that appears to match Data Set 2. Ideally, Data Set 2 should match the static scale weight tickets, which, however, is not the case.

Both Data Set 1 and Data Set 2 contain individual axle left and axle right (wheel) weights that are produced by the sensors. For each vehicle, while the individual left and right axle weights are identical between the two data sets, the sum of the left and right weights matches (within 1 pound error).

We obtained actual weights from three sources. A summary is shown in Table 2.1.

Table 2.1 Actual Weights

	Truck Tickets		Intercomp's Report	Data Set 2
Vehicle #	**Measurements**	**Average**		
1	47,600/47,400	47,500	47,500	47,500
2	48,700/48,700	48,700	48,700	45,900
3	73,680	73,680	73,690	73,700
4	79,200/79,220	79,210	79,210	79,210
5	70,920/70,900	70,910	70,910	70,910
6	NA	NA	79,800	79,900
7	(81,540-200)/81,300	81,320	81,400	81,300
8	47,600/47,400	47,500	47,500	47,500
9	28,140/28,140	28,140	28,140	28,040

Note that vehicle number 6 is a MNDOT truck and the actual weights are: 11,960 lbs for the steer axle, 34,360 lbs for the drive tandem, and 33,580 lbs for the trailer tandem for a total gross vehicle weight (GVW) of 79,900 lbs. There are no truck tickets for vehicle number 6. The first ticket weight of vehicle number 7 is reduced by 200 lbs to account for the weight of a staff member who was standing on the scale at that time.

There are two data records in Data Set 1 and Data Set 2 that do not have matching records. One is from Data Set 1 at 08:26:27 on August 29, 2013, which has no matching record in Data Set 2. This vehicle is a 5-axle vehicle traveling at a speed of 62.9 mph. The measured GVW is 79,216 lbs. Vehicle number is not recorded. The second data record is from Data Set 2 on December 12, 2013, which does not have a matching record in Data Set 1. This vehicle is vehicle number 6 (Semi Flatbed Trailer 5 Axle), with measured GVW of 80,811 lbs. We drop both of these observations from further analysis. This leaves us with 301 valid observations.

Statistical summaries of these data are shown in Figure 2.1 and Table 2.2.

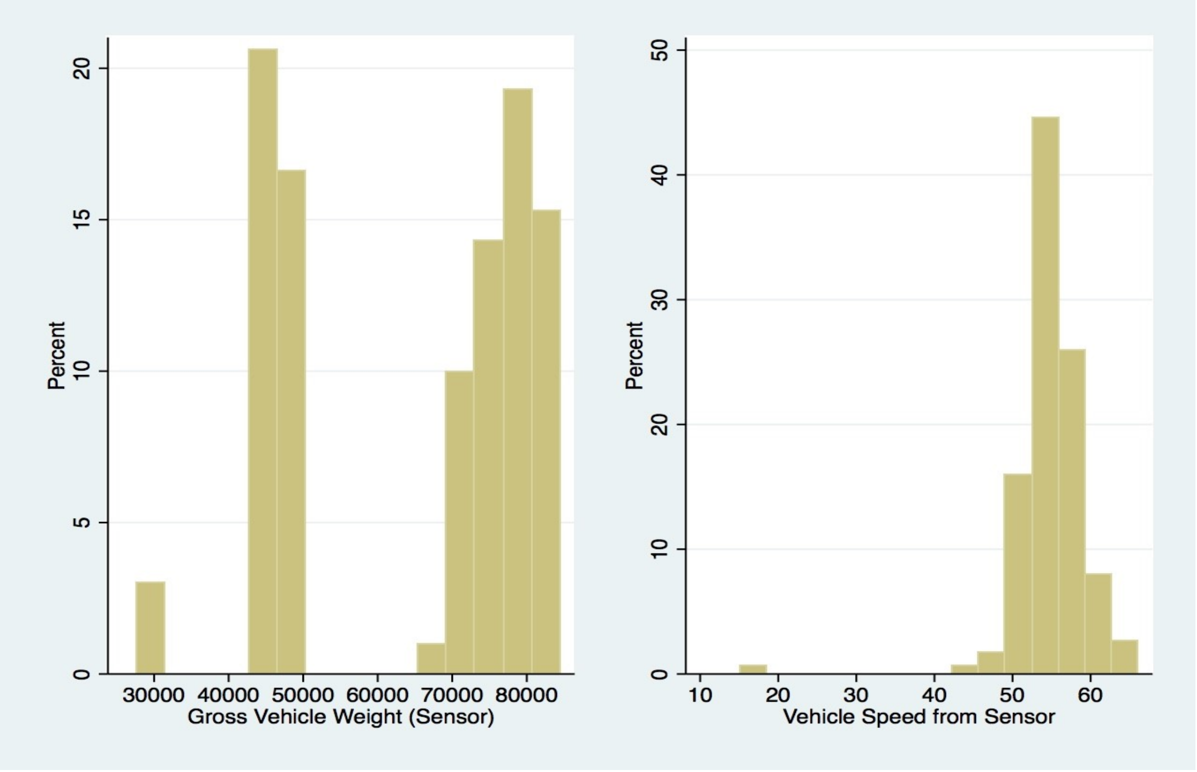

Figure 2.1 Frequency (in Percentage) of GVW and Speed

Table 2.2 Statistical Summary of Axle Count, GVW and Speed

Axle Count Sensor Detected	Freq.	mean(GVW)	mean(Speed)
3	29	46865.10345	53.23448
4	83	46321.61446	54.55542
5	127	75468.5748	56.60866
7	62	74568.5	53.19516

2.1.3 Overall Approach

Our approach to the analysis consists of three steps. First, we perform the same analyses and tests that are documented in Intercomp's report (Kroll, Young, & Kroll, 2015). The analyses include regression models with error percentage as the response variable, and observation number as the explanatory variable. Note that the observation number is taken as a proxy for time and ambient conditions (temperature, humidity, and barometric pressure). The authors of that report also examine the relationship between dynamic weights and static weights. In this section, we will first use actual weights from Data Set 2, which are used to generate the analyses of the report. Then, we redo the analysis with truck-ticket weights treated as actual weights, which are the true vehicle weights. We also discuss the appropriateness of the models used in Intercomp's report.

Second, we use truck-ticket weights to perform additional analyses in which we consider the effects of speed, axle count, vehicle number, temperature, and axle weights (steer, drive, and trailer) on error percentage. This part of our analysis goes beyond what is in Intercomp's report (Kroll, Young, & Kroll, 2015).

Finally, because we observed a discrepancy in the actual weights of vehicle number 2 between Data Set 2 and the truck tickets, we repeated all of the analyses after deleting the data pertaining to vehicle number 2.

Note: All the measurements of vehicle number 2 in Data Set 1 are less than the actual weights from truck tickets. Therefore, it is possible that there is a systematic error in the data for vehicle number 2.

2.1.4 Summary of Results

We performed a statistical analysis of sensor accuracy by using two different sets of actual weights. One set of weights came from Intercomp, and the other from truck-ticket weights. When performing the analysis with the first set of weights, the vast majority of our results matched those documented in Intercomp's report. However, we obtained different results when using truck-ticket weights. In both cases, we could not verify the conclusion in Intercomp's report regarding the effect of temperature and environmental factors because of limited data. Finally, we analyzed the correlation between other factors, such as axle count, axle type, speed, and vehicle class, and error percentage. Such analysis is not presented in the Intercomp report.

The results of the additional analysis mentioned above depend on which actual weight values are used in the analysis. If we use truck-ticket weights, we find that observation number, axle count, and vehicle number positively correlate with error percentage. Similarly, the measured and actual weights are statistically different, with the sensor underestimating the weights for Class 6 and 7 vehicles, and overestimating weights for Class 10. Even though weights are statistically different except for Class 9 and 10, they are not practically different and the differences are within tolerable bounds.

The remainder of this chapter is organized as follows. In section 2.2 we evaluate the performance of Intercomp's sensor and present additional analyses considering factors that are not considered in the latter. In section 2.3 we summarize our findings and conclusions.

2.2 ANALYSIS

2.2.1 Verification of Intercomp's Analyses

Intercomp uses the actual weights from Data Set 2 to generate analyses in their report (Kroll, Young, & Kroll, 2015). We will redo their analyses and tests to verify the accuracy of their conclusions. Because the actual weights in Data Set 2 are different from the truck tickets, we will also perform analyses using truck tickets as the actual weights.

2.2.1.1 Part 1: Use actual weights from Data Set 2

The error percentage was calculated as (Measured weight – Actual weight)/Actual weight, where measured weight is the Gross Vehicle Weight (GVW) estimated by the Intercomp sensor. Figure 2.2 shows the scatter-plot by observation number, which are observations sorted in chronological order.

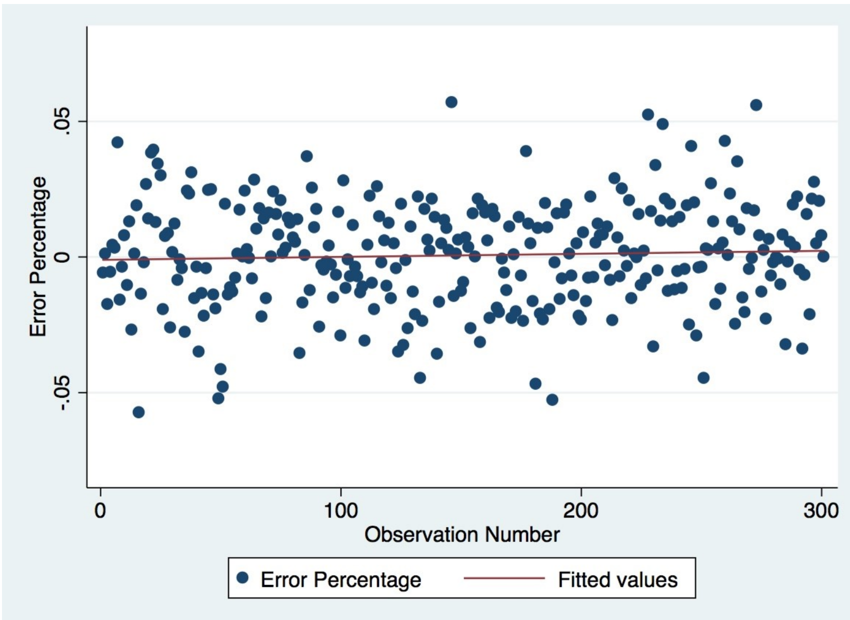

Figure 2.2 Error Percentage Scatter-plot and Best-fit Line by Observation Number

The solid line is the fitted regression line (best fit, least square). Figure 2.2 is very similar to Table 5 in Intercomp's report (Kroll, Young, & Kroll, 2015), in terms of both the scatter distribution and the regression line. This suggests that the two analyses yield the same results. This is not surprising because we used the same data that was used in the Intercomp's report.

The regression result for Model (1.1) is shown in Table 2.3.

$$\text{Error Percentage} = \text{Cons} + \text{Coef.} \times \text{Observation Number} + \text{Error Term} \qquad (1.1)$$

Table 2.3 Regression Analysis (Error Percentage = Cons + Coef. × Observation Number + Error Term)

Source	SS	df	MS
Model	.000281077	1	.000281077
Residual	.114745609	299	.000383765
Total	.115026686	300	.000383422

Number of obs	= 301
F(1, 299)	= 0.73
Prob > F	= 0.3928
R-squared	= 0.0024
Adj R-squared	= -0.0009
Root MSE	= .01959

errpercentage_d2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNumber	.0000111	.000013	0.86	0.393	-.0000145	.0000367
_cons	-.001048	.0022639	-0.46	0.644	-.0055032	.0034073

The regression result shows that error percentage does not exhibit a pattern with respect to observation number (Prob > F is greater than 0.05). This means that error percentage did not change in a statistically significant way with observation number. These conclusions are similar to those in the report (Kroll, Young, & Kroll, 2015), although the numerical values in our regression table are not exactly the same because the data we used are slightly different - only 301 data points are used in our analyses.

When comparing the dynamic (measured) weights and static (actual) weights, the report (Kroll, Young, & Kroll, 2015) assumes a linear model with no intercept, which means Y=aX, where Y = fitted value, x = static weight. With this same assumption, we could get the result shown in Figure 2.3.

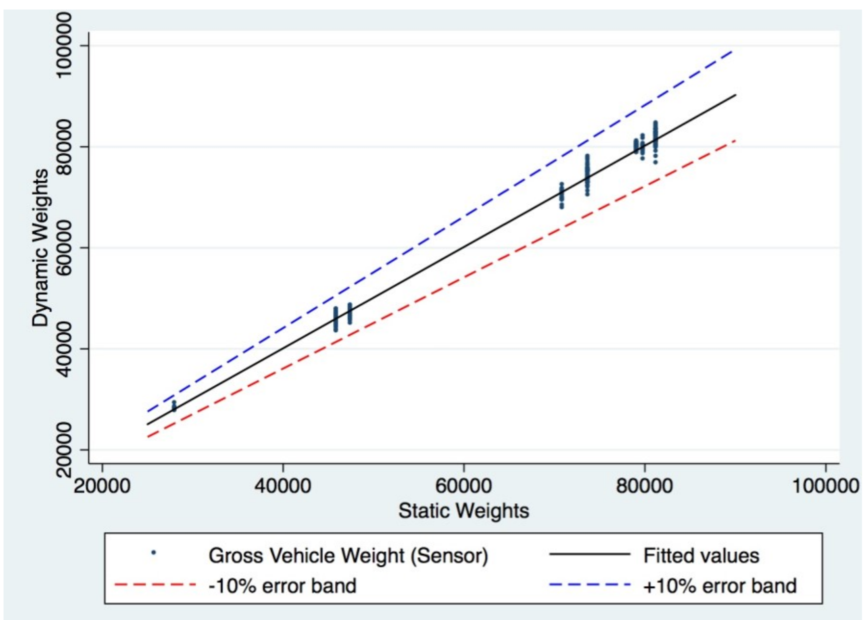

Figure 2.3 Dynamic and Static Weights with Error Bands

The dots are the scatter-plots of the data. The solid line is the fitted regression line. The upper and lower dash lines are the 10% error bands. Figure 2.3 is very similar to Table 6 in Intercomp's report (Kroll, Young, & Kroll, 2015). All the data points are within the 10% error band. The regression result for Model (1.2) is shown in Table 2.4.

$$\text{Dynamic Weight} = \text{Coef.} \times \text{Static Weight} + \text{Error Term} \tag{1.2}$$

Table 2.4 Regression Analysis (Dynamic Weight = Coef. × Static Weight + Error Term)

Source	SS	df	MS
Model	1.3334e+12	1	1.3334e+12
Residual	483609051	300	1612030.17
Total	1.3339e+12	301	4.4316e+09

Number of obs =	301
F(1, 300) =	.
Prob > F =	0.0000
R-squared =	0.9996
Adj R-squared =	0.9996
Root MSE =	1269.7

GVW	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aGVW_d2	1.002818	.0011026	909.49	0.000	1.000648	1.004987

The response variable is the measured weight from the sensor, which is also called the dynamic weight. The explanatory variable aGVW_d2 is the actual gross vehicle weight from Data Set 2. The estimated coefficient is 1.002818, which means there is a slope error of 0.28%, with its 95% confidence interval between 0.06% and 0.49%. This result is the same in Intercomp's report (Kroll, Young, & Kroll, 2015). The dynamic weights are quite close to the static weights. This result is similar to the conclusions reached in Intercomp's report.

We obtained the temperature data on the dates of the test runs from external sources (Weather Underground, n.d.). Table 2.5 shows a summary of these temperature values.

Table 2.5 Temperature Summary

Date	temp_~an	temp_min	temp_max
08/29/2013	85	75	95
10/01/2013	65	53	77
10/02/2013	63	48	77
10/18/2013	46	38	53
10/22/2013	33	26	40
10/23/2013	36	33	39
12/12/2013	10	1	19

The temperature unit is Fahrenheit. Mean, minimum and maximum of the day are shown in Table 2.5, represented by variable names temp_~an, temp_min and temp_max, respectively. We perform our analyses based on the average temperature of each day, which is the second column of Table 2.5.

The regression result for Model (1.3) is shown in Table 2.6.

$$\text{Error Percentage } = \text{Cons} + \text{Coef.} \times \text{ Average Temperature } + \text{ Error Term} \qquad (1.3)$$

Table 2.6 Regression Analysis (Error Percentage = Cons + Coef. × Average Temperature + Error Term)

Source	SS	df	MS
Model	.000026273	1	.000026273
Residual	.115000412	299	.000384617
Total	.115026686	300	.000383422

Number of obs =	301
F(1, 299) =	0.07
Prob > F =	0.7940
R-squared =	0.0002
Adj R-squared =	-0.0031
Root MSE =	.01961

errpercent~2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp_mean	-.0000193	.000074	-0.26	0.794	-.000165	.0001263
_cons	.0015011	.0035145	0.43	0.670	-.0054153	.0084174

From Table 2.6, we see that the error percentage is not correlated with temperature, which is also the conclusion stated in Intercomp's report. The Intercomp's report uses observation number as a proxy for environmental factors and concludes that there is negligible impact due to weather or environmental conditions. The dates in the data set range from August to December, and the temperature ranges from 95 F to 1 F. That being said, the dates are not evenly spread out, and there are very few observations. Specifically, there is only one day in the data with average temperature below freezing, and only one day with average temperature above 80 F. Therefore, the data is insufficient to support any conclusions regarding environmental factors.

In summary, we find that when using the actual weights from Data Set 2, which are also used by Intercomp to generate their report (Kroll, Young, & Kroll, 2015), all of the conclusions could be verified by performing the same statistical analyses and tests, except for the conclusion regarding the effect of environmental conditions.

2.2.1.2 Part 2: Use actual weights from truck tickets

In this section, the actual weights are the truck-ticket weights. The same analyses are performed. First, Figure 2.4 shows the scatter-plot of the error percentage with respect to the observation number. The fitted line shows an overweighting trend over time.

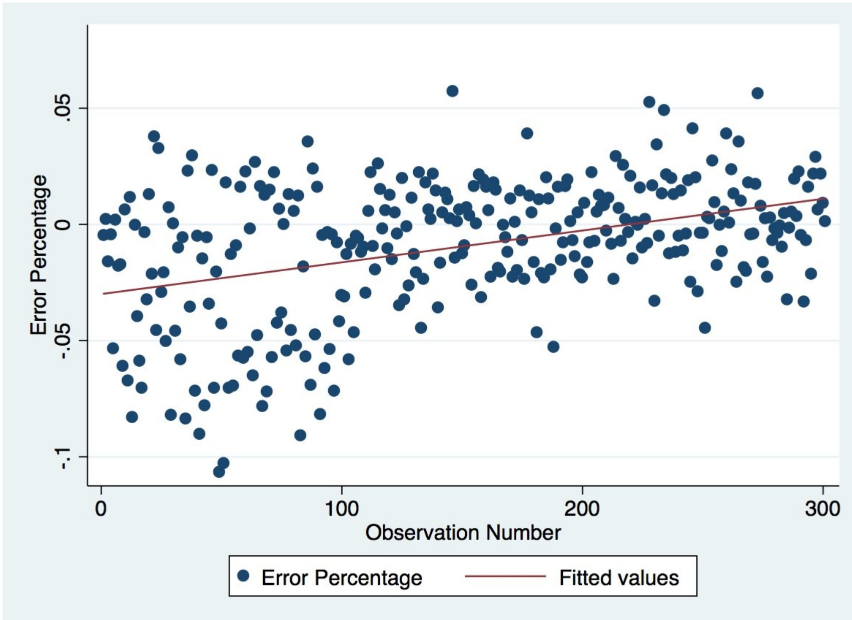

Figure 2.4 Error Percentage by Observation Number

The regression result for Model (1.4) is shown in Table 2.7.

$$\text{Error Percentage } = \text{Cons} + \text{Coef.} \times \text{ Observation Number } + \text{ Error Term} \tag{1.4}$$

Table 2.7 Regression Analysis (Error Percentage = Cons + Coef. × Observation Number + Error Term)

Source	SS	df	MS
Model	.042469004	1	.042469004
Residual	.22011939	299	.000736185
Total	.262588394	300	.000875295

Number of obs =	301
F(1, 299) =	57.69
Prob > F =	0.0000
R-squared =	0.1617
Adj R-squared =	0.1589
Root MSE =	.02713

errpercentage_t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNumber	.0001367	.000018	7.60	0.000	.0001013	.0001721
_cons	-.0300407	.0031356	-9.58	0.000	-.0362114	-.02387

We see that the error percentage is correlated with the observation number, and this effect is significant. The coefficient is positive, which means there is significant evidence that the sensor overestimates weights over time (p-value is zero).

We assume the same linear model:

$$\text{Dynamic Weight} = \text{Coef.} \times \text{Static Weight} + \text{Error Term} \qquad (1.5)$$

when comparing the dynamic (measured) weights and static (actual) weights. A summary plot is shown in Figure 2.5.

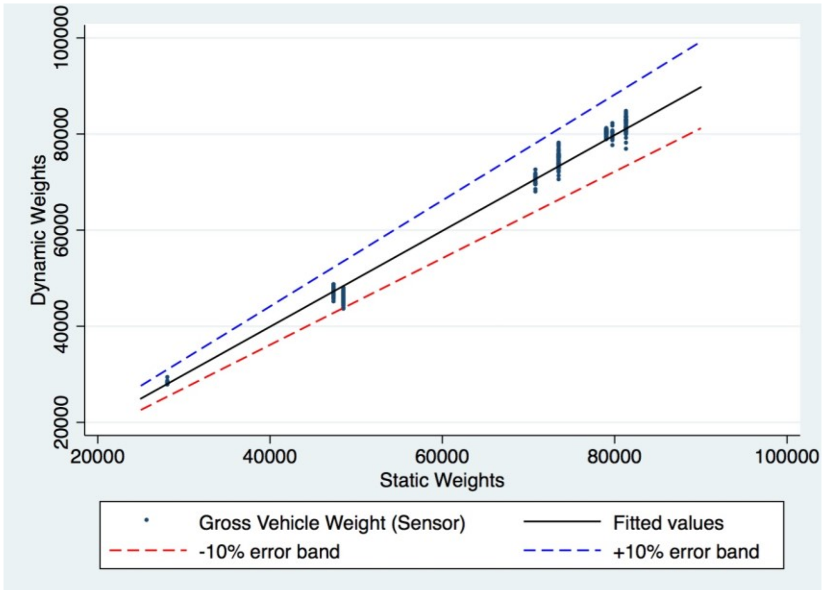

Figure 2.5 Dynamic and Static Weights with Error Bands

As we can see from Figure 2.5, some data points fall out of the 10% error band. This is because the slope estimate is smaller than what we observed in the previous section. The reason is that the sensor sometimes underestimates the weights, especially for vehicle number 2, so that the coefficient is less than 1. The regression result for Model (1.5) is in Table 2.8.

Table 2.8 Regression Analysis (Dynamic weight = Coef. × Static Weight + Error Term)

Source	SS	df	MS
Model	1.3330e+12	1	1.3330e+12
Residual	906750624	300	3022502.08
Total	1.3339e+12	301	4.4316e+09

Number of obs =	301
F(1, 300) =	.
Prob > F =	0.0000
R-squared =	0.9993
Adj R-squared =	0.9993
Root MSE =	1738.5

GVW	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aGVW_t	.9975748	.0015022	664.10	0.000	.9946187	1.000531

The variable aGVW_t is the actual gross vehicle weight from truck tickets. We can see that when using the truck-ticket weights as the actual weights, the result is not as good as in Intercomp's report. Overall, the sensor underestimates the weights, although it has a tendency to overestimate across time. Furthermore, the time effect is statistically significant.

We use the same temperature data for the analysis of the error percentage versus temperature.

The regression result for Model (1.6) is shown in Table 2.9.

$$\text{Error Percentage} = \text{Cons} + \text{Coef.} \times \text{Average Temperature} + \text{Error Term} \qquad (1.6)$$

Table 2.9 Regression Analysis (Error percentage = Cons + Coef. × Average Temperature + Error Term)

Source	SS	df	MS
Model	.046640502	1	.046640502
Residual	.217545267	299	.000727576
Total	.264185769	300	.000880619

Number of obs =	301
F(1, 299) =	64.10
Prob > F =	0.0000
R-squared =	0.1765
Adj R-squared =	0.1738
Root MSE =	.02697

errpercent~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp_mean	-.0008148	.0001018	-8.01	0.000	-.0010151	-.0006145
_cons	.0274559	.0048338	5.68	0.000	.0179433	.0369686

As shown in Table 2.9, the average daily temperature significantly affects the error percentage even though we only use limited data, which is different from Part 1 and Intercomp's report. One possible explanation for this observation is that daily mean temperature serves as a proxy for the observation number, giving rise to the same overall trend as we observed with the observation number. The sign of the coefficient is negative because temperature drops over time.

The analysis is different from that in Part 1 because of the large discrepancy in the truck-ticket weights of vehicle number 2 (Data Set 2 assumes 45,900, whereas truck tickets amount to 48,700).

2.2.2 Additional Performance Analyses

The most relevant criterion for assessing accuracy is the difference between the actual and the dynamic weights over time. As we have already analyzed in Part 1, when using the actual weights from truck tickets, the error is not independent of time. In particular, it shows that sensors overestimate over time. In this section, we will analyze other factors that may affect the error in addition to observation number (or time).

As shown in Table 2.10 and Table 2.5, the test run data are from 7 different days with the following properties:

- The numbers of observations are not the same across all days.
- The intervals between dates are not uniform.

- The numbers of observations for an important vehicle class (#9) are not the same.
- The average temperatures are not the same.

For these reasons, the observation number is not a reasonable representative of either time or environmental factors.

Table 2.10 Data Summary of Test Runs

MM/DD/YYYY format	MNclass 6	7	9	10	Total
08/29/2013	0	0	4	0	4
10/01/2013	0	26	26	0	52
10/02/2013	0	25	25	0	50
10/18/2013	0	0	5	0	5
10/22/2013	29	0	33	29	91
10/23/2013	0	32	31	31	94
12/12/2013	0	0	5	0	5
Total	29	83	129	60	301

Next, we will look at factors such as axle count, vehicle number, speed, MN classification, and axle type. In this section, actual weights are the truck-ticket weights.

2.2.2.1 Individual Factor Analyses – Axle count, Vehicle number, and Speed

Intercomp's report does not contain an analysis of performance by speed or axle spacing. However, it mentions that this performance is good [Page 20 of the Intercomp's report (Kroll, Young, & Kroll, 2015)]. Our goal in this section is to verify this claim.

First we will plot the data to observe the relationship between the error percentage and other factors. Figure 2.6, Figure 2.7, and Figure 2.8 contain scatter plots of error percentage versus axle count, vehicle number and speed.

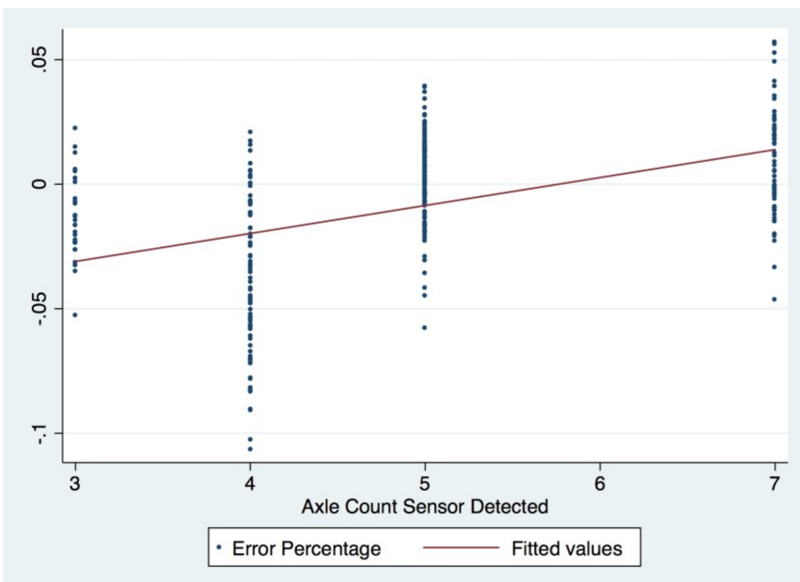

Figure 2.6 Error Percentage and Axle Count

Figure 2.6 shows that error percentages are quite different for different axle numbers, although they remain within the tolerable error bounds. A possible reason could be that errors add up as the number of axle increases.

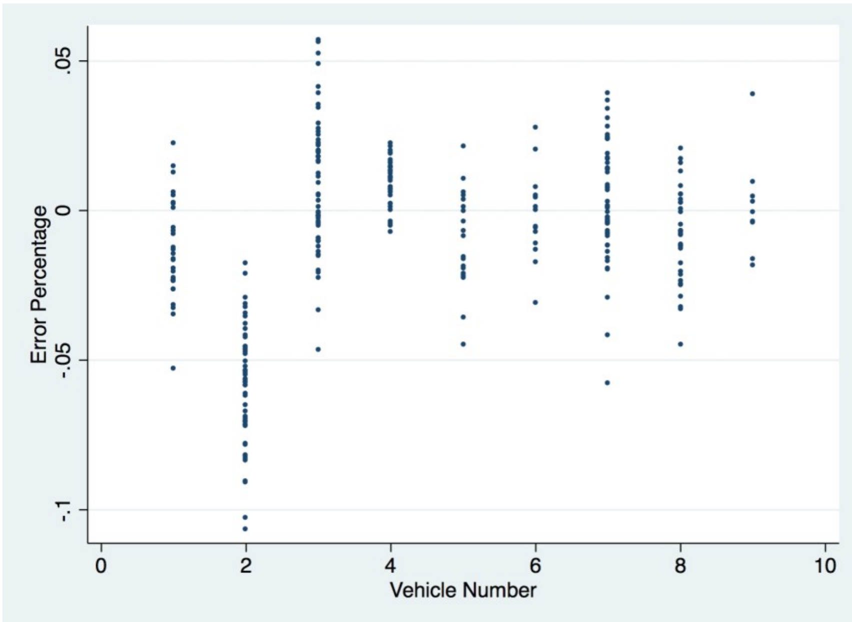

Figure 2.7 Error Percentage and Vehicle Number

Figure 2.7 shows that Vehicle # 2's weights are underestimated. Reasons should be investigated. The actual weight from truck tickets and the report is 48,700, but Intercomp used 45,900 in its data analysis.

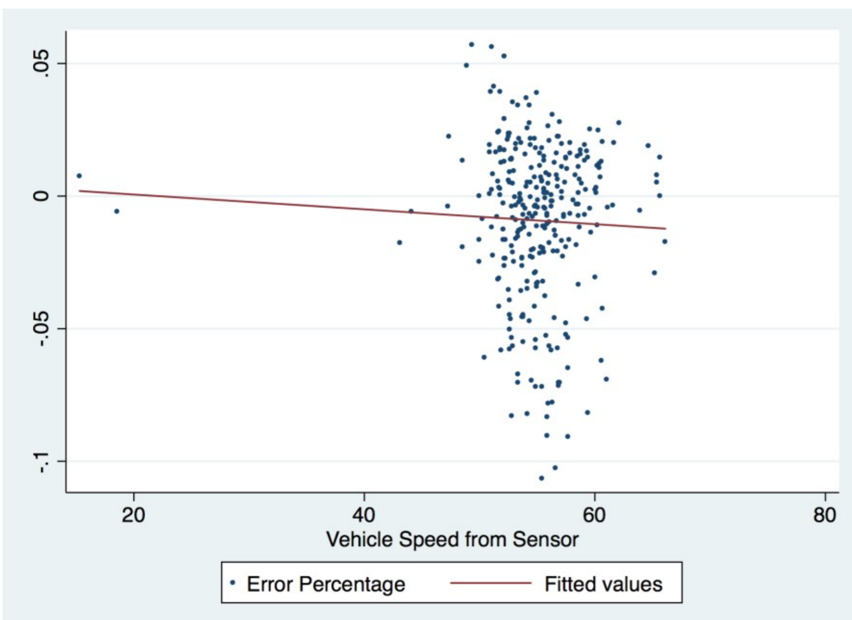

Figure 2.8 Error Percentage and Speed

As shown in Figure 2.8, the error percentages appear to be distributed randomly around zero. Truck speeds are recorded between 15 and 66 mph, but most observations are between 45 and 60 mph. Because of the small range of vehicle speeds, observations we make regarding the effect of speed cannot be extrapolated to speed values that lie outside the range in the test data. We do not have axle spacing data in the given data sets.

Now we will consider Model (1.7) described below:

$$\text{Error Percentage} = \text{Cons} + \text{Coef.}\,1 \times \text{Observation Number} + \text{Coef.}\,2 \times \text{Axle Count} \quad (1.7)$$
$$+\text{Coef.}\,3 \times \text{Vehicle Number} + \text{Coef.}\,4 \times \text{Speed} + \text{Error term}$$

The regression result is shown in Table 2.11.

Table 2.11 Regression Analysis of Other Factors (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.098266941	4	.024566735
Residual	.165918828	296	.000560537
Total	.264185769	300	.000880619

Number of obs	=	301
F(4, 296)	=	43.83
Prob > F	=	0.0000
R-squared	=	0.3720
Adj R-squared	=	0.3635
Root MSE	=	.02368

errpercentage_t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNum~r	.0000916	.0000166	5.52	0.000	.0000589	.0001243
AxleCount	.0092909	.0011576	8.03	0.000	.0070127	.0115692
VN	.0032568	.0005754	5.66	0.000	.0021244	.0043892
Speed	-.0004317	.0003035	-1.42	0.156	-.001029	.0001656
_cons	-.0595777	.0179238	-3.32	0.001	-.0948519	-.0243035

Table 2.11 indicates that observation number, axle count and vehicle number are correlated with the error percentage. The coefficients for all three factors are all positive, which means larger variable value is associated with larger error percentage. Speed is not a significant factor, with a p-value of 0.16.

2.2.2.2 MN Classification Analysis

Since axle count is a good representative of MNDOT classification, we do not include MNDOT classification into the previous model. Instead, we analyze this factor separately.

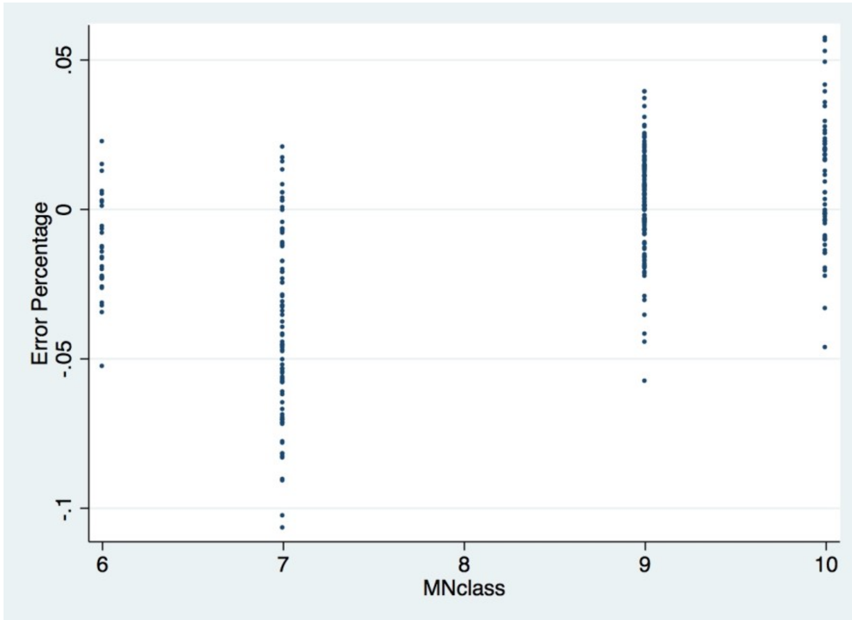

Figure 2.9 Error Percentage and MNDOT Classification

Figure 2.9 indicates that the distributions of error percentage are different among different vehicle classes. The under weights of class 7 are due to the large discrepancy of vehicle number 2 between Data Set 2 and truck tickets. The error percentage of class 9 is evenly distributed around 0. Hypothesis tests about the mean of the error percentage for each class are given in Table 2.12, Table 2.13, Table 2.14, and Table 2.15. In each test, the null hypothesis is that the mean of the error percentage is zero. We use a critical value of 0.05 to interpret the results.

Table 2.12 Hypothesis Test of Class 6 (Hypothesis: Mean Error Percentage in Estimating Weight of Mnclass 6 = 0)

```
-> MNclass = 6

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |      29   -.0133662    .0031257    .0168323   -.0197689   -.0069636
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =  -4.2763
Ho: mean = 0                                     degrees of freedom =       28

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0001         Pr(|T| > |t|) = 0.0002          Pr(T > t) = 0.9999
```

There is significant evidence that the error percentage is less than 0, which means the measured weights are smaller than the actual weights of class 6.

Table 2.13 Hypothesis Test of Class 7 (Hypothesis: Mean Error Percentage in Estimating Weight of Mnclass 7 = 0)

```
-> MNclass = 7

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |      83   -.0394166    .0033809    .0308013   -.0461423    -.032691
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t = -11.6587
Ho: mean = 0                                     degrees of freedom =       82

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Class 7 has the same result as class 6, as shown in Table 2.13. The measured weights are significantly smaller than the actual weights.

Table 2.14 Hypothesis Test of Class 9 (Hypothesis: Mean Error Percentage in Estimating Weight of Mnclass 9 = 0)

```
-> MNclass = 9

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |     129    .0026005    .0014775    .0167812   -.0003229     .005524
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =   1.7601
Ho: mean = 0                                     degrees of freedom =      128

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9596         Pr(|T| > |t|) = 0.0808          Pr(T > t) = 0.0404
```

The null hypothesis that the error percentage of class 9 is zero is accepted. The measured weight is statistically the same as the actual weight (p-value = 0.08). This is important because class 9 is an important vehicle class for MNDOT.

Table 2.15 Hypothesis Test of Class 10 (Hypothesis: Mean Error Percentage in Estimating Weight of Mnclass 10 = 0)

```
-> MNclass = 10

One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
errper~t	60	.0092942	.0027942	.0216442	.003703	.0148855

```
    mean = mean(errpercentage_t)                                  t =   3.3262
Ho: mean = 0                                     degrees of freedom =       59

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9992         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.0008
```

As shown in Table 2.15, the error percentage of Class 10 is significantly greater than 0, which means the measured weights are statistically greater than the actual weights. This result is opposite of what we observe for Classes 6 and 7.

