# Model based Methodology for Validation of Traffic Flow-detectors by Minimizing Human Bias in Video Data Processing

Pushkin Kachroo, *Senior Member, IEEE*, Neveen Shlayan, *Member, IEEE*, Alexander Paz, Shankar Sastry, *Fellow, IEEE*, and Shital Patel

***Abstract*—This paper provides a model-based method for analysis and hypothesis testing for paired data where one source of data has to be validated against another source of data that contains subjective and dynamic error. This study deals with human-observed flow counts collected from traffic videos of freeway cameras. The available videos are mainly used for the purpose of manual observation by transportation personnel in case of emergency. This amounts to a varying inconsistency of the quality of the videos which presents an additional challenge when analyzing the data. Video processing can not be performed due to the mentioned issues with regards to the video quality. The processing has to be performed manually by humans who unfortunately have an inherent bias. If the video data has to be used for validating flow detector sensors, then a technique that performs validation with subjective and dynamic erroneous data as a result of the human bias is needed. This paper presents a methodology to deal with this issue. It is based on statistical testing with heteroscedasticity which is demonstrated through a case study using data from traffic flow detectors and traffic cameras installed on highways in the Southern Nevada Region. A model for the relationship between the video ratings and the distribution of the human-errors is developed taking into consideration the human bias. A method for identification of faulty detectors is also demonstrated based on the developed technique.**

## I. Introduction

Detection is inevitably an essential element of Intelligent Transportation Systems (ITS). ITS relies on traffic flow detectors in order to evaluate the state of the transportation system in many of its aspects, for instance reliability, travel time, congestion, structural health, and safety. Detectors measurements and data are then used to support the making of decisions at various operational, financial, and jurisdictional levels. Moreover, most Nonintrusive Advanced Remote Sensors (NARS) systems entail calibration and/or validation to minimize errors, probably due to limitations of the technology in handling different vehicle characteristics, different combinations and mixtures of vehicles in space, and complex lane changing manoeuvres. Thus, the ability to validate the detectors and quantify their confidence levels is vital and of a high priority. Generally, if detectors are installed and are operational, detector data is readily available for flexible temporal and spacial ranges.

However, in order to validate the detector data, another set of data corresponding to the identical time period and location must be obtained and must be considered to be a "true" measurement for each and every detector. Obtaining data that is as close as possible to the true measurement is difficult. Regardless of the technique being used to collect the ground truth, the data is always going to be noisy. In this study, however, we particularly deal with traffic volume readings that were manually verified from traffic videos. This data is referred to as video data in the rest of this paper. The available videos were mainly used by the agency for the purpose of manual observation of the highways in case of emergency or any unexpected events such as incidents. For that purpose, the videos were tilted, zoomed in or out, and rotated based on the operational needs. This amounted to a varying inconsistency of the quality of the videos which presents an additional challenge since a dynamic error rate is associated with each video based on the difficulty of view and human errors. The dynamic nature of the error is due to the subjectivity present not only across the different human raters but also within the individual human rater. In this paper, we develop a statistical model-based technique to deal with this issue.

The problem encountered is very specific to the nature of the work, thus literature is very limited and does not directly target the specific methods used in this paper. However, various papers on general detector verification methods were studied. For instance, the work done by Lui [5] detects the traffic flow and considers the temporal and spatial bias using a mobility-based clustering model. In general, when two sets of data are compared where the population's standard deviation is unknown, paired student t- test is used. However, in this work every data point has a unique distribution that is different than the rest of the data points due to the dynamical error involved in each. It extends beyond Welch's t-test since not only standard deviation is different for the two data sets but also it varies among the data points. Therefore, a modified Welch's t-test is proposed and developed in this study to accommodate

P. Kachroo is a Professor, University of Nevada Las Vegas, Department of Electrical and Computer Engineering (e-mail: pushkin@unlv.edu).

N. Shlayan is an Assistant Professor, Department of Electrical Engineering, State University of New York, Maritime (e-mail: nshlayan@sunymaritime.edu).Corresponding author.

A. Paz is an Assistant Professor, Department of Civil Engineering, University of Nevada Las Vegas (e-mail: apaz@unlv.edu).

S. Sastry is with University of California Berkeley (e-mail: sastry@coe.berkeley.edu).

S. Patel is a Principal ITS/Traffic Engineer, Regional