**An Adaptive Algorithm to Compare Travel Time Reliability on Arterial Road Corridors**

**Subhadipto Poddar**

Department of Civil, Construction, and Environmental Engineering

Iowa State University

394 Town Engineering, Ames, IA 50011

Tel: 515-294-2140; Email: [spoddar@iastate.edu](mailto:spoddar@iastate.edu)

**Tongge Huang**

Department of Civil, Construction, and Environmental Engineering

Iowa State University

394 Town Engineering, Ames, IA 50011

Tel: 515-294-2140; Email: [tongge@iastate.edu](mailto:tongge@iastate.edu)

**Trevor Kirsch**

Department of Civil, Construction, and Environmental Engineering

Iowa State University

394 Town Engineering, Ames, IA 50011

Tel: 515-294-2140; Email: [tjkirsch@iastate.edu](mailto:tjkirsch@iastate.edu)

**Dr. Anuj Sharma (Corresponding Author)**

Associate Professor, Iowa State University

Department of Civil, Construction, and Environmental Engineering

Iowa State University

394 Town Engineering, Ames, IA 50011

Tel: 515-294-2140; Email: [anujs@iastate.edu](mailto:anujs@iastate.edu)

Keywords: Travel Time Reliability, Anomaly Detection, Local Outlier Factor, Elbow Method, Linear Regression

Submitted for publication on April 2, 2017

**Abstract**

Travel time reliability is an important metric for corridor progression as it can be used to determine the variability of the experienced travel time at an aggregate level. In this study, the travel time reliability performance of 4 arterial corridors in West Des Moines was analyzed for a duration of two years. First, an adaptive, data-driven algorithm was proposed to remove probe data anomalies using the density-based Local Outlier Factor (LOF) methodology. In the next step, a linear regression analysis was implemented to determine the performance of the different corridors temporally. The reliability was estimated based on the orientation of the best-fit-line through the probe data point cloud for each intersection. The resultant frequency distribution of experienced variability was combined with other intersections to analyze the performance of the entire corridor. Ultimately, results determined that the travel time reliability was an accurate proxy for the performance of the corridors. Known locations where delay was expected due to roadway or roadside construction had highly variable data. Lastly, corridors that were outfitted with adaptive signal control experienced improvements in reliability and a resultant decrease in average delay variability.

**1.** **Introduction**

Travel time reliability has become a very important metric when estimating the performance of a signalized arterial corridor. Unreliable travel time estimates can lead to poorly timed signal offsets, which may result in excessive queues. With unreliable information, increased delay becomes inevitable as arterial congestion increases. Traditionally, signal timing optimization and arterial corridor progression were computed based on historical travel trends. With the rise of big data and the ubiquitous availability of vehicular probe information, performance measures can be generated to model the interaction between near real-time platoon progressions and signal timing. Data availability and new data sources have spawned updated methodologies for corridor analyses, while constant data streams can provide up-to-date analytics on performance. The purpose of this analysis is to utilize vehicular probe data to estimate the travel time reliability on arterial corridors in West Des Moines. Additionally, a data screening method is introduced to filter out unreasonable probe estimates at an aggregate level as a form of anomaly detection. Linear regression techniques are also applied to the travel time data to measure the reliability of the travel time experienced along a corridor.

This paper is organized into six sections. The literature review discusses past research focused on travel time reliability estimation and methods of anomaly detection which reflect the processes conducted within this study. It is followed by the project methodology, which gives an elaborate description about the adaptive algorithm used for anomaly detection. The results and discussion sections analyze the temporal travel time reliability on different types of corridors. From this analysis, conclusions are drawn. The paper concludes with the future scope of the study, including information and additional methodologies that could be used to refine the analysis.

**2.** **Literature Review**

## 2.1 Travel Time Reliability

Recent literature has focused on the development of new methodologies to estimate the reliability of travel time data. Research by Day et al. (2010) focused on the offset timing of signals along State Route 37 (SR 37) in Indiana. The purpose of this research was to demonstrate a new tool that can assist in arterial signal progression management: the Purdue Coordination Diagram (PCD). The PCD provides a means to visualize both the controller and detector event data to assist in the analysis of vehicular arrival patterns during coordinated movements. During a case study, the PCD was able to identify the causes of poor progression. Results from signal timing adjustments indicated a 1.7 min (28%) reduction in the mean travel time for the optimized direction after signal offset adjustment. Additional methodologies for travel time reliability estimates have focused on combining past historical trends with current vehicular probe information to augment real-time traffic conditions. Research conducted by Feng et al. (2014) first estimated the initial corridor travel time based on historical travel patterns and signal timing alone. Following this, probe data were integrated into the analysis to improve the travel time predictions. Results from two case studies indicated that the iterative process of including real-time probe data with the historical predictions made travel time distributions more accurate when compared to traditional methods alone. Both studies determined that new methodologies exist to quantify the performance of a corridor by using travel time reliability data.

As demonstrated by recent research, the vehicular travel time through a signalized corridor is often measured with probe data due to its large quantity and availability. A study by Liu and Ma (2008) sought to determine if substituting probe data for traditional data collection methods resulted in an accurate depiction of the traffic flow through a network. The travel time along the corridor was estimated with a virtual probe vehicle that navigated the arterial based on the surrounding traffic conditions. The algorithm was tested on an 11 intersection corridor and the study determined that the probe data methodology was capable of generating accurate corridor travel times under a variety of traffic conditions. The use of travel time reliability as a quantitative performance measure has been demonstrated in research as well. One such study was conducted by Day et al. (2014) on arterials SR 37 and State Route 931 (SR 931) in Indiana. The analysis investigated the use of vehicle probe data as a viable means of measuring the experienced travel times on arterial corridors. The developed method considers both central tendency and reliability of the travel time data. The study ultimately determined that reliability can be used to generate an estimate of traffic performance. Both studies found that probe data provides an accurate alternative to traditional traffic flow measuring techniques.

Another study on the SR 37 corridor by Lavrenz et al. (2015) focused on the long-term benefits of signal optimization by using corridor performance measures as an estimate of effectiveness. A 6-year analysis period was utilized, with signal offset optimization occurring every 2-3 years. Following the removal of outliers, both the travel time and travel time reliability were calculated before and after signal optimization. Although the total volume of traffic increased by 36% during the study period, the number of vehicles arriving on green increased by 41%. Ultimately, the re-timing of the traffic signals resulted in a benefit-cost ratio of 52, representing a Net Present Value of $3.7 million dollars. Results from this study and the previous literature demonstrate the value in calculating travel time reliability. Work has also been completed internationally to improve the quantification of corridor progression. A study in India by Gopi et al. (2014) and in Germany by Wünsch et al. (2014) both demonstrated that travel time reliability provides an accurate and objective evaluation of network progression.