To compare the mean of each MN class, we use Tukey pairwise comparison tests, which are shown in Table 2.16.

Table 2.16 Pairwise Comparison between MN Classes

errpercent~t	Contrast	Std. Err.	Tukey t	Tukey P>|t|	Tukey [95% Conf.	Interval]
MNclass						
7 vs 6	-.0260504	.0048382	-5.38	0.000	-.0385507	-.0135501
9 vs 6	.0159668	.0046095	3.46	0.003	.0040576	.027876
10 vs 6	.0226605	.0050727	4.47	0.000	.0095545	.0357664
9 vs 7	.0420172	.0031561	13.31	0.000	.033863	.0501714
10 vs 7	.0487109	.0038008	12.82	0.000	.0388911	.0585307
10 vs 9	.0066937	.0035049	1.91	0.226	-.0023617	.0157491

From the pairwise comparison results we could see that only the mean error percentages of Class 9 and 10 are statistically the same.

To summarize, the measurement errors are practically small for vehicle classes other than 6 and 7, although statistically different for all vehicle classes except Class 9 and 10. For Class 6 and 7 vehicles, the sensor underestimates weight.

2.2.2.3 Axle Type Analysis

Because we are interested in the performance of axles, we will do additional analyses based on different axle types, which are steer, drive and trailer. Note that vehicle number 1, 2 and 8 do not have trailer axles.

First, we will look at three types of axles separately to see whether there is any difference between axle types. Again, the error percentage is calculated as (measured weight – actual weight)/actual weight. Scatter-plots of three axle types are shown in Figure 2.10, and Figure 2.11 and Figure 2.12.

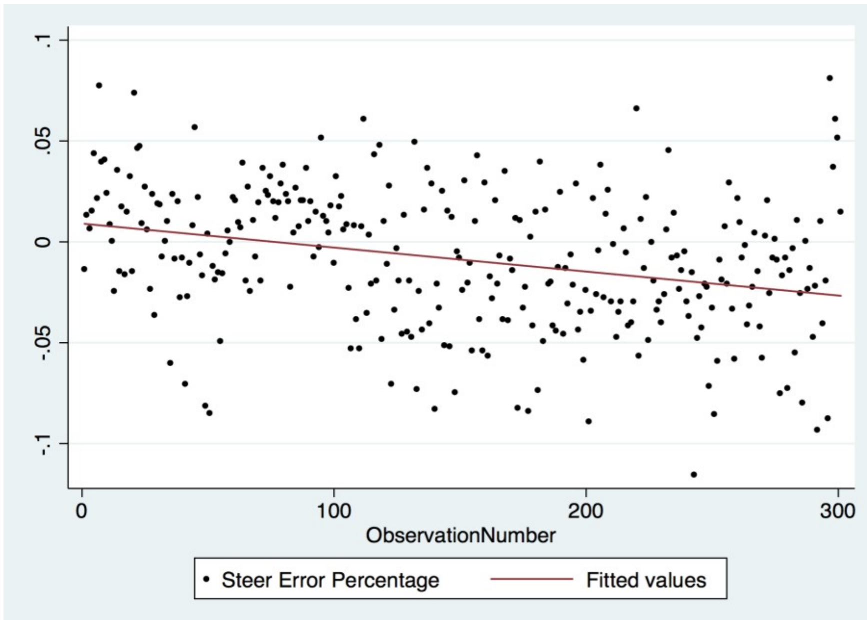

Figure 2.10 Steer Axle Error Percentage and Observation Number

As we can see from Figure 2.10, the steer error percentage has a decreasing trend over time. The slope of the fitted line is negative. The regression result for Model (1.8) is shown in Table 2.17.

$$\begin{aligned} \text{Steer Error Percentage} = {} & \text{Cons} + \text{Coef. } 1 \times \text{Observation Number} \\ & + \text{Coef. } 2 \times \text{Axle Count} + \text{Coef. } 3 \times \text{Vehicle Number} \\ & + \text{Coef. } 4 \times \text{Speed} + \text{Error term} \end{aligned} \tag{1.8}$$

Table 2.17 Steer Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.101875449	4	.025468862
Residual	.25110958	296	.000848343
Total	.352985029	300	.001176617

Number of obs	=	301
F(4, 296)	=	30.02
Prob > F	=	0.0000
R-squared	=	0.2886
Adj R-squared	=	0.2790
Root MSE	=	.02913

errpercentag~eer	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNum~r	-.0000981	.0000204	-4.80	0.000	-.0001383	-.0000579
AxleCount	-.0112918	.0014242	-7.93	0.000	-.0140946	-.0084891
VN	.0030449	.0007079	4.30	0.000	.0016519	.004438
Speed	.0002243	.0003734	0.60	0.548	-.0005105	.0009591
_cons	.0359684	.0220502	1.63	0.104	-.0074267	.0793635

Observation number, axle count and vehicle number are all statistically significant in this model. Speed is not significant. The differences are that the coefficients of Observation number and axle count were positive in the previous model, but negative in this model. That is to say, the steer error percentage has opposite trend compared with overall measured weight error percentage.

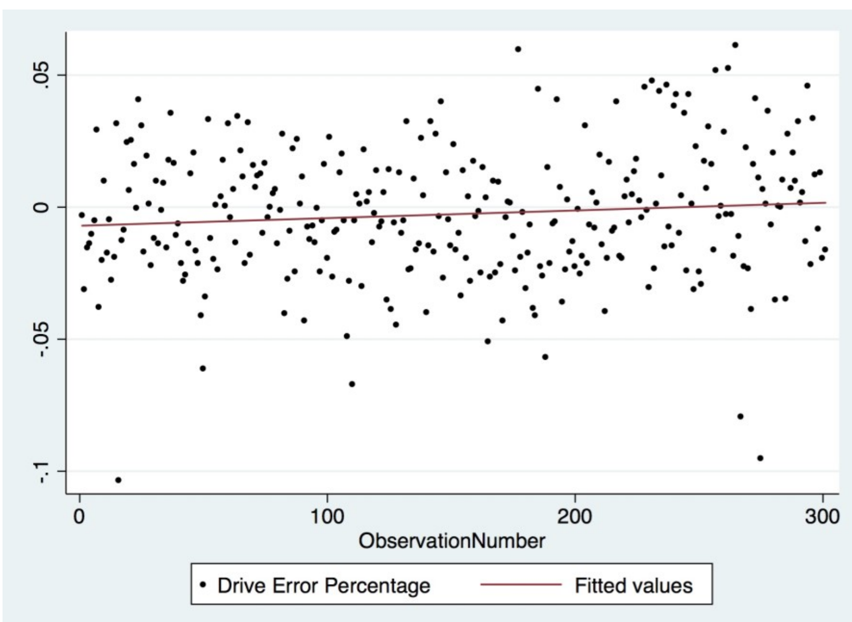

Figure 2.11 Drive Axle Error Percentage and Observation Number

Figure 2.11 also shows the fitted line for the drive error percentage, which is almost horizontal across time. The data points are randomly distributed over the plot. The regression result for Model (1.9) is shown in Table 2.18.

$$\text{Drive Error Percentage} = \text{Cons} + \text{Coef. 1} \times \text{Observation Number} + \text{Coef. 2} \times \text{Axle Count} + \text{Coef. 3} \times \text{Vehicle Number} + \text{Coef. 4} \times \text{Speed} + \text{Error term} \quad (1.9)$$

Table 2.18 Drive Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.042201262	4	.010550316
Residual	.141338255	296	.000477494
Total	.183539517	300	.000611798

Number of obs	= 301
F(4, 296)	= 22.10
Prob > F	= 0.0000
R-squared	= 0.2299
Adj R-squared	= 0.2195
Root MSE	= .02185

errpercentage_~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNum~r	9.23e-06	.0000153	0.60	0.547	-.0000209	.0000394
AxleCount	.0090757	.0010685	8.49	0.000	.006973	.0111784
VN	-.0011285	.0005311	-2.13	0.034	-.0021736	-.0000834
Speed	-.0005922	.0002801	-2.11	0.035	-.0011435	-.0000409
_cons	-.0113698	.0165429	-0.69	0.492	-.0439264	.0211868

As we can see in Table 2.18, axle count, vehicle number and speed are significant in this model (p-value < 0.05). But observation number is not significantly important (p-value = 0.55).

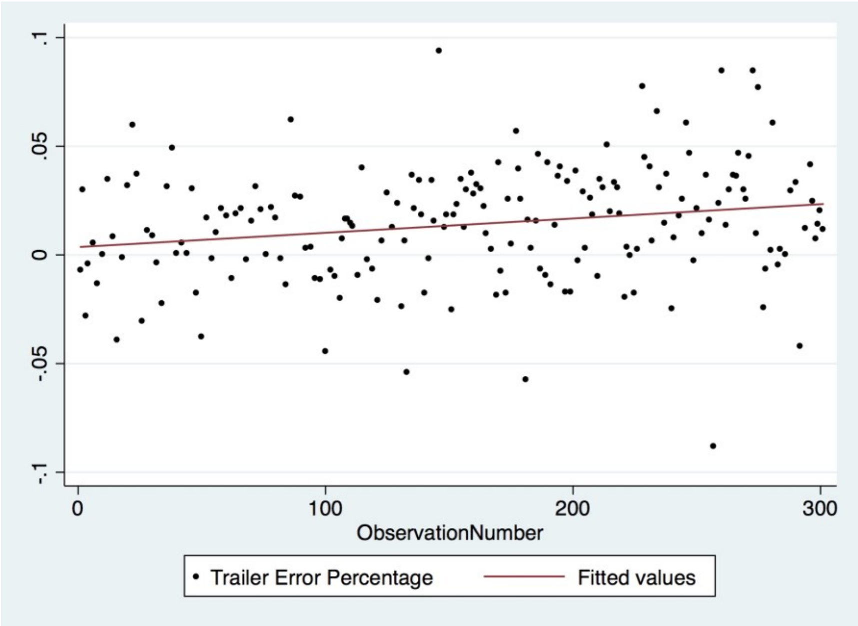

Figure 2.12 Trailer Axle Error Percentage and Observation Number

Figure 2.12 shows the scatter-plot of trailer error percentage and observation number. The fitted line is above zero in this model, which indicates that the sensor seems to overweight trailer weights. The regression result for Model (1.10) is shown in Table 2.19.

$$\text{Trailer Error Percentage} = \text{Cons} + \text{Coef.}\,1 \times \text{Observation Number} \\ + \text{Coef.}\,2 \times \text{Axle Count} + \text{Coef.}\,3 \times \text{Vehicle Number} \\ + \text{Coef.}\,4 \times \text{Speed} + \text{Error term} \tag{1.10}$$

Table 2.19 Trailer Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.008343447	4	.002085862
Residual	.125461301	184	.000681855
Total	.133804748	188	.000711727

Number of obs	=	189
F(4, 184)	=	3.06
Prob > F	=	0.0180
R-squared	=	0.0624
Adj R-squared	=	0.0420
Root MSE	=	.02611

errpercentag~ler	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNum~r	.0000757	.000026	2.91	0.004	.0000244	.0001269
AxleCount	-.0050012	.0032316	-1.55	0.123	-.011377	.0013745
VN	-.0016814	.0016302	-1.03	0.304	-.0048978	.0015349
Speed	-.0006856	.0004155	-1.65	0.101	-.0015054	.0001342
_cons	.0767238	.03997	1.92	0.056	-.0021347	.1555823

Table 2.19 shows that the trailer error percentage becomes statistically larger over time. All other factors are not significant in this model. However, because only 5 and 7 axle vehicles have trailer weights and the axle count discrepancy comes from vehicle number 2, the axle count factor is not significant in this analysis.

2.2.3 Repeated Analyses without Vehicle Number 2

In this section, the analyses are based on the test vehicle data without vehicle number 2, due to the large discrepancy of the actual weights from Data Set 2 and truck tickets. In the interest of brevity, we present only those analyses that are significantly different from our previous results. That is, for the most part, we summarize our results, and present complete details of the analysis in Appendix A.

2.2.3.1 Verification of Intercomp's Analyses

Without the data of vehicle number 2, the results we obtain in this section match those reported in the Intercomp's report. This is also true when we use actual weights from Data Set 2. The observation number is not correlated with the error percentage. The dynamic weights and static weights are very close to each other and all the data points are within the 10% error bands. However, because of the incompleteness of the data, conclusions about environmental conditions are not reliable.

2.2.3.2 Additional Performance Analyses

The distribution of error percentage is shown in Figure 2.13. The solid line is the best-fit line. Without the data of vehicle number 2, the mean of the error percentage for each vehicle is very close to zero, which is different as compared to Figure 2.7.

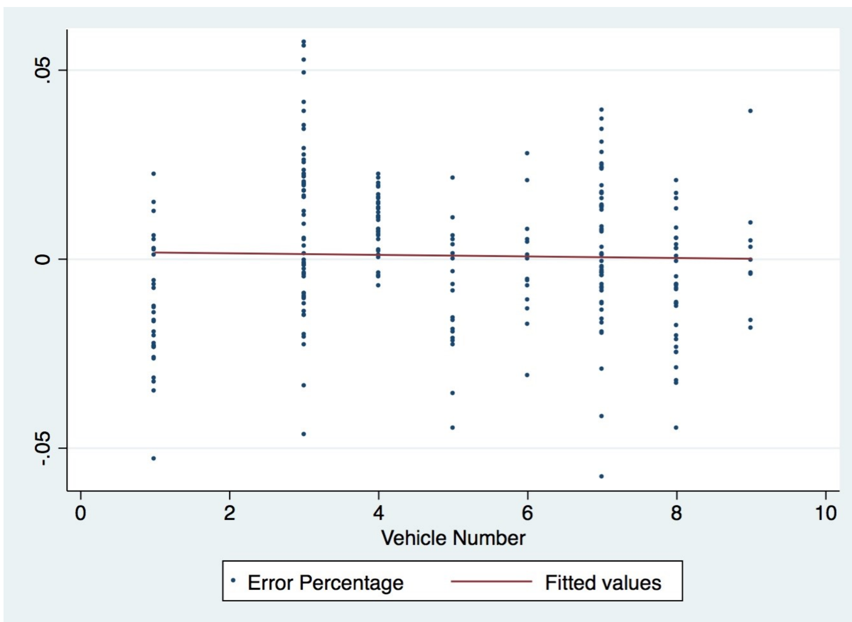

Figure 2.13 Error Percentage by Vehicle Number

Next, observation number, axle count, vehicle number and speed are considered for further analysis. We used the same model:

$$\text{Error Percentage} = \text{Cons} + \text{Coef. 1} \times \text{Observation Number} + \text{Coef. 2} \times \text{Axle Count} \quad (1.11)$$

$$+\text{Coef. 3} \times \text{Vehicle Number} + \text{Coef. 4} \times \text{Speed} + \text{Error term}$$

The regression analysis result is shown in Table 2.20.

Table 2.20 Other Factor Regression Analysis (Error Percentage =Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.011616659	4	.002904165
Residual	.081115144	245	.000331082
Total	.092731803	249	.000372417

Number of obs =	250
F(4, 245) =	8.77
Prob > F =	0.0000
R-squared =	0.1253
Adj R-squared =	0.1110
Root MSE =	.0182

errpercentage_t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNumber	2.31e-06	.0000144	0.16	0.872	-.000026	.0000307
AxleCount	.0054638	.0009313	5.87	0.000	.0036294	.0072982
VN	.0001715	.0004936	0.35	0.729	-.0008008	.0011437
Speed	-.0000179	.0002414	-0.07	0.941	-.0004935	.0004576
_cons	-.0273563	.0144967	-1.89	0.060	-.0559104	.0011978

Compared with Table 2.11, we see that the only factor that affects error percentage is axle count, while in the analysis containing vehicle number 2, observation number, axle count and vehicle number all significantly affect error percentage. That is, the inclusion or exclusion of vehicle number 2 affects accuracy estimates.

For MN classification, the results are the same as previous analysis. The mean of the error percentage is zero for class 9, but not for other classes. Steer, drive and trailer axle have the same trend and statistical significance as in previous analysis.

2.3 CONCLUSIONS

With the limited test data available to perform the analysis, the performance of Intercomp sensor appears to be similar to what is documented by Intercomp in its report. Our findings are somewhat different when we use truck-ticket weights rather than the weights used in the Intercomp report. The number of observations with different average daily temperatures and vehicle speeds are very small. Therefore, any conclusions about the association between these factors and sensor accuracy are not reliable. Measurement errors depend on axle type. For some axles, the sensor overestimates weights, and for some others it underestimates. However, the total estimated weights appear to be within tolerable bounds.

CHAPTER 3: COMPARISON OF THE INTERCOMP AND THE IRD WIM SYSTEMS' PERFORMANCE

3.1 INTRODUCTION

3.1.1 Research Goal

Our research goal in this chapter is to compare and contrast reported performance of the two systems, and lay the foundation for analysis that leads to the selection of a superior system in terms of better accuracy and lower life cycle cost in the future. We analyzed three months of CSV-formatted data provided by OTSM, which was produced by the installed sensors and controllers. We also developed a methodology for comparing and contrasting the two sensor systems' performances, which was then used to compare their performances with limited data.

3.1.2 Data Description

The data used in this chapter are from January 22^{nd}, 2016 to March 28^{st}, 2016 for Intercomp sensor and from October 1^{st}, 2015 to March 28^{st}, 2016 for IRD sensor. All data provided by OTSM were in CSV-format. A summary of the variables in these files is shown in Table 3.1.

Table 3.1 Variable Summary of Original CSV-format Files

Variable Name	Explanation	Variable Name	Explanation
veh	Unique Vehicle Number	aw3	Axle Weight
lane	Lane Number	aw4	Axle Weight
time	Time that passing the sensor	aw5	Axle Weight
axlec	Axle Count	aw6	Axle Weight
speed	Speed that passing the sensor	aw7	Axle Weight
as1feet	Axle Space	aw8	Axle Weight
as2	Axle Space	aw9	Axle Weight
as3	Axle Space	aw10	Axle Weight
as4	Axle Space	aw11	Axle Weight
as5	Axle Space	aw12	Axle Weight
as6	Axle Space	gvw	Gross Vehicle Weight
as7	Axle Space	class	MN Classification
as8	Axle Space	err	Error Code
as9	Axle Space	imagef	Indicator of Image Capture
as10	Axle Space	license	License Number
as11	Axle Space	temp	Temperature

Variable Name	Explanation	Variable Name	Explanation
aw1kips	Axle Weight	calfac	Calibration Indicator
aw2	Axle Weight		

To produce statistical analyses, summary tables, and figures, we combined all data files and analyzed the entire dataset using STATA (version 14.1). A summary of the variables that we created is shown in Table 3.2.

Table 3.2 Summary of Created Variables

Variable Name	Explanation	Variable Name	Explanation
date	Date	IRD_err_20	IRD Error Indicator
month	Month	IRD_err_21	IRD Error Indicator
month_numeric	Numeric Representation of Month	IRD_err_22	IRD Error Indicator
sensor	Numeric Representation of Sensor	IRD_err_23	IRD Error Indicator
sensorName	Name of Sensor	IRD_err_24	IRD Error Indicator
indv1	Individual Error Code	IRD_err_25	IRD Error Indicator
indv2	Individual Error Code	IRD_err_26	IRD Error Indicator
indv3	Individual Error Code	IRD_err_27	IRD Error Indicator
indv4	Individual Error Code	IRD_err_28	IRD Error Indicator
indv5	Individual Error Code	IRD_err_29	IRD Error Indicator
indv6	Individual Error Code	IRD_err_30	IRD Error Indicator
good	Indicator of Error-free Data	IRD_err_31	IRD Error Indicator
intercomp_err_3	Intercomp Error Indicator	IRD_err_32	IRD Error Indicator
intercomp_err_8	Intercomp Error Indicator	IRD_err_33	IRD Error Indicator
intercomp_err_13	Intercomp Error Indicator	IRD_err_34	IRD Error Indicator
intercomp_err_17	Intercomp Error Indicator	IRD_err_35	IRD Error Indicator
intercomp_err_31	Intercomp Error Indicator	IRD_err_36	IRD Error Indicator
intercomp_err_33	Intercomp Error Indicator	IRD_err_37	IRD Error Indicator
IRD_err_1	IRD Error Indicator	IRD_err_38	IRD Error Indicator
IRD_err_2	IRD Error Indicator	IRD_err_39	IRD Error Indicator
IRD_err_3	IRD Error Indicator	IRD_err_40	IRD Error Indicator
IRD_err_4	IRD Error Indicator	IRD_err_41	IRD Error Indicator
IRD_err_5	IRD Error Indicator	IRD_err_42	IRD Error Indicator
IRD_err_6	IRD Error Indicator	IRD_err_43	IRD Error Indicator

Variable Name	Explanation	Variable Name	Explanation
IRD_err_7	IRD Error Indicator	IRD_err_44	IRD Error Indicator
IRD_err_8	IRD Error Indicator	IRD_err_45	IRD Error Indicator
IRD_err_9	IRD Error Indicator	IRD_err_46	IRD Error Indicator
IRD_err_10	IRD Error Indicator	IRD_err_47	IRD Error Indicator
IRD_err_11	IRD Error Indicator	IRD_err_48	IRD Error Indicator
IRD_err_12	IRD Error Indicator	IRD_err_49	IRD Error Indicator
IRD_err_13	IRD Error Indicator	IRD_err_50	IRD Error Indicator
IRD_err_14	IRD Error Indicator	IRD_err_51	IRD Error Indicator
IRD_err_15	IRD Error Indicator	IRD_err_52	IRD Error Indicator
IRD_err_16	IRD Error Indicator	IRD_err_53	IRD Error Indicator
IRD_err_17	IRD Error Indicator	IRD_err_54	IRD Error Indicator
IRD_err_18	IRD Error Indicator	IRD_err_57	IRD Error Indicator
IRD_err_19	IRD Error Indicator	IRD_err_65	IRD Error Indicator

Statistical summaries of the gross vehicle weight and speed are shown in Table 3.3 and Table 3.4, respectively.

Table 3.3 GVW Statistical Summary from All Data

sensorName	Freq.	mean(gvw)	sd(gvw)	min(gvw)	max(gvw)
IRD	4934264	9.9	15.88637	-131.1	236.6
Intercomp	1677457	128.4	5145.146	0.0	3,258,588.8

The summaries are for all recorded observations including all classes and errors. Note that Intercomp does not have any negative gross vehicle weight but has extremely large maximum weight. The variance of gross vehicle weight for Intercomp is also much larger than that of IRD. We conjecture that these differences might have been caused by the differences in the algorithms they used to convert the raw sensor data into weights.

Table 3.4 Speed Statistical Summary from All Data

sensorName	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	4934264	70.95401	6.265212	3	124
Intercomp	1677457	89.39383	2134.049	.5	2272727

For speed as well, Intercomp has a very large maximum value, which does not appear to be practically possible. The variance of speed for Intercomp is also much larger than that for IRD.

Because Class 9 is the most important class for MNDOT, we separately calculated statistical summaries of vehicle weight and speed for Class 9 only, which are shown in Table 3.5 and Table 3.6, respectively.

Table 3.5 GVW Statistical Summary for Class 9

sensorName	Freq.	mean(gvw)	sd(gvw)	min(gvw)	max(gvw)
IRD	464,183	51.04181	16.95232	5.96	177.35
Intercomp	75,237	56.6875	20.17785	5.908	163.31

In general, the Class-9 weight measurements from Intercomp and IRD sensors are more similar as compared to weight measurements for all classes. Intercomp has fewer observations and larger mean and variance compared to IRD. The differences in frequency of observations are explained to a certain extent by the fact that Intercomp data pertains to a 3-month period, as opposed to a 6-month period for IRD data.

Table 3.6 Speed Statistical Summary for Class 9

sensorName	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	464,183	66.28343	5.383505	9	124
Intercomp	75,237	65.3552	8.42032	4.6	111.4

In terms of speed, Intercomp has a slightly larger mean and a much larger standard deviation.

The two systems have different (and individual) error reporting systems, which indicate either no error, or error (not valid), or warning (valid but falls out of manually specified range). Data with warnings are generally treated as “good data”. As of the completion of this analysis, the Intercomp team did not provide a detailed explanation of their error codes. Therefore, all error-coded data are treated as not valid in this chapter for Intercomp. Summaries of the errors generated by the two systems are shown in

Table 3.7 and Table 3.8. Note that because a record may contain more than one error, the sum of the error percentages is more than 100%.

Observe that 96.21% of the IRD-system data are error free. Among the error codes, "significant weight differences" is the most common error type, which accounts for 0.93% of all the records.

Table 3.7 Intercomp Error Code Summary

Error Code	Description	Percentage
0	No error	16.87%
3	No loop/Loop violation	13.42%
8	Axle count mismatch	8.75%
13	Under speed	3.04%
17	Over speed	76.37%
31	Weight balance	45.21%
33	Acceleration/Deceleration violation	0.01%

Table 3.8 IRD Error Code Summary

Error Code	Description	Percentage
0	Normal, no error	96.21%
1	Upstream loop failure (Downstream loop only)	0.70%
2	Downstream loop failure (Upstream loop only)	0.13%
7	Zero axles detected (failure of both axles)	0.01%
8	Unequal axle counts (difference of up and down-stream axle counts)	0.01%
13	Vehicle too slow (indicated by loop activation)	0.01%
17	Vehicle too fast	0.02%
19	One axle detected	0.01%
34	Significant speed change	0.13%
35	Significant weight differences	0.93%
37	Unequal axle counts on sensors	0.53%
38	Tailgating	0.29%
43	Overweight (good data)	0.54%

Error Code	Description	Percentage
44	Over GVW (good data)	0.41%
46	Drastic speed change	0.86%

All other error codes have a frequency of 0% in IRD data. A detailed error codebook can be found in the Bullconverter manual (Kwon, 2015) provided by Professor Taek Kwon.

Because Intercomp only contains 16.87% of valid data, we compare the system both with and without the errors. Further explanation of the error codes are needed from Intercomp, otherwise the fact that only 16.87% of the observations are considered valid is in and of itself an indication of poor performance of the Intercomp sensor. In the remainder of this chapter, we refer to IRD data as valid data when it has no errors. However, for Intercomp, this terminology has a different meaning. When we say that Inercomp data is without error, we mean that the data that has "0" in the error code field. This data is potentially a subset of without-error data.

Statistical summaries of gross vehicle weight and speed without errors are shown in Table 3.9 and Table 3.10.

Table 3.9 GVW Summary for Class 9 without Error

sensorName	Freq.	mean(gvw)	sd(gvw)	min(gvw)	max(gvw)
IRD	411,192	50.0	15.37133	7.8	80.0
Intercomp	47,596	57.8	17.07525	19.9	110.2

Table 3.10 Speed Summary for Class 9 without Error

sensorName	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	411,192	65.94135	4.458091	9	76
Intercomp	47,596	64.21599	4.263861	32.2	81.2

The Intercomp sensor has fewer observations by a factor of approximately 8.6. The mean and standard deviation are close for speed, but not for the gross vehicle weights. The Intercomp sensor measures higher weights with greater variance. Equally importantly, the minimum and maximum of weights and speeds are different for the two sensors, with Intercomp sensor reporting a greater minimum and maximum among without-error data.

3.1.3 Overall Approach

The overall approach of our analyses consists of two parts. First, self-consistency analysis of each sensor will be performed based on daily observation count, gross vehicle weight, and speed. Second, comparisons of the two sensors are given in terms of daily observation count, gross vehicle weight and speed. For each criterion, statistical mean, variance and distribution analysis will be analyzed.

3.1.4 Summary

In the self-consistency analyses we found out that Intercomp's system performance is not consistent over time. There are significant differences in daily observation count, gross vehicle weight and speed before and after February 19th, 2016 (when the Intercomp sensors failed). IRD's system appears self-consistent over the data time frame. The differences in the two systems' observations are statistically significant for all criteria – daily observation count, gross vehicle weight and speed. Therefore, at least one, or both sensors are reporting incorrect measurements. Without test vehicle data, it is not possible to ascertain which sensor is more accurate.

The remainder of this chapter is organized as follows. Section 3.2 gives self-consistency analysis of Intercomp sensor and Section 3.3 gives self-consistency analysis of IRD sensor. Section 3.4 compares and contrasts the two systems. In Section 3.5 we summarize our findings and recommendations.

3.2 SELF-CONSISTENCY ANALYSIS: INTERCOMP SENSOR

An important criterion of sensor performance is self-consistency. A sensor's measurements may change gradually over time, or exhibit a sudden and sharp turning point. Whereas the former may be caused by a variety of reasons, it is typically the case the latter occurs when the sensor fails and requires recalibration. In this chapter, we consider three types of consistencies – daily observation count, gross vehicle weight and speed. We focus exclusively on Class-9 vehicles, which are of greatest interest to MNDOT. Valid data of Intercomp used in this chapter are from January 22nd to March 28th of 2016.

All of the factors we analyzed, including daily observation count, gross vehicle weight and speed, indicate that Intercomp's measurements underwent a sudden and significant change at about the February 19th mark.

3.2.1 Daily Observation Count

Daily observation counts are the frequencies of the data records from each day. The counts equal to the volumes of daily traffic captured by the sensor. Average daily traffic counts are important for MNDOT because these counts are reported to the FHWA. We do not assume the same number of vehicle counts for each day, because counts vary by the day of week, season and holidays. But we do believe that a sudden and significant change in the pattern of average daily traffic count is a signal of sensor system failure and possible need for recalibration. We consider the effect of errors and lanes (driving or passing) from which data originated and perform separate analysis for Class-9 vehicles.

Figure 3.1 shows the scatter-plot of the daily observation count of Intercomp for all data and the data without errors (with an error code of "0"). The daily observation count of all data started at about 20,000 and increased gradually over time, reached about 30,000 counts at the end of March. For error-free data, daily counts started with an average of 6,000 counts per day, but decreased suddenly around mid-February to about 3,000 average daily count. This indicates that although the general traffic increased, the systems generated a lot more errors from mid-February, which made the error-free data significantly fewer.

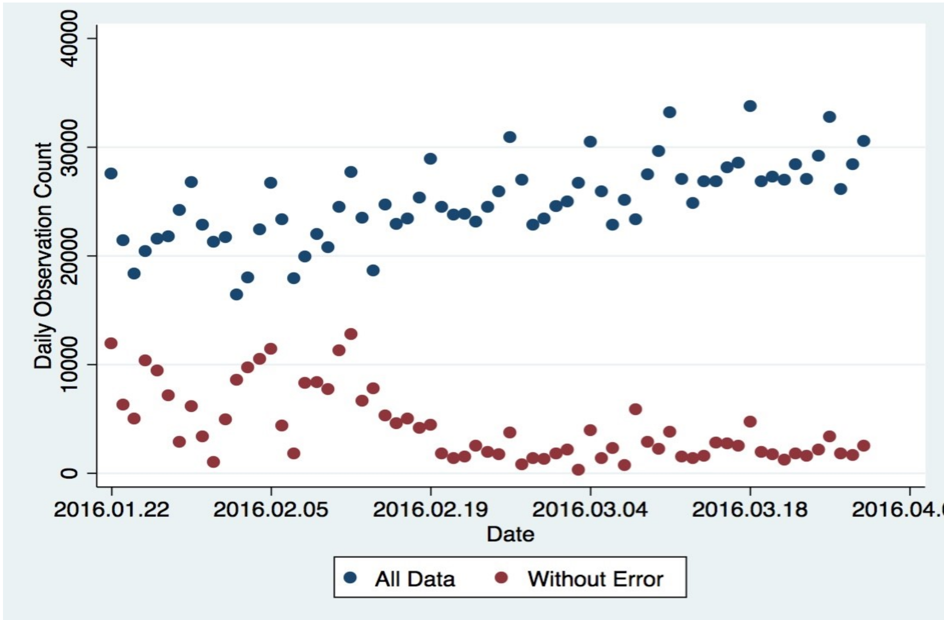

Figure 3.1 Scatter-plot of Daily Observation Count - Intercomp

Figure 3.2 shows scatter plots of daily observation counts for Class-9 vehicles only. The blue dots are Class-9 vehicles with errors; the red dots are Class-9 vehicles without errors. It is clear that there are two different patterns in both with and without-error data – one before and the other after February 19th. Before February 19th, the counts are consistently distributed with a very large range (from about 100 to 3,000). But after February 19th, the counts for both with and without errors went down significantly (around 600 for data with error and 200 for data without error).

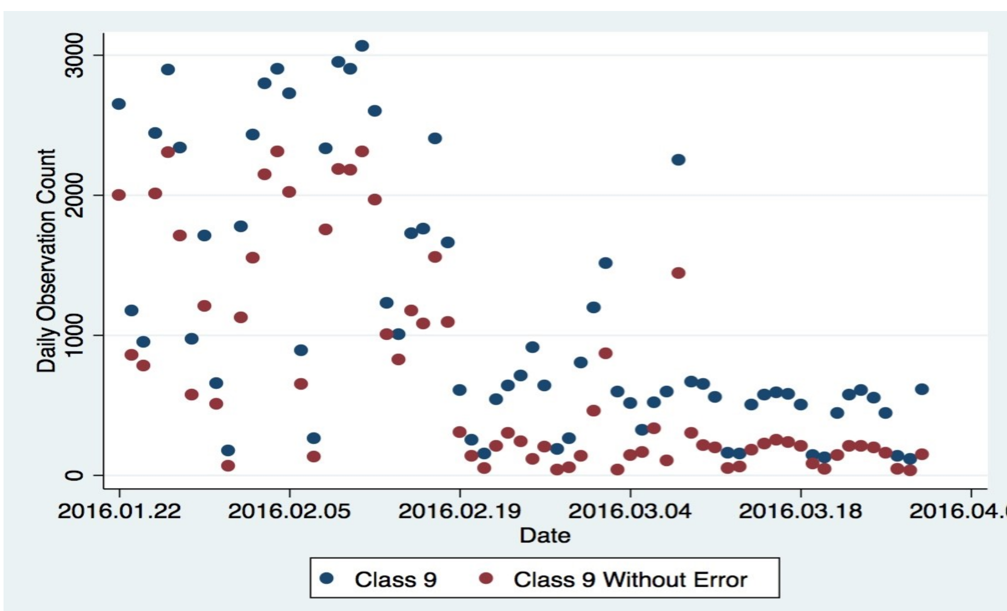

Figure 3.2 Scatter-plot of Daily Observation Count - Intercomp Class 9

Figure 3.3 shows the daily observation counts by lane. Lane 1 is the driving lane and Lane 2 is the passing lane. Lane 2 has a very consistent weekly pattern over time. Lane 1 data exhibits a significant and sharp change after February 19th. Before that, the counts range from 500 to 2,500 with a consistent day-of-week effect. After February 19th, except for three outliers, all the observation counts are below 500, which is very different compared to the counts before February 19th. This means that the changes in daily observation count are primarily from Lane 1.

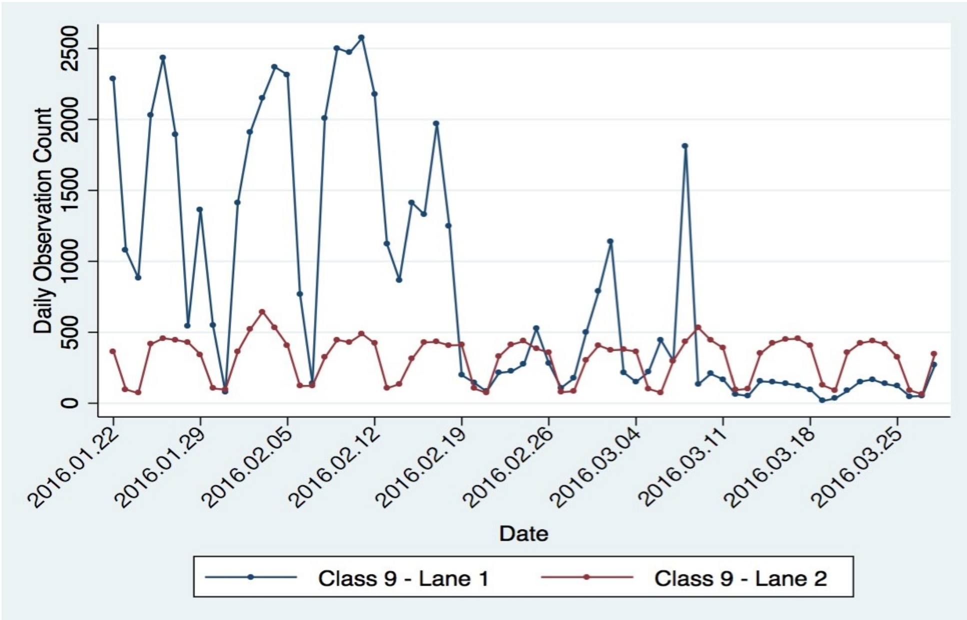

Figure 3.3 Daily Observation Count by Lane - Intercomp Class 9

The change in daily observation counts before and after Feb 19th is significant and the change is mainly from Lane 1. Further investigation is needed to identify the reason for the significant count change. This significant count drop might be an indicator that recalibration is required.

3.2.2 Gross Vehicle Weight

Because different vehicle classes have different gross vehicle weight distributions and our primary focus is on Class-9 vehicles, all the analyses of gross vehicle weights and speed are for Class-9 vehicle data only. Analyses of other class can be performed in a similar fashion.

We evaluate gross vehicle weight consistency in three aspects – daily mean of GVW, daily standard deviation of GVW and distributions of GVW. After accounting for day-of-week and seasonal variations, and average traffic counts, consistency of these three measurements should indicate good performance.

3.2.2.1 Mean of Gross Vehicle Weight

Figure 3.4 shows daily average gross vehicle weight for Class 9 from January 22nd to March 28th. A clear weekly pattern is observed for both with- and without-error data. However, there is significant jump in the mean of gross vehicle weights in early March for Class-9 data with errors. The mean of GVW of Class 9 without errors is consistent over time. This indicates that Intercomp sensor generated significantly larger gross-vehicle-weight observations, but those data are labeled as errors in their system, leaving the remaining data reasonably consistent over time.

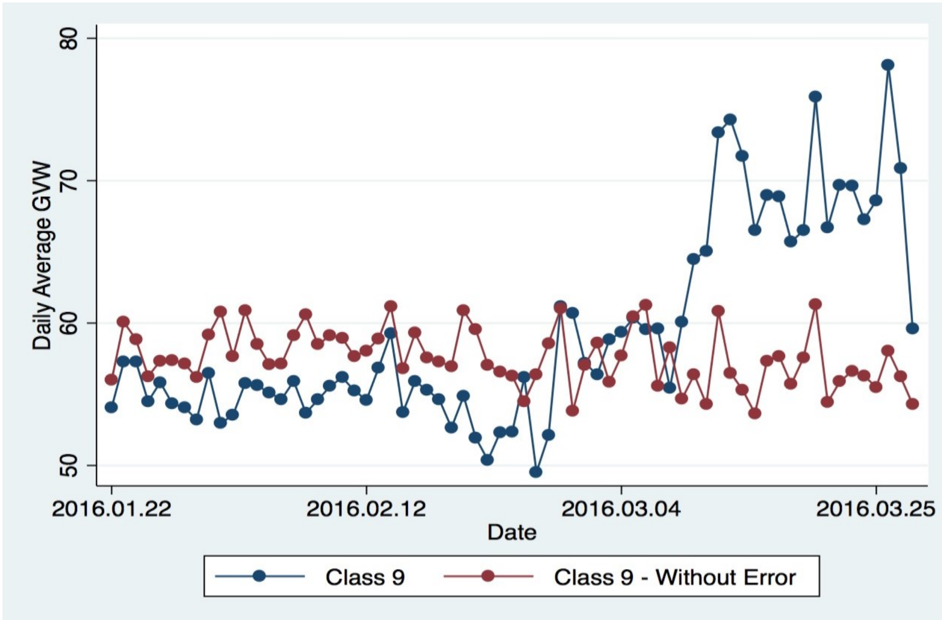

Figure 3.4 Mean of GVW - Intercomp Class 9

Figure 3.5 shows the average gross vehicle weight by separate lanes. Similar to what we observed for daily traffic count, the mean of GVW is quite consistent before mid February for both lanes. After around Feb 19th, the average GVW of Lane 1 began to fluctuate and increased significantly in early March, from about 60,000 lbs.to over 100,000 lbs. The jump in the value of the average GVW indicates that something may be wrong with the Lane-1 sensor of Intercomp. However, we are not able to offer the reason behind this suddenly increase.

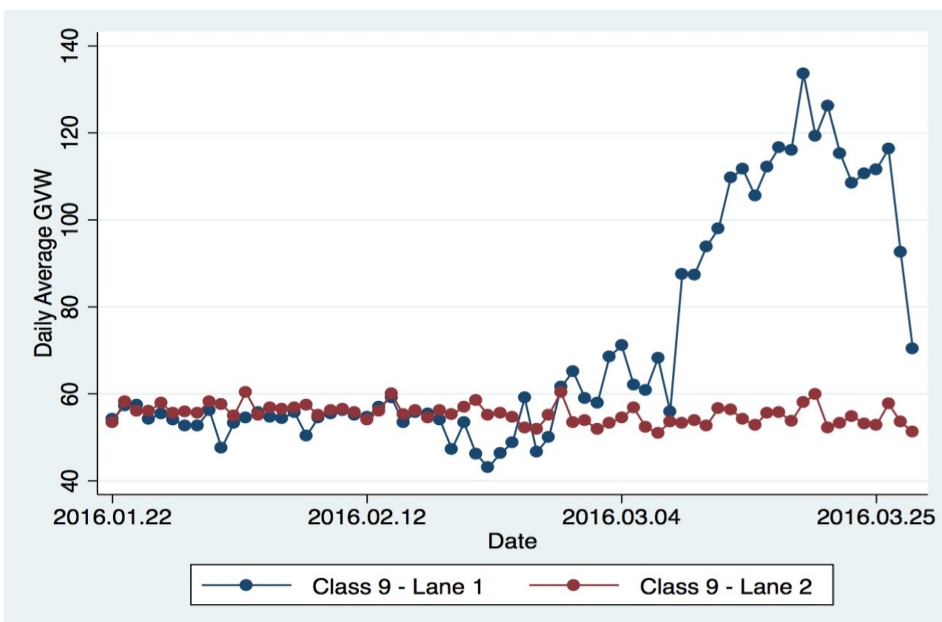

Figure 3.5 Mean of GVW - Intercomp Class 9 by Lane

In conclusion, there is significant change in the Class-9 mean GVW from Intercomp sensor beginning around Feb 19th. The change comes primarily from Lane 1. Interestingly, The Intercomp controller identifies these large gross-vehicle-weight observations as errors.

3.2.2.2 Variance of Gross Vehicle Weight

In addition to the average gross vehicle weight, variance of GVW is another important determinant of consistency. Figure 3.6 shows the daily standard deviation of GVW of Class-9 vehicles. Similar to mean of GVW, there is a weekly pattern for both with and without error data. There is a significant jump from 18 to over 30 of the standard deviation of GVW for Class-9 data (including errors) in early March. The standard deviation for error free Class-9 data remains reasonably consistent.

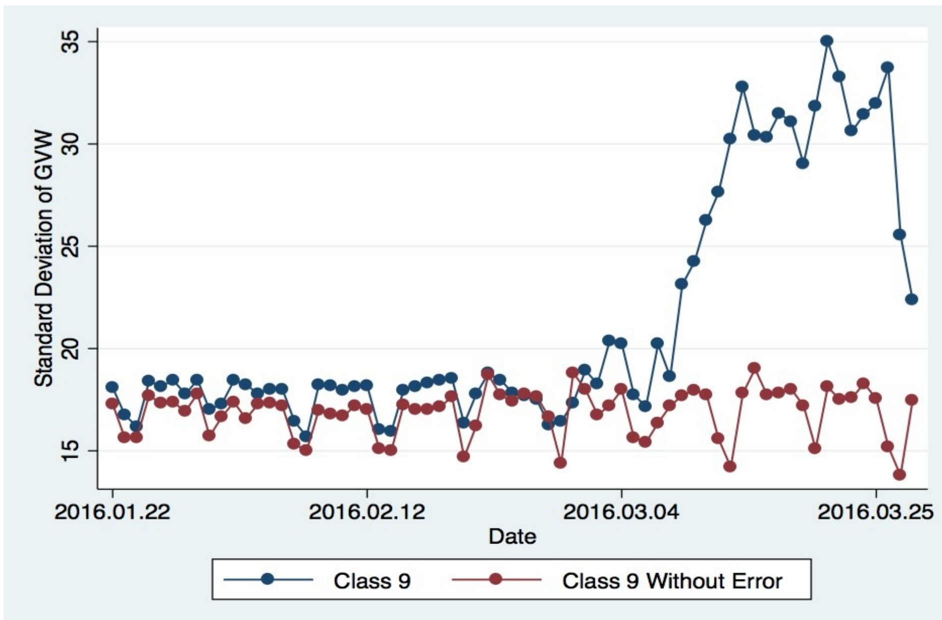

Figure 3.6 Standard Deviation of GVW - Intercomp Class 9

Figure 3.7 gives us the standard deviation of GVW for Intercomp of Class-9 vehicles by lane. Lane 2 has a consistent weekly pattern over time. Lane 1 has a slightly higher standard deviation after mid March and the fluctuation is large compared with the data before. This is another indication that the significant change in the GVW data of Imtercomp is mainly from Lane 1.

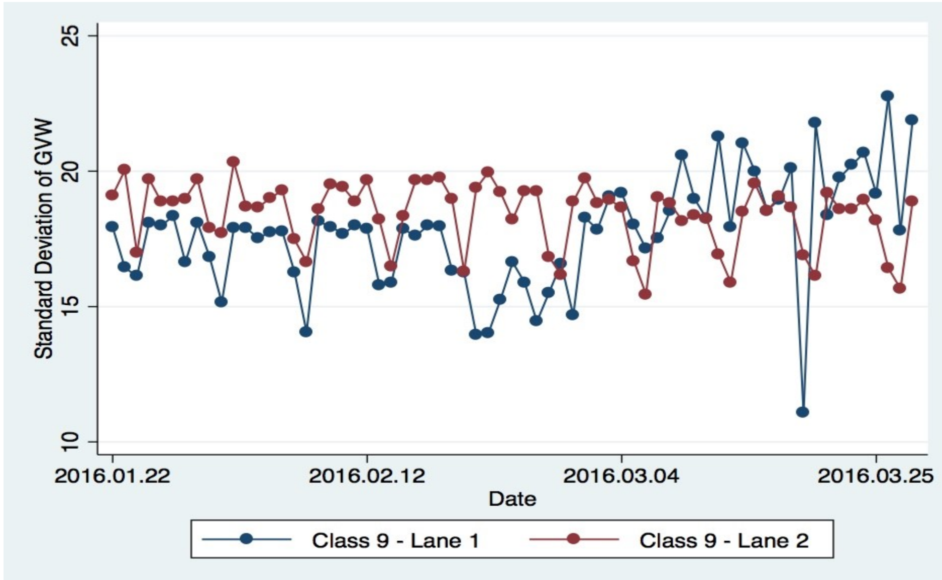

Figure 3.7 Standard Deviation of GVW by Lane - Intercomp Class 9

3.2.2.3 Distribution of Gross Vehicle Weight

Distribution of GVW is another important criterion we consider for consistency. Daily or monthly consistent distributions are indicators of sensor calibration remaining intact because we do not expect the distributions of gross vehicle weight to change significantly across consecutive days or months.

Figure 3.8 shows the box-plot of GVW of Class 9 vehicles of Intercomp. Middle lines in the boxes are medians of GVW for each day. Bottom and top of the boxes represent the first and third quartiles of GVW for each day. The extended solid lines are within 1.5 IQR (third quartile – first quartile). Individual dots beyond the solid lines are considered outliers in the data. Daily box plots are consistent before early March, which includes similar median and the range of the box-plot. There are no outliers before March. However, the system generated larger GVW data in March. Box-plots have larger ranges and larger medians. Many data points are identified as outliers. The distribution of GVW of Class 9 has changed in March, with larger values and more outliers among observations.

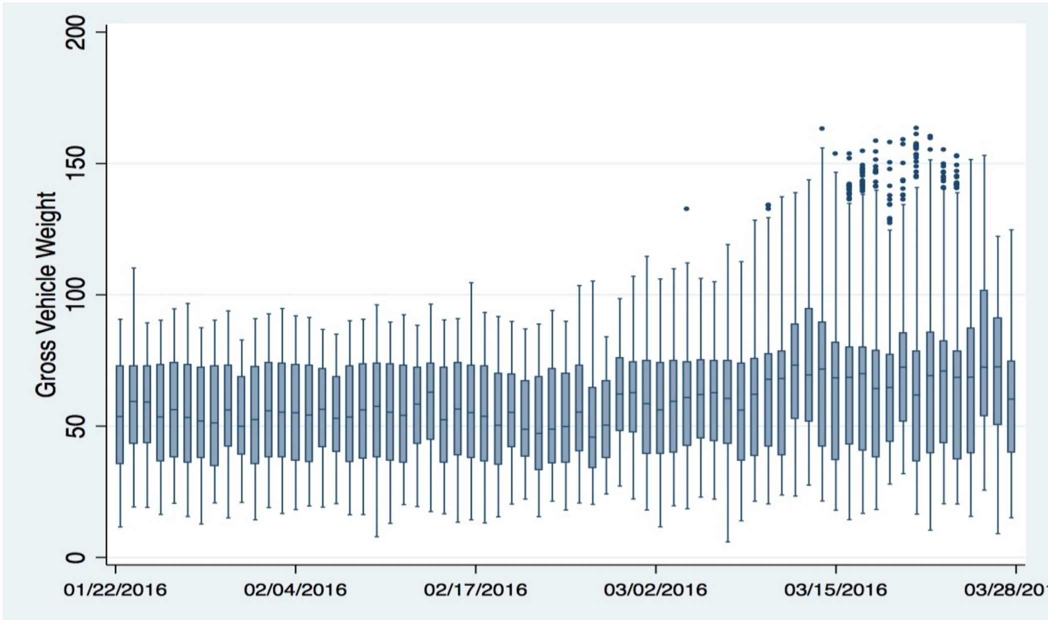

Figure 3.8 Daily Box-plot of GVW - Intercomp Class 9 With Error

Similarly, Figure 3.9 shows the box-plot of GVW for Class 9 without errors. Similar to the mean and variance, the distributions without errors are consistent with time, which means that the unusual GVW data was labeled as with-error data by the system.

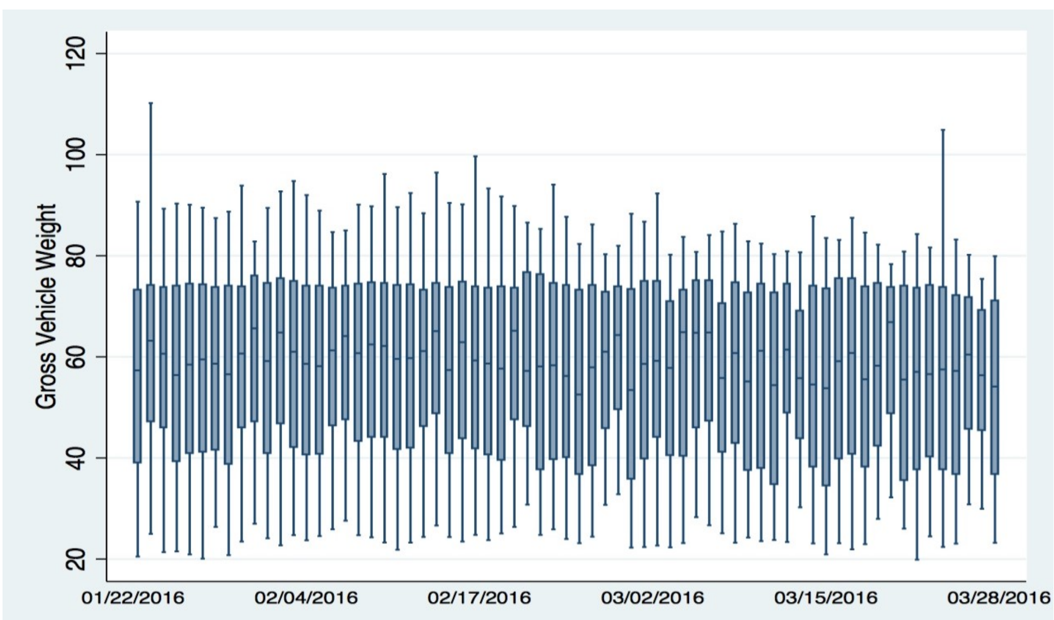

Figure 3.9 Daily Box-plot of GVW - Intercomp Class 9 without Error

In addition to daily distributions, we also considered monthly distributions. Figure 3.10 shows the monthly comparison of GVW distributions for Class 9. Distributions of January and February are quite consistent, with two peaks around 30,000 lbs. and 75,000 lbs. But the distribution of March is different from the other two. It has more vehicles with GVW in excess of 80,000 lbs.

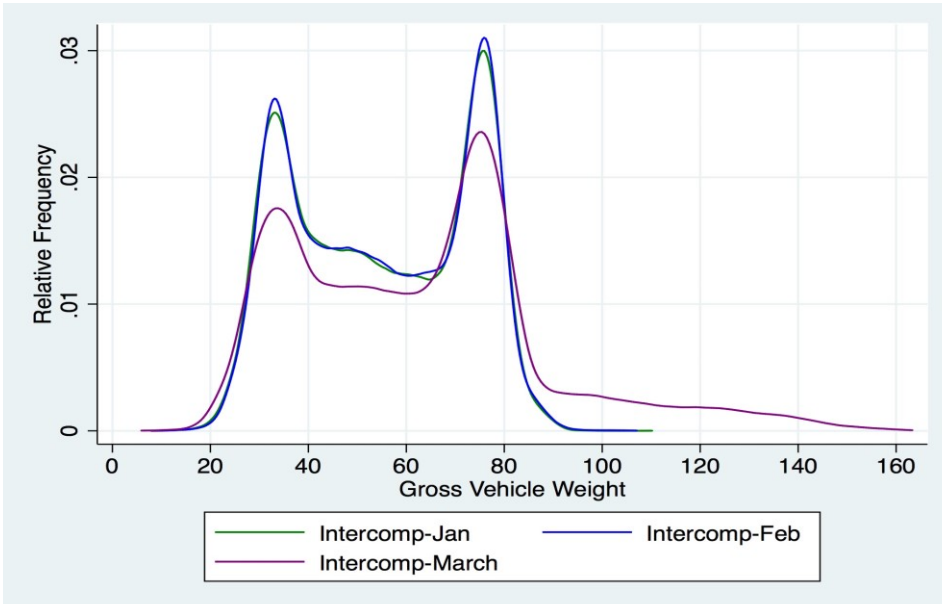

Figure 3.10 Month-to-month Comparison of Class 9 GVW - Intercomp

To test the distributional differences statistically, we use Kolmogorov-Smirnov tests for monthly data. Kolmogorov-Smirnov test is a nonparametric test of the equality of continuous distributions. It quantifies a distance between the empirical distribution functions of two data samples. The null distribution of the statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. The quantified distance is called *deviance.* It is a numeric value that represents the difference in distributions of the two samples. We also recommend that deviance be used to as an alternative criterion to modify sensor systems and determine recalibration time. Deviance is more appropriate than the mean for capturing the distributional change in GVW measurements.

Table 3.11 shows the Kolmogorov-Smirnov tests of Intercomp Class 9 for different months. Only consecutive months are compared to check whether GVW distributions change significantly from one month to the next. The distributions of January and February are statistically the same with a p-value of 0.39. But the distributions of February and March are significantly different with a p-value of 0. This indicates that Intercomp-reported GVW distribution changed significantly in March.

Table 3.11 Monthly Two-sample Kolmogorov-Smirnov Test - Intercomp

Month	Deviance	p-value
January versus February	0.0084	0.39
February versus March	0.1224	0.00

Similar to the mean and variance, Figure 3.11 shows the monthly distribution of GVW for separate lanes. It is clear that GVW distributions of Lane 2 are quite consistent for the three months. But the GVW distribution of March is significantly different from the other two months for Lane 1. This again indicates that the significant difference of GVW in March is mainly from Lane 1.

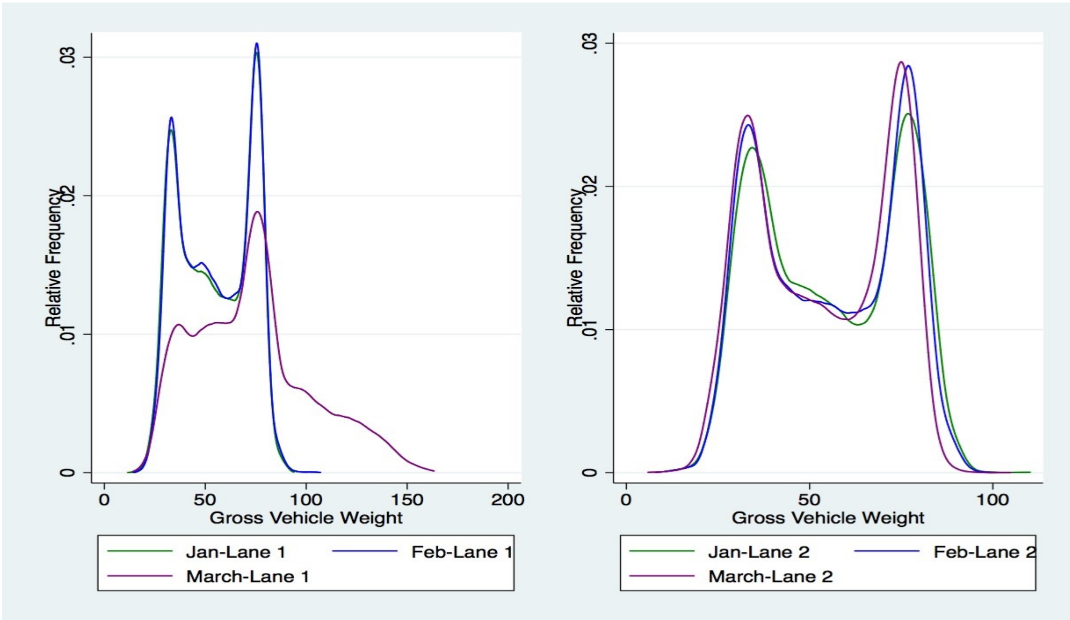

Figure 3.11 Comparison of Intercomp Monthly Distribution by Lane – Class 9

3.2.3 Speed

Speed is another important performance metric both for traffic monitoring purposes and because the speed affects the accuracy of vehicle-class detection, axle spacing detection and error code generation. Because there is no speed limit change from January to March, our null hypothesis is that Class-9 speed distribution remains the same during this period of time. What we mean by that statement is that although individual vehicle speeds vary significantly, the overall speed pattern (across all vehicles) remains the same over time. We also acknowledge a limitation of our analysis of speed data. Because we did not have data on weather events such as snowstorms, it is possible that the distribution of speed changes in response to weather conditions rather than as a result of measurement inconsistency.

Similar to gross vehicle weight, time consistency of mean, variance and distribution of speed are considered in the sequel.

3.2.3.1 Mean of Speed

Figure 3.12 shows the daily average speed of Class-9 vehicles with and without errors. Average speed of Class-9 data with errors has two different patterns before and after early March. Before early March, the average speeds for most of days are around 65 mph, with some lower-speed days around 55 mph. Beginning in March, the average speed jumps significantly to around 70 mph. Average speeds of Class-9 without error has a slightly increasing trend over time but the change is both small and gradual, hence it does not suggest consistency. One may conclude from the speed data that the Intercomp system changes significantly in March and that the unusual data are tagged by its controller as with-error data.

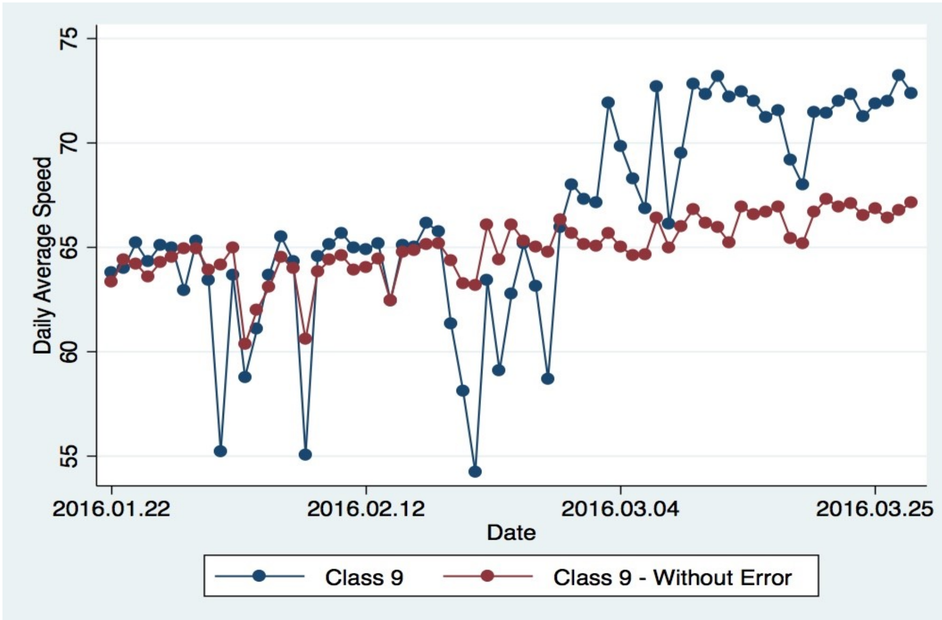

Figure 3.12 Mean of Speed - Intercomp Class 9

Figure 3.13 shows the daily average speed of Class-9 vehicles by lane. There is a clear change in Lane-1 data in March (from 65 mph to 80 mph), whereas the average speed in Lane 2 remains consistent over time (around 70 mph). Because in general the average speed of driving lane is lower than that of the passing line, it is likely that something went wrong with the sensor in Lane 1 in March.

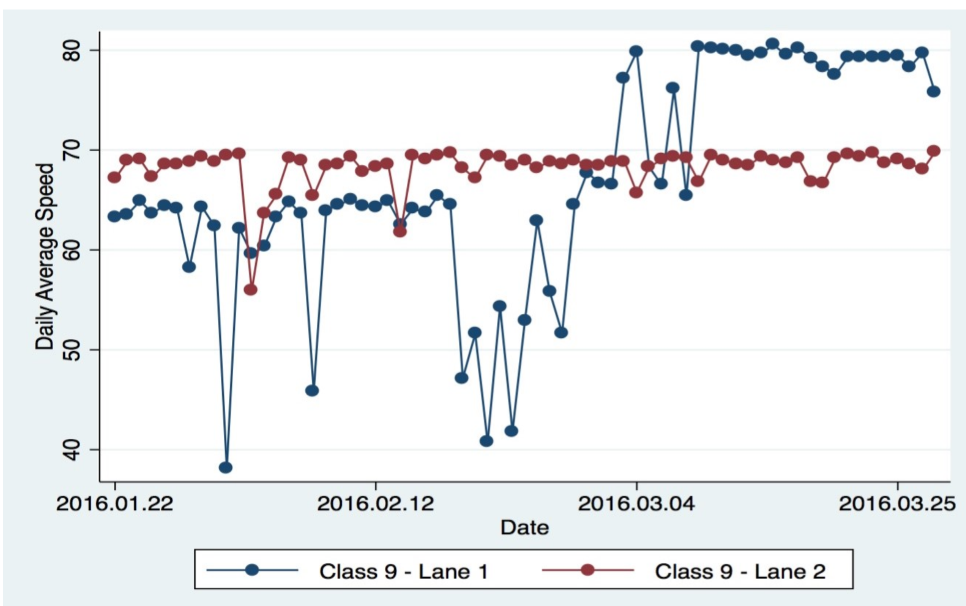

Figure 3.13 Mean of Speed by Lane - Intercomp Class 9

3.2.3.2 Variance of Speed

While the variance of speed is another important determinant of system consistency, it is more likely to be impacted by the traffic itself rather than the accuracy of the sensor. For example, weekdays may have higher speed variance than weekends because weekdays are likely to have traffic jams during peak hours. Therefore, the interpretation of the comparison of variances must be done more carefully.

Figure 3.14 shows daily standard deviations of speed for Class-9 vehicles with or without errors. Speed standard deviation has a large range for Class-9-with-error data from 5 mph to 15 mph, but becomes more consistent in March. Recall that speeds are on average recorded as being higher in March. The standard deviation of speed of Class-9-without-error data is considered consistent over time with a slightly decreasing trend. But this may be because the amount of valid data becomes smaller over time. It is possible that Intercomp sensor fails to capture the variability of the speed in March because the system incorrectly generates much larger but less variable speed data.

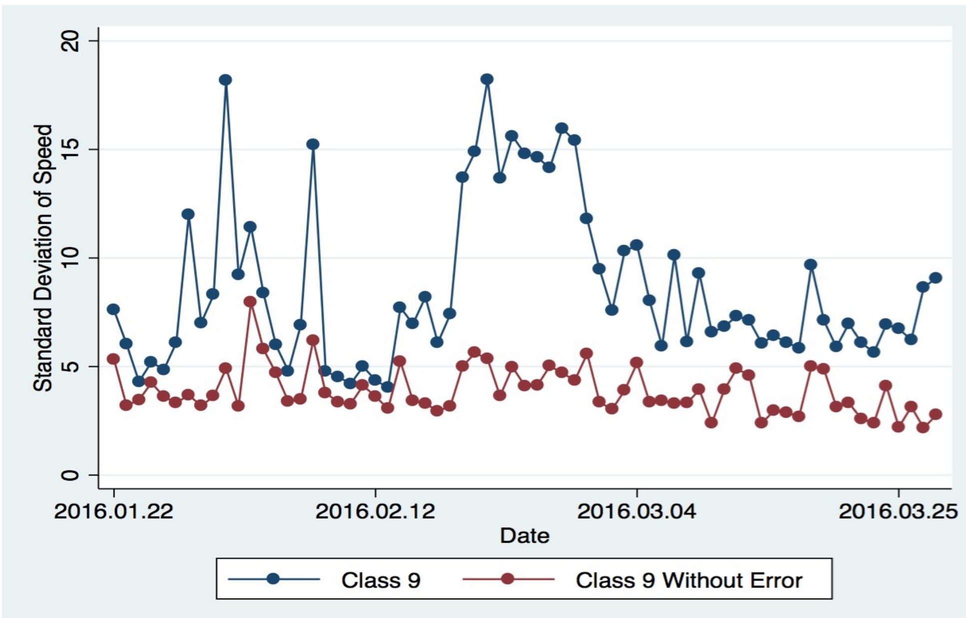

Figure 3.14 Standard Deviation of Speed - Intercomp Class 9

Figure 3.15 shows speed standard deviation of Class 9 by lane. Statistics for both lanes fluctuate over time but Lane 2 has a more consistent pattern than Lane 1. Overall, it is difficult to draw conclusions about consistency, or lack thereof, from the variance of speed data.

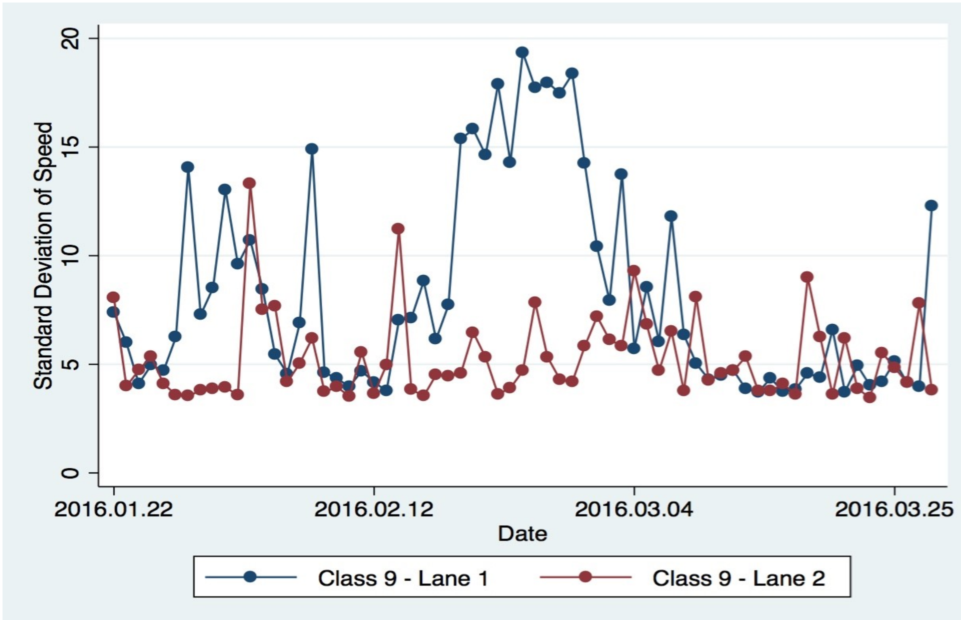

Figure 3.15 Standard Deviation of Speed by Lane - Intercomp Class 9

3.2.3.3 Distribution of Speed

Similar to gross vehicle weight, we compare distributions of speed for different months. Figure 3.16 shows the monthly distributional comparison of speed. It is clear that January and February seem consistent. But the speed distribution of March is significantly different from January and February.

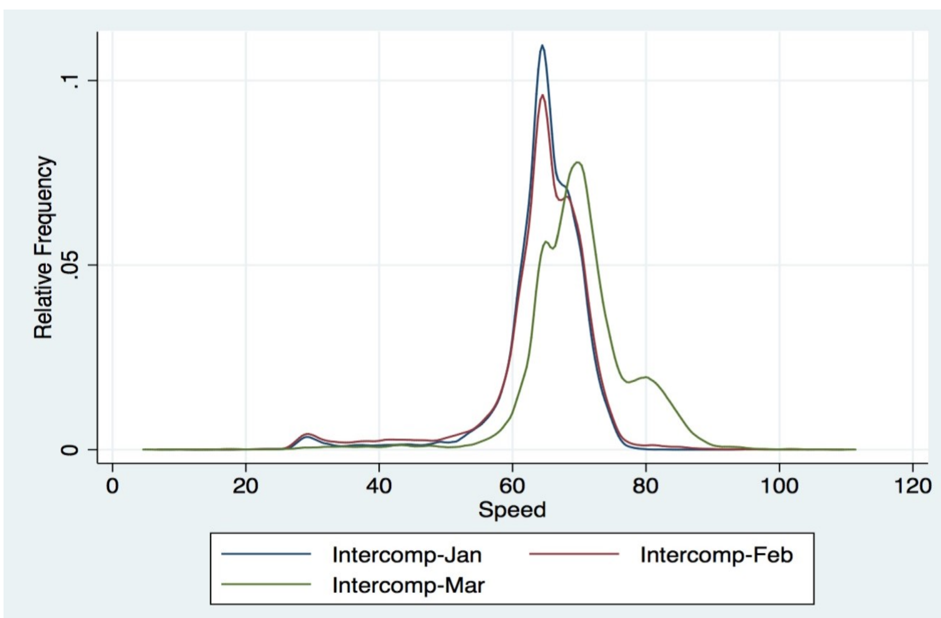

Figure 3.16 Monthly Distribution of Speed – Intercomp Class 9

To test the distribution difference statistically, we again use the Kolmogorov-Smirnov test; results are shown in Table 3.12. Although the *p*-values of both tests are 0, the deviance of the difference of January and February is much smaller than the deviance of February and March (0.032 versus 0.326). Therefore, the distribution shift of speed from January to February is much smaller than the shift from February to March.

Table 3.12 Monthly Two-sample Kolmogorov-Smirnov Test - Intercomp

Month	Deviance	p-value
January versus February	0.0320	0.00
February versus March	0.3269	0.00

Figure 3.17 shows the distribution differences of speed by lane. It is clear that the distributional difference of March comes mainly from Lane 1.

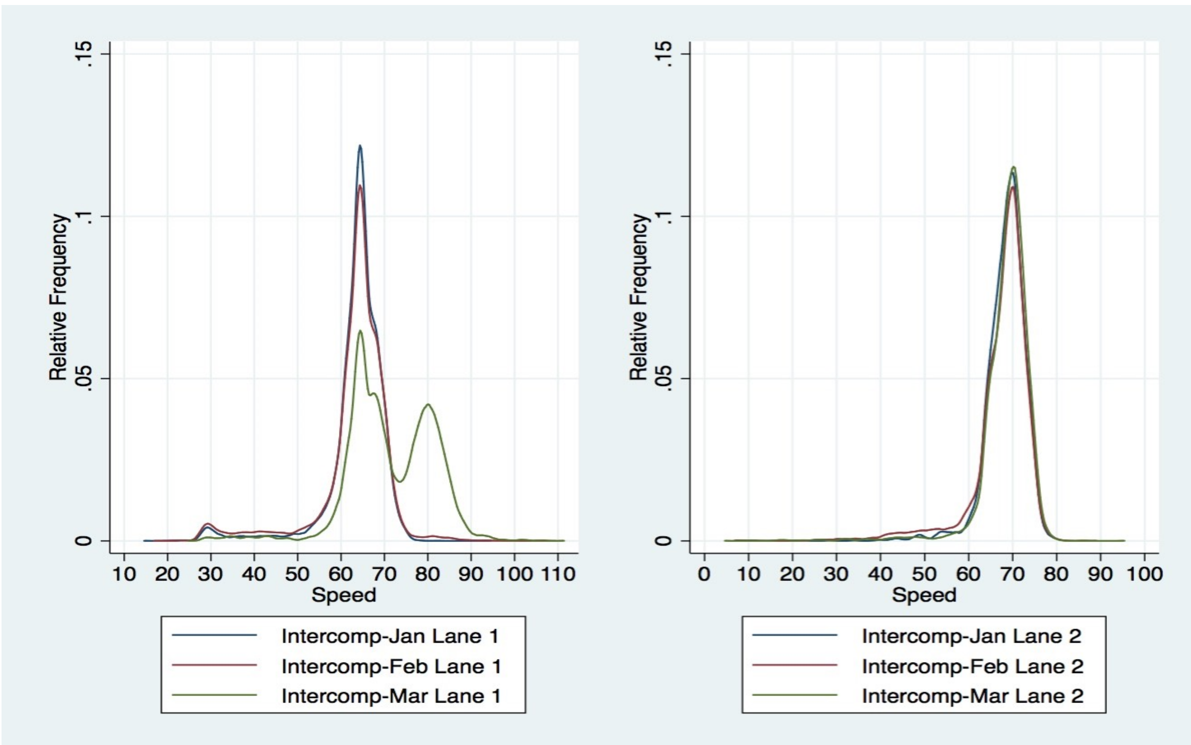

Figure 3.17 Monthly Distribution of Speed by Lane - Intercomp Class 9

To summarize, in section 3.2 we compared the self-consistency of Intercomp in daily observation count, gross vehicle weight and speed. The analyses of all three factors give similar results – that the Intercomp sensor became inconsistent sometime after February 19th and much worse in March. The inconsistency is mainly in observations from Lane 1 (driving lane) and many such observations were tagged as data with error by the Intercomp controller. Overall, the Intercomp system generated too many observations with errors, resulting in only 17% of data as valid data.

3.3 SELF-CONSISTENCY ANALYSIS: IRD SENSOR

In this section, we repeat the Section 3.2 analyses for the IRD sensor. That is, the mean, variance and distribution of daily observation counts, gross vehicle weights and speed are analyzed. Valid IRD data are from October 1st to March 28th. Note that the two sensors have different sizes of data. This is not a problem so long as we compare each sensor with itself, which is what we do in Section 3.2 and 3.3 .

Our findings in terms of daily observation count, gross vehicle weight and speed indicate that the IRD-sensor data is self-consistent over 6 months.

3.3.1 Daily Observation Count

We first look at the daily observation count of all classes, which is shown in Figure 3.18. Daily observation counts range from 15,000 to 40,000 with a slight decrease from October to the end of December and a slight increase from January. This appears to be a reasonable reflection of real traffic because of the weather condition in Minnesota. There are no sudden changes and no reason to suspect an inconsistency.