## 2.2 Anomaly Detection

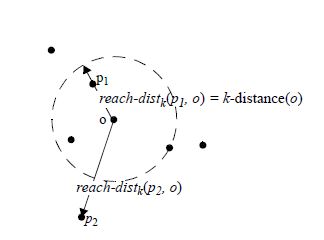
Although probe data has been demonstrated as a reliable proxy for traditional traffic monitoring methods, the data may initially be erratic or “noisy” in nature. Because of this, data cleaning (or the removal of outliers) must be performed to ensure an accurate analysis. Many supervised, semi-supervised, and unsupervised techniques have been generated over the years to detect and remove anomalies. An elaborate comparative study by Chandola et al. (2009) discussed the following data cleansing methodologies:

* Classification based models are mainly a supervised type of method were a point is identified as an anomaly because it is distinct from the other points in a feature space. The main disadvantage of this method is that it cannot assign an anomaly score to any points within the space.
* The statistical method of outlier analysis identifies a point as an outlier based upon a generated distribution. It assumes that the sample is drawn from some underlying distribution as well (Gaussian, Poisson, etc.) and identifies outliers based upon this base distribution assumption.
* Clustering based models are unsupervised in nature, but they lack the proper methodology to detect an anomaly. A data point in the cluster must be either an outlier or a non-outlier within the dataset. Further anomalies are identified as byproducts of the original clustering analysis.
* Nearest neighbor is an unsupervised method of modeling which considers the concentration and distance of a group of points around a single point. This method is the most suitable for transportation data analysis. It has been used widely by Munz et al. (2007), Chen et al. (2016), and Myung et al. (2014) for flow prediction and travel time prediction alike. The method used in this travel time reliability study was generated from the density-based Local Outlier Factor (LOF).

### 2.2.1 Density-Based Local Outlier Factor

Density based algorithms can detect the presence of surrounding points around a specific point in the domain space. As determined by Breunig et al. (2000), most of the algorithms like DBSCAN, CLIQUE, or Wavecluster can conduct clustering, but are not robust enough in terms of outlier detection. Additionally, research by Chandola et al. (2010) has demonstrated that the Kth nearest neighbor method has the power to identify clusters as well, but is incapable of detecting outliers when the clusters are of variable density. Because of this, there is a need to identify the points based upon their location relative to their nearest neighboring points. The method which quantifies the points in this manner is known as the Local Outlier Factor (LOF).

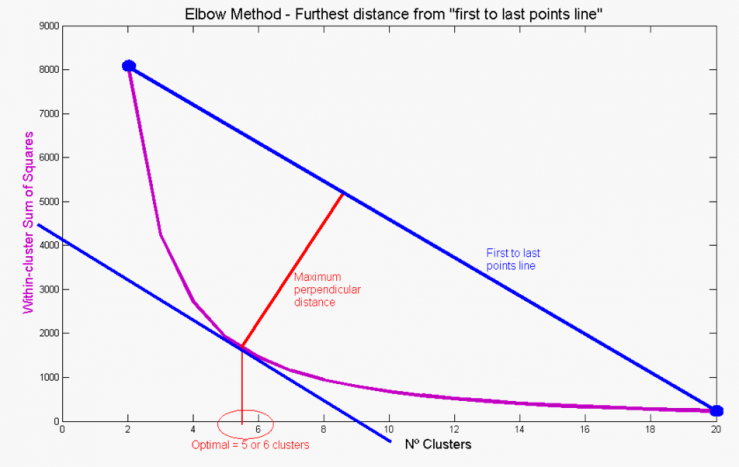
The LOF method identifies a group of k-objects, known as the minimum points, around each point of interest and denotes it as k-distance from that object. For instance, in Figure 2.1, the k is assumed to be 4 and k-distance of the object is shown. The reachability distance of an object “*o”* with respect to another object “*p2*” is measured as the maximum of the distance of *o* from *p2* and *o*’s k-distance. After this, the local reachability density of *o* is determined as the inverse of the reachability distance of object *o* with respect to its minimum nearest neighboring points. Finally, the Local Outlier Factor is defined as the sum of the ratio of the object *o* with respect to its neighboring data points.



*Figure 2.1: Reachability Distance and K-Distance with K=4*

### 2.2.2 Elbow Method

One of the main challenges of an unsupervised learning method is how to establish a “cutoff” between normal data and anomalous data. Research has been conducted on a multitude of methods, including K-means clustering by Thorndike (1953) as well as principal component analysis (PCA) by Jackson (2016). Despite this, the elbow method is one of the most popular methods utilized to date when deciding how to establish a cutoff threshold. The elbow method cutoff sorts the parameter which is to be cutoff and then joins consecutive points. An imaginary line is then drawn joining the start and the end points within the domain space. The cutoff point is proposed as that point which has the maximum perpendicular distance from this imaginary line. Figure 2.2 displays a visualization of the elbow method.



*Figure 2.2: Description of the Elbow method*

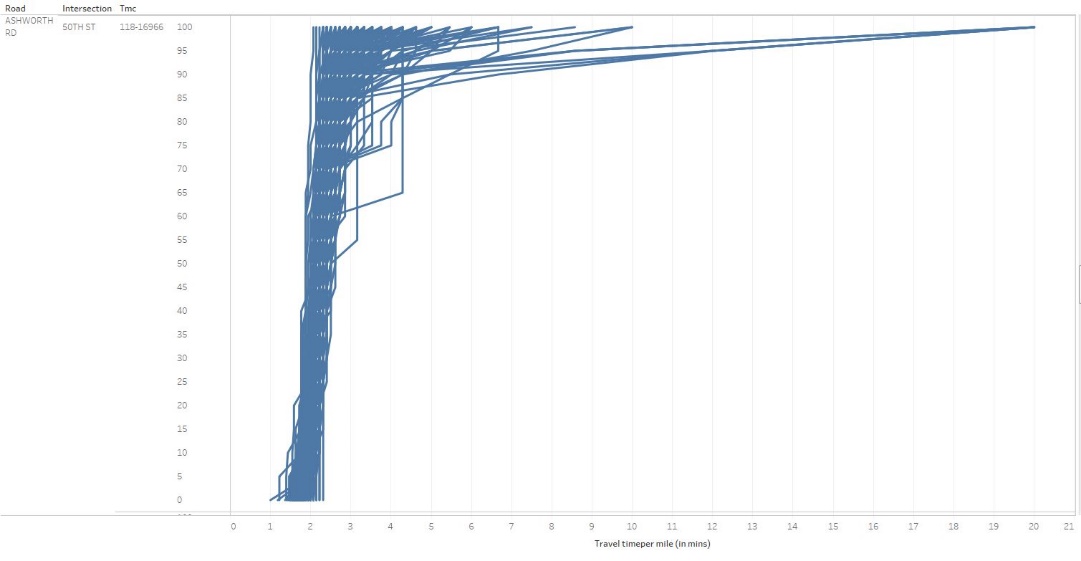
3. Methodology

## 3.1 Data Description

The provider of the probe data for this project was INRIX, which is a commercial vendor of crowd-sourced data commonly used for understanding real-time traffic flow characteristics. The data consisted of two years of vehicular probe measurements on 4 major arterial corridors within West Des Moines: Ashworth Road., Grand Avenue, Mills Civic Parkway, and University Avenue. There was a total of 70 individual segments between all 4 corridors. The data extraction process began with the procurement of the probe data for each segment for both 2015 and 2016 using a map-reduce script in Apache Pig.