The observation count without error has a pattern that is similar to that calculated from the data with error, which suggests that the errors as percent of total vehicle counts remain stable over the 6 months.

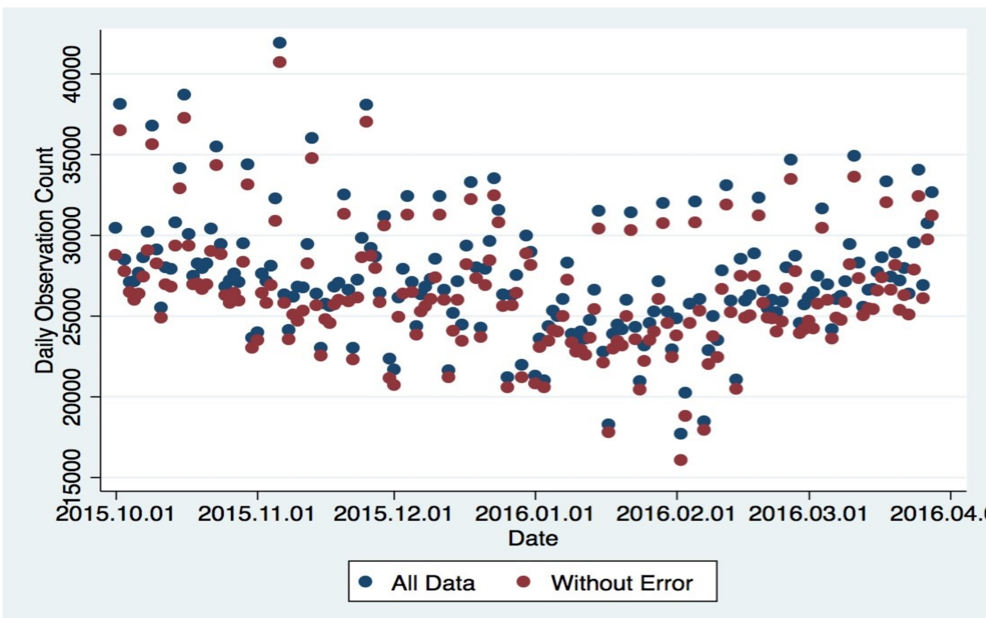

Figure 3.18 Scatter-plot of Daily Observation Count - IRD

Class-9 daily observation counts are shown in Figure 3.19. There is a clear pattern of daily observation counts with some higher counts (≈ 3,000) and some lower counts (≈ 1,000). This is due to the day-of-week effect. Class-9 counts are lower during weekends. There is also a drop in counts surrounding the Christmas and the New-Year day, which also seems reasonable. Class-9 without-error counts are similar to the data with errors and are generally consistent over time.

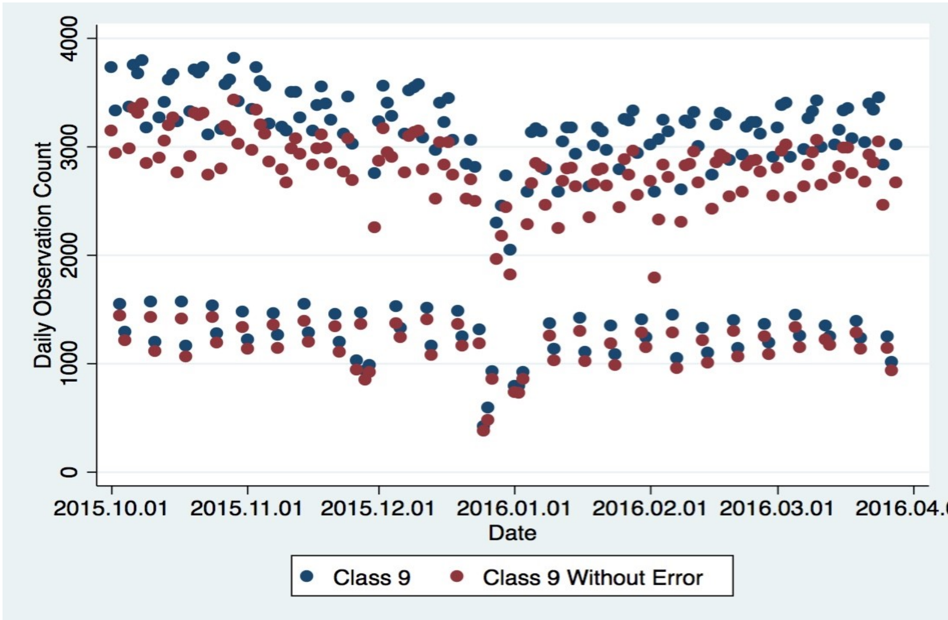

Figure 3.19 Scatter-plot of Daily Observation Count - IRD Class 9

Figure 3.20 shows the daily observation count of Class 9 by lane. Lane 1 has a larger range (1,000 to 3,000) compared with Lane 2 (100 to 600). Both lanes have consistent day-of-week patterns over time. Counts of weekdays are much larger than those of weekends, which appear to be representative of the anticipated weekday versus weekend traffic pattern.

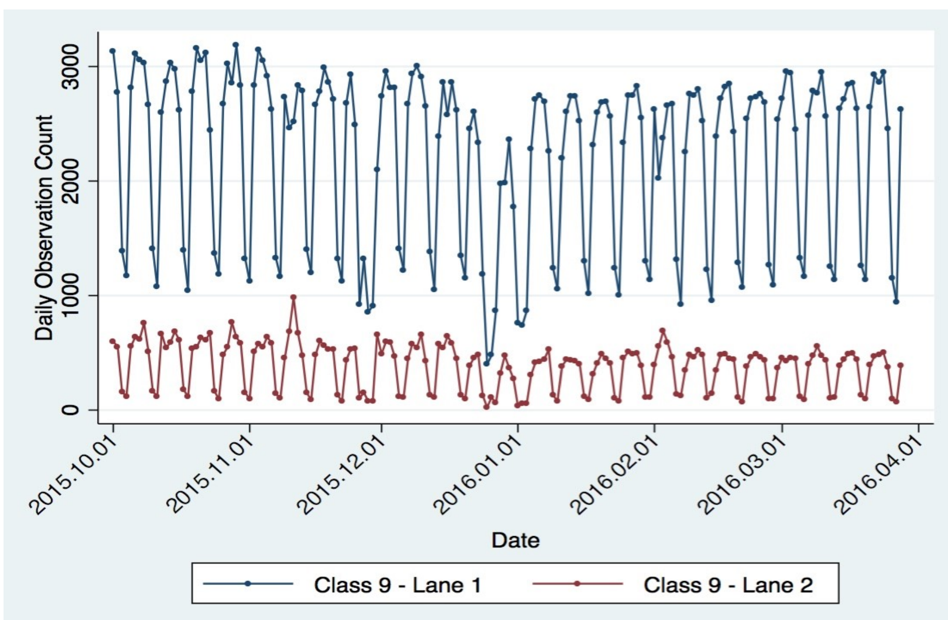

Figure 3.20 Daily Observation Count by Lane - IRD Class 9

In general, after considering the effect of system errors, lanes and the effect of day of week, daily observation counts of the IRD sensor are self-consistent from October to March.

3.3.2 Gross Vehicle Weight

To analyze consistency of gross vehicle weight of IRD, we consider mean of GVW, standard deviation of GVW and distribution of GVW. Because Class 9 is the most important vehicle class for MNDOT, all the analyses for gross vehicle weight and speed are for Class 9 only.

3.3.2.1 Mean of Gross Vehicle Weight

Daily average mean of GVW of Class 9 vehicles with and without errors is shown in Figure 3.21. Means without errors are very close to those with errors because as we mentioned, 96% of the data are error free for IRD. There are some fluctuations in mean GVW observations, but there are no sudden significant changes or changing patterns during the observation period.

Figure 3.21 Mean of GVW - IRD Class 9

Figure 3.22 shows the average GVW for Class 9 vehicles by lane. Both lanes are consistent over time. The two lanes have clearly different average GVW, which are around 50,000 lbs. for Lane 1 and 60,000 lbs. for Lane 2. A priori, we find no explanation why Lane-2 vehicles are heavier than Lane-1 vehicle. We have two conjectures: either the differences are caused by different speeds and different sensor calibration at different speeds, or the algorithm that IRD uses to report weights depends on the observed speed. Vehicle speeds in the two lanes are indeed different because Lane 1 is the driving lane and Lane 2 is the passing lane.

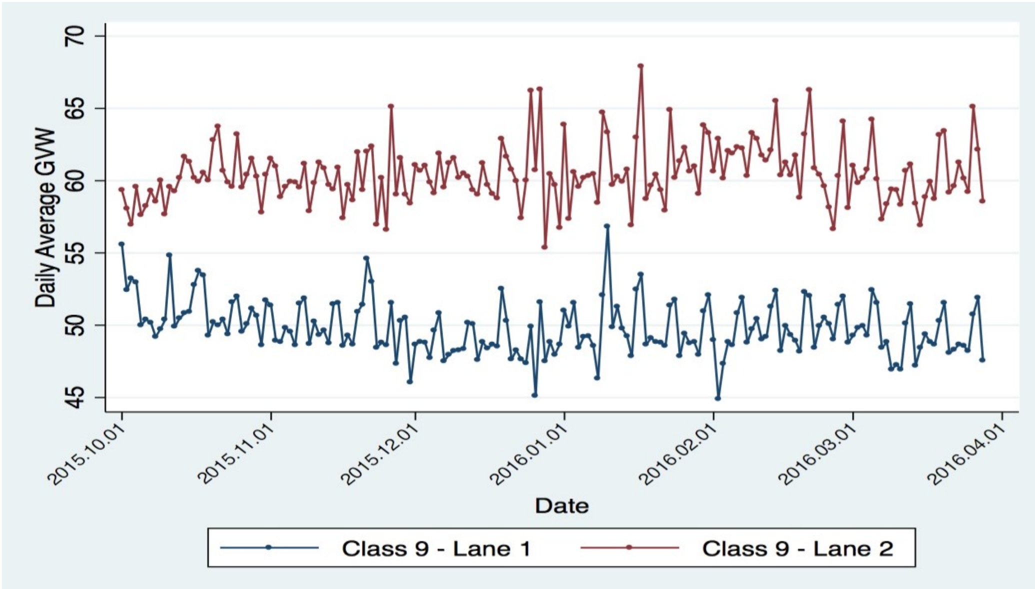

Figure 3.22 Mean of GVW - Class 9 by Lane

3.3.2.2 Variance of Gross Vehicle Weight

Figure 3.23 shows the standard deviation of GVW for Class 9. The without-error data have lower standard deviation. This is not surprising because IRD controller labels many outliers as with-error data. That being said, the differences between with-error and without-error data are small. They both have the same day-of-week patterns. For this reason, we conclude that the variance of GVW for Class 9s is consistent over time.

Figure 3.23 Standard Deviation of GVW - IRD Class 9

Figure 3.24 are the standard deviations of Class 9 by lane. Similar to the mean of GVW, two lanes have different levels of variability, with standard deviation hovering around 15 kips. for Lane 1 and around 20 kips. for Lane 2. Both lanes' day-of-week patterns are reasonably consistent over time. We do not have an explanation why the standard deviations of GVW are different in the two lanes. There is no obvious explanation as to why they should be different other than algorithm design and speed sensitivity of weight measurement by the sensor.

Figure 3.24 Standard Deviation of GVW by Lane - IRD Class 9

3.3.2.3 Distribution of Gross Vehicle Weight

To analyze the consistency of distribution of GVW, we look at box-plot of daily Class-9 GVW as well as monthly distributions. Statistical tests of distribution equality are performed for monthly comparisons.

Figure 3.25 gives daily box-plots of Class-9 GVW. The box-plots are consistent over time. There is no significant change in patterns. The median of GVW is around 50,000 lbs. The range of first and third quartiles is from 40,000 lbs. to 75,000 lbs. Different from Intercomp, IRD has a lot of outliers as defined in box-plots, which means the GVW data of IRD exhibits a greater range than that of Intercomp.

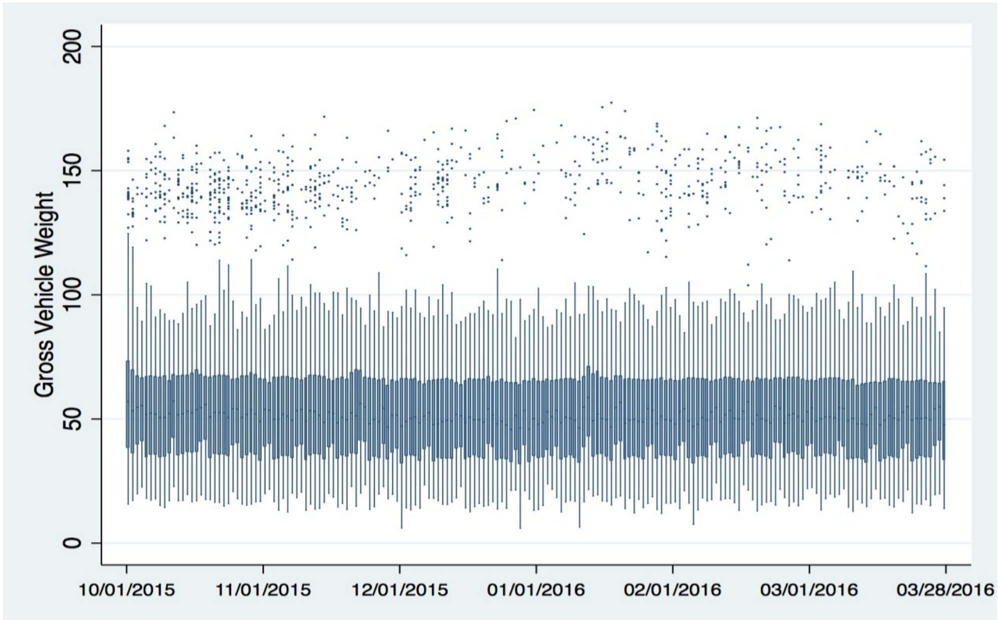

Figure 3.25 Daily Box-plot of GVW - IRD Class 9 with Error

Figure 3.26 shows the daily box-plots of Class 9 without error. The plots are very similar to Figure 3.25 except there are only two outliers, which means IRD's error reporting system identified the bulk of GVW outliers as errors.

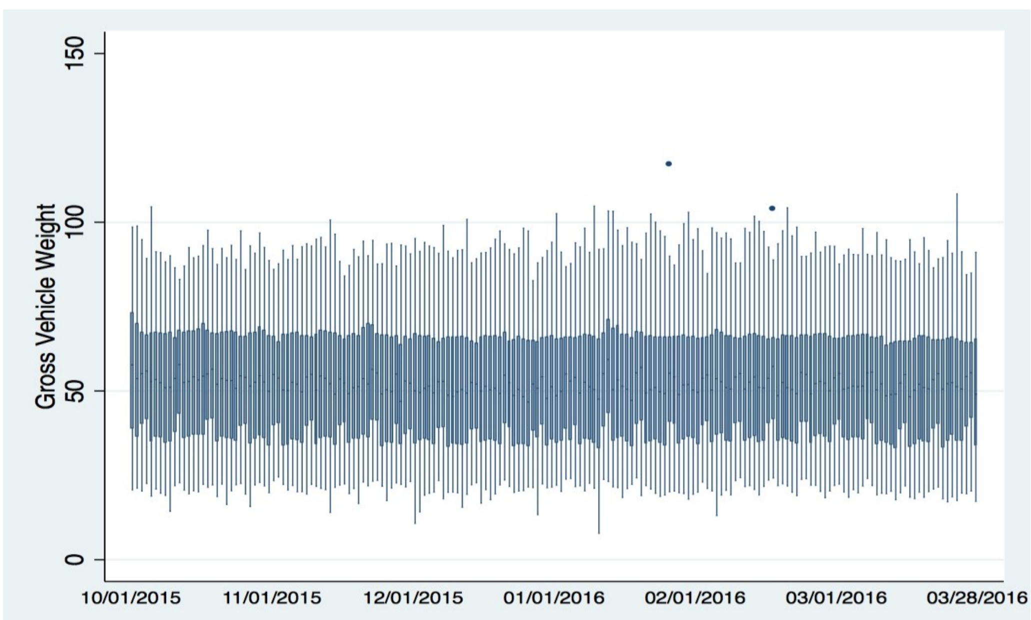

Figure 3.26 Daily Box-plot of GVW – IRD Class 9 without Error

Monthly distributional comparison of Class-9 GVW is given in Figure 3.27. The relative frequency distributions for each month are very close to each other except for the uptick that occurred around 80,000 lbs. This is caused because we combined Lane 1 and Lane 2 data in this figure. We find that the uptick disappears when we plot each lane's data separately.

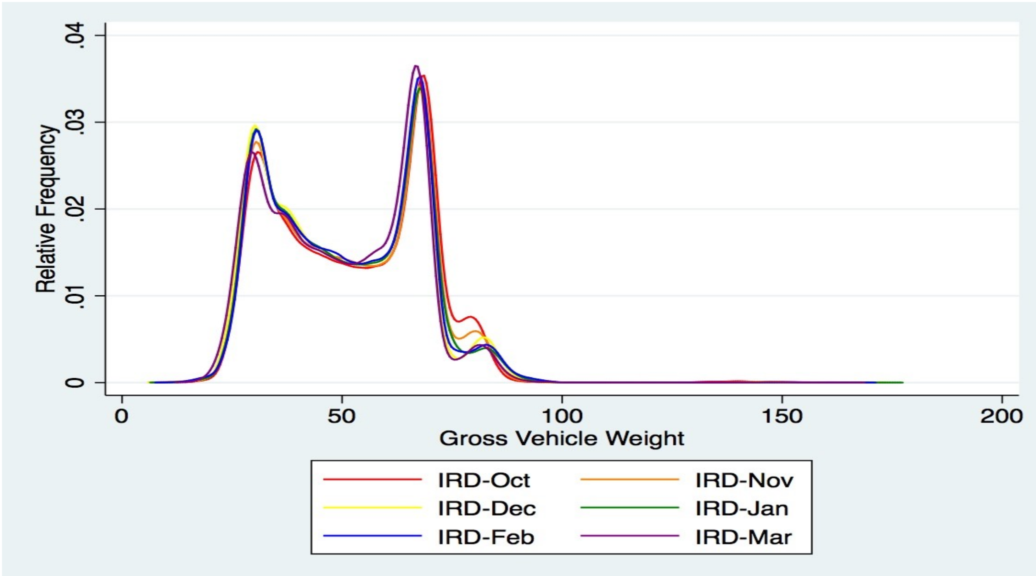

Figure 3.27 Month-to-month Comparison of Class 9 GVW – IRD

Table 3.13 shows the Kolmogorov-Smirnov tests of GVW measurements reported by the IRD sensor for Class-9 vehicles for different months. Only consecutive months are compared to see whether the patterns change from one month to the next. Because all *p*-values are less than 0.05, we conclude that the distributions of GVW of IRD for each month are statistically different. However, the absolute value of the deviance is very small (see Table 4.14), which permits us to conclude that although the distributions are statistically different because we have a large amount of data for each month, they are not practically different. We conclude therefore that IRD data are consistent across different months.

Table 3.13 Monthly Two-sample Kolmogorov-Smirnov Test - IRD

Month	Deviance	p-value
October versus November	0.0306	0.00
November versus December	0.0390	0.00
December versus January	0.0207	0.00
January versus February	0.0079	0.02
February versus March	0.0384	0.00

Monthly distribution comparison of Class-9 GVW by lane is shown in Figure 3.28. Similar to Figure 3.27, distributions are consistent from month over month, and there is no uptick around 80,000 lbs. The reason for the uptick around 80,000 lbs. in the combined data is because the uptick is the second peak of GVW distribution of Lane 2. The first peak of GVW distributions for Lane 1 and Lane 2 are both around 35,000 lbs. But for the second peak, Lane 1 appears around 70,000, which is the second peak in the distribution of combined data. Because the data frequency of Lane 2 (100 to 600) is much smaller than Lane 1 (1,000 to 3,000), the second peak of Lane 2 becomes a small peak in the distributions of combined data.

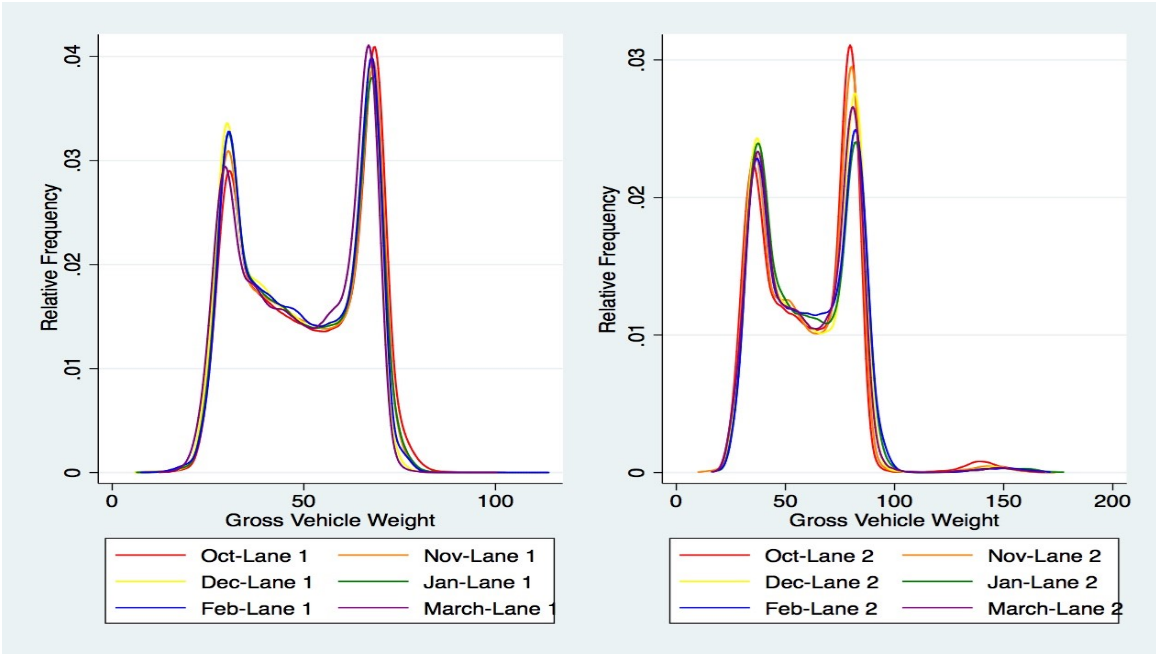

Figure 3.28 Comparison of IRD Monthly Distribution by Lane - Class 9

3.3.3 Speed

In this section, the mean, standard deviation and distribution of speed are analyzed for Class-9 vehicles.

3.3.3.1 Mean of Speed

Figure 3.29 shows the daily average speed for Class-9 vehicles. The difference between data with error and data without error is small. They both have consistent average speeds except for some low observations, which may be attributed to traffic and/or weather.

Figure 3.29 Mean of Speed - IRD Class 9

Figure 3.30 shows the daily average speed of Class-9 vehicles by lane. There is a clear difference between the two lanes' data. The mean of Lane 1 around 65 mph and Lane 2 around 70 mph. The fact that passing lane's mean speed is higher than that of the driving lane appears reasonable on an intuitive level. Speed patterns of both lanes are consistent over time.

Figure 3.30 Mean of Speed by Lane - IRD Class 9

3.3.3.2 Variance of Speed

Standard deviation is used to measure the variability of speed for Class 9 vehicle. Speed data without error has a much smaller standard deviation than data with error. Some individual days have very large standard deviations, but those days are evenly spread out across time. Therefore, we may conclude that both with and without error data have consistent speed variability patterns over time.

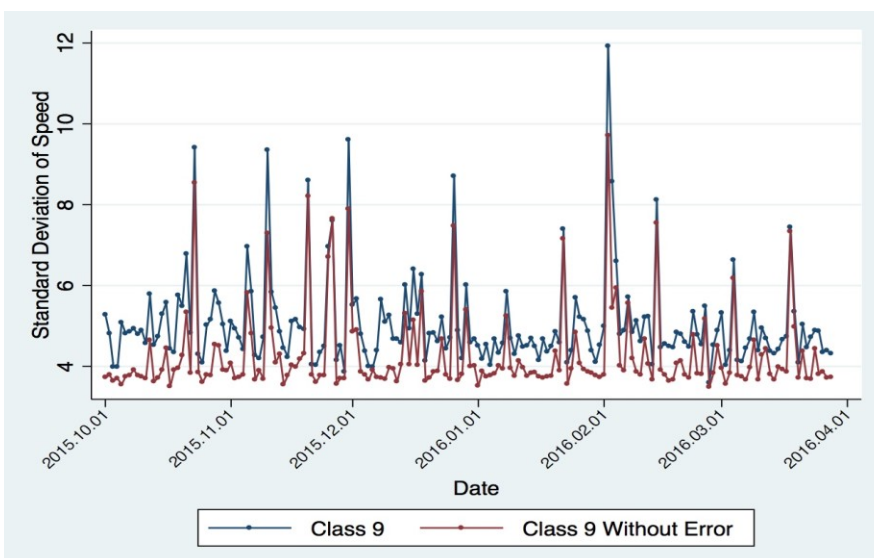

Figure 3.31 Standard Deviation of Speed - IRD Class 9

In Figure 3.32 we observe that the speed variability of Lane-1 data is smaller than that of Lane-2 data. The variability patterns of both lanes are consistent over time.

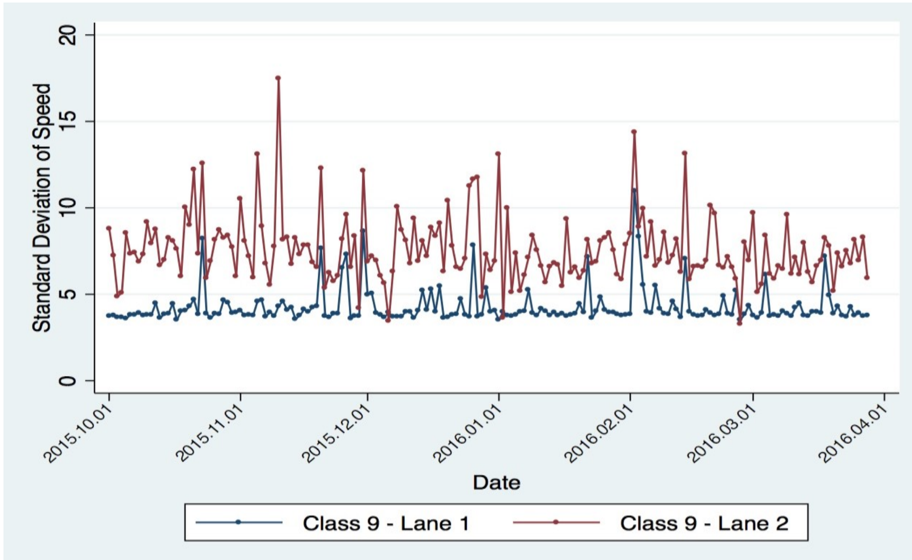

Figure 3.32 Standard Deviation of Speed by Lane - IRD Class 9

3.3.3.3 Distribution of Speed

Similar to the analyses of speed distribution of Intercomp, we perform distributional analyses of speed by month for IRD. As shown in Figure 3.33, speed distributions of IRD are consistent from October 2015 to March, 2016. Statistical results of Kolmogorov-Smirnov test are shown in Table 3.14. Although all the *p*-values are significant for distributional difference, the absolute values of deviances for each comparison are quite small.

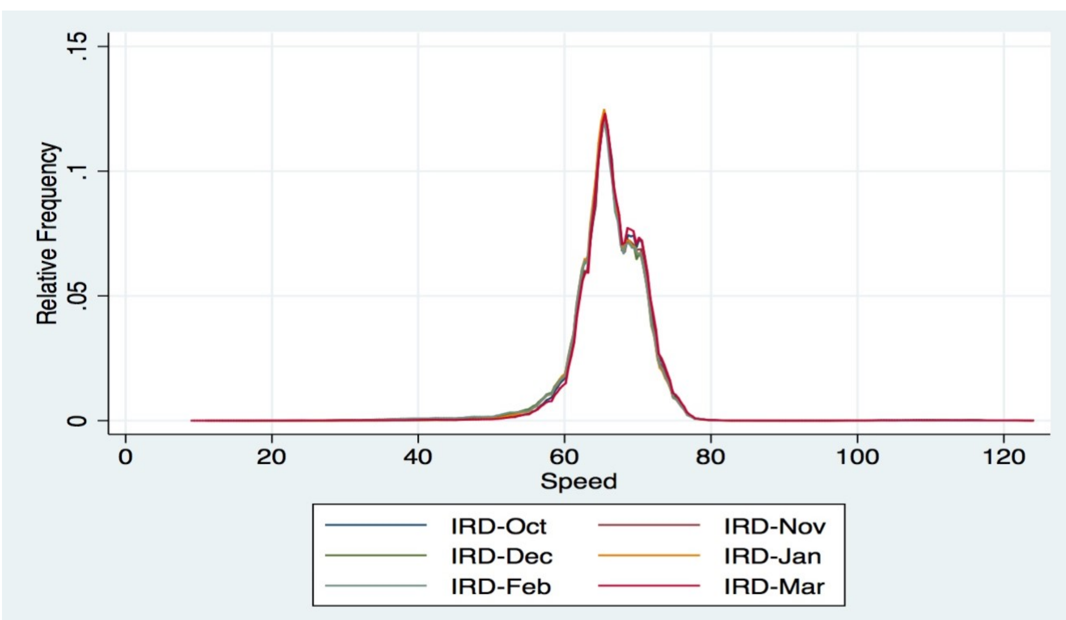

Figure 3.33 Monthly Distribution of Speed - IRD Class 9

Table 3.14 Monthly Two-sample Kolmogorov-Smirnov Test - IRD

Month	Deviance	p-value
October versus November	0.0284	0.00
November versus December	0.0112	0.00
December versus January	0.0076	0.03
January versus February	0.0175	0.00
February versus March	0.0539	0.00

Figure 3.34 shows the monthly distributions of speed by lane. As we can see, the distributions are consistent for both lanes.

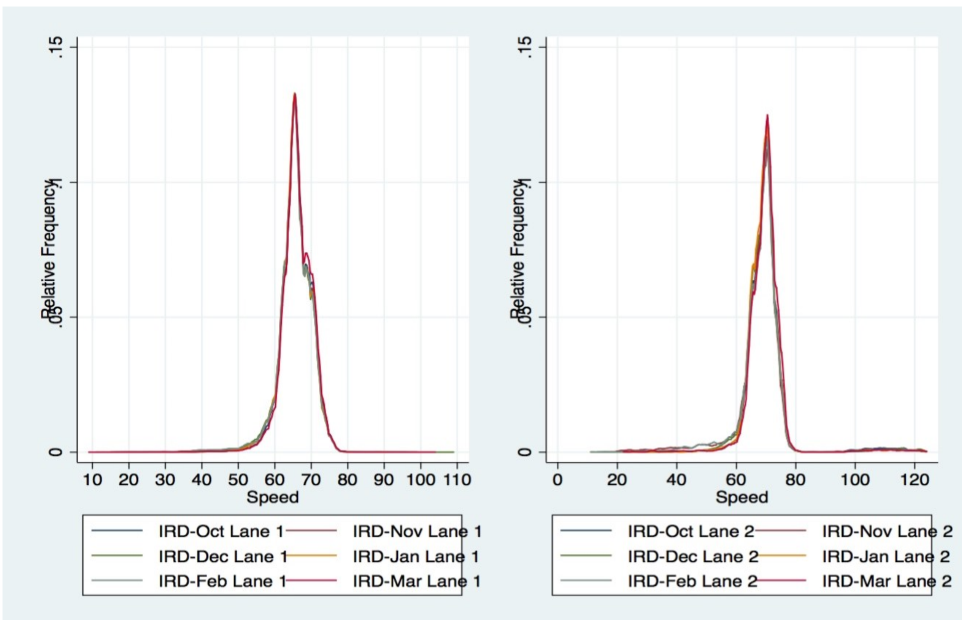

Figure 3.34 Monthly Distribution of Speed by Lane - IRD Class 9

In summary, IRD is considered consistent in terms of the distributions of daily observation count, gross vehicle weight and speed distribution.

3.4 COMPARISON OF INTERCOMP AND IRD

Intercomp and IRD sensors are installed near each other and there is no entry to or exit from the highway between the locations of the two sensor systems, which means they measure the same vehicles during the same time interval. Therefore, we may compare their sensor observations for the same time interval to assess differences. Recall that the valid Intercomp data starts on January 22nd, 2016 and ends March 28th. It is a proper subset of dates for which we have IRD data. Therefore, we use only the January 22nd to March 28th data for both sensors in these comparisons. Our analysis does not indicate which sensor is more accurate. We can know only whether their observations are similar or not. Accuracy comparisons will be possible if we had true vehicle weights.

3.4.1 Daily Observation Count

Because the two sensor systems are located close to each other, the actual number of vehicles passing through the two sensors during the same time interval should be nearly identical. Figure 3.35 shows the daily observation count of all classes for the two sensors. The two sensors have the same count pattern but for most of the days Intercomp counts fewer vehicles than IRD. It is unclear to us what may cause such difference in vehicle counts.

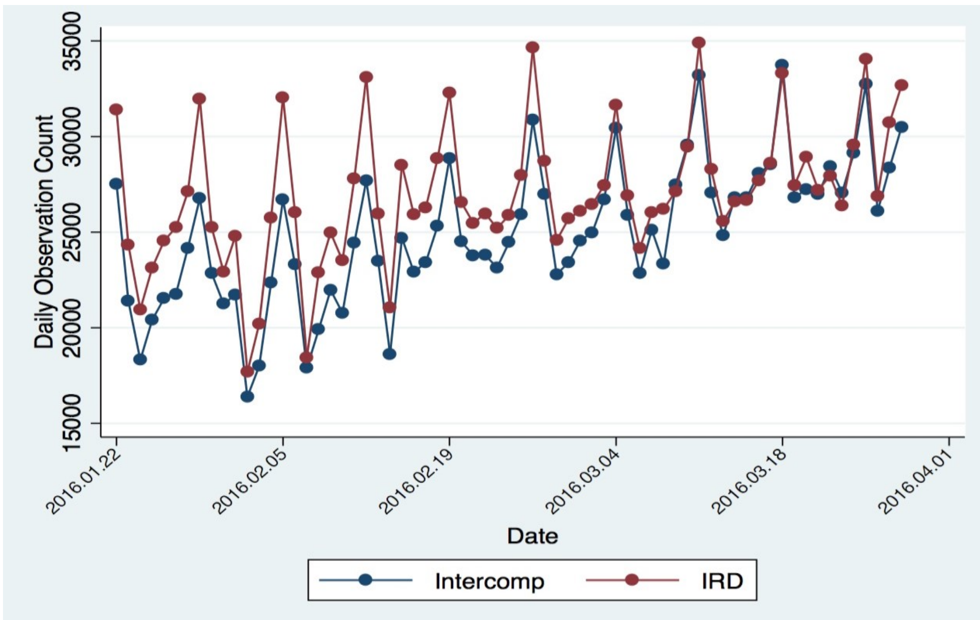

Figure 3.35 Comparison of Daily Observation Count - All Data

Figure 3.36 shows the comparison of the two sensors with and without error for Class-9 vehicles. Before mid February, the two sensors have similar patterns although IRD counts are more than those of Intercomp. Beginning around Feb 19th, Intercomp significantly undercounts Class-9 vehicles compared with IRD, especially for the error-free counts of Class-9 vehicles.

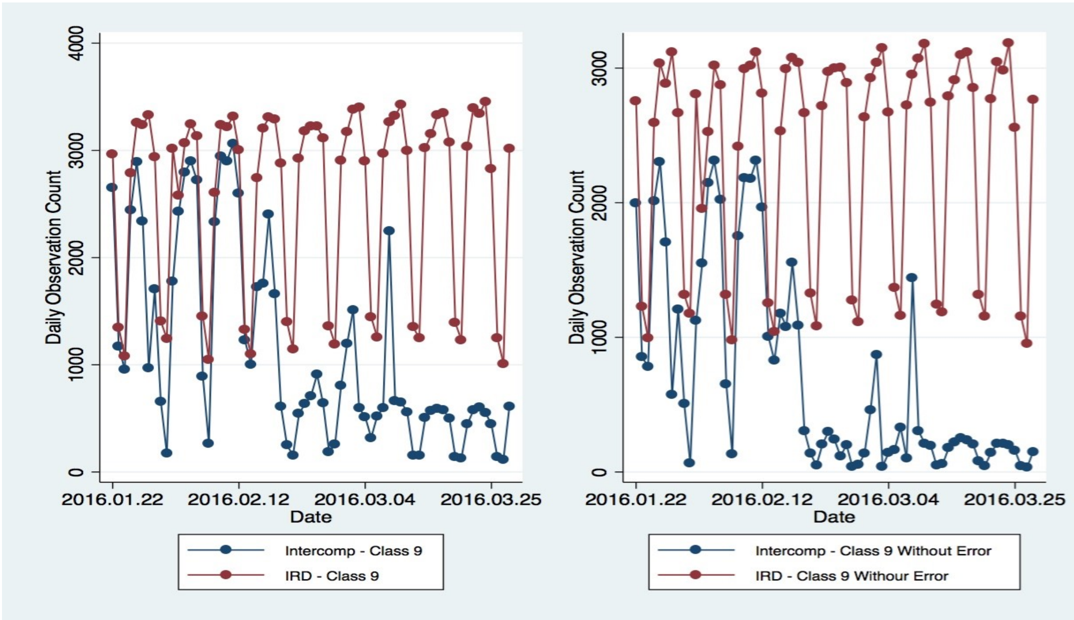

Figure 3.36 Comparison of Daily Observation Count - Class 9 with and without Error

Figure 3.37 shows daily observation count comparison by lane. Counts for Lane 2 are very close between IRD and Intercomp, but there is significant difference in Lane-1 counts. IRD counts more than Intercomp in Lane 1 and the difference becomes large after around Feb 19th.