## 3.2 Anomaly Detection

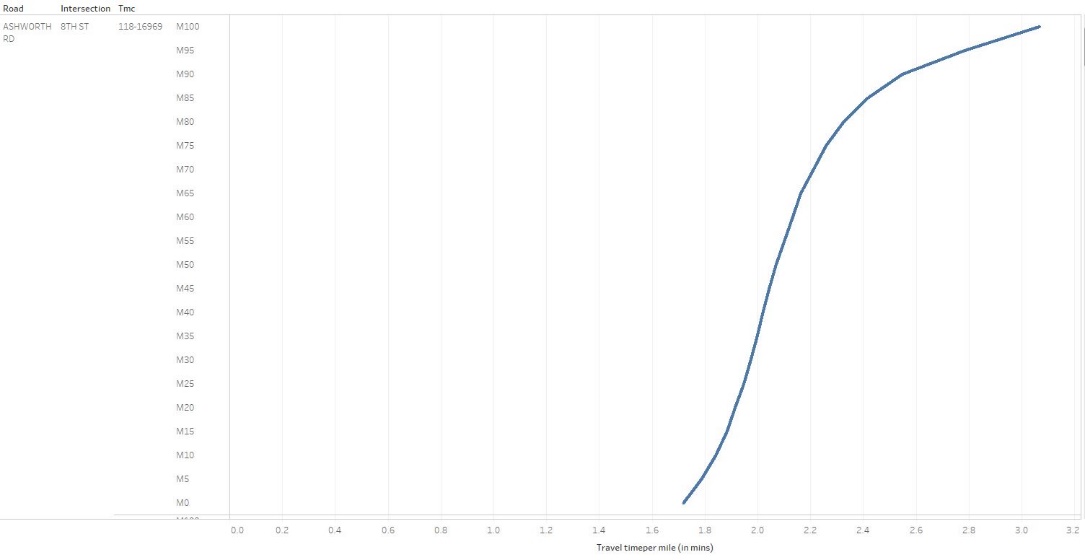
After data retrieval, travel time reliability plots were estimated. The travel time reliability plots displayed the corridor travel rate (travel time per unit mile of the segment) for each day of each year in 5-percentile increments as shown in Figure 3.1.1:



*Figure 3.1.1: Quantile Plot of Travel Rate*

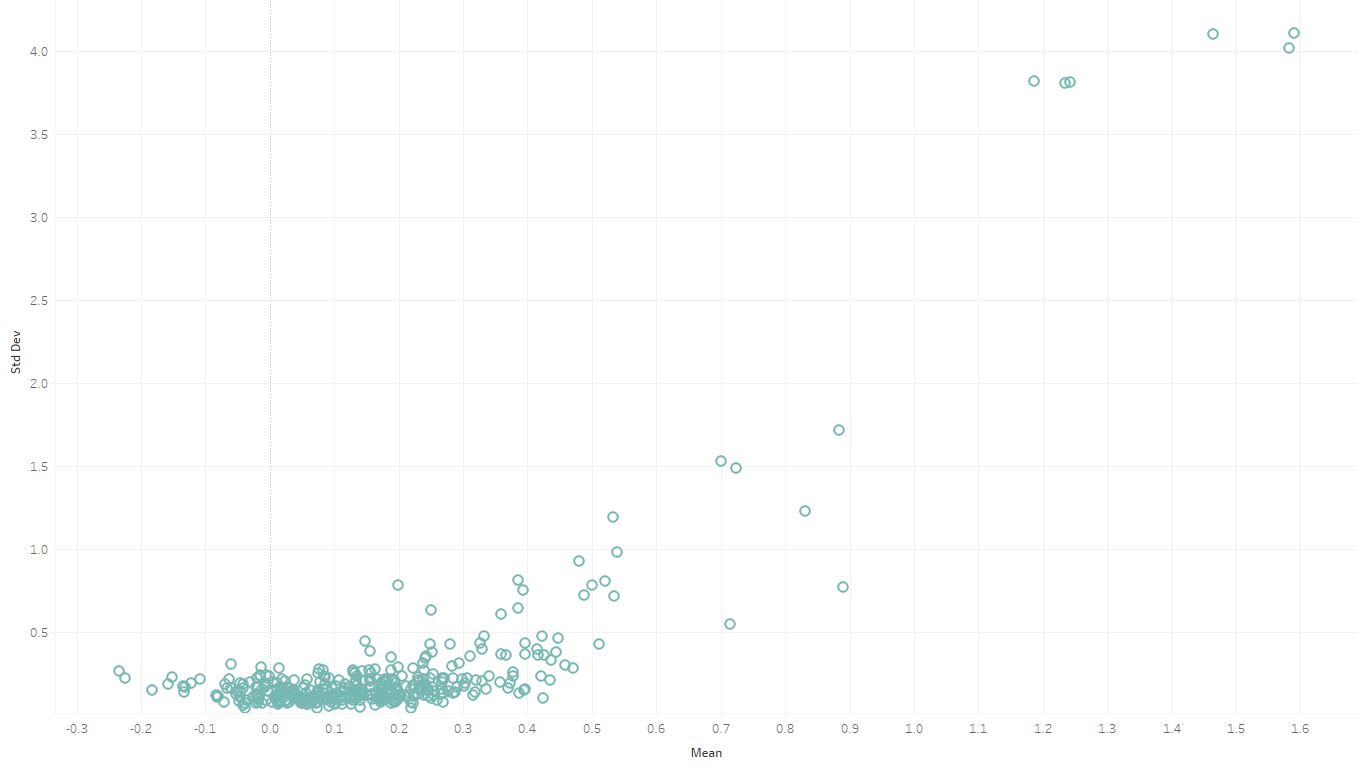
The reason for conducting anomaly detection was to remove outliers from the initial probe data. To conduct anomaly detection, the following process was applied:

Step 1: For each day during 2015 and 2016, the average travel rate was computed for each quantile (in 5-percentile increments). This was plotted for each corridor respectively. The purpose of generating these plots was to establish the average travel time estimate for that corridor within each year. For example, the average travel time reliability plot for 2016 for the intersection of 50th St. and Ashworth Rd. in the westbound direction is shown in Figure 3.1.2.