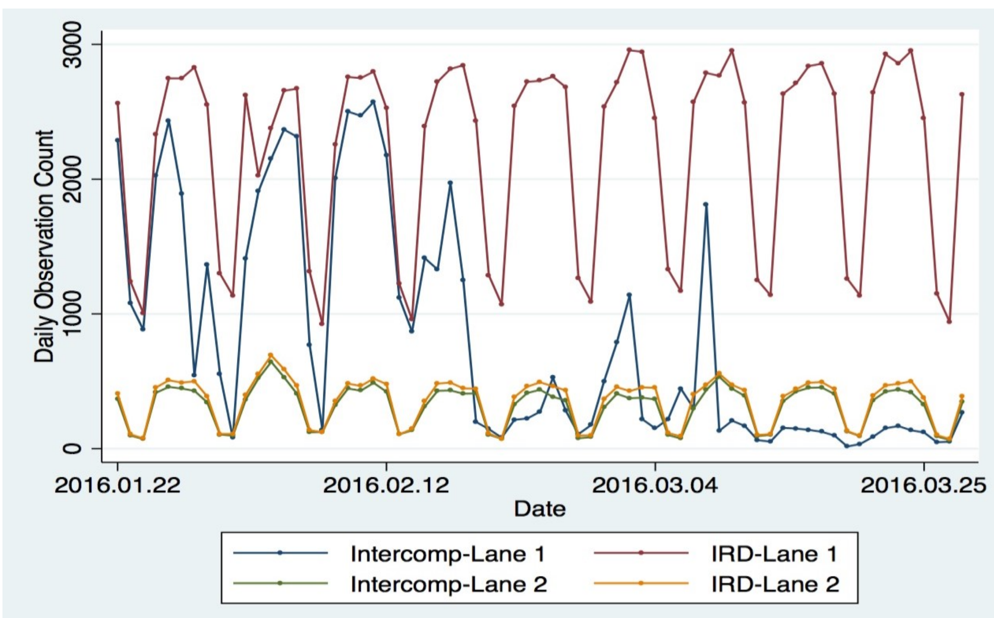

Figure 3.37 Comparison of Daily Observation Count by Lane - Class 9

It appears from this data that something might have gone wrong with Intercomp Lane-1 sensors around February 19th. It is not self-consistent and counts much less than IRD in general.

3.4.2 Gross Vehicle Weight

3.4.2.1 Mean of Gross Vehicle Weight

To test the equality of the mean of gross vehicle weight of the two sensors, we use two-sample T tests, which are shown in Table 3.15. The second column is the mean difference of IRD and Intercom. The third column is the T statistic calculated for each category. The null hypothesis is that the means of GVW for the two sensors are same. Column 4 gives *p*-values of different alternative hypotheses.

For Class-9 data, with or without error, the mean of GVW for IRD is statistically smaller than Intercomp. Same results are obtained for Class 9 Lane 1 (with or without error). But for Class 9 data from Lane 2, with or without errors, opposite results are achieved. The mean of GVW for IRD is statistically larger than Intercomp.

Table 3.15 Two-sample T test of Mean of GVW

GVW	diff=mean(IRD)-mean(Intercomp)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	-6.01	-71.63	0.00	0.00	1.00
Class 9 Without Error	-6.69	-75.89	0.00	0.00	1.00
Class 9 Lane 1	-8.25	-84.86	0.00	0.00	1.00
Class 9 Lane 2	5.45	28.62	1.00	0.00	0.00
Class 9 Without Error Lane 1	-8.33	-86.39	0.00	0.00	1.00
Class 9 Without Error Lane 2	3.38	15.26	1.00	0.00	0.00

Because in Section 3.2 we analyzed the self-consistency of Intercomp and found out that there might be something wrong with the Intercomp system after February 19th. Therefore, we repeated the two-sample T tests of GVW for the two sensors using data from Jan 22nd to Feb 19th only, which are shown in Table 3.16.

Compared with Table 3.15, all the results and significance are the same.

Table 3.16 Two-sample T test of Mean of GVW – Jan 22nd to Feb 19th

GVW	diff=mean(IRD)-mean(Intercomp)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	-4.24	-43.14	0.00	0.00	1.00
Class 9 Without Error	-6.48	-61.61	0.00	0.00	1.00
Class 9 Lane 1	-5.81	-57.18	0.00	0.00	1.00
Class 9 Lane 2	5.19	18.35	1.00	0.00	0.00
Class 9 Without Error Lane 1	-8.11	-74.97	0.00	0.00	1.00
Class 9 Without Error Lane 2	3.35	10.48	1.00	0.00	0.00

A natural question that arises is the following. Why do IRD sensors report smaller weights in Lane 1 and larger weights in Lane 2? Without further investigation of the sensor systems, it is difficult to explain these differences. One conjecture we have is that the two sensors calibrate differently based on vehicle speeds. That being said, the reasons for these differences are worth investigating either in a laboratory or field-test setting.

3.4.2.2 Variance of Gross Vehicle Weight

We use F tests to check if the two sensors have same GVW variability - results are shown in Table 3.17. The second column is the ratio of the standard deviations of IRD and Intercomp GVW measurements. The third column is the F statistic calculated for each category. The null hypothesis is that the ratio of the standard deviation of GVW for the two sensors is 1. *P*-values of different alternative hypothesis are given in the fourth column.

Similar to the results of means of GVW, IRD has statistically more variability than Intercomp for Class-9 vehicles in Lane 2. But in Lane 1 and combined data, Intercomp has statistically more variability than IRD.

Table 3.17 F Test of Variance Ratio of GVW

	ratio=sd(IRD)/ sd(Intercomp)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	0.83	0.69	0.00	0.00	1.00
Class 9 Without Error	0.95	0.90	0.00	0.00	1.00
Class 9 Lane 1	0.74	0.55	0.00	0.00	1.00
Class 9 Lane 2	1.12	1.25	1.00	0.00	0.00
Class 9 Without Error Lane 1	0.91	0.82	0.00	0.00	1.00
Class 9 Without Error Lane 2	1.08	1.16	1.00	0.00	0.00

The same analyses are also performed for data between Jan 22nd and Feb 19th only, which are shown in Table 3.18. All results and significance are the same.

Table 3.18 F Test of Variance Ratio of GVW – Jan 22nd to Feb 19th

	ratio=sd(IRD)/ sd(Intercomp)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	0.94	0.89	0.00	0.00	1.00
Class 9 Without Error	0.96	0.93	0.00	0.00	1.00
Class 9 Lane 1	0.87	0.75	0.00	0.00	1.00
Class 9 Lane 2	1.12	1.26	1.00	0.00	0.00
Class 9 Without Error Lane 1	0.91	0.83	0.00	0.00	1.00
Class 9 Without Error Lane 2	1.07	1.14	1.00	0.00	0.00

3.4.2.3 Distribution of Gross Vehicle Weight

To compare the distribution of GVW for Class-9 vehicles, we use cumulative distribution function to see the frequencies of GVW for each sensor. As shown in Figure 3.38, GVW cumulative distribution of Class

9 vehicle in Lane 2 of IRD is stochastically larger than that of Intercomp, which locates on the bottom right of the figure. But for other three cases, the opposite result is obtained – Intercomp is stochastically larger than IRD.

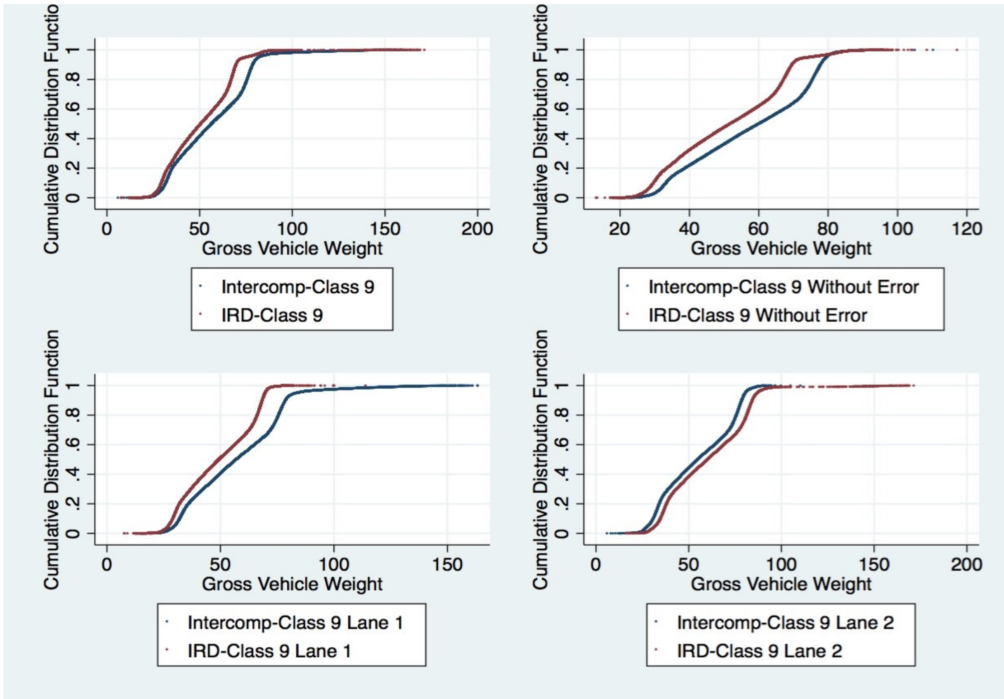

Figure 3.38 CDF of GVW - Class 9 without Error and by Lane

Same results are also obtained for data between Jan 22nd and Feb 19th, as shown in Figure 3.39.

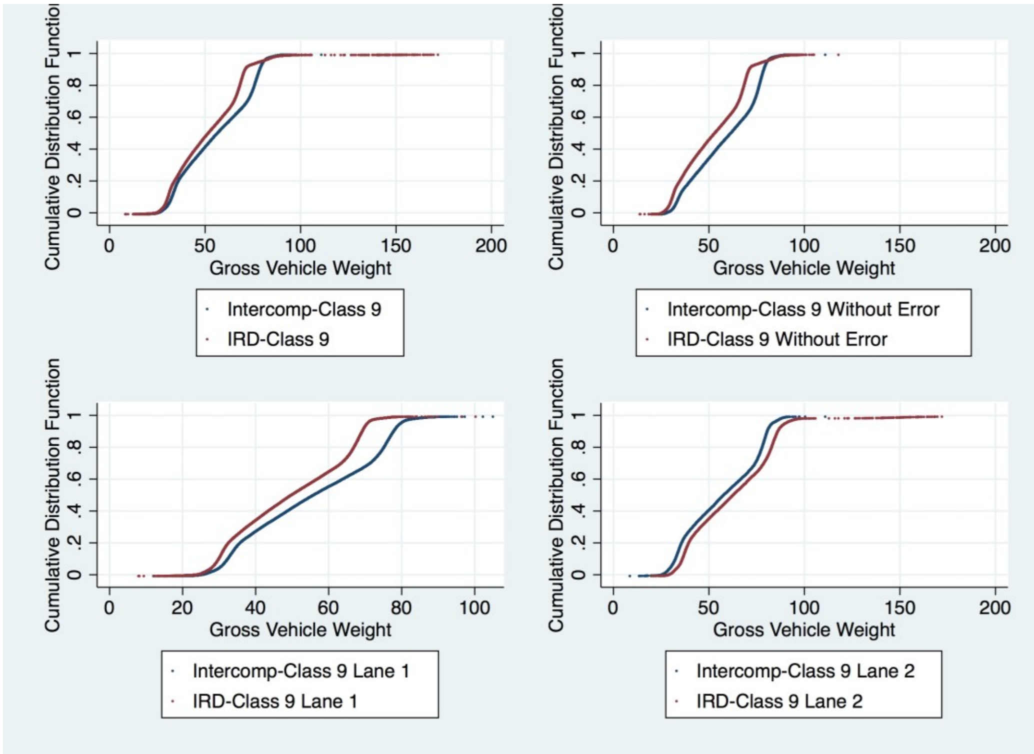

Figure 3.39 CDF of GVW - Class 9 without Error and by Lane Jan 22nd to Feb 19th

To test the distribution differences statistically, we also perform Kolmogorov-Smirnov tests of Class 9 GVW for the two sensors, as shown in Table 3.19. With all p-values of 0, we conclude that the distributions of GVW for the two sensors are not statistically the same for any of the categories.

Table 3.19 Two-sample Kolmogorov-Smirnov Test of GVW

	Deviance	p-value
Class 9	0.25	0.00
Class 9 Without Error	0.27	0.00
Class 9 Lane 1	0.30	0.00
Class 9 Lane 2	0.18	0.00
Class 9 Without Error Lane 1	0.32	0.00
Class 9 Without Error Lane 2	0.19	0.00

Same tests are performed for the data between Jan 22nd and Feb 19th only, as shown in Table 3.20. All the results are the same.

Table 3.20 Two-sample Kolmogorov-Smirnov Test of GVW Jan 22nd to Feb 19th

	Deviance	p-value
Class 9	0.22	0.00
Class 9 Without Error	0.26	0.00
Class 9 Lane 1	0.27	0.00
Class 9 Lane 2	0.17	0.00
Class 9 Without Error Lane 1	0.31	0.00
Class 9 Without Error Lane 2	0.18	0.00

3.4.3 Speed

In what follows, we present comparisons of mean, variance and distributions of speeds recorded by the two sensors. Since our approach taken is similar to that in Sections 3.4.1 and 3.4.2 we will not restate the methodology in this section. Instead, we summarize our findings first and then present the results of our analysis in tables and figures in the ensuing subsections.

Intercomp generates many extremely large speed data in Lane 1, but its error reporting system tags them as errors. If those extreme data are excluded, IRD in general has a larger mean and variance of speed than Intercomp.

3.4.3.1 Mean of Speed

The means of speed for Intercomp and IRD are statistically different for Class-9 vehicles regardless error and lane. Results of the statistical tests are shown in Table 3.21. Mean of speed of IRD is statistically larger than Intercomp.

Table 3.21 Two-sample T test of Mean of Speed

Speed	diff=mean(IRD)-mean(Intercomp)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	0.88	26.48	1.00	0.00	0.00
Class 9 Without Error	1.84	78.95	1.00	0.00	0.00
Class 9 Lane 1	1.47	48.23	1.00	0.00	0.00
Class 9 Lane 2	1.14	16.69	1.00	0.00	0.00
Class 9 Without Error Lane 1	2.01	81.44	1.00	0.00	0.00
Class 9 Without Error Lane 2	2.24	35.77	1.00	0.00	0.00

Same result could be obtained for the data between Jan 22nd and Feb 19th, which is shown in Table 3.22

Table 3.22 Two-sample T test of Mean of Speed – Jan 22nd to Feb 19th

Speed	diff=mean(IRD)-mean(Intercomp)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	1.58	43.20	1.00	0.00	0.00
Class 9 Without Error	1.79	61.64	1.00	0.00	0.00
Class 9 Lane 1	1.85	49.90	1.00	0.00	0.00
Class 9 Lane 2	1.00	9.08	1.00	0.00	0.00
Class 9 Without Error Lane 1	1.75	59.89	1.00	0.00	0.00
Class 9 Without Error Lane 2	2.07	20.66	1.00	0.00	0.00

3.4.3.2 Variance of Speed

The variances of speed of the two sensors are also statistically different, which is shown in Table 3.23. Except for all Class-9 vehicles and Class-9 vehicles in Lane 1 with errors, standard deviation of speed of IRD is statistically larger than that of Intercomp. But for all Class-9 vehicles and Class-9 vehicles in lane 1 with errors, standard deviation of speed of IRD is statistically smaller than Intercomp. This is because Intercomp generated a lot of extremely large speed data in Lane 1 and those data are captured by the error system.

Table 3.23 F Test of Variance Ratio of Speed

Speed	ratio=sd(IRD)/sd(Intercomp)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	0.64	0.41	0.00	0.00	1.00
Class 9 Without Error	1.06	1.12	1.00	0.00	0.00
Class 9 Lane 1	0.52	0.27	0.00	0.00	1.00
Class 9 Lane 2	1.30	1.69	1.00	0.00	0.00
Class 9 Without Error Lane 1	1.05	1.11	1.00	0.00	0.00
Class 9 Without Error Lane 2	1.21	1.47	1.00	0.00	0.00

Again, as shown in Table 3.24, for the data between Jan 22nd and Feb 19th, the results are the same.

Table 3.24 F Test of Variance Ratio of Speed – Jan 22nd to Feb 19th

Speed	ratio=sd(IRD)/sd(Intercomp)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	0.83	0.69	0.00	0.00	1.00
Class 9 Without Error	1.11	1.22	1.00	0.00	0.00
Class 9 Lane 1	0.72	0.52	0.00	0.00	1.00
Class 9 Lane 2	1.34	1.79	1.00	0.00	0.00
Class 9 Without Error Lane 1	1.08	1.16	1.00	0.00	0.00
Class 9 Without Error Lane 2	1.20	1.44	1.00	0.00	0.00

3.4.3.3 Distribution of Speed

Figure 3.40 shows the histogram of speed for each sensor under different data set. For all cases, the histograms seem different across sensors, especially for Lane 1. To test the difference statistically, we use Kolmogorov-Smirnov tests, which are shown in Table 3.25. All the distributions of speed are significantly different between the two sensors. Among all the cases, Class-9 Lane 2 has the closest distribution with a deviance of 0.0992. The differences may come from the sensors, as well as the algorithms their controllers use to convert raw data into measurements and tag certain observations as errors.

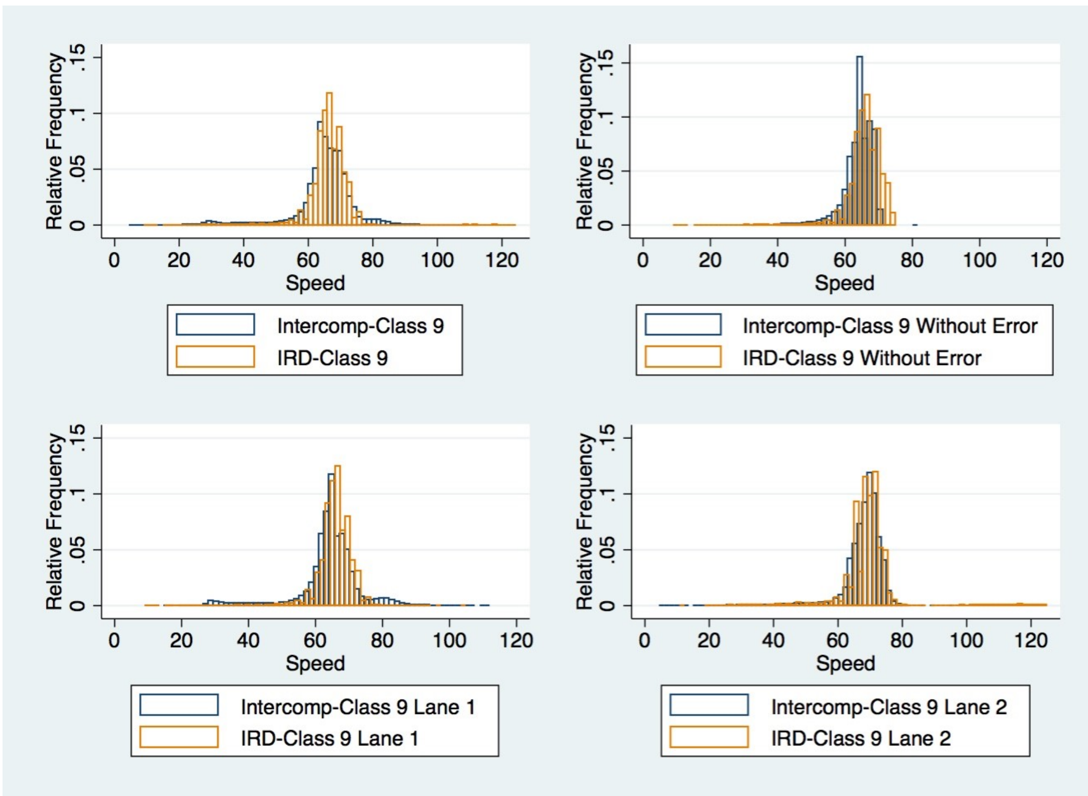

Figure 3.40 Comparison of Histogram of Speed - Class 9 without Error and by Lane

Table 3.25 Two-sample Kolmogorov-Smirnov Test of Speed

Speed	Deviance	p-value
Class 9	0.1562	0.00
Class 9 Without Error	0.2676	0.00
Class 9 Lane 1	0.2292	0.00
Class 9 Lane 2	0.0992	0.00
Class 9 Without Error Lane 1	0.3191	0.00
Class 9 Without Error Lane 2	0.4677	0.00

As shown in Table 3.26, same results are obtained for data between Jan 22nd and Feb 19th.

Table 3.26 Two-sample Kolmogorov-Smirnov Test of Speed – Jan 22nd to Feb 19th

Speed	Deviance	p-value
Class 9	0.2003	0.00
Class 9 Without Error	0.2784	0.00
Class 9 Lane 1	0.2408	0.00
Class 9 Lane 2	0.0931	0.00
Class 9 Without Error Lane 1	0.2989	0.00
Class 9 Without Error Lane 2	0.4263	0.00

In conclusion, we compared Intercomp sensor with IRD sensor in terms of daily observation count, gross vehicle weight and speed. The two systems do not have the same performance for each criterion. The main difference comes from Lane 1 observations from the two systems. This is true even if we compare the two sensors after excluding the data after February 19th.

3.5 CONCLUSIONS

3.5.1 Findings

Main findings of this chapter can be summarized as follows:

1. The Intercomp sensor is not self-consistent during the time period Jan 22nd to March 28th. There is a sudden and significant difference between the patterns of daily observation count, GVW and speed before and after Feb 19th.
2. The inconsistency of Intercomp sensor is mainly from Lane 1, the driving lane.
3. IRD sensor is considered self-consistent from October 1st to March 28th.
4. Intercomp and IRD have statistically different daily observation count, GVW distribution and speed distribution. Because we do not have the actual weights, speeds and vehicle counts, the accuracy of the two systems is not known. However, the lack of self-consistency suggests that the Intercomp sensor system stopped performing according to specifications around mid February.

3.5.2 Recommendations

Based on the analyses in this chapter, we recommend actions that MNDOT can take both to improve performance evaluation and detect the need for recalibration.

Improving Performance Evaluation

1. Collect more data for both sensors. The data used in this chapter is from Jan 22nd, 2016 to March 28th, 2016 for Intercomp and October 1st, 2015 to March 28th, 2016 for IRD. Whereas we observe a potential failure of Intercomp sensor in Lane 1, we cannot tell how long the IRD sensor will continue to perform consistently. That is, more data will help shine light on the time between required recalibration of the two sensors.

2. Investigate the reasons for the significant change around Feb 19th for Intercomp.

3. Get more information of the error codes used by Intercomp so that we may separate errors (unusable data) and warnings (usable data). On this topic, it will be helpful to learn what thresholds are used to generate tags – good, warning (but usable), and bad data.

4. Perform more test runs so we can compare measured weights against true weights.

Recalibration

MNDOT uses the mean of front axle weight to monitor gross vehicle weight consistency over time. Recalibration threshold is plus or minus 5% of the baseline. We suggest using variance and deviance of the distribution of weights as additional criteria, in addition to the mean of front axle weight. Deviance measures the distribution difference of two samples of data. Such comparisons could be made biweekly or monthly to detect shifts in the distribution. Under normal circumstances, we will not expect the weight distributions to change much over time. Similarly, a comparison of variances could inform MNDOT if observations are beginning to exhibit a different pattern. Note that the mean weight (single axle or combined) may not change even though both the distribution and variance may change.

CHAPTER 4: COMPARISON OF THE KISTLER AND THE IRD WIM SYSTEMS' PERFORMANCE

4.1 INTRODUCTION

4.1.1 Research Goal

The goal of this chapter is to compare and contrast reported performance of the Kistler and the IRD WIM systems, and lay the foundation for analysis that leads to the selection of a superior system in terms of accuracy and life-cycle cost. We analyzed ten months of the IRD-system data and four months of the Kistler-system data in CSV format provided by the OTSM, which was produced by the installed sensors and controllers. We also developed a methodology for comparing and contrasting the two sensor systems' performances, which was then used to compare their performances based on the available data.

4.1.2 Data Description

The data used in this chapter are from April 1st, 2016 to July 28th, 2016 for the Kistler system and from October 1st, 2015 to July 28th, 2016 for the IRD system. The data for the Kistler system only contains Lane-1 (driving lane) measurements. To compare the two systems on an equal footing, we only consider Lane-1 data of the IRD system as well. All data provided by the OTSM were in CSV-format. A summary of the data fields in these files is shown in Table 4.1.

Table 4.1 Variable Summary of Original CSV-format Files

Variable Name	Explanation	Variable Name	Explanation
veh	Unique Vehicle Number	aw3	Axle Weight
lane	Lane Number	aw4	Axle Weight
time	Time that passing the sensor	aw5	Axle Weight
axlec	Axle Count	aw6	Axle Weight
speed	Speed that passing the sensor	aw7	Axle Weight
as1feet	Axle Space	aw8	Axle Weight
as2	Axle Space	aw9	Axle Weight
as3	Axle Space	aw10	Axle Weight
as4	Axle Space	aw11	Axle Weight
as5	Axle Space	aw12	Axle Weight
as6	Axle Space	GVW	Gross Vehicle Weight
as7	Axle Space	class	MN Classification
as8	Axle Space	err	Error Code
as9	Axle Space	imagef	Indicator of Image Capture

Variable Name	Explanation	Variable Name	Explanation
as10	Axle Space	license	License Number
as11	Axle Space	temp	Temperature
aw1kips	Axle Weight	calfac	Calibration Indicator
aw2	Axle Weight		

To produce statistical analyses, summary tables, and figures, we combined all data files and analyzed the entire dataset using STATA (version 14.1). A summary of the variables that we created is shown in Table 4.2.

Table 4.2 Summary of Created Variables

Variable Name	Explanation	Variable Name	Explanation
date	Date	IRD_err_19	IRD Error Indicator
month	Month	IRD_err_20	IRD Error Indicator
month_numeric	Numeric Representation of Month	IRD_err_21	IRD Error Indicator
sensor	Numeric Representation of Sensor	IRD_err_22	IRD Error Indicator
sensorName	Name of Sensor	IRD_err_23	IRD Error Indicator
indv1	Individual Error Code	IRD_err_24	IRD Error Indicator
indv2	Individual Error Code	IRD_err_25	IRD Error Indicator
indv3	Individual Error Code	IRD_err_26	IRD Error Indicator
indv4	Individual Error Code	IRD_err_27	IRD Error Indicator
indv5	Individual Error Code	IRD_err_28	IRD Error Indicator
indv6	Individual Error Code	IRD_err_29	IRD Error Indicator
good	Indicator of Error-free Data	IRD_err_30	IRD Error Indicator
kistler_err_7	Kistler Error Indicator	IRD_err_31	IRD Error Indicator
kistler_err_14	Kistler Error Indicator	IRD_err_32	IRD Error Indicator
kistler_err_16	Kistler Error Indicator	IRD_err_33	IRD Error Indicator
kistler_err_19	Kistler Error Indicator	IRD_err_34	IRD Error Indicator
kistler_err_34	Kistler Error Indicator	IRD_err_35	IRD Error Indicator
kistler_err_35	Kistler Error Indicator	IRD_err_36	IRD Error Indicator
kistler_err_67	Kistler Error Indicator	IRD_err_37	IRD Error Indicator
kistler_err_69	Kistler Error Indicator	IRD_err_38	IRD Error Indicator

Variable Name	Explanation	Variable Name	Explanation
kistler_err_70	Kistler Error Indicator	IRD_err_39	IRD Error Indicator
IRD_err_1	IRD Error Indicator	IRD_err_40	IRD Error Indicator
IRD_err_2	IRD Error Indicator	IRD_err_41	IRD Error Indicator
IRD_err_3	IRD Error Indicator	IRD_err_42	IRD Error Indicator
IRD_err_4	IRD Error Indicator	IRD_err_43	IRD Error Indicator
IRD_err_5	IRD Error Indicator	IRD_err_44	IRD Error Indicator
IRD_err_6	IRD Error Indicator	IRD_err_45	IRD Error Indicator
IRD_err_7	IRD Error Indicator	IRD_err_46	IRD Error Indicator
IRD_err_8	IRD Error Indicator	IRD_err_47	IRD Error Indicator
IRD_err_9	IRD Error Indicator	IRD_err_48	IRD Error Indicator
IRD_err_10	IRD Error Indicator	IRD_err_49	IRD Error Indicator
IRD_err_11	IRD Error Indicator	IRD_err_50	IRD Error Indicator
IRD_err_12	IRD Error Indicator	IRD_err_51	IRD Error Indicator
IRD_err_13	IRD Error Indicator	IRD_err_52	IRD Error Indicator
IRD_err_14	IRD Error Indicator	IRD_err_53	IRD Error Indicator
IRD_err_15	IRD Error Indicator	IRD_err_54	IRD Error Indicator
IRD_err_16	IRD Error Indicator	IRD_err_57	IRD Error Indicator
IRD_err_17	IRD Error Indicator	IRD_err_65	IRD Error Indicator
IRD_err_18	IRD Error Indicator		

Statistical summaries of the gross vehicle weight and speed are shown in Table 4.3 and Table 4.4, respectively. The unit of GVW in Table 4.3 is kips and the unit of vehicle speed in Table 4.6 is mph.

Table 4.3 GVW Statistical Summary from All Data

System Name	Freq.	mean(GVW)	sd(GVW)	min(GVW)	max(GVW)
IRD	4,311,947	13.22	18.94	0.05	197.1
Kistler	1,800,245	14.99	22.09	0	215.44

The summaries are for all recorded observations including all classes and errors. The means of GVW of the two systems are statistically different. The difference is also large enough to be of practical significance. The variance of gross vehicle weight for the Kistler system is a little bit larger than that of the IRD system. We conjecture that these differences might have been caused either by sensor accuracy or by the differences in the algorithms that the two systems use to convert the raw sensor data into weights.

Table 4.4 Speed Statistical Summary from All Data

System Name	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	4,311,947	68.89	6.53	3	124
Kistler	1,800,245	66.63	10.25	-1	1676.4

The variance of speed for the Kistler system is also much larger than that for the IRD system. The Kistler system also generates negative speed values, which does not happen in the IRD system. The speed of -1 only appears for Class-14 vehicles. All negative and extreme large (>120 mpg) speed data records are captured by error codes. Although different combinations of error codes are generated for different records, all negative speed records are labeled by error codes 7 and 14 of the Kistler system. Extreme large speed records are most likely to be labeled by error code 7, 19 and 70. Explanations of error codes are given in Table 4.7 and the Bullconverter manual (Kwon, 2015).

Because Class 9 is the most important class for MNDOT, we separately calculated statistical summaries of vehicle weight and speed for Class 9 only, which are shown in Table 4.5 and Table 4.6, respectively.

Table 4.5 GVW Statistical Summary for Class 9

System Name	Freq.	mean(GVW)	sd(GVW)	min(GVW)	max(GVW)
IRD	693,551	49.60	15.58	5.96	113.88
Kistler	285,236	57.90	18.46	9.84	121.04

In general, the Class-9 weight measurements from the Kistler and the IRD sensors are statistically different. They are also sufficiently different in magnitude to make a practical difference. The Kistler system has larger mean and variance compared to the IRD system. The differences in the frequency of observations are explained to a certain extent by the fact that the Kistler-system data pertains to a 4-month period, as compared to a 10-month period for the IRD system.

Table 4.6 Speed Statistical Summary for Class 9

System Name	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	693,551	65.77	4.74	9	109
Kistler	285,236	64.78	5.02	1.6	86.1

In terms of speed, the two systems are practically close, although they are statistically different. The Kistler system has a smaller minimum and maximum speed value than the IRD system.

The two systems have different error-reporting systems. Broadly speaking, all observations can be grouped into one of the three categories: no error, error (not valid), and warning (valid but outside of the specified range). We treated the data with either no errors or warnings as "good data". Summaries of the errors generated by the two systems are shown in Table 4.7 and Table 4.8. Note that because a record may contain more than one error, the sum of the error percentages is more than 100%.

Table 4.7 The Kistler System Error Code Summary

Error Code	Description	Percentage
null	No error	83.75%
67	High vehicle dynamics	9.09%
35	High weight imbalance	6.21%
69	Single-track vehicle	2.23%
7	Force record missing	2.00%
14	Cannot process vehicle	1.16%
34	Strong change in acceleration	0.79%
70	Sensor missing	0.51%
19	Single axle vehicle	0.46%
16	Vehicle stopped while driving through	0.14%

Observe that 83.75% of the Kistler-system data are error free. Among the error codes, high vehicle dynamics is the most common error type, which accounts for 9.09% of all the records.

Table 4.8 The IRD system Error Code Summary

Error Code	Description	Percentage
0	Normal, no error	83.65%
1	Upstream loop failure (Downstream loop only)	0.79%
2	Downstream loop failure (Upstream loop only)	0.10%

Error Code	Description	Percentage
7	Zero axles detected (failure of both axles)	0.01%
8	Unequal axle counts (difference of up and down-stream axle counts)	0.01%
13	Vehicle too slow (indicated by loop activation)	0.01%
17	Vehicle too fast	0.01%
19	One axle detected	0.01%
34	Significant speed change	0.06%
35	Significant weight differences	0.96%
37	Unequal axle counts on sensors	0.35%
38	Tailgating	0.59%
43	Overweight (good data)	0.16%
44	Over GVW (good data)	0.07%
46	Drastic speed change	0.41%
null	No error code	13.11%

In total, 96.76% of the IRD data are error free, which includes the "Code 0" and the "null" codes. The IRD system did not generate code "0" for error-free data from June 22nd to July 28th, which is 13.11% of all the records. Upon further investigation, it was determined that this was not an error. Starting 6/22/16, a new version of the BullConverter software (V4.18) was used. In that version the Err# column for a vehicle record that does not contain an error or waning is shown blank instead of "0". Therefore, blanks (or null) values are the same as normal observations with no errors, and are treated as such in our analysis. That is the reason why the total error-free data are 96.76%. The error code summary in Table 4.8 is slightly different from what we reported in Table 3.8 although both tables report the IRD system's error codes. There are three reasons for this. First, Table 3.8 used data from October 1st, 2015 to March 28th, 2016. During this time, there were no "null" error codes. Therefore, Table 3.8 reports the total percent of error-free observations. In contrast, Table 4.8 contains data for which "null" error codes were reported. In this table, the "null" codes are reported separately. Second, the interval of time covered by data used for Table 4.8 is larger. Third, in Table 4.8, we only consider Lane 1 data to allow comparisons with Table 4.7. In contrast, Table 3.8 uses data from all lanes.

Among the error codes in Table 4.8, significant weight difference is the most common error type, which accounts for less than 1% of all records. There are error codes in the IRD-system and the Kistler-system data that we do not report because those error codes have a frequency of 0% in the data. A complete error codebook can be found in the Bullconverter manual (Kwon, 2015) provided by Professor Taek Kwon.

In the remainder of this chapter, we refer to data as valid data when it has no errors. Note that this includes data with warnings. Statistical summaries of gross vehicle weight and speed without errors are shown in Table 4.9 and Table 4.10, respectively.

Table 4.9 GVW Summary for Class 9 without Error

System Name	Freq.	mean(GVW)	sd(GVW)	min(GVW)	max(GVW)
IRD	649,729	50.18	15.48	7.80	102.52
Kistler	169,957	63.69	16.52	14.94	121.04

Table 4.10 Speed Summary for Class 9 without Error

System Name	Freq.	mean(speed)	sd(speed)	min(speed)	max(speed)
IRD	649,729	65.69	4.57	9	76
Kistler	169,957	64.69	5.22	1.6	86.1

The Kistler system has fewer observations by a factor of approximately 3.8. The mean and standard deviation of speed are practically close, although they are statistically different. The Kistler system reports higher measured weights with greater variance. Equally importantly, the minimum and maximum of weights and speeds are different for the two sensors, with the Kistler system reporting a greater minimum and maximum of gross vehicle weights and a larger range of speed among error-free data.

4.1.3 Overall Approach

Our analyses consist of two parts. First, we perform self-consistency analysis of each system based on daily observation count, gross vehicle weight, and speed. Second, we compare the two sensors in terms of daily observation count, gross vehicle weight and speed. For each criterion, statistical mean, variance and distribution are analyzed.

The remainder of this section is organized as follows. Section 4.2 gives self-consistency analysis of the Kistler system and Section 4.3 repeats a similar analysis for the IRD system. Section 4.4 compares and contrasts the two systems. In Section 4.5 we summarize our findings and recommendations.

4.2 SELF-CONSISTENCY ANALYSIS: KISTLER SENSOR

An important criterion of sensor performance is self-consistency. A sensor's measurements may change gradually over time, or exhibit a sudden and sharp turning point. While the former may be caused by a variety of reasons, it is typically the case the latter occurs when the sensor fails or requires recalibration. In this analysis, we evaluate consistency using three metrics – daily observation counts, gross vehicle weights and speeds. We focus exclusively on Class-9 vehicles, which are of the greatest interest to MNDOT.

4.2.1 Daily Observation Count

Daily observation counts inform the daily traffic volume as recorded by the sensor. Average daily traffic counts are important for MNDOT because these counts are reported to the FHWA. We do not assume the same number of vehicle counts for each day because counts vary by the day of week, season, weather, and holidays. However, we do believe that a sudden and significant change in the pattern of average daily traffic count is likely to be a signal of sensor system failure and represents possible need for recalibration. We consider the effect of errors from which data originated and perform separate analysis for Class-9 vehicles.

Figure 4.1 shows the scatter-plot of the daily observation counts of the Kistler system for all data and data without errors (but including warnings' data). The daily observation counts of all data started at about 15,000 and increased gradually over time, reaching about 17,000 counts at the end of July. For error-free data, daily counts started with an average of 12,000 counts per day and increased to about 13,000 average daily count. In general the daily observation counts with or without error have the same increasing trend and it appears that error frequencies are consistent over time.