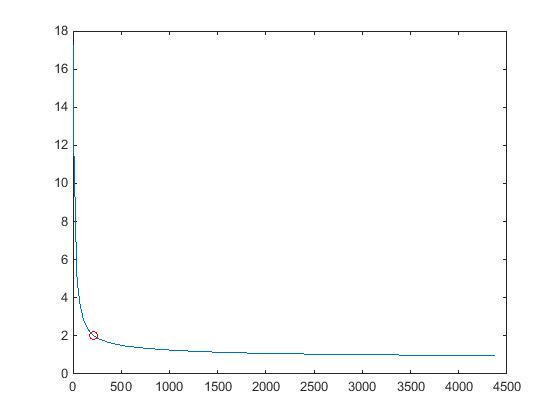
*Figure 3.1.2: Average Travel Time Reliability Curve*

Step 2: A typical travel time was defined as the median reliability curve as determined by the travel time quantile plots. This median value established the “Base Day” for that corridor for each year. In Figure 3.1.3, the Base Day for 2016 for the Ashworth Rd. corridor is shown.



*Figure 3.1.3: Average Daily Travel as Determined by Quantile Analysis*

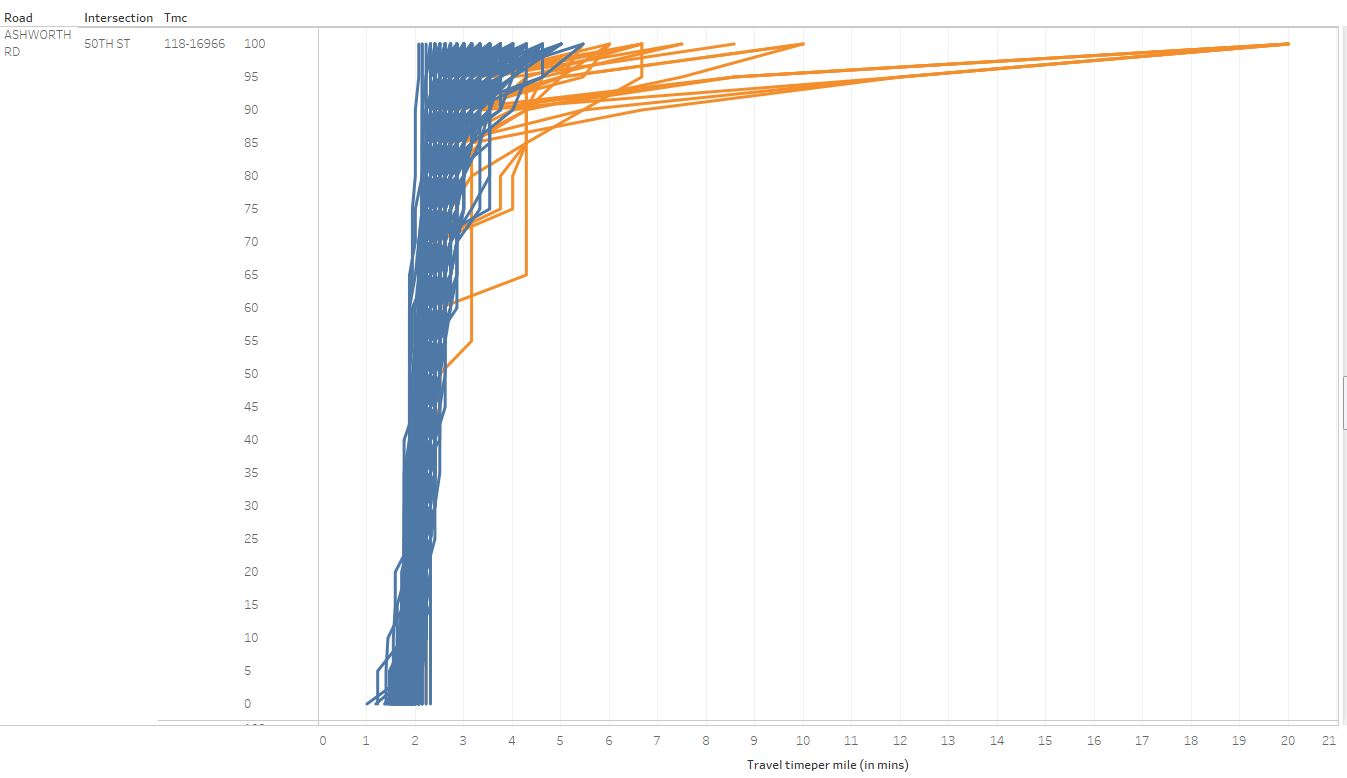
Step 3: The difference between each 5th percentile quartile value with respect to the Base Day was calculated, as well as the mean and standard deviation between the two curves. An example is shown in Figure 3.1.4 for the same intersection as previously mentioned:



*Figure 3.1.4: Difference between Average Daily Travel and Other Percentiles*

Step 4: Based upon the computed mean and standard deviation, the LOF score was calculated for each segment separately. The LOF was calculated by considering the minimum points from 10 to 100 (in intervals of 10). Here, the average LOF was computed for each point as proposed by Chen et al. (2016). The days in which the intersection behaves abnormally were assigned a higher average LOF score.

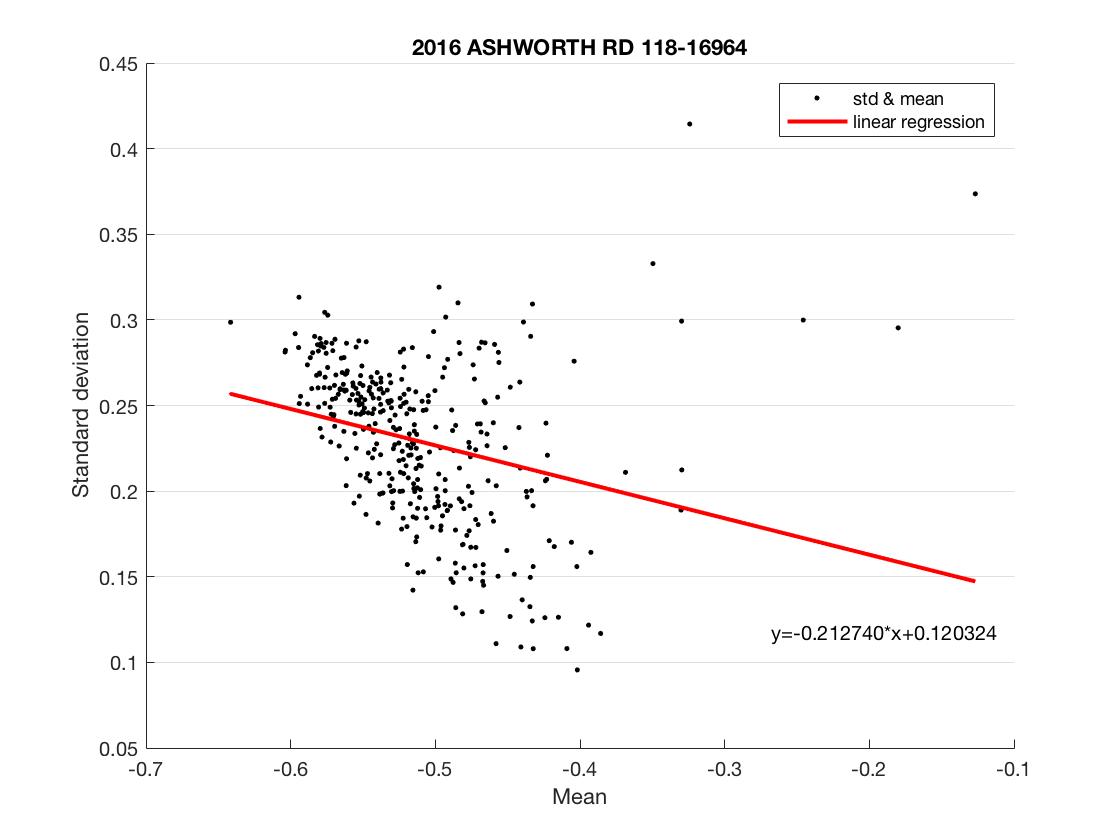
Step 5: All the average LOF values from Step 4 for each corridor were assembled in descending order and a line was drawn to join them. Next, an elbow method cutoff was proposed to locate the “higher than average” LOF values as shown in Figure 3.1.5. The days which had a positive mean and a higher LOF value (beyond the cutoff) were flagged as anomalies and removed from the dataset to ensure greater accuracy.