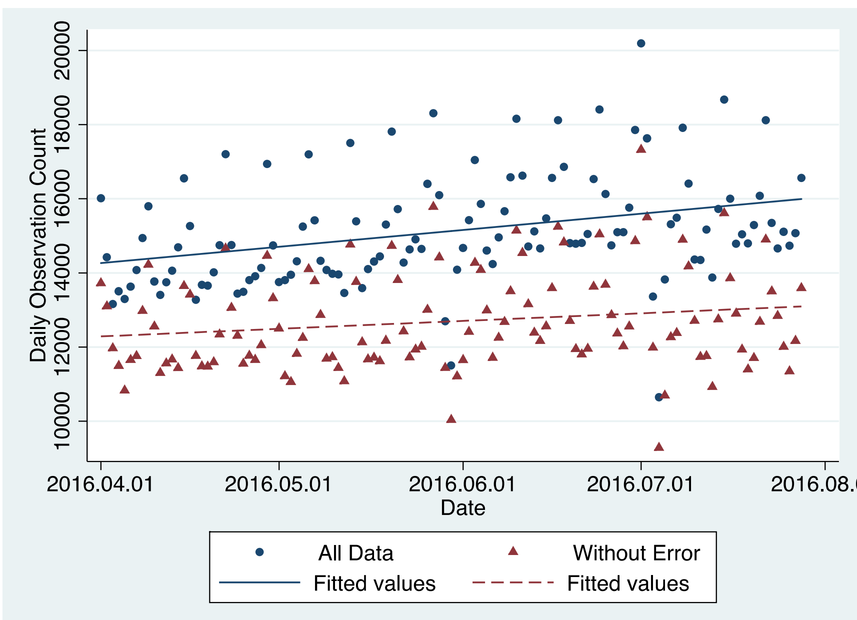

Figure 4.1 Scatter-plot of Daily Observation Count - the Kistler System

Figure 4.2 shows scatter plots of daily observation counts for Class-9 vehicles only. The dots are Class-9 vehicles with errors; the triangles are Class-9 vehicles without errors. It is clear that the daily observation count follows a weekly pattern – both with and without errors for Class-9 vehicles. The percentage of error free data is consistent over time. For all Class-9 vehicle data, the weekday counts are around 3,000 and the weekend counts are around 1,100, whereas for the Class-9 data without error, the counts for weekday and weekends are 1,900 and 900, respectively. The counts without error have a slightly decreasing trend for weekdays and weekends.

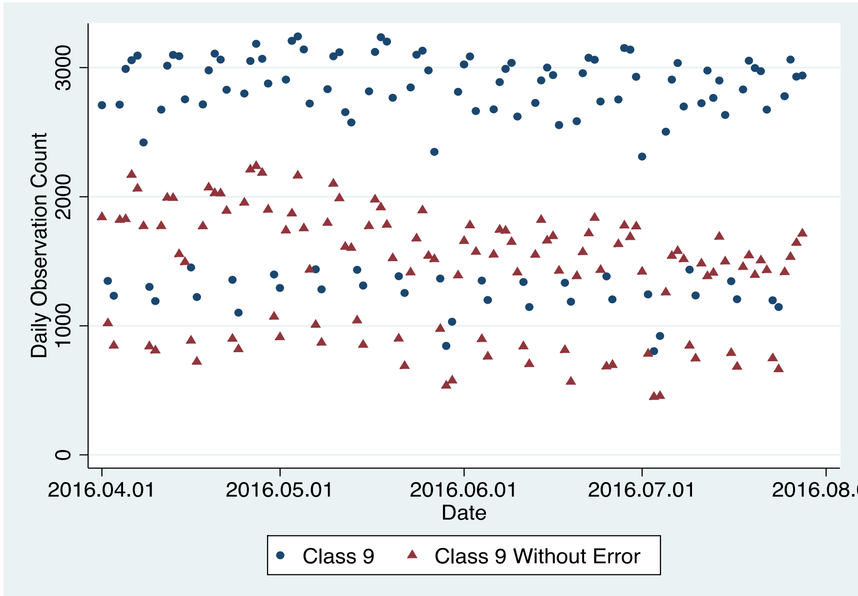

Figure 4.2 Scatter-plot of Daily Observation Count – the Kistler Class-9 Data

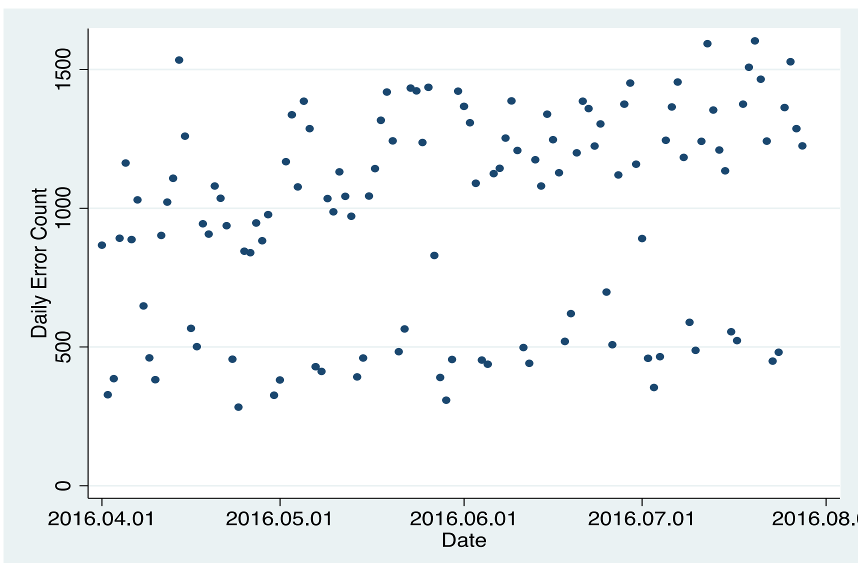

Figure 4.3 Scatter-plot of Daily Error Count – the Kistler Class-9 Data

Figure 4.3 is the daily error count for Class-9 data from the Kistler system. We see an increasing trend for weekdays (from 1,000 to 1,500), while the error counts of weekends appear to be consistent over time. This may be an indication that the system generates more errors over time. It is possible to extrapolate this trend to identify a future point in time when the system may need to be recalibrated. We will illustrate this methodology by using the Kistler Class-9 vehicle error percentage data. The error percentage is calculated as

$$\text{Daily Error Percentage} = \frac{\text{Daily Error Code Count}}{\text{Daily Observation count}} \tag{3.1}$$

We use a regression model by setting date as the explanatory variable and daily error percentage as the response variable, which could be represented as

$$\text{Daily Error Percentage} = \text{Cons} + \text{Coef.} \times \text{Date} + \text{Error Term} \quad (3.2)$$

We set April 1st, 2016 as date 1 of the data set and use ordinary least squares regression to estimate the coefficients of the model. The regression output is shown in Table 4.11.

Table 4.11 Regression Output of Error Percentage (Daily Error Percentage = Cons + Coef. × Date + Error Term)

Source	SS	df	MS			
				Number of obs	=	119
				F(1, 117)	=	87.71
Model	.245387525	1	.245387525	Prob > F	=	0.0000
Residual	.327348451	117	.00279785	R-squared	=	0.4284
				Adj R-squared	=	0.4236
Total	.572735976	118	.004853695	Root MSE	=	.05289

errpercent~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
obs	.0013219	.0001412	9.37	0.000	.0010424	.0016015
_cons	.3196145	.0097591	32.75	0.000	.300287	.3389419

The coefficients for both the date and the intercept are statistically significant. In particular, the 95% confidence level for the coefficient of Date is (0.00104, 0.00160), which means there is a clear trend of increase for the error percentage over time. The regression output yields the model as follows.

$$\text{Daily Error Percentage} = 0.32 + 0.00132 \times \text{Date} + \text{Error Term} \quad (3.3)$$

A positive coefficient of Date means there is an increasing trend of daily error percentage as time goes by. After obtaining the model coefficient, we may extrapolate to a date at which the recalibration threshold will be crossed. As an illustration, assume the recalibration threshold is x from the baseline (which is the fitted error percentage after last recalibration). We could project the Date of the next recalibration as

$$\text{Date} = \frac{x - 0.32}{0.00132} \quad (3.4)$$

The projection is not an accurate prediction of the error percentage, but rather an estimation of the expected error percentage based on linear regression. This projection date could serve as a planning guide when recalibration actions involve significant lead times. It is also possible to update the estimate to recalibration time as more data becomes available.

Overall, the daily observation count of the Kistler system is considered consistent. But when the total number of observation count was increasing over time, the Class-9 vehicle observation count was decreasing. This may be due to either real traffic pattern change or inaccuracy of the system. We will examine this issue in more detail in Section 4.4

4.2.2 Gross Vehicle Weight

Because different vehicle classes have different gross-vehicle-weight distributions and our primary focus is on Class-9 vehicles, we analyze gross vehicle weights and speed for Class-9 vehicles only. Analyses of other classes can be performed in a similar fashion.

We evaluate gross vehicle weight consistency in three aspects – daily mean of GVW, daily standard deviation of GVW and distributions of GVW. After accounting for day-of-week and seasonal variations, and average traffic counts, consistency of these three measurements should indicate good performance.

4.2.2.1 Mean of Gross Vehicle Weight

Figure 4.4 shows daily average gross vehicle weight for Class-9 vehicles from April 1st to July 28th. A clear weekly pattern is observed for both with- and without-error data. While the mean of GVW of class 9 is increasing over time, the mean of GVW of Class 9 without errors is consistent over the same time period. This indicates that the Kistler system generated larger gross-vehicle-weight observations, but those data are labeled as errors in their system, leaving the remaining data reasonably consistent over time.

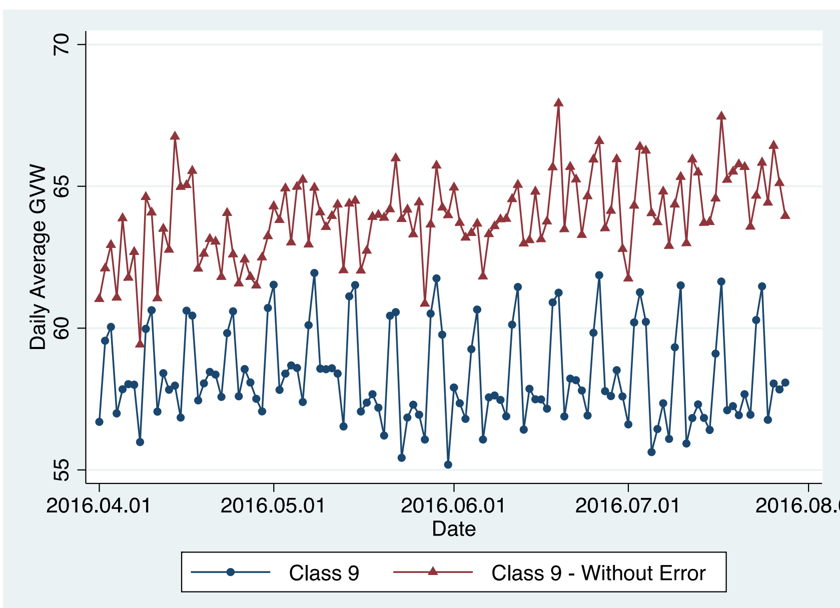

Figure 4.4 Mean of GVW – the Kistler Class-9 Data

4.2.2.2 Variance of Gross Vehicle Weight

In addition to the average gross vehicle weight, variance of GVW is another important determinant of consistency. Figure 4.5 shows the daily standard deviation of GVW of Class-9 vehicles. Similar to mean of GVW, there is a weekly pattern for both with and without error data. There is a slight increase of the standard deviation of GVW for Class-9 data (including errors) over time. The standard deviation for error free Class-9 data remains reasonably consistent around 16.

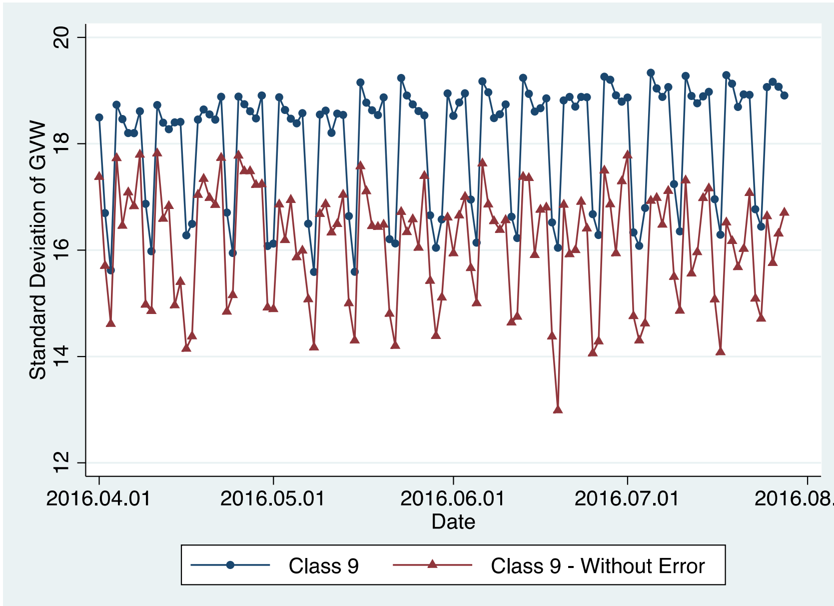

Figure 4.5 Standard Deviation of GVW – the Kistler Class-9 Data

4.2.2.3 Distribution of Gross Vehicle Weight

Distribution of GVW is another important criterion we consider for consistency. Daily or monthly consistent distributions are indicators of sensor calibration remaining intact because we do not expect the distributions of gross vehicle weight to change significantly across consecutive days or months.

Figure 4.6 shows the box-plot of GVW of Class-9 vehicles of the Kistler's system. Middle lines in the boxes are medians of GVW for each day. Bottom and top of the boxes represent the first and third quartiles of GVW for each day. The extended solid lines are within 1.5 IQR (third quartile – first quartile). Individual dots beyond the solid lines are considered outliers in the data. Daily box plots are consistent over time, with a median around 60,000 lbs. There is no outlier in the data set.

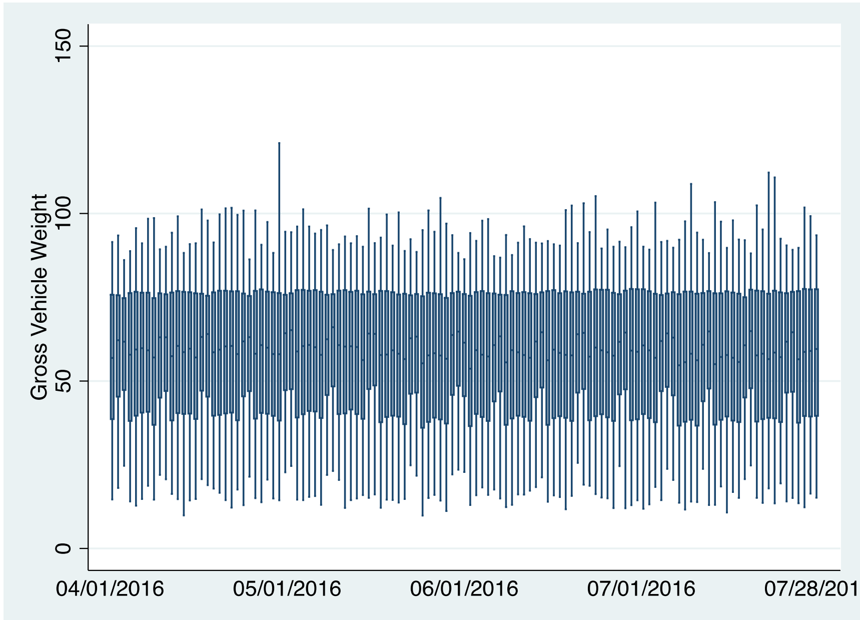

Figure 4.6 Daily Box-plot of GVW - the Kistler Class-9 Data

Similarly, Figure 4.7 shows the box-plot of GVW for Class 9 without errors. Compared to the distributions of all class 9 weights, the distributions without error have a slightly increasing trend over time. The median of GVW is around 70,000 lbs, which is higher than the mean of GVW with errors. This indicates that the error-coding algorithm in the Kistler system may be eliminating some low weights and that may be unreasonable.

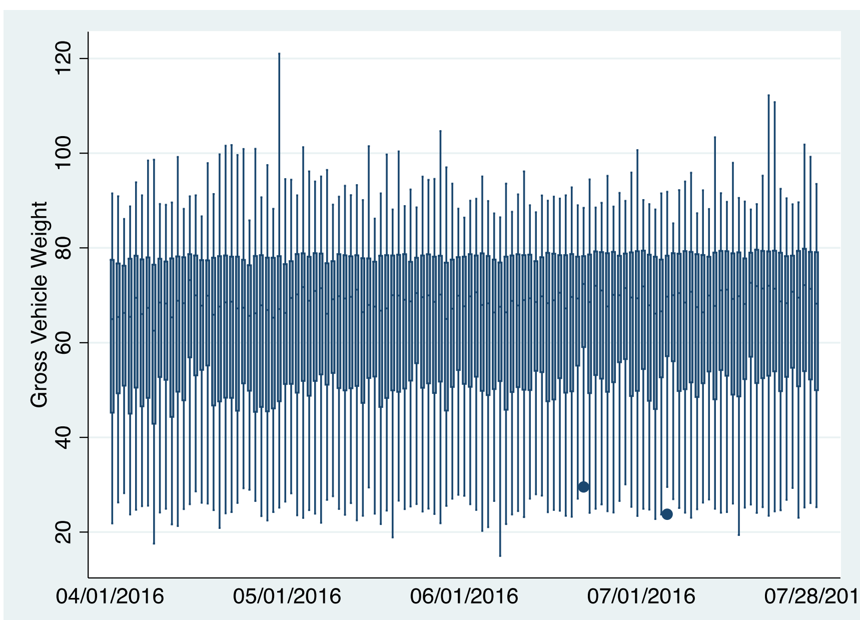

Figure 4.7 Daily Box-plot of GVW – the Kistler Class-9 Data without Error

In addition to daily distributions, we also considered monthly distributions. Figure 4.8 shows the monthly comparison of GVW distributions for Class 9. Distributions of all four months are quite consistent, with two peaks at about 30,000 lbs. and 75,000 lbs.

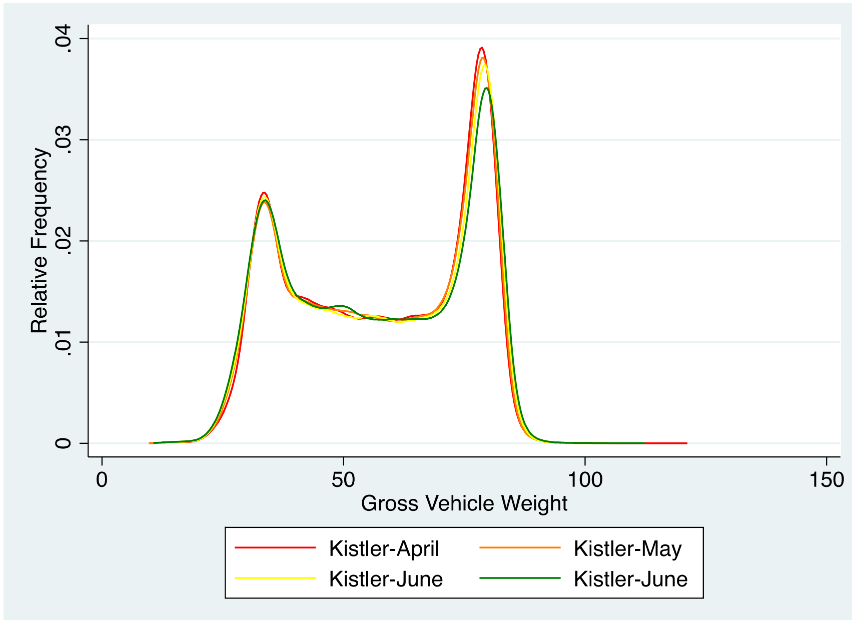

Figure 4.8 Month-to-month Comparison of Class 9 GVW - the Kistler System

To test the distributional differences statistically, we use Kolmogorov-Smirnov tests for monthly data. The Kolmogorov-Smirnov test is a nonparametric test of the equality of continuous distributions. It quantifies a distance between the empirical distribution functions of two data samples. The test statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. The quantified distance is called *deviance.* It is a numerical value that represents the difference in distributions of the two samples. We recommend that deviance be used to as an additional criterion (in addition to mean GVW) to identify system failures and recalibration times. Deviance is more appropriate than the mean because it captures distributional change in GVW measurements.

Table 4.12 shows the Kolmogorov-Smirnov tests of the Kistler Class-9 GVWs for different months. Only consecutive months are compared to check whether GVW distributions change significantly from one month to the next. The distributions of consecutive months are statistically different with all *p*-values less than 0.05. However, the deviances of each pair of months are not large. In particular, the deviance of April versus May is 0.0099, which means there is only 0.99% shift of the GVW distributions between the two months. This means that the GVW distributions are practically not different, although with the large amount of data available for each month, they are found to be statistically different.

Table 4.12 Monthly Two-sample Kolmogorov-Smirnov Test of GVW - the Kistler System

Month	Deviance	p-value
April versus May	0.0099	0.002
May versus June	0.0165	0.000
June versus July	0.0131	0.000

4.2.3 Speed

Speed is another important performance metric both for traffic monitoring purposes and because the speed affects the accuracy of vehicle-class detection, axle spacing detection and error code generation. Because there is no speed limit change from April to July, our null hypothesis is that Class-9 speed distribution remains the same during this period of time. What we mean by that statement is that although individual vehicle speeds vary significantly, the overall speed pattern (across all vehicles) remains the same over time.

Similar to gross vehicle weight, time consistency of mean, variance and distribution of speed are considered in the sequel.

4.2.3.1 Mean of Speed

Figure 4.9 shows the daily average speed of Class-9 vehicles with and without errors. For most of the days, the mean of speed ranges from 62 mpg to 66 mpg, with some exceptions when speed is much lower. The low average speeds may be due to traffic jams (road work, accidents, etc.) on certain days. Except for those unusual days, the pattern of average speed of the Kistler system is considered consistent over time. Moreover, the average speed without errors is almost the same as with error average speed.

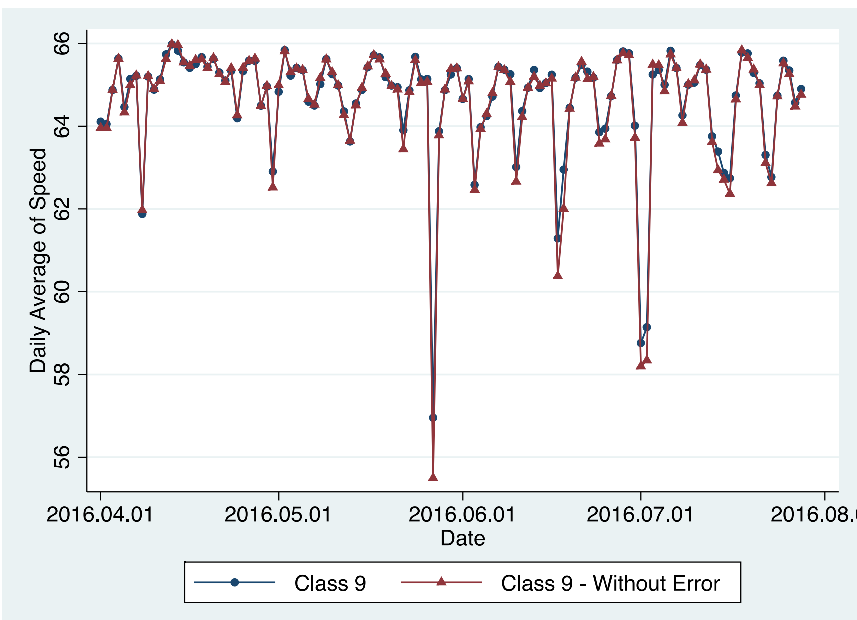

Figure 4.9 Mean of Speed - the Kistler Class-9 Data

4.2.3.2 Variance of Speed

While the variance of speed is another important determinant of system consistency, it is more likely to be impacted by the traffic itself rather than the accuracy of the sensor. For example, weekdays may have higher speed variance than weekends because weekdays are likely to have varying levels of traffic congestion at different times of the day. Therefore, the interpretation of the comparison of variances must be done more carefully.

Figure 4.10 shows daily standard deviations of speed for Class-9 vehicles with or without errors. Speed standard deviation has a large range for Class-9-with-error data from 3 mph to 12 mph, but becomes less consistent in over time. Similar to the mean of speed, the standard deviation of speed of Class-9-without-error data is very close to Class-9-with-error data.

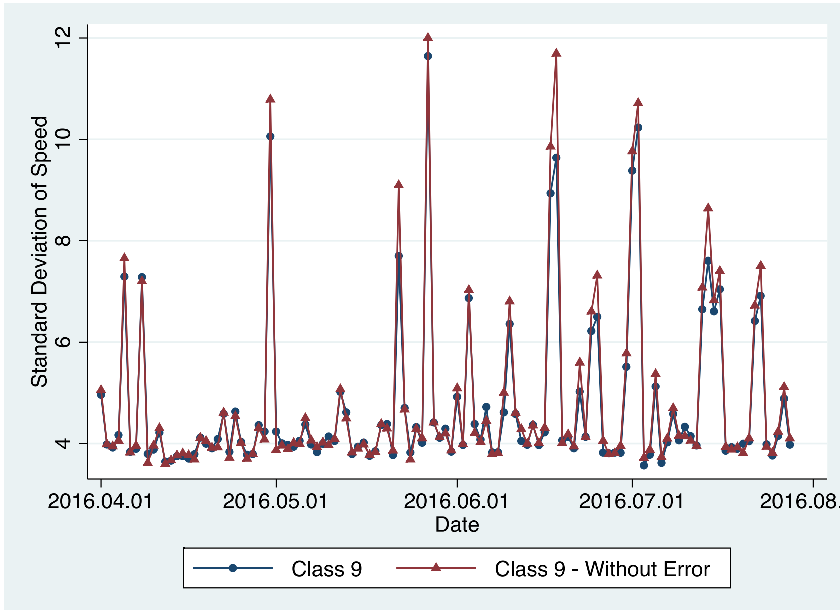

Figure 4.10 Standard Deviation of Speed - the Kistler Class-9 Data

4.2.3.3 Distribution of Speed

Similar to gross vehicle weight, we compare distributions of speed for different months. Figure 4.11 shows the monthly distributional comparison of speed. It is clear that all four months seem consistent.

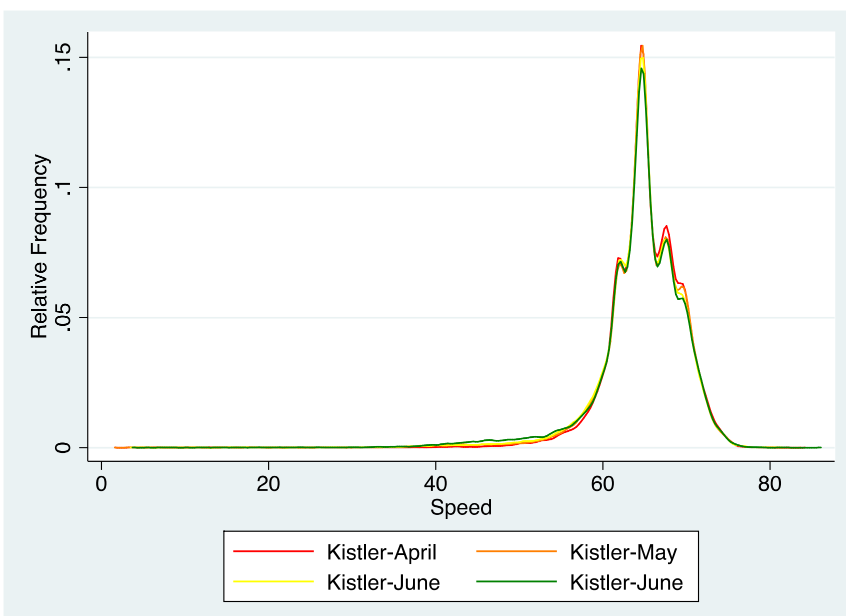

Figure 4.11 Monthly Distribution of Speed – the Kistler Class-9 Data

To test the distribution difference statistically, we again use the Kolmogorov-Smirnov test; results are shown in Table 4.13. The *p*-values of all tests are less than the significance level (0.05), which means the distributions of speed are consider statistically different. However, the deviances of between the distributions are quite small. All distributional shifts are within 2% for any two months. Therefore, we conclude that the distributions are not practically different.

Table 4.13 Monthly Two-sample Kolmogorov-Smirnov Test of Speed - the Kistler System

Month	Deviance	p-value
April versus May	0.0178	0.000
May versus June	0.0125	0.000
June versus July	0.0166	0.000

To summarize, we compared the self-consistency of the Kistler system in terms of daily observation counts, gross vehicle weights and speeds. All three measurements indicate that the Kistler system performs consistently over the given period. Moderate amount of errors are generated. Most of these errors have low GVW. This may need further investigation because errors do not appear to be unbiased. Monthly distributions of GVW and speed are considered consistent and deviance is recommended as a metric of distributional shifts.

4.3 SELF-CONSISTENCY ANALYSIS: THE IRD SYSTEM

In this Section, we repeat the Section 4.2 analyses for the IRD-system data. That is, the mean, the variance and the distribution of daily observation counts, gross vehicle weights and speeds are analyzed. Valid IRD-system data are from October 1st to July 28th. Note that the two systems have different sizes of data. This is not a problem so long as we compare each system with itself, which is what we do in this section.

Our findings in terms of daily observation count, gross vehicle weight and speed indicate that the IRD-system data is self-consistent over the 10 months.

4.3.1 Daily Observation Count

We first analyze daily observation count of all classes, which is shown in Figure 4.12. Daily observation counts range from 10,000 to 20,000 with a slight decrease from October to the end of January and a clear increase from February. This appears to be a reasonable reflection of real traffic because of the weather conditions in Minnesota. There are no sudden changes and no reason to suspect an inconsistency.

The observation count without error has a pattern that is similar to that calculated from the data with error, which suggests that the errors as percent of total vehicle counts remain stable over the 10 months.

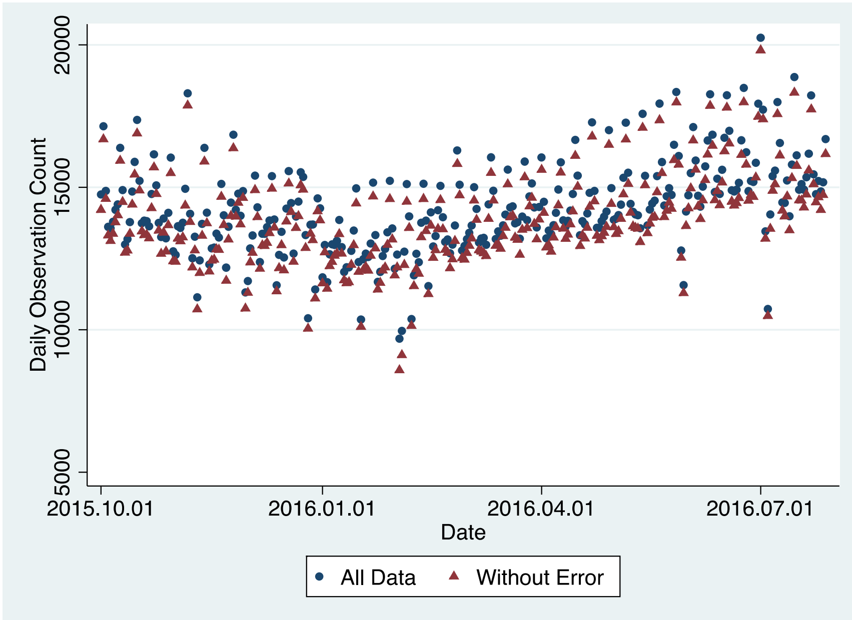

Figure 4.12 Scatter-plot of Daily Observation Count - the IRD system

Class-9 daily observation counts are shown in Figure 4.13. There is a pattern of daily counts with some higher counts (≈ 3,000) and some lower counts (≈ 1,100). This is due to the day-of-week effect. Class-9 counts are lower during weekends. There is also a drop in counts surrounding the Christmas and the New-Year's day, which also seems reasonable. Class-9 without-error counts are similar to the data with errors and are generally consistent over time.

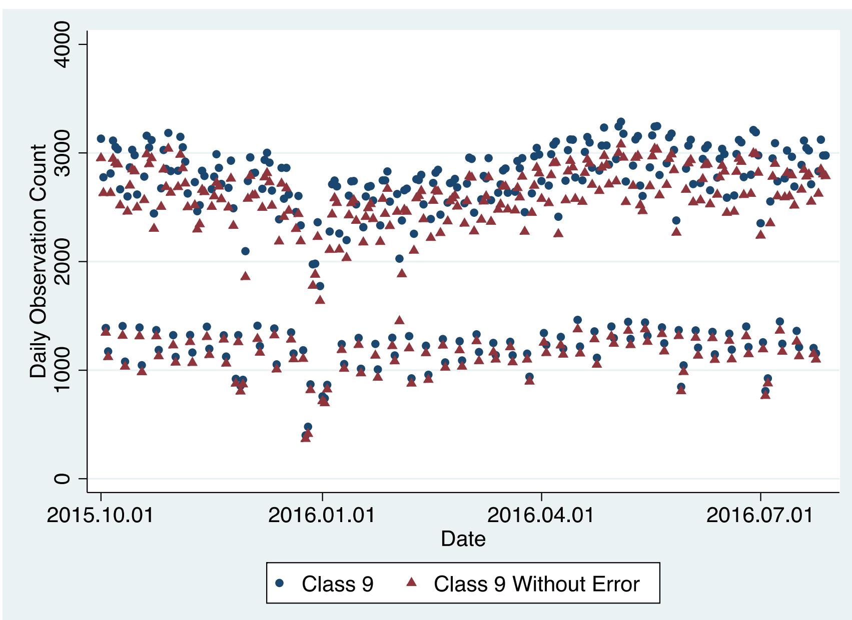

Figure 4.13 Scatter-plot of Daily Observation Count - the IRD system Class 9

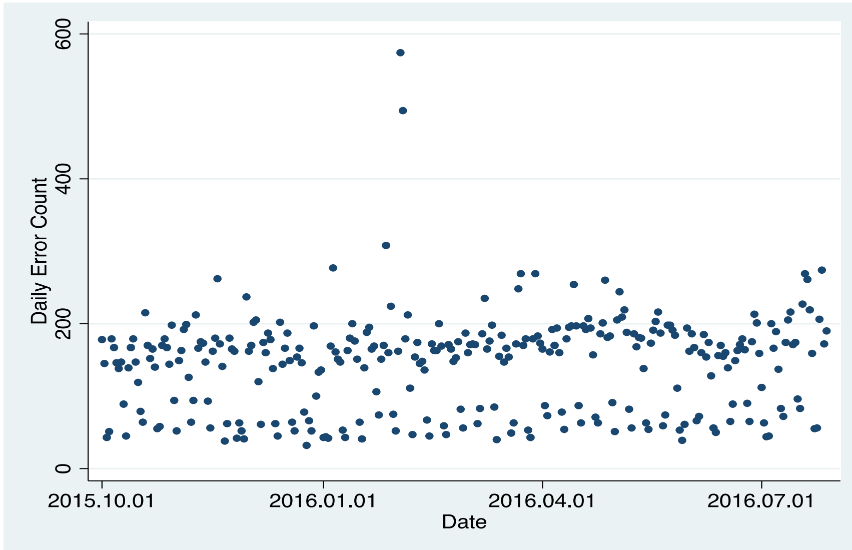

Figure 4.14 Scatter-plot of Daily Error Count - the IRD system Class 9

Figure 4.14 shows the daily error count of the IRD system Class-9 vehicle. Although there are a few outliers, the pattern of counts appears to be consistent over time. There is also a clear weekday and weekend effect because the total vehicle counts are different.

In general, after considering the effect of system errors and the effect of day of week, daily observation counts of the IRD system are self-consistent from October to July.

4.3.2 Gross Vehicle Weight

To analyze consistency of gross vehicle weight of the IRD system, we consider mean of GVW, standard deviation of GVW and distribution of GVW. Because Class 9 is the most important vehicle class for MNDOT, all the analyses for gross vehicle weight and speed are for Class-9 data only.

4.3.2.1 Mean of Gross Vehicle Weight

Daily average mean of GVW of Class 9 vehicles with and without errors is shown in Figure 4.15. Means without errors are very close to those with errors because 97% of the data are error free for the IRD system. There are some fluctuations in mean GVW observations, but there are no sudden significant changes or changing patterns during the observation period.

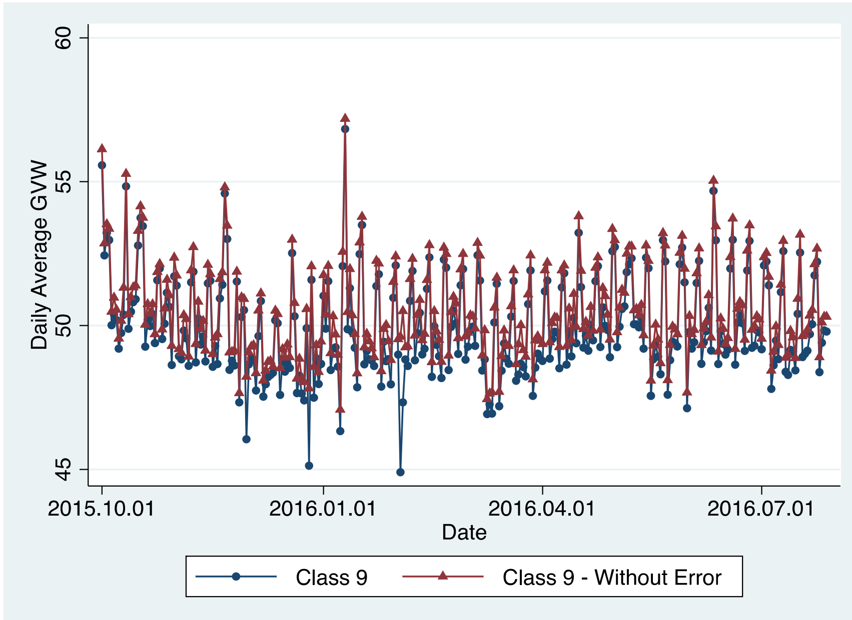

Figure 4.15 Mean of GVW - the IRD system Class 9

4.3.2.2 Variance of Gross Vehicle Weight

Figure 4.16 shows the standard deviation of GVW for Class 9. The without-error data have a slightly lower standard deviation, but are considered very close to with-error data. This is not surprising because 97% of the IRD-system data are error free. That being said, the differences between with-error and without-error data are small. They both have the same day-of-week patterns. For this reason, we conclude that the variance of GVW for Class 9 is consistent over time.