*Figure 3.1.5: Elbow Method Cutoff*

## 3.3 Linear Regression

Linear regression is a statistical methodology used to model the relationship between two variables by estimating a “best fit” line through a scatter plot of point data. By using the mean and standard deviation values calculated previously, a linear regression method was used to analyze the corridor performance for each year. The data from the same intersection as previous is displayed in Figure 3.2:



*Figure 3.2: Linear Regression Example Plot*

The points on the scatter plot represent the daily mean and standard deviation data for the intersection. The mean and standard deviation were measured from the Base Day as mentioned previously. By using the following equation, the slope of the linear line was converted to an angle measurement (in degrees):

Where:

s = slope

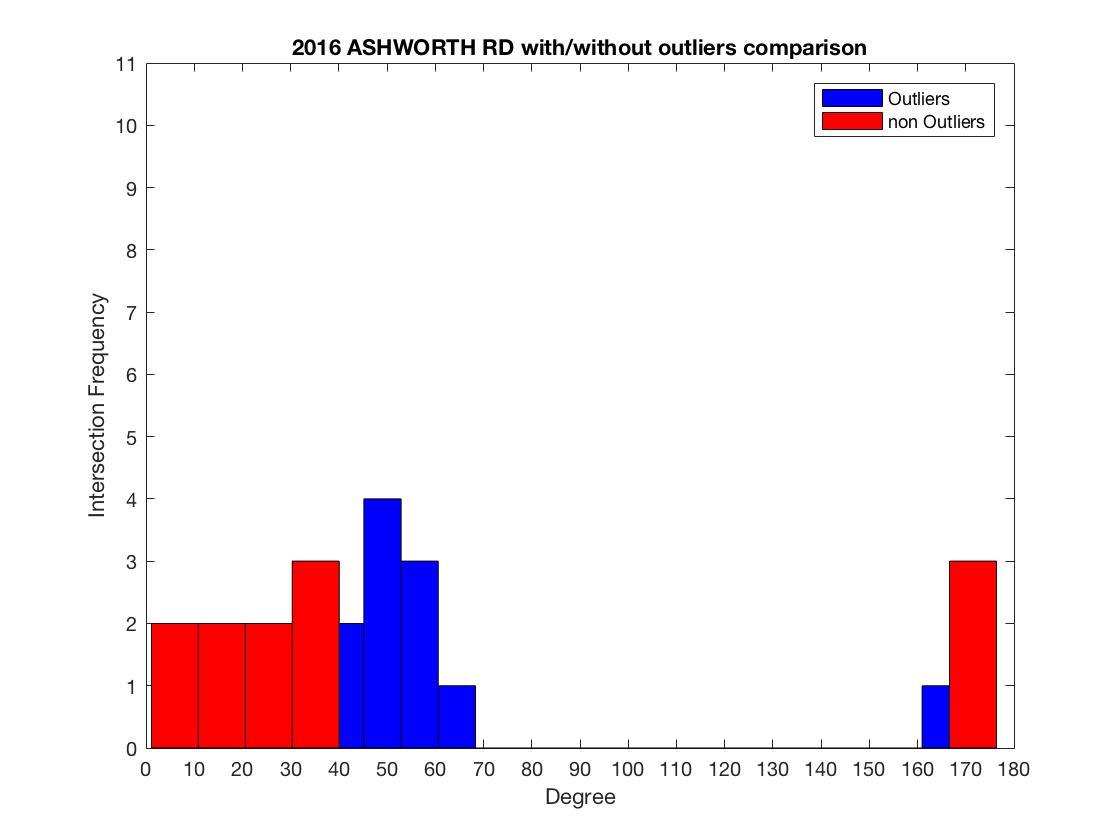
α = angle (degrees)

The slope of the linear regression line in the previous figure was -0.21274, which translates to 168 degrees. The negative slope indicated that most of the plotted mean values were negative. Because of this, the best fit line for this intersection began in the left quadrant of the point cloud, demonstrating that most mean and standard deviation values were low. A low value of both mean and stand deviation from the Base Day was indicative of excellent reliability (small variance) of the travel time estimates throughout the year. On the contrary, a linear regression line with a positive slope indicated grater mean and standard deviation measurements, indicating poor reliability. Similar plots and analyses have been conducted for each intersection for 2015 and 2016 on the corridors of interest.

4. Results and Discussion

## 4.1 Comparison between Outlier and Non-Outlier Data

Based on the travel time reliability analysis, there were several abnormal daily observations for each intersection which were identified as outliers. Although this analysis uses historical probe data, one key emphasis considered during the analysis was to generate a process that could be mirrored in real-time applications. The inclusion of outliers would affect real-time travel time reliability estimations. Because of this, it was necessary to compare the abnormal situations and normal situations to create an abnormality filter to allow for up-to-date travel time reliability processing. In Figure 4.1, the 2016 Ashworth Rd. corridor data was graphed as an example to show the differences between the data when including and removing the identified outliers. Note that the x-axis displays the analysis results in degree, where a 0 degree and 180-degree result represent accurate results (more reliable estimates), and a degree measure closer to 90 degrees demonstrates less accurate results (less reliable estimates).



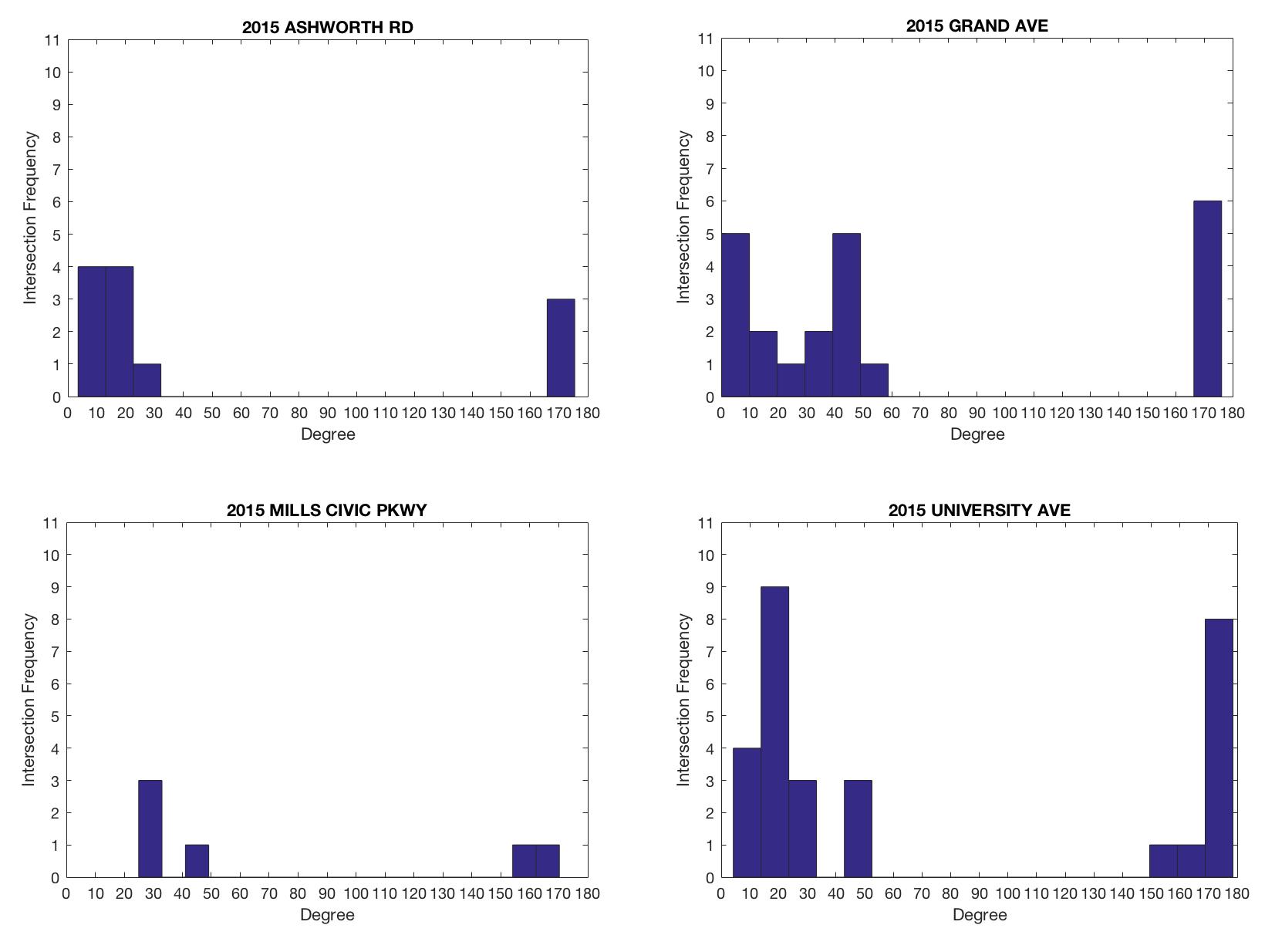
*Figure 4.1: Comparison between Outlier and Non-Outlier Data*

When including the outliers in the analysis, the diagram shifts to the middle (or the less reliable portion of the graph) due to introduced inaccuracies. The red colored bars above indicate the data without outliers, while the blue bars incorporate the outliers for demonstrative purposes. The linear regression analysis would also be impacted by outliers. Including outliers effects the slope of the best-fit-line. Because the outliers would cause an increase in the mean and standard deviation of all data points, the slope generated for each intersection would become steeper, indicating that the traffic situations for these intersections were worse than experienced.

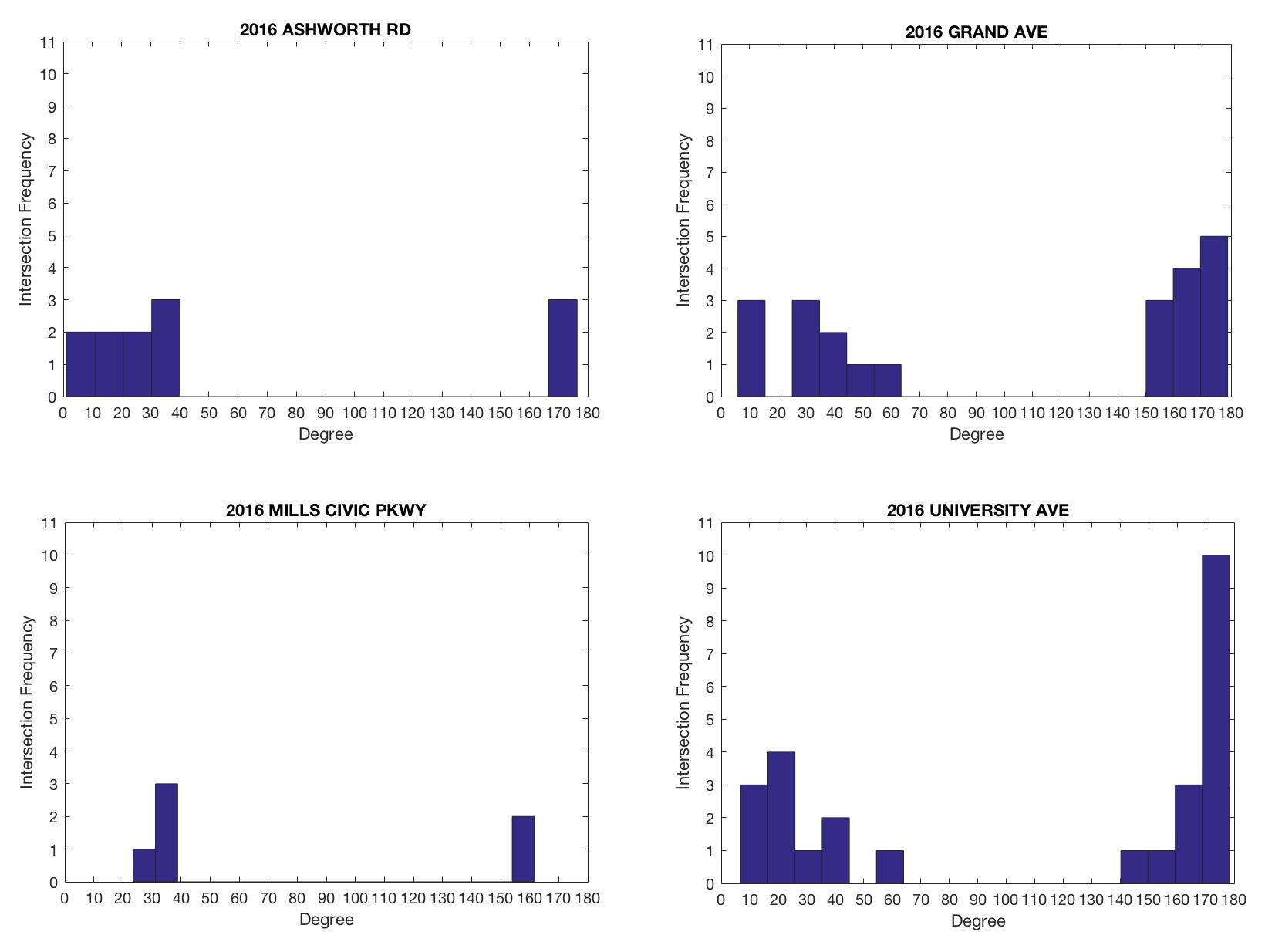
## 4.2 Temporal Comparison of Individual Corridors

The corridor travel time reliability was measured by the slope calculated with the linear regression methodology for each intersection. Based on the frequencies depicted in the histogram, the traffic situation can be evaluated annually for each corridor at an aggregate level.