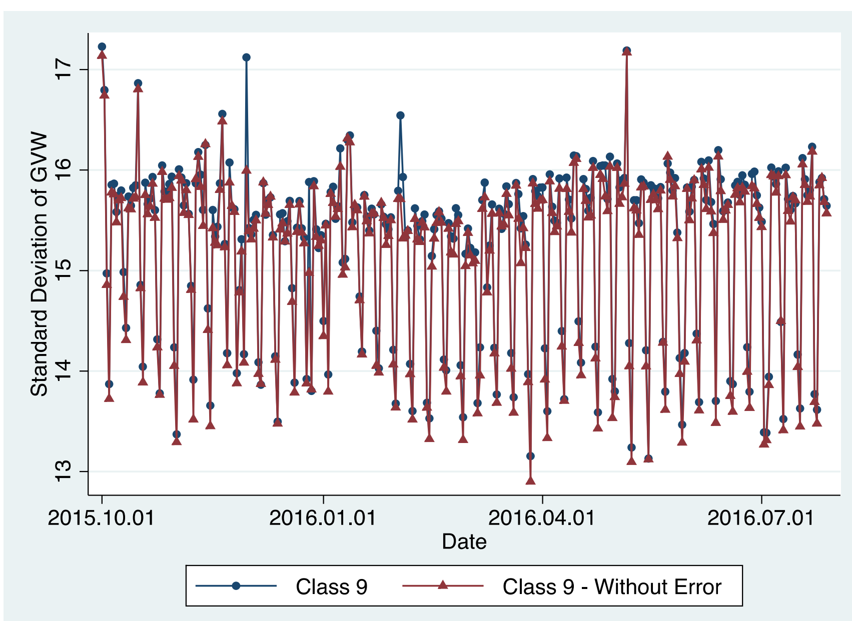

Figure 4.16 Standard Deviation of GVW - the IRD system Class 9

4.3.2.3 Distribution of Gross Vehicle Weight

To analyze the consistency of the distributions of GVW, we look at box-plots of daily Class-9 GVW as well as monthly distributions. Statistical tests of distribution equality are performed for monthly comparisons.

Figure 4.17 gives daily box-plots of Class-9 GVW. The box-plots are consistent over time. There is no significant change in patterns. The median of GVW is around 50,000 lbs. The range of first and third quartiles is from 40,000 lbs. to 75,000 lbs. Similar to the Kistler system, the IRD system has few outliers in box-plots.

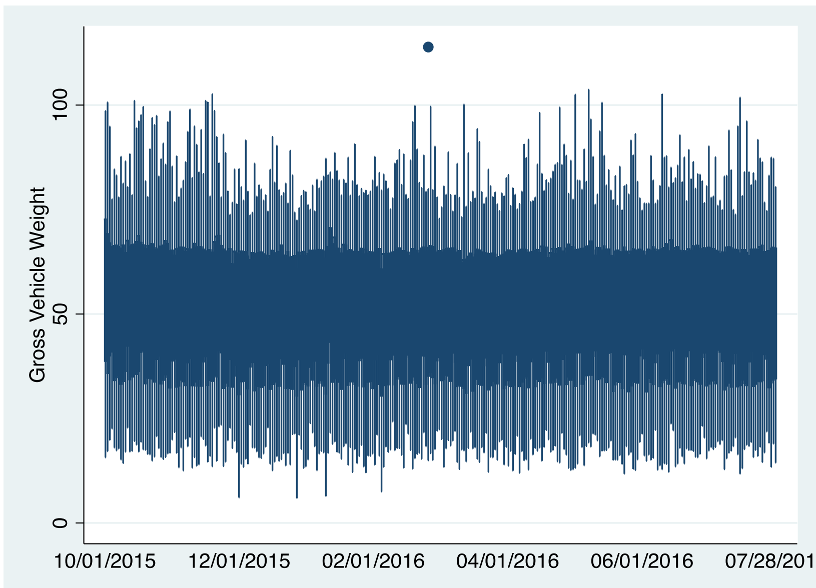

Figure 4.17 Daily Box-plot of GVW - the IRD system Class 9

Figure 4.18 shows the daily box-plots of Class 9 without error. The box-plots are very similar to Figure 4.17 except that there are no outliers.

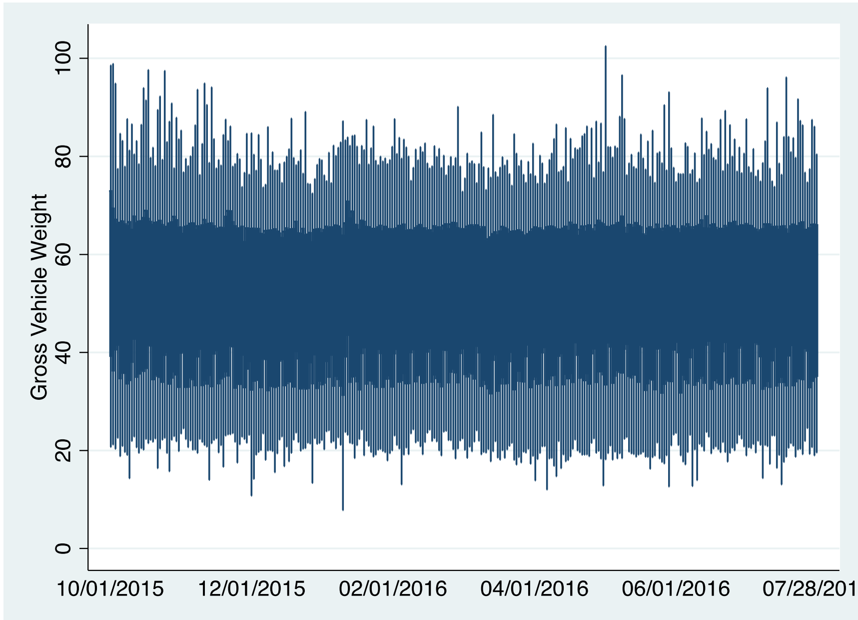

Figure 4.18 Daily Box-plot of GVW – the IRD system Class 9 without Error

Monthly distributional comparison of Class-9 GVW is given in Figure 4.19. The relative frequency distributions for each month are very close to each other.

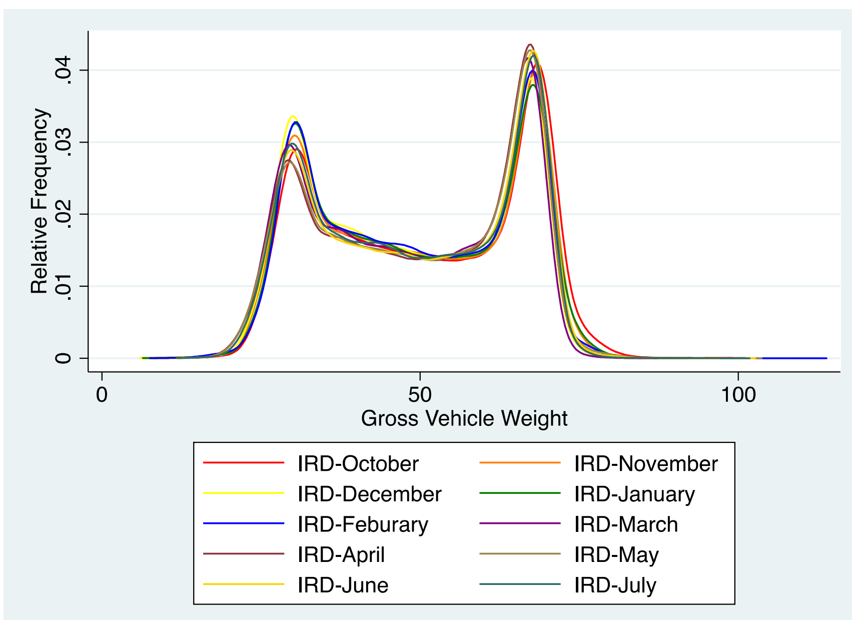

Figure 4.19 Month-to-month Comparison of Class 9 GVW – the IRD system

Table 4.14 shows the Kolmogorov-Smirnov tests of GVW measurements reported by the IRD sensor for Class-9 vehicles for different months. Only consecutive months are compared to see whether the patterns change from one month to the next. Except the distributions of April and May (with a p-value of 0.169), the *p*-values of all other pairs of months are less than 0.05. Therefore, the distributions of GVW for April and May are statistically the same. All other distributional comparisons over two consecutive months are statistically different. However, the absolute value of the deviance is very small (see Table 4.14, all within 5%) and that permits us to conclude that although the distributions are statistically different because we have a large amount of data for each month, they are not practically different. That is, the IRD-system data are consistent across different months.

Table 4.14 Monthly Two-sample Kolmogorov-Smirnov Test of GVW - the IRD system

Month	Deviance	p-value
October versus November	0.0356	0.000
November versus December	0.0432	0.000
December versus January	0.0333	0.000
January versus February	0.0168	0.000
February versus March	0.0413	0.000
March versus April	0.0319	0.000
April versus May	0.0059	0.169
May versus June	0.0208	0.000
June versus July	0.0141	0.000

4.3.3 Speed

In this section, the mean, standard deviation and distribution of speed are analyzed for Class-9 vehicles.

4.3.3.1 Mean of Speed

Figure 4.20 shows the daily average speed for Class-9 vehicles. The difference between data with error and data without error is small. They both have consistent average speeds except for some days that have lower speeds, which may be attributed to weather and traffic congestion.

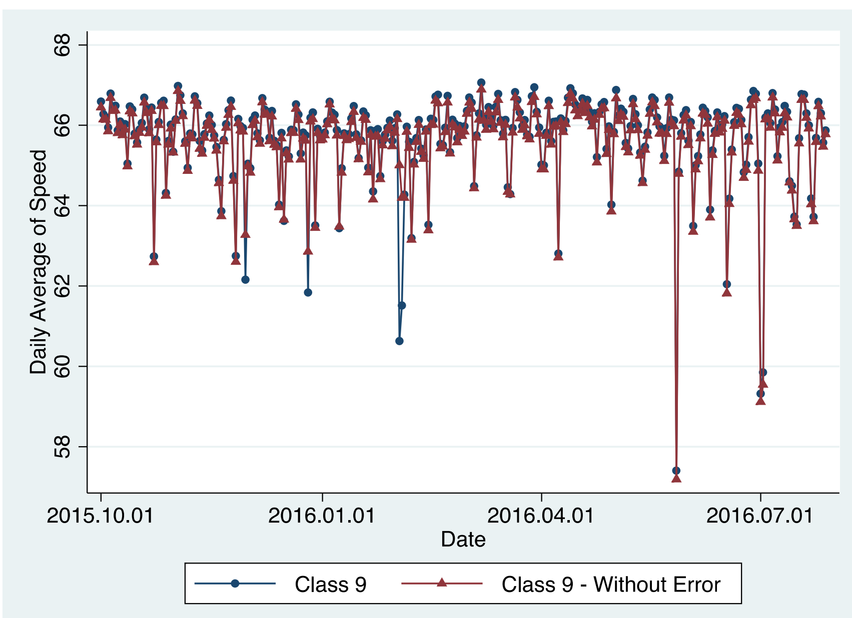

Figure 4.20 Mean of Speed - the IRD system Class 9

4.3.3.2 Variance of Speed

Standard deviation is used to measure the variability of speed for Class 9 vehicle. Speed data without error has a much smaller standard deviation than data with error. Some days have very large standard deviations, but those days are evenly spread out across time. Therefore, we may conclude that both with and without error data have consistent speed variability patterns over time.

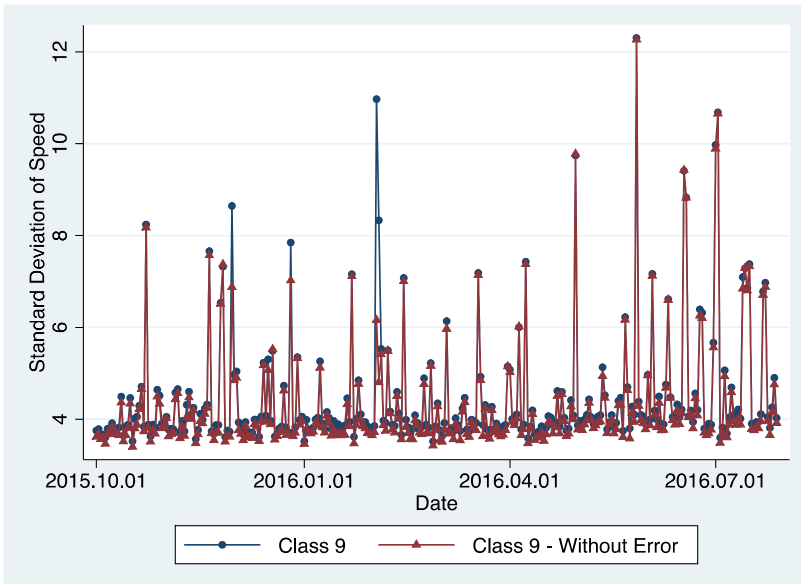

Figure 4.21 Standard Deviation of Speed - the IRD system Class 9

4.3.3.3 Distribution of Speed

Similar to the analyses of speed distribution of the Kistler system, we perform distributional analyses of speed by month for the IRD system. As shown in Figure 4.22, speed distributions of IRD are consistent from October 2015 to July 2016. Statistical results of Kolmogorov-Smirnov test are shown in Table 4.15. With *p*-values of 0.074 and 0.36, the distributional differences are not statistically significant for December versus January, and January versus February. Although all other *p*-values are significant for distributional difference, the absolute values of deviances for each comparison are quite small.

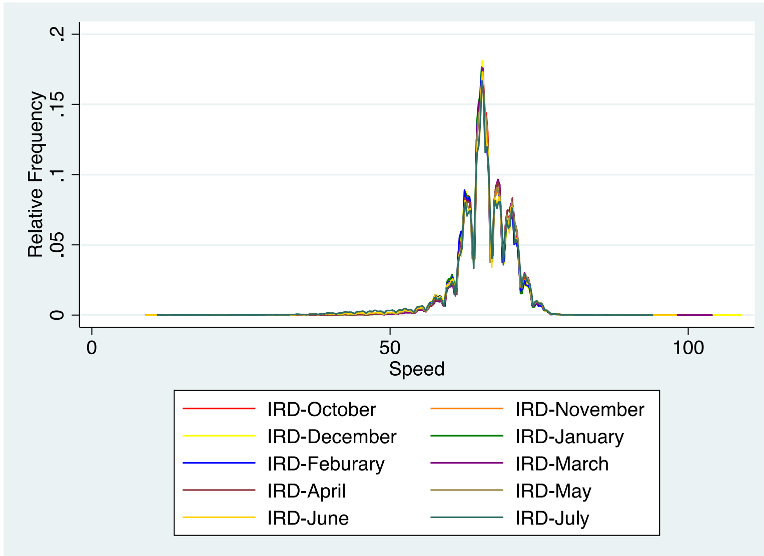

Figure 4.22 Monthly Distribution of Speed - the IRD system Class 9

Table 4.15 Monthly Two-sample Kolmogorov-Smirnov Test of Speed - the IRD system

Month	Deviance	p-value
October versus November	0.0199	0.000
November versus December	0.0195	0.000
December versus January	0.0074	0.074
January versus February	0.0053	0.369
February versus March	0.0499	0.000
March versus April	0.0094	0.004
April versus May	0.0177	0.000
May versus June	0.0119	0.000
June versus July	0.0173	0.000

In summary, similar to the Kistler system, the IRD system is considered consistent in terms of the distributions of daily observation counts, gross vehicle weights and speed distributions. To further evaluate whether the Kistler system and the IRD system have similar measurements, we compare the two systems in Section 4.4

4.4 COMPARISON OF THE KISTLER AND THE IRD SYSTEMS

The Kistler and the IRD sensors are installed near each other and there is no entry to or exit from the highway between the locations of the two sensor systems, which means that they approximately measure the same vehicles during the same time interval. Some differences could come from vehicles changing lanes in between these two sensors. Note that the Kistler WIM system is downstream from the IRD WIM system. We assume that the number of vehicles changing lanes is not significant and proceed to compare their reported observations for the same time interval. Recall that the valid Kistler-system data starts on April 1st, 2016 and ends July 28th. It is a proper subset of dates for which we have the IRD-system data. Therefore, we use only the April 1st to July 28th data for both systems in these comparisons. Our analysis does not indicate which system is more accurate. We can determine only whether their observations are similar or not. Accuracy comparisons will be possible if we have true vehicle weights.

4.4.1 Daily Observation Count

Because the two sensor systems are located close to each other, the actual number of vehicles passing through the two sensors during the same time interval should be nearly identical. Figure 4.23 shows the daily observation counts of all classes for the two sensors. It is clear that the counts of the two systems are quite close to each other. There is a slightly increasing trend of both the Kistler and the IRD systems.

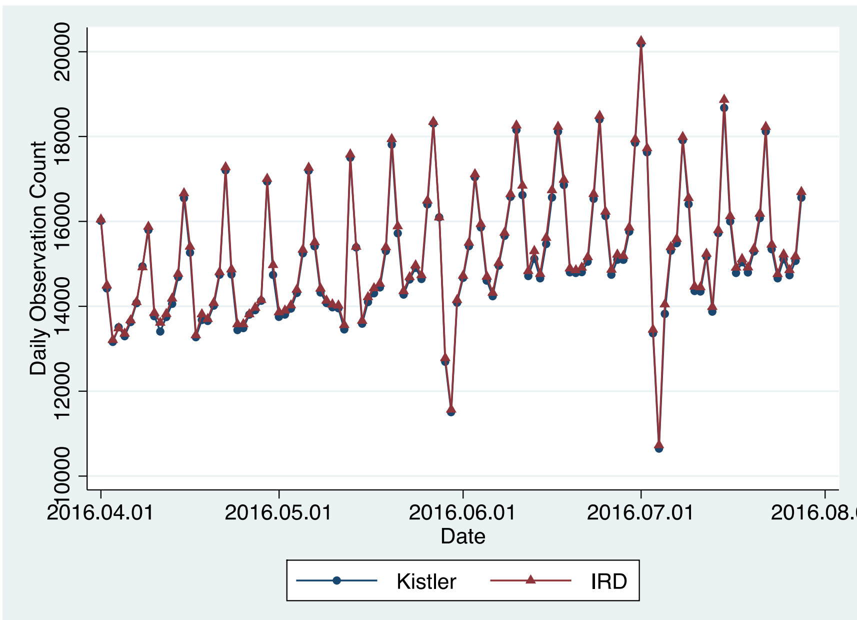

Figure 4.23 Comparison of Daily Observation Count - All Data

Figure 4.24 shows the comparison of the two sensors without error for all vehicle classes. Because the Kistler system generates more errors than the IRD system, the error-free daily observation count of the Kistler system is smaller than that of the IRD system. Similar to with-error counts, there is a slightly increasing trend for both the Kistler and the IRD systems.

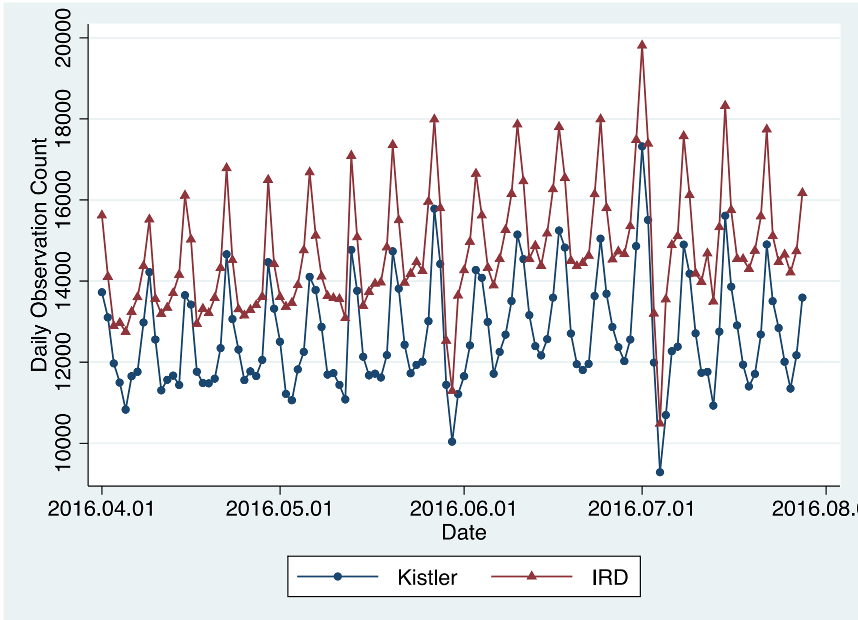

Figure 4.24 Comparison of Daily Observation Count - All Classes without Error

Figure 4.25 shows daily observation count comparison of Class-9 vehicles. Similar to the all-class data, daily observation counts of the Kistler and the IRD systems are quite close to each other.

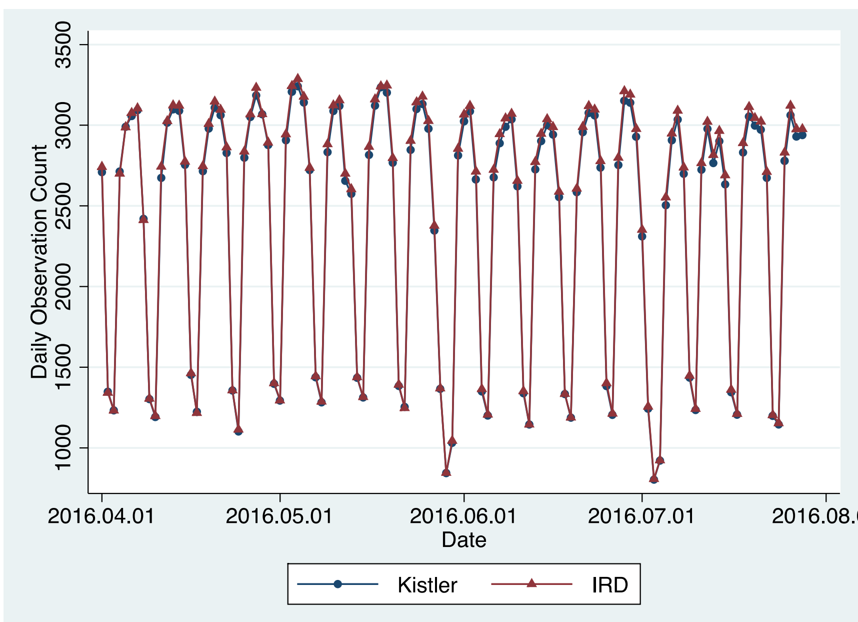

Figure 4.25 Comparison of Daily Observation Count - Class 9

Figure 4.26 shows the daily observation counts of Class-9 vehicle without error. Because the Kistler system generates more errors than the IRD system, the daily observation count of Class-9 vehicle without error for the Kistler system is less than that of the IRD system. Moreover, while the daily observation count of the IRD system is quite consistent over time, the Class-9 counts from the Kistler system exhibit a slightly decreasing trend, which indicates increasing number of observations with errors over time. Both systems have similar weekly patterns.

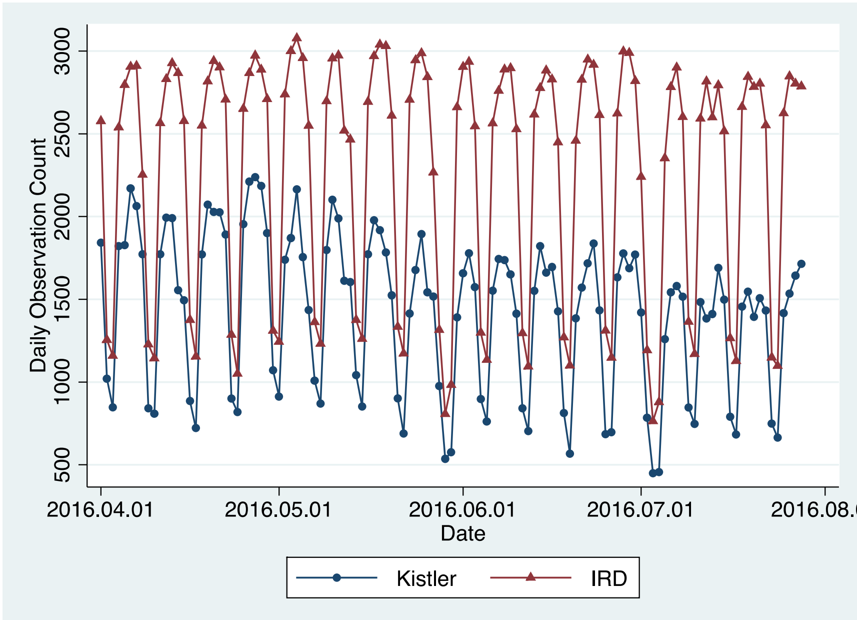

Figure 4.26 Comparison of Daily Observation Count - Class 9 without Error

4.4.2 Gross Vehicle Weight

4.4.2.1 Mean of Gross Vehicle Weight

To test the equality of the mean of gross vehicle weight of the two sensors, we use two-sample T tests, which are shown in Table 4.16. The second column is the mean difference between the GVWs measured by the IRD system and the Kistler system. The third column is the T statistic calculated for each category. The null hypothesis is that the means of GVW for the two systems are same. Column 4 gives *p*-values of different alternative hypotheses.

For Class-9 data, with or without error, the mean of GVW for the IRD system is statistically smaller than that for the Kistler system.

Table 4.16 Two-sample T test of Mean of GVW

GVW	diff=mean(IRD)-mean(Kistler)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	-8.11	-0.018	0.00	0.00	1.00
Class 9 Without Error	-13.36	-0.027	0.00	0.00	1.00

4.4.2.2 Variance of Gross Vehicle Weight

We use F tests to check if the two sensors have same GVW variability - results are shown in Table 4.17. The second column is the ratio of the standard deviations of the IRD system and the Kistler system GVW measurements. The third column is the F statistic calculated for each category. The null hypothesis is that the ratio of the standard deviation of GVW for the two sensors is 1. *P*-values of different alternative hypothesis are given in the fourth column.

Both for with- and without-error data, the IRD system GVWs are less variable than those of the Kistler system.

Table 4.17 F Test of Variance Ratio of GVW

	ratio=sd(IRD)/sd(Kistler)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	0.85	0.72	0.00	0.00	1.00
Class 9 Without Error	0.94	0.88	0.00	0.00	1.00

4.4.2.3 Distribution of Gross Vehicle Weight

To compare the distribution of GVW for Class-9 vehicles, we use cumulative distribution of GVWs for each system. As shown in Figure 4.27 and Figure 4.28, for both with- and without- error data, GVW cumulative distribution of Class 9 vehicle of the Kistler system is stochastically larger than that of the IRD system. This indicates that the Kistler system generates stochastically larger GVW than the IRD system.

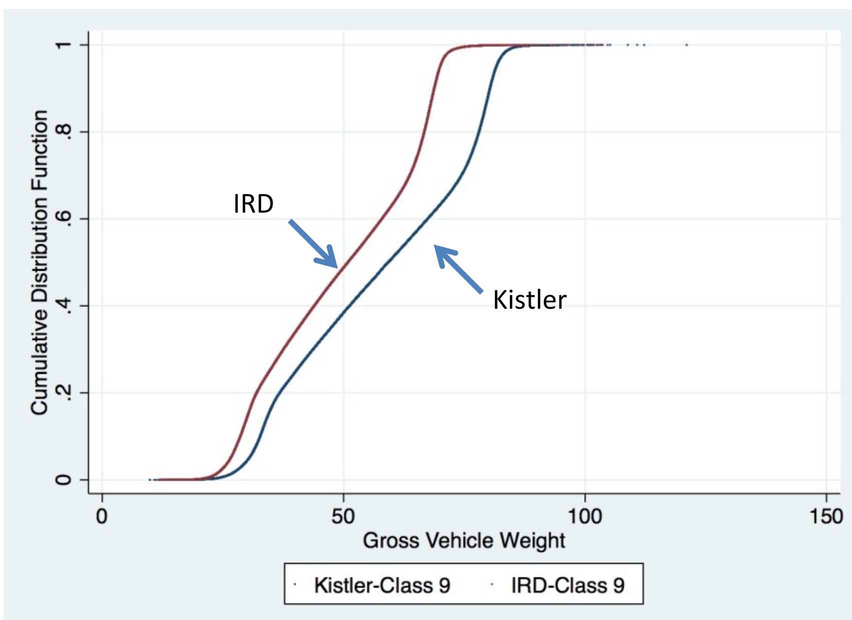

Figure 4.27 CDF of GVW - Class 9

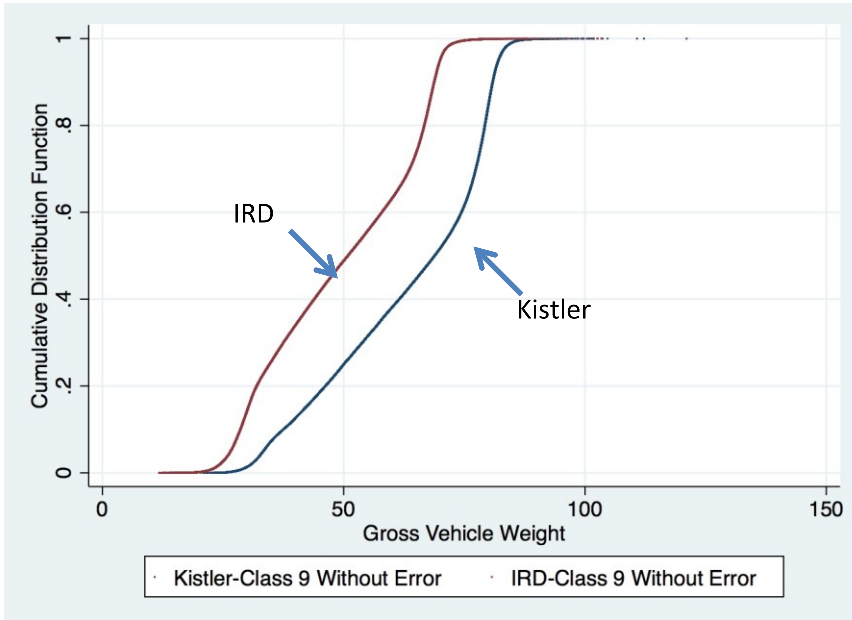

Figure 4.28 CDF of GVW - Class 9 without Error

To test the distribution differences statistically, we also perform Kolmogorov-Smirnov tests of Class 9 GVW for the two sensors, as shown in Table 4.18. With all *p*-values of 0, we conclude that the distributions of GVW for the two systems are not statistically the same for both with- and without error data. Moreover, the deviances of the two distributions are relatively large (0.33 and 0.44), which means there is more than 30% in difference for the two distributions.

Table 4.18 Two-sample Kolmogorov-Smirnov Test of GVW

	Deviance	**p-value**
Class 9	0.3284	0.000
Class 9 Without Error	0.4407	0.000

4.4.3 Speed

In what follows, we present comparisons of mean, variance and distributions of speeds recorded by the two sensors. Our approach is similar to that in Sections 4.4.1 and 4.4.2 . Therefore, we do not explain our methodology. Instead, we summarize our findings first and then present the results of our analysis in tables and figures in the ensuing subsections.

The speed distributions of the two systems are not the same. The IRD system in general has a larger mean of speed than the Kistler system. For variance, the IRD system has a larger variance for with-error data than the Kistler system. But for without-error data, the opposite holds.

4.4.3.1 Mean of Speed

The means of speed for the Kistler system and the IRD system are statistically different for Class-9 vehicles both with and without error. Results of the statistical tests are shown in Table 4.19. Mean of speed of the IRD system is statistically larger than the Kistler system.

Table 4.19 Two-sample T test of Mean of Speed

Speed	diff=mean(IRD)-mean(Kistler)	T statistic	p-value of Ha		
			diff < 0	diff != 0	diff > 0
Class 9	0.98	73.30	1.00	0.00	0.00
Class 9 Without Error	0.94	59.88	1.00	0.00	0.00

4.4.3.2 Variance of Speed

The variances of speed of the two sensors are also statistically different, which is shown in Table 4.20. While the variances of the two systems are very close, the ratios of the standard deviations with or without error are materially different. For all Class-9 data, this ratio is greater than 1 (i.e. the IRD system has the larger SD), whereas for error-free Class-9 data, this ratio is less than 1 (i.e. the IRD system has lower SD). Both results are statistically significant.

Table 4.20 F Test of Variance Ratio of Speed

Speed	ratio=sd(IRD)/sd(Kistler)	F statistic	p-value of Ha		
			ratio < 1	ratio != 1	ratio > 1
Class 9	1.01	1.01	1.00	0.00	0.00
Class 9 Without Error	0.95	0.91	0.00	0.00	1.00

4.4.3.3 Distribution of Speed

Figure 4.29 shows the histogram of speed for each system under different data set. For all cases, the histograms seem different across sensors. To test the difference statistically, we use Kolmogorov-Smirnov tests, which are shown in Table 4.21. All the distributions of speed are significantly different between the two sensors. There is about 20% difference in distribution for the two systems. The differences may come from the sensors, as well as the algorithms their controllers use to convert raw data into measurements and tag certain observations as errors.

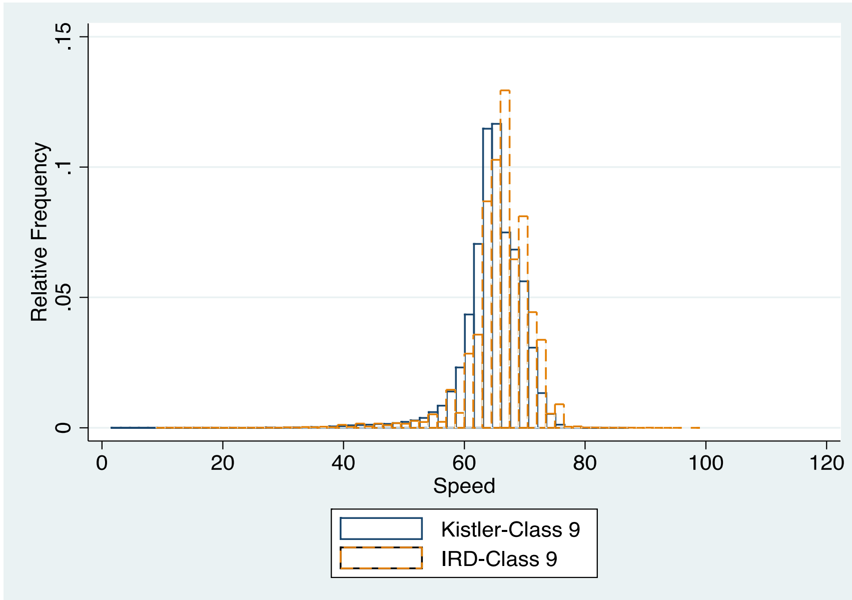

Figure 4.29 Comparison of Histogram of Speed - Class 9

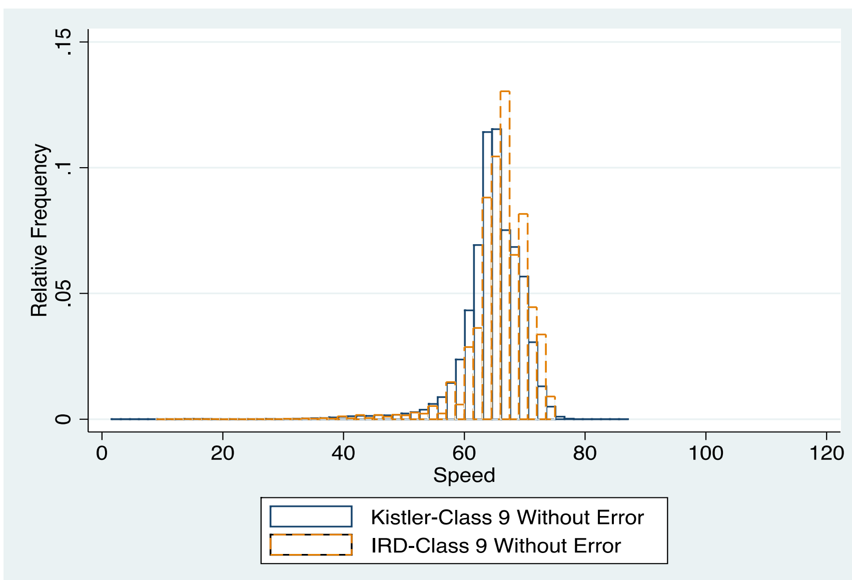

Figure 4.30 Comparison of Histogram of Speed - Class 9 without Error

Table 4.21 Two-sample Kolmogorov-Smirnov Test of Speed

Speed	Deviance	p-value
Class 9	0.2080	0.000
Class 9 Without Error	0.2046	0.000

In conclusion, in section 4.4 we compared the Kistler system with the IRD sensor in terms of daily observation count, gross vehicle weight and speed. Daily observation number is quite close for the two systems, except that the Kistler system generates more error codes than the IRD system. The two systems have significantly different distributions of GVW and speed. It is not possible to conclude which system is more accurate because of the lack of true GVW and speed data.

4.5 CONCLUSIONS

4.5.1 Findings

We have obtained the following main findings as a result of the analyses performed in this chapter:

5. The Kistler system appears to be self-consistent from April 1st to July 28th .

6. The IRD system appears to be self-consistent from October 1st to July 28th.

7. The Kistler system and the IRD system have similar daily observation counts, but statistically different GVW and speed distributions. Because we do not have the actual weights, speeds and vehicle counts, the accuracy of the two systems is not known.

8. The Kistler system generates more errors than the IRD system. It may be that either the IRD system is more reliable than the Kistler system, or that the Kistler system has a stronger capability of identifying errors.

4.5.2 Recommendations

Based on the analyses, we recommend actions that MNDOT can take both to improve performance evaluation and detect the need for recalibration.