Per the reliability plots and linear regression analysis, a greater positive mean and standard deviation will be represented by a greater slope. An intersection with a greater positive mean and standard deviation indicates poor performance. The degree of difference was again utilized to examine the results. For the histograms shown below (Figure 4.2.1 and Figure 4.2.2), the Grand Ave. corridor had the worst reliability (or most unreliable estimates). For the most part, University Ave. had the best reliability, except for occasional unique intersection which had additional variance. For Ashworth Rd. and Mills Civic Pkwy., the reliability was mostly good.

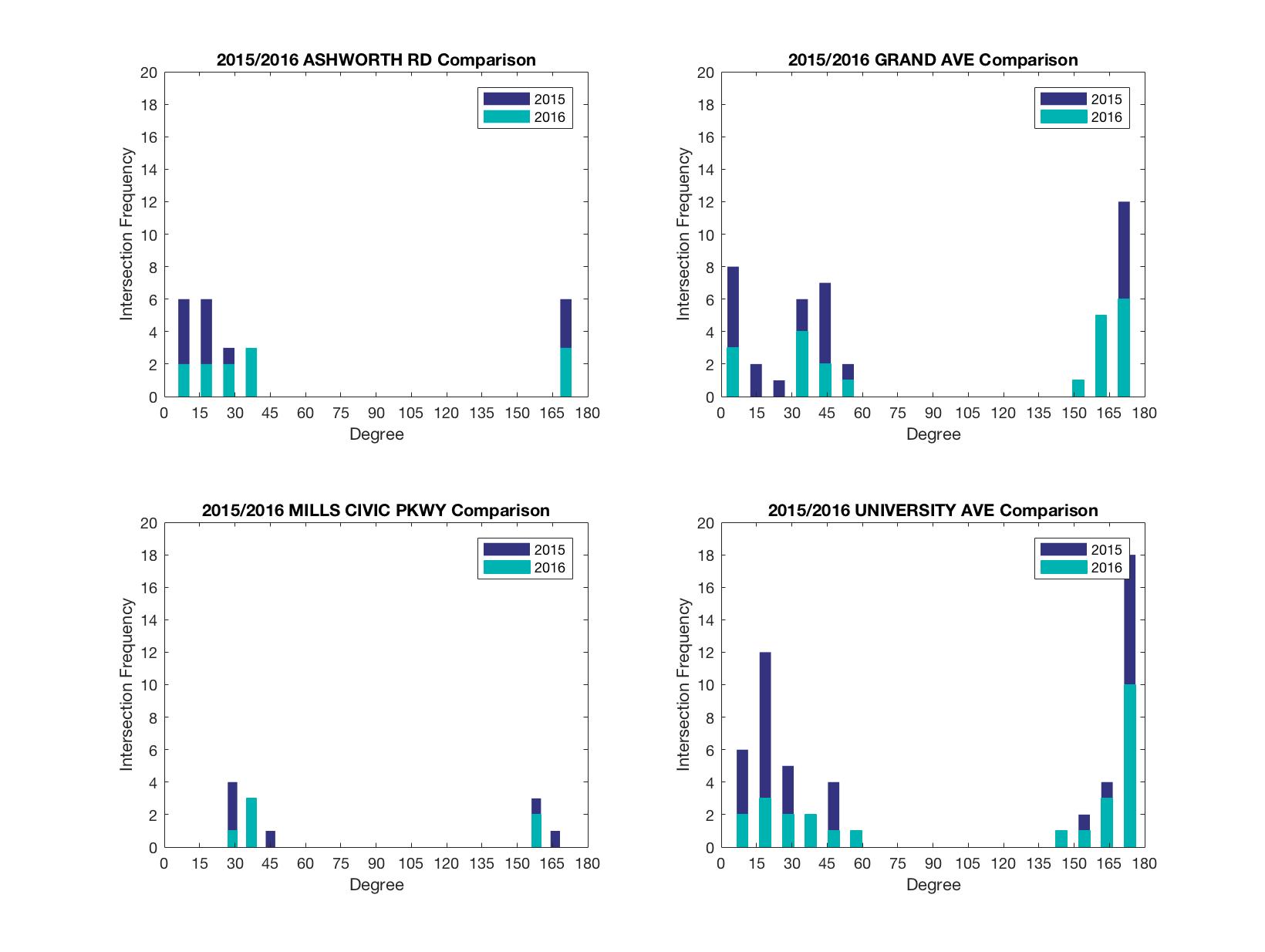


*Figure 4.2.1: Corridor Performance in 2015*



*Figure 4.2.2: Corridor Performance in 2016*

## 4.3 Annual Comparison of Corridors

The histograms below in Figure 4.3 show the annual corridor performance between 2015 and 2016. The traffic conditions changed along the Ashworth Rd. corridor from 2015 to 2016 due to an increase in traffic volume, which may explain the decrease in reliability in 2016. The travel time reliability increase from 2015 to 2016 on the University Ave. corridor. Adaptive signal control was implemented during this time along the corridor, which may be the reason behind the streamlined travel time performance between the years. For the Grand Ave. corridor, the reliability had changed for the better, although the traffic volume increased during 2016. In 2016, this corridor was expanded into a 4-lane road. Because of this, the work zones in the area during 2015 may had led to increased delays and therefore generated variability and inconsistencies. Also, according to the Des Moines Register, the parking garage that crosses 7th St. north on the Grand Ave. corridor was redeveloped and replaced during 2016. This construction event may have caused the abnormal traffic phenomenon displayed in the data. 