For each WIM system, MNDOT will need to establish criteria to determine when recalibration is appropriate. We recommend that MNDOT use multiple criteria including daily observation counts, daily error percentages, daily mean and standard deviation of gross vehicle weights, daily mean and standard deviation of speeds, and monthly distributional drifts of gross vehicle weights and speed. MNDOT could set the recalibration threshold for each criterion (e.g, 5% change from baseline). Moreover, a linear model could be used to project the estimated time to next recalibration for each criterion. An example is given in Section 4.2.1 for error percentage. For other monthly-calculated criteria, e.g., deviance, the data were insufficient to trust results from our linear-regression model, although that model will provide a rough estimate of the time to the next recalibration.

Overall, a combination of linear-regression models to predict future recalibration time for each criterion, and threshold-based triggers for important metrics such as the mean GVW may work well.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

There are currently three Weigh-In-Motion (WIM) systems installed at MnROAD. One of these WIM systems (including sensor and controller) is manufactured by Intercomp, a Minnesota company headquartered in Medina. The second WIM system has sensors manufactured by Kistler and controllers manufactured by IRD. The Kistler WIM sensors use a piezo-quartz technology to determine the weight of the vehicle as it passes over the sensors, whereas the Intercomp WIM sensors are a strain gauge type of technology for determining the weight of the vehicle. The third system has Kistler sensors and controller. The Kistler sensors cost about $28,000 per lane in materials cost. For the purpose of this research study, the Intercomp sensors cost $21,340 per lane in materials cost. Installation cost is in the range of $15,000 to $18,000 per lane for both sensors. This research project aim to utilize statistical inference and comparison techniques to compare the performance of the Kistler and Intercomp WIM sensors. To this end, we perform three sets of analyses.

We first analyzed test-vehicle data to perform an accuracy check of the results that were reported by the sensor-vendor Intercomp. We performed this analysis in two formats – using the actual weights data that Intercomp used and using truck-ticket weights as actual weights. It was shown that, for the most part, the results reported by Intercomp match what we obtained from our analysis. With the limited test data available to perform the analysis, the performance of Intercomp sensor appeared to be similar to what was reported by Intercomp in its report. Our findings were somewhat different when we used truck-ticket weights rather than the weights used in the Intercomp report. The number of observations with different average daily temperatures and vehicle speeds were very small. Based on the existing data, any conclusions about the association between these factors and sensor accuracy were not reliable. Measurement errors depended on axle type. For some axles, the sensor overestimated weights, and for others it underestimated. However, the total weights appeared to be within tolerable bounds.

In Chapter 2, we analyzed limited data from the two sensor systems (Intercomp and IRD) to lay the foundation for analysis that would lead to the selection of a superior system in terms of better accuracy and lower life-cycle cost in the future. Our analysis included tests of self-consistency and comparisons of measurements provided by the two sensor systems. By examining vehicle counts, vehicle weights (especially Class-9 vehicles), and vehicle speeds, we concluded that Intercomp sensor data was not self-consistent during the period from Jan 22 to March 28. There is a sudden and significant difference between the patterns of daily observation count, GVW and speed before and after Feb 19. The inconsistency of Intercomp sensor was mainly from Lane 1, the driving lane. IRD sensor was considered self-consistent from October 1 to March 28. Intercomp and IRD had statistically different daily observation count, GVW distribution and speed distribution. Because we did not have the actual weights, speeds and vehicle counts, the accuracy of the two systems was not known. However, the lack of self-consistency suggested that the Intercomp sensor system stopped performing according to specifications around mid-February.

Based on the obtained results, we recommend actions that MNDOT can take both to improve performance evaluation and detect the need for recalibration. These actions to improve performance evaluation include: (1) Collect more data for both sensors, which will help shine light on the time between required recalibration of the two sensors; (2) Investigate the reasons for the significant change around Feb 19 for Intercomp; (3) Obtain more information of the error codes used by Intercomp so that we may separate errors (unusable data) and warnings (usable data). Specifically, it would be helpful to learn what thresholds are used to generate tags – good, warning (but usable), and bad data; (4) Perform more test runs so we can compare measured weights against true weights.

As for recalibration, MNDOT uses the mean of front axle weight to monitor GVW consistency over time. Recalibration threshold is plus or minus 5% of the baseline. We suggest using variance and deviance of the distribution of weights as additional criteria, in addition to the mean of front axle weight. Deviance measures the distribution difference of two samples of data. Such comparisons could be made biweekly or monthly to detect shifts in the distribution. Under normal circumstances, we would not expect the weight distributions to change much over time. Similarly, a comparison of variances could inform MNDOT if observations are beginning to exhibit a different pattern. Note that the mean weight (single axle or combined) may not change even though both the distribution and variance may change.

During the course of this project, the Intercomp sensor failed in late February 2016. Without Intercomp data, we had to rely on alternative data from Kistler to perform a third set of analyses. Specifically, we analyzed ten-month data from the IRD WIM system and four-month data from the Kistler WIM system to evaluate sensor accuracy in Chapter 3. We examined the self-consistency of the two systems and compared them in terms of daily observations and error counts, GVWs (especially for Class-9 vehicles) and speed. Our analyses indicated that both systems were self-consistent within the given time frame of each data set. However, the measurements provided by the two systems were different. The Kistler system and the IRD system had similar daily observation counts, but statistically different GVW and speed distributions. Because we did not have the actual weights, speeds and vehicle counts, the accuracy of the two systems remained unknown. The Kistler system generated more errors than the IRD system. It may be that either the IRD system was more reliable than the Kistler system, or that the Kistler system had a stronger capability of identifying errors. Without additional data, we were unable to conclude which system was more accurate.

Based on the analyses, we recommend actions that MNDOT can take both to improve performance evaluation and detect the need for recalibration. For each WIM system, MNDOT will need to establish criteria to determine when recalibration is appropriate. We recommended that MNDOT use multiple criteria including daily observation counts, daily error percentages, daily mean and standard deviation of GVWs, daily mean and standard deviation of speeds, and monthly distributional drifts of GVWs and speed. MNDOT could set the recalibration threshold for each criterion (e.g, 5% change from baseline). Moreover, a linear model could be used to project the estimated time to next recalibration for each criterion. For other monthly-calculated criteria, e.g., deviance, the data are insufficient to trust the results from our linear-regression model, although that model provides a rough estimate of the time to the next recalibration. Overall, a combination of linear-regression models to predict future recalibration time for each criterion, and threshold-based triggers for important metrics, such as the mean GVWs may work well.

REFERENCES

Kroll, K., Young, M., & Kroll, K. (2015). *Strain Gauge Strip Sensor for Precision Weigh-In-Motion.* Medina, MN: Intercomp.

Kwon, T. M. (2015). *BullConverter: User Manual.* Saint Paul, MN: Minnesota Department of Transportation.

Weather Underground. (n.d.). Retrieved 02 22, 2016, from Weather Underground: https://www.wunderground.com

APPENDIX A

COMPLETE RESULTS OF ANALYSES BASED ON TEST VEHICLE DATA WITHOUT VEHICLE NUMBER 2

In Section 2.2.3 we verified Intercomp's analyses based on the test vehicle data without vehicle number 2. However, in the interest of brevity, we present in that section only those analyses that are significantly different from our previous results. This appendix details the complete set of results obtained from the analyses.

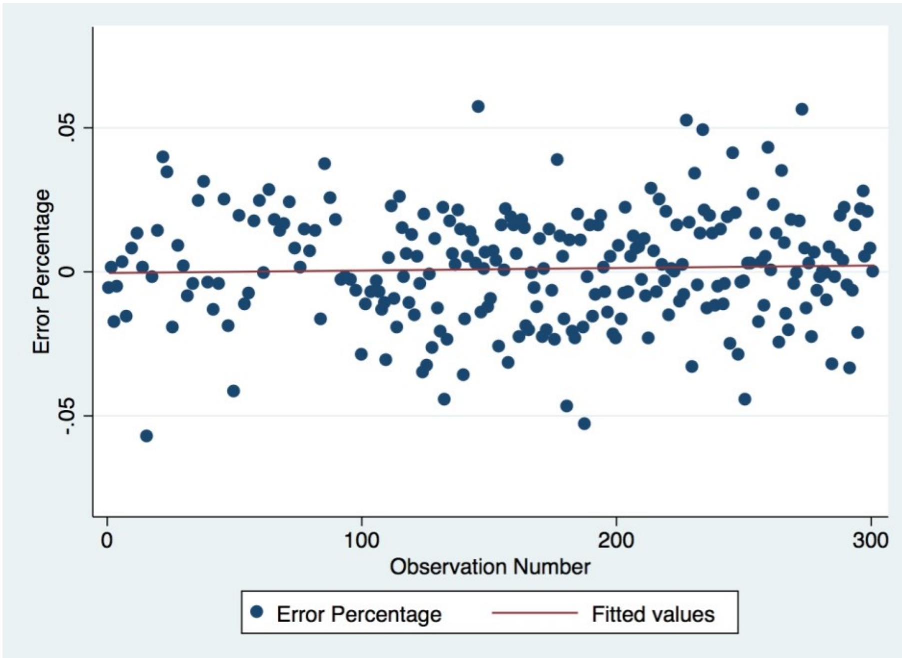

Figure A.1 Error Percentage Scatter-plot and Best-fit by Observation Number- Data Set 2

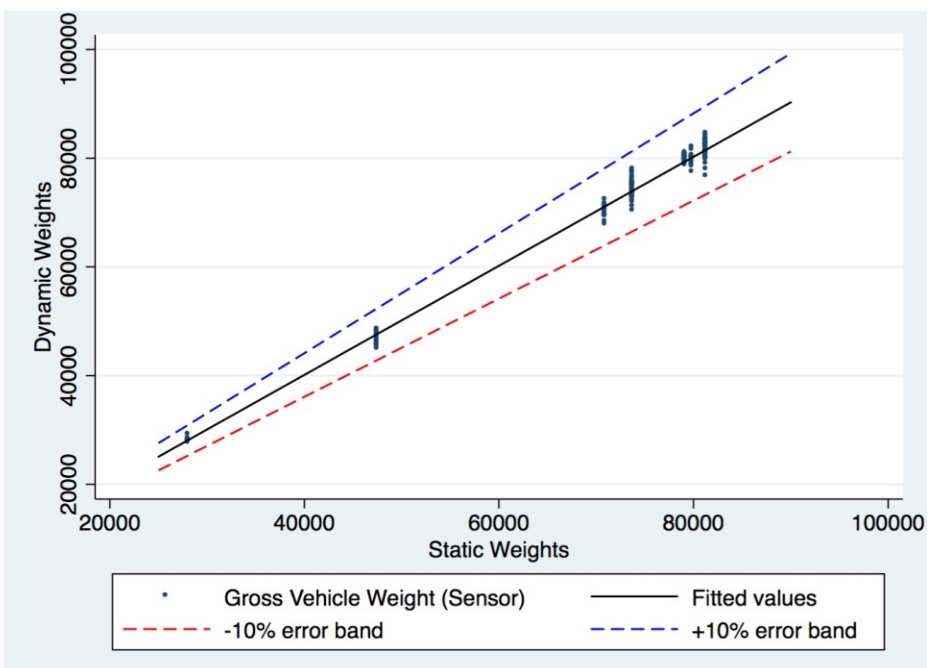

Figure A.2 Dynamic and Static Weights with Error Bands-Data Set 2

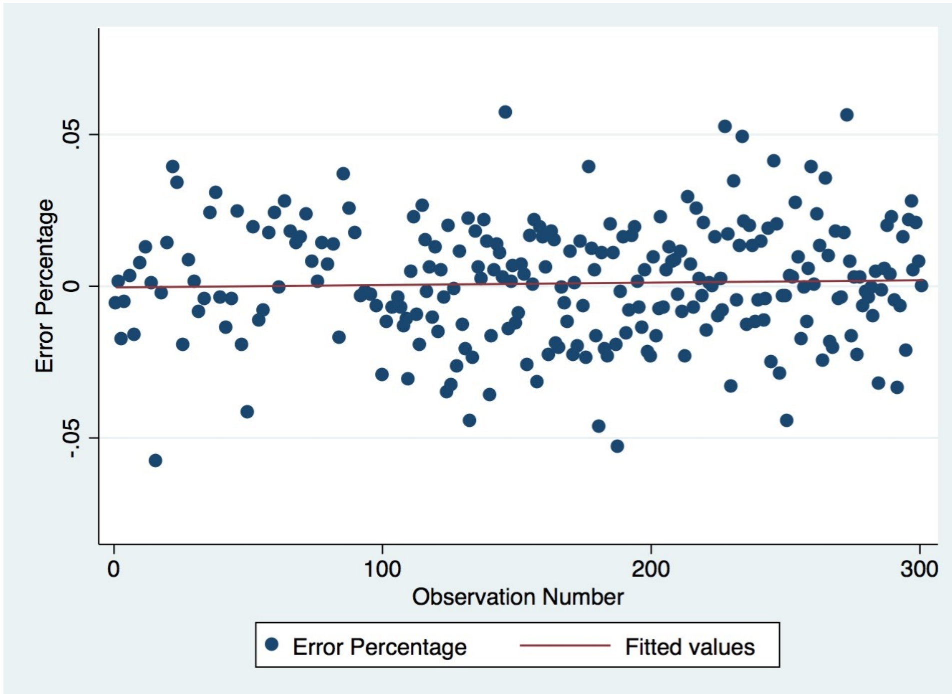

Figure A.3 Error Percentage by Observation Number – Truck Tickets

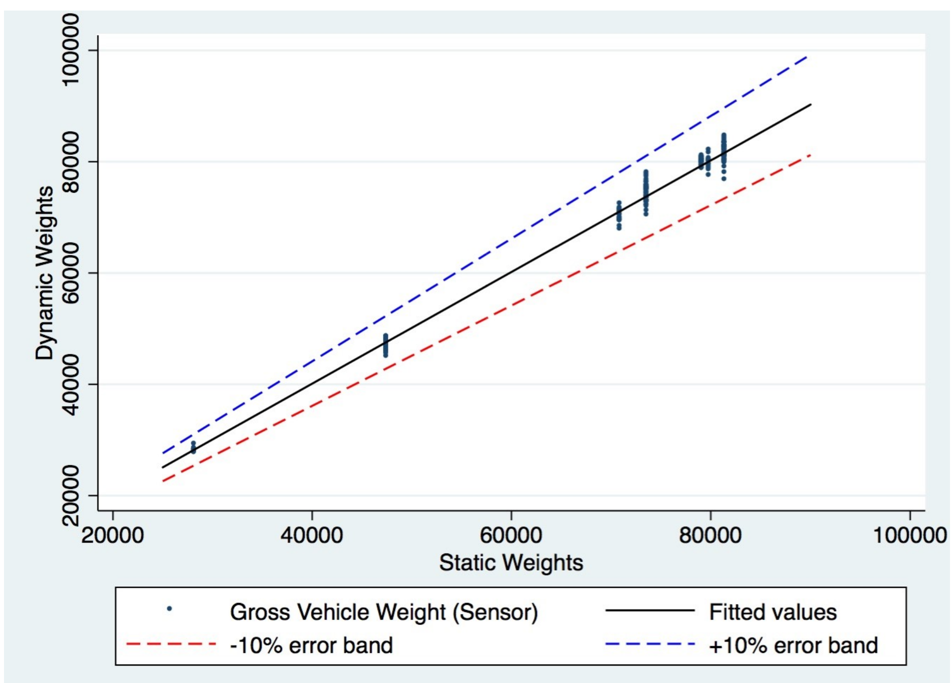

Figure A.4 Dynamic and Static Weights with Error Bands – Truck Tickets

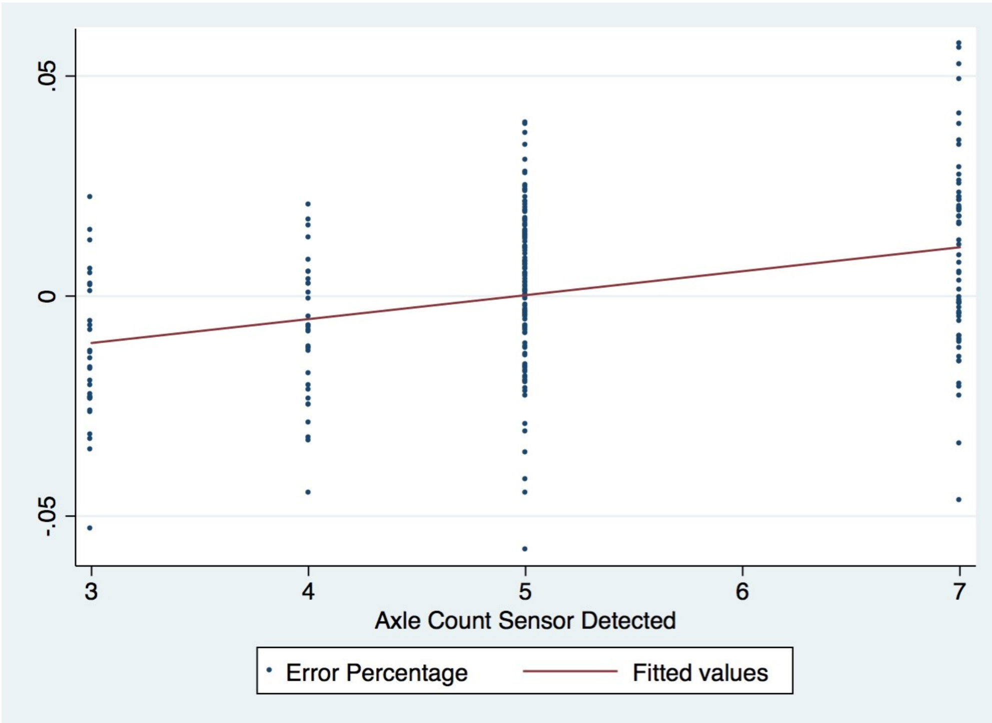

Figure A.5 Error Percentage by Axle Count

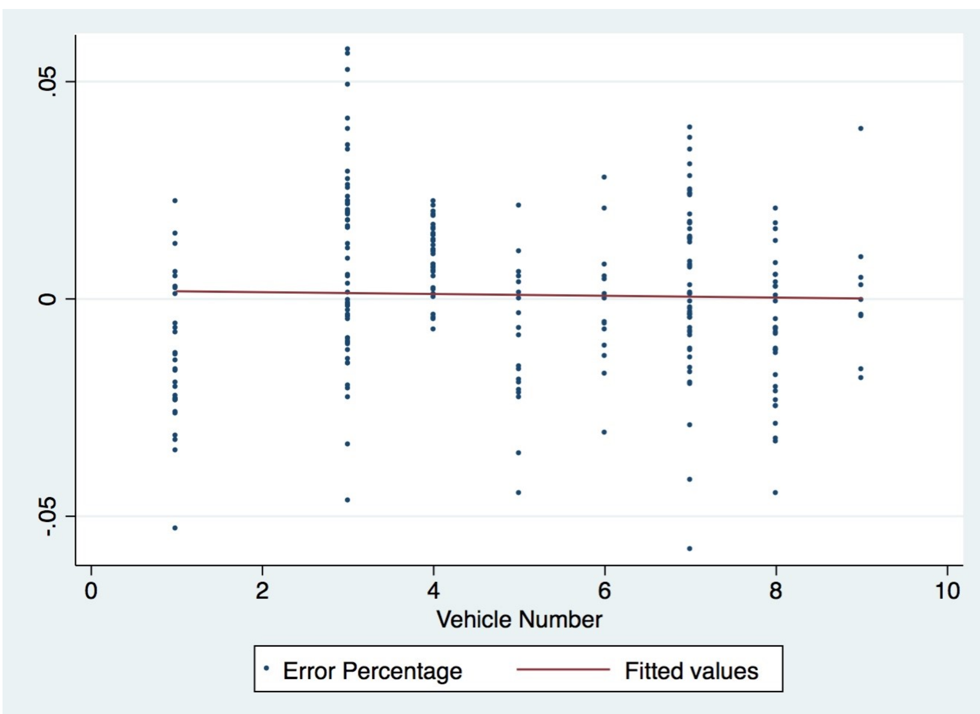

Figure A.6 Error Percentage by Vehicle Number

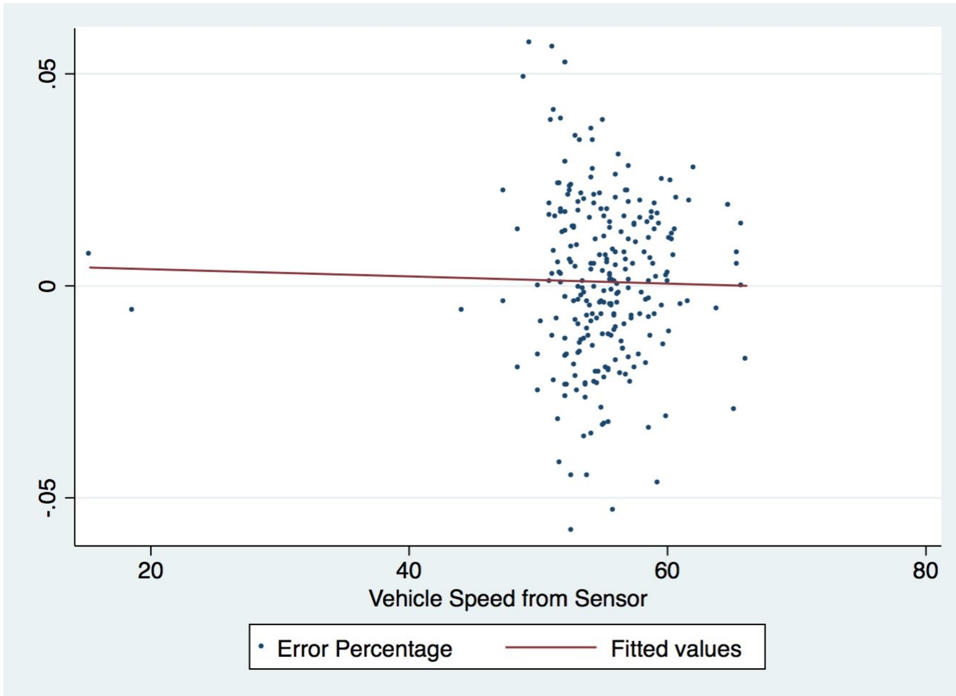

Figure A.7 Error Percentage by Speed

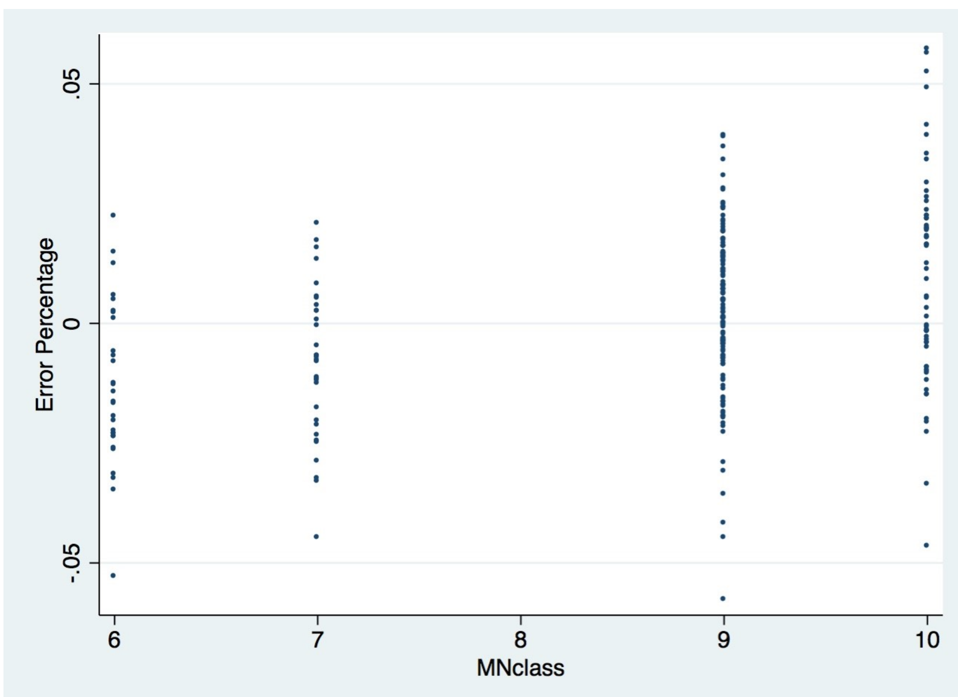

Figure A.8 Error Percentage by MNDOT Classification

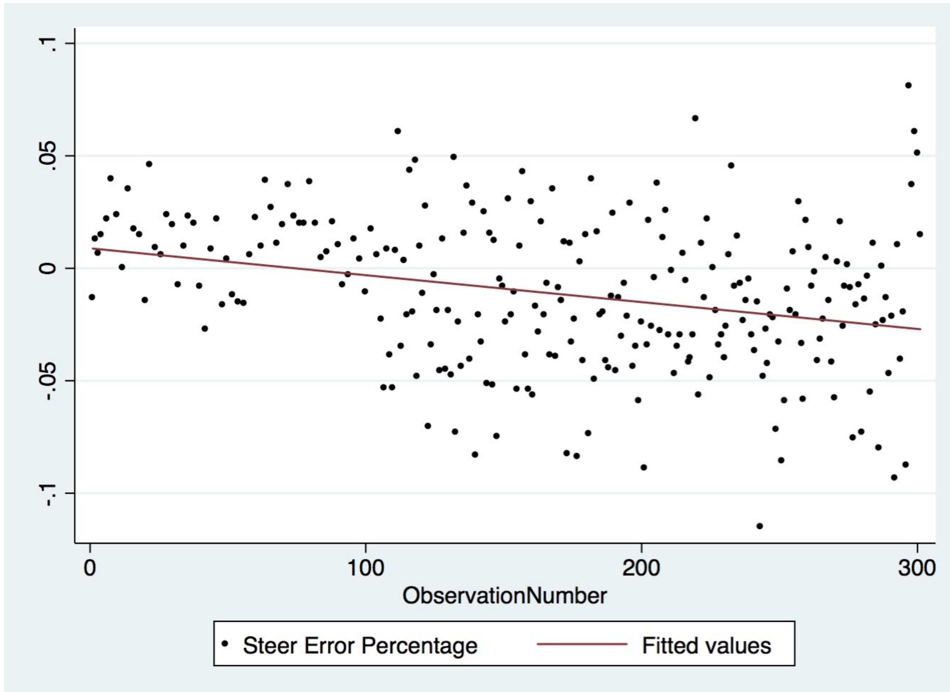

Figure A.9 Steer Axle Error Percentage and Observation Number

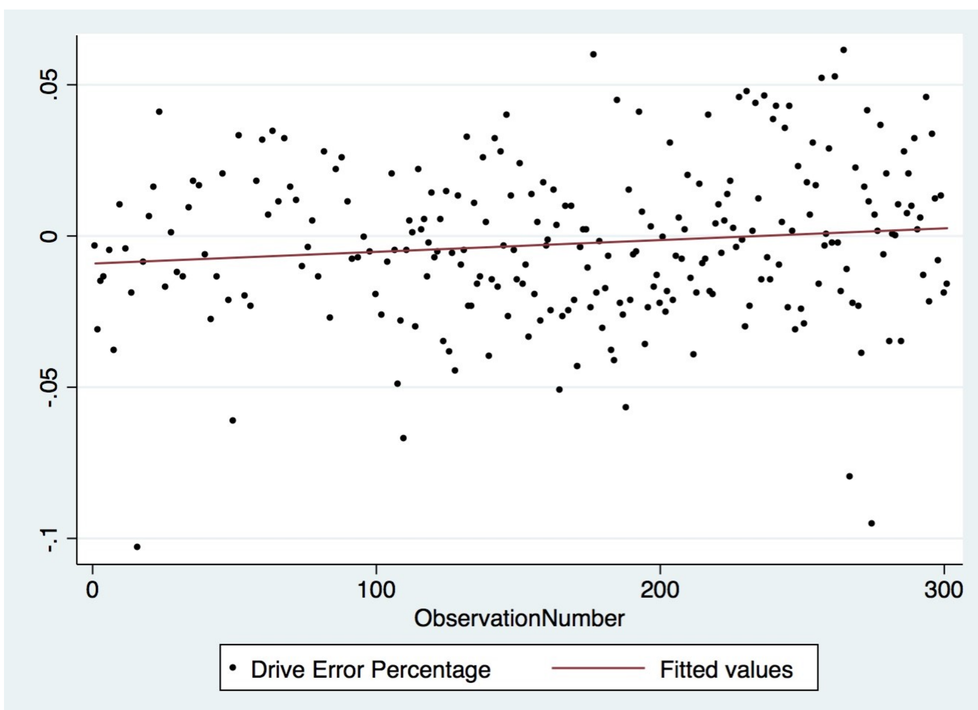

Figure A.10 Drive Axle Error Percentage and Observation Number

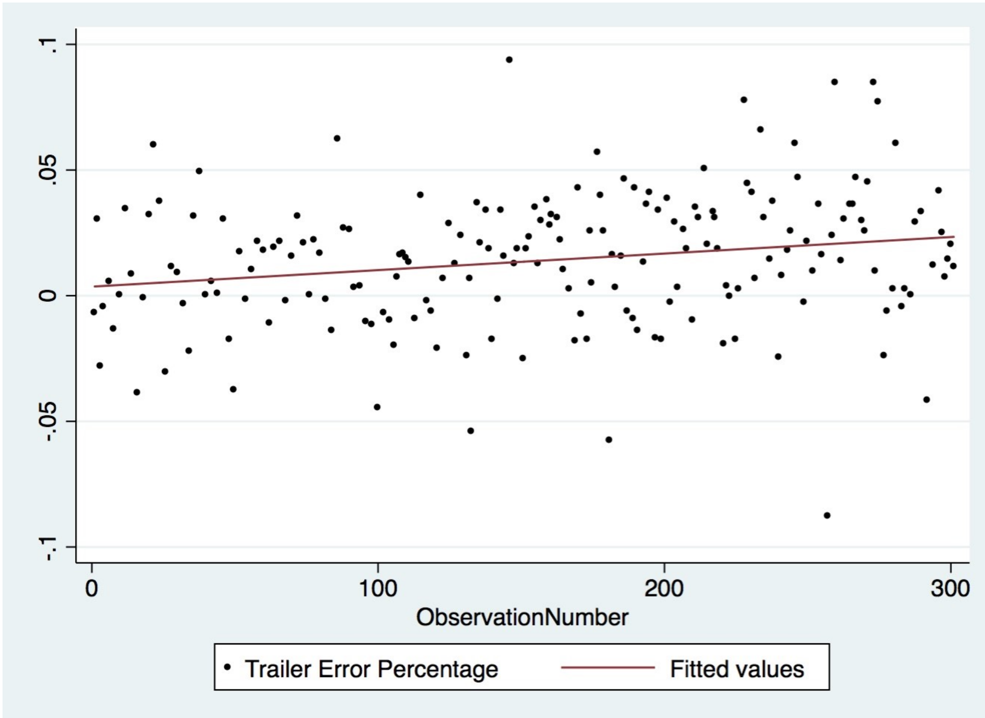

Figure A.11 Trailer Axle Error Percentage and Observation Number

Table A.1 Regression Analysis (Error Percentage = Cons + Coef. × Observation Number + Error Term)

Source	SS	df	MS
Model	.000125426	1	.000125426
Residual	.092561519	248	.000373232
Total	.092686944	249	.000372237

Number of obs =	250
F(1, 248) =	0.34
Prob > F =	0.5626
R-squared =	0.0014
Adj R-squared =	-0.0027
Root MSE =	.01932

errpercentage_d2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNumber	8.69e-06	.000015	0.58	0.563	-.0000208	.0000382
_cons	-.0004131	.002833	-0.15	0.884	-.0059929	.0051667

Table A.2 Regression Analysis (Dynamic Weight = Coef. × Static Weight + Error Term)

Source	SS	df	MS
Model	1.2263e+12	1	1.2263e+12
Residual	434945844	249	1746770.46
Total	1.2267e+12	250	4.9069e+09

Number of obs =	250
F(1, 249) =	.
Prob > F =	0.0000
R-squared =	0.9996
Adj R-squared =	0.9996
Root MSE =	1321.7

GVW	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aGVW_d2	1.003199	.0011973	837.88	0.000	1.000841	1.005558

Table A.3 Regression Analysis (Error Percentage = Cons + Coef. × Average Temperature + Error Term)

Source	SS	df	MS
Model	.000031664	1	.000031664
Residual	.09265528	248	.00037361
Total	.092686944	249	.000372237

Number of obs =	250
F(1, 248) =	0.08
Prob > F =	0.7712
R-squared =	0.0003
Adj R-squared =	-0.0037
Root MSE =	.01933

errpercent~2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp_mean	.0000257	.0000883	0.29	0.771	-.0001482	.0001996
_cons	.0000124	.0038283	0.00	0.997	-.0075278	.0075526

Table A.4 Regression Analysis (Error Percentage = Cons + Coef. × Observation Number + Error Term)

Source	SS	df	MS
Model	.000105465	1	.000105465
Residual	.092626338	248	.000373493
Total	.092731803	249	.000372417

Number of obs =	250
F(1, 248) =	0.28
Prob > F =	0.5956
R-squared =	0.0011
Adj R-squared =	-0.0029
Root MSE =	.01933

errpercentage_t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationNumber	7.96e-06	.000015	0.53	0.596	-.0000216	.0000375
_cons	-.0004033	.002834	-0.14	0.887	-.005985	.0051785

Table A.5 Regression Analysis (Dynamic Weight = Coef. × Static Weight + Error Term)

Source	SS	df	MS
Model	1.2263e+12	1	1.2263e+12
Residual	435904111	249	1750618.92
Total	1.2267e+12	250	4.9069e+09

Number of obs =	250
F(1, 249) =	.
Prob > F =	0.0000
R-squared =	0.9996
Adj R-squared =	0.9996
Root MSE =	1323.1

GVW	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aGVW_t	1.003183	.0011986	836.95	0.000	1.000822	1.005543

Table A.6 Regression Analysis (Error Percentage = Cons + Coef. × Average Temperature + Error Term)

Source	SS	df	MS
Model	.00002084	1	.00002084
Residual	.092710964	248	.000373835
Total	.092731803	249	.000372417

Number of obs =	250
F(1, 248) =	0.06
Prob > F =	0.8135
R-squared =	0.0002
Adj R-squared =	-0.0038
Root MSE =	.01933

errpercent~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp_mean	.0000209	.0000883	0.24	0.814	-.0001531	.0001948
_cons	.0000986	.0038295	0.03	0.979	-.0074439	.007641

Table A.7 Hypothesis Tests by MN Classification (Hypothesis: Mean Error Percentage in Estimating Weight of Mnclass = 0)

```
-> MNclass = 6

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |      29   -.0133662    .0031257    .0168323   -.0197689   -.0069636
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =  -4.2763
Ho: mean = 0                                     degrees of freedom =       28

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0001         Pr(|T| > |t|) = 0.0002          Pr(T > t) = 0.9999

---------------------------------------------------------------------------------
-> MNclass = 7

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |      32   -.0083329     .002819    .0159464   -.0140822   -.0025836
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =  -2.9560
Ho: mean = 0                                     degrees of freedom =       31

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0030         Pr(|T| > |t|) = 0.0059          Pr(T > t) = 0.9970

---------------------------------------------------------------------------------
-> MNclass = 9

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |     129    .0026005    .0014775    .0167812   -.0003229     .005524
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =   1.7601
Ho: mean = 0                                     degrees of freedom =      128

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9596         Pr(|T| > |t|) = 0.0808          Pr(T > t) = 0.0404

---------------------------------------------------------------------------------
-> MNclass = 10

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
errper~t |      60    .0092942    .0027942    .0216442     .003703    .0148855
------------------------------------------------------------------------------
    mean = mean(errpercentage_t)                                  t =   3.3262
Ho: mean = 0                                     degrees of freedom =       59

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9992         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.0008
```

Table A.8 Steer Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.093563025	4	.023390756
Residual	.184868247	245	.000754564
Total	.278431272	249	.001118198

Number of obs	=	250
F(4, 245)	=	31.00
Prob > F	=	0.0000
R-squared	=	0.3360
Adj R-squared	=	0.3252
Root MSE	=	.02747

errpercent~eer	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationN~r	-.0001043	.0000217	-4.80	0.000	-.0001471	-.0000615
AxleCount	-.0112241	.001406	-7.98	0.000	-.0139934	-.0084548
VN	.0030357	.0007452	4.07	0.000	.001568	.0045034
Speed	.0003815	.0003645	1.05	0.296	-.0003364	.0010995
_cons	.0280668	.0218852	1.28	0.201	-.0150403	.0711739

Table A.9 Drive Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.045035248	4	.011258812
Residual	.117879554	245	.000481141
Total	.162914802	249	.000654276

Number of obs	=	250
F(4, 245)	=	23.40
Prob > F	=	0.0000
R-squared	=	0.2764
Adj R-squared	=	0.2646
Root MSE	=	.02193

errpercenta~ve	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationN~r	.0000315	.0000173	1.81	0.071	-2.69e-06	.0000656
AxleCount	.0100441	.0011227	8.95	0.000	.0078327	.0122554
VN	-.0003802	.000595	-0.64	0.523	-.0015522	.0007918
Speed	-.00049	.0002911	-1.68	0.094	-.0010633	.0000833
_cons	-.0306299	.0174758	-1.75	0.081	-.065052	.0037922

Table A.10 Trailer Axle Regression Analysis (Error Percentage = Cons + Coef.1 × Observation Number + Coef.2 × Axle Count + Coef.3 × Vehicle Number + Coef.4 × Speed + Error term)

Source	SS	df	MS
Model	.008343447	4	.002085862
Residual	.125461301	184	.000681855
Total	.133804748	188	.000711727

Number of obs	=	189
F(4, 184)	=	3.06
Prob > F	=	0.0180
R-squared	=	0.0624
Adj R-squared	=	0.0420
Root MSE	=	.02611

errpercent~ler	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ObservationN~r	.0000757	.000026	2.91	0.004	.0000244	.0001269
AxleCount	-.0050012	.0032316	-1.55	0.123	-.011377	.0013745
VN	-.0016814	.0016302	-1.03	0.304	-.0048978	.0015349
Speed	-.0006856	.0004155	-1.65	0.101	-.0015054	.0001342
_cons	.0767238	.03997	1.92	0.056	-.0021347	.1555823