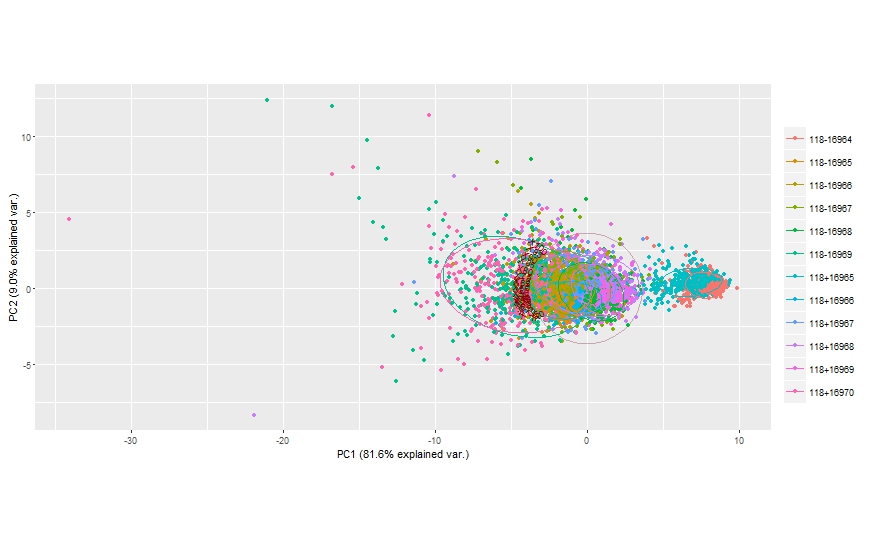
*Figure 4.3: Annual Comparison of Corridors*

5. Conclusion

The purpose of this analysis was to estimate the reliability of the travel time experienced along 4 corridors in West Des Moines. The travel time was estimated using probe data provided by INRIX for 2015 and 2016. The probe data was initially partitioned into percentile categories to efficiently estimate travel time performance and detect aggregate outliers. Following the filtering of incorrect data, a “Base Day” was established to represent the typical travel time along the corridor throughout the year. The variance of all the probe data travel times compared to the Base Day were calculated annually and a linear regression methodology determined the travel time reliability for each corridor. Results from the analysis determined that the reliability could be easily estimated through data visualization. General trends indicated that the corridor travel time became more reliable from 2015 to 2016. Construction impacts and work zone delays were also easily discoverable within the probe data; large variance in the data was evident when a known traffic flow suppressant occurred. Lastly, the addition of adaptive signal control on the University Ave. corridor was found to improve the reliability of the travel time estimates and reduce the variance in day to day operations.

6. Future Scope

Although the conducted analysis of travel time reliability for the 4 corridors of interest in West Des Moines was comprehensive, further research can be performed to refine the comparison methodology and increase the accuracy of the reliability metrics. By including pre-programmed signal timings into the corridor progression analysis, an ideal travel time could be computed based on the known length of the segment. This parameter could then be used to generate parameters such as the Travel Time Index (TTI), the Buffer Time Index (BTI), or the Planning Time Index (PTI). The benefit to computing the travel time reliability using an index is that the information can be easily understood by the system users. Additionally, the accuracy of the conclusions derived from the probe data analysis could be improved if vehicular trajectory data from other roadside sensors were included. This information could be incorporated into the data screening process to ensure travel times are only collected from through-moving vehicles. This would further refine the accuracy of the reliability estimates as only vehicles traveling through the entire corridor would be included in the analysis. Lastly, using a more robust methodology for analysis (such as the Principal Component Analysis) would allow for a multi-dimensional analysis of the percentile probe data. A linear regression analysis does not accommodate this dimensionality. A potential plot using the Principal Component Analysis method with the existing percentile probe data is shown in Figure 6.1



*Figure 6.1: Corridor Probe Data Analyzed with the Principal Component Analysis*

7. References

Agovic, A., Arindam, B., Auroop, G., Vladimir, P., 2009. Anomaly detection using manifold embedding and its applications in transportation corridors. Intell. Data Anal. 13, 435–455. doi:10.3233/IDA-2009-0375

Andersen, T. (2016, October 19). Part of Grand Avenue being expanded into 4-Lane Road. Retrieved March 22, 2017, from <http://www.kcci.com/article/part-of-grand-avenue-being-expanded-into-4-lane-road/6909458>

Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. LOF: Identifying Density-Based Local Outliers. Proc. 2000 ACM SIGMOD Int. Conf. Manag. data - SIGMOD ’00 29, 93–104. doi:10.1145/342009.335388

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 1–58. doi:10.1145/1541880.1541882

Chen, Z., Liu, X.C., Zhang, G., 2016. Non-recurrent congestion analysis using data-driven spatiotemporal approach for information construction. Transp. Res. Part C Emerg. Technol. 71, 19–31. doi:10.1016/j.trc.2016.07.002

Clustering: "Two's company, three's a crowd". https://quantdare.com/2016/07/clustering-two-company-three-crowd/Jackson, D.A., 2016.

Day, C., Haseman, R., Premachandra, H., Brennan, T., Wasson, J., Sturdevant, J., Bullock, D., 2010. Evaluation of Arterial Signal Coordination. Transp. Res. Rec. J. Transp. Res. Board 2192, 37–49. doi:10.3141/2192-04

Day, C.M., Remias, S.M., Li, H., Mekker, M.M., McNamara, M.L., Cox, E.D., Bullock, D.M., 2015. Performance Ranking of Arterial Corridors Using Travel Time and Travel Time Reliability Metrics. Transp. Res. Rec. J. Transp. Res. Board 23p. doi:10.3141/2487-04

Feng, Y., Hourdos, J., Davis, G.A., 2014. Probe vehicle based real-time traffic monitoring on urban roadways. Transp. Res. Part C Emerg. Technol. 40, 160–178. doi:10.1016/j.trc.2014.01.010

Gopi, P., Sachdeva, S.N., Bharati, A.K., Professor, E.-A., 2014. Evaluation of Travel Time Reliability on Urban Arterial. Artigo 3, 797–803.

J. (2015, January 01). 12 Developments to Watch this Year. Retrieved March 22, 2017, from <http://www.desmoinesregister.com/story/money/business/development/2014/12/31/des-moines-developments-watch/21126315/>

Lavrenz, S.M., Smith, W.B., Bullock, D.M., 2016. Assessing Longitudinal Arterial Performance and Traffic Signal Retiming Outcomes. Transp. Res. Board 95th Annu. Meet. 47906, 1–19. doi:10.3141/2558-07

Liu, H.X., Ma, W., 2009. A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. Transp. Res. Part C Emerg. Technol. 17, 11–26. doi:10.1016/j.trc.2008.05.002

Münz, G., Li, S., Carle, G., 2007. Traffic Anomaly Detection Using K-Means Clustering. GI/ITG Work. MMBnet.

Myung, J., Kim, D.-K., Kho, S.-Y., Park, C.-H., 2014. Travel Time Prediction Using k Nearest Neighbor Method with Combined Data from Vehicle Detector System and Automatic Toll Collection System. Transp. Res. Rec. J. Transp. Res. Board, 51–59. doi:10.3141/2256-07

Stopping Rules in Principal Components Analysis : A Comparison of Heuristical and Statistical Approaches Stable. pp. 2204–2214.

Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18, 267–276. doi:10.1007/BF02289263

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Statistical Methodol. doi:10.1111/1467-9868.00293

Wünsch, G., 2015. Bavarian Road Administration uses Probe Data for Large - Scale Traffic Signal Evaluation Support. doi:10.3141/2487-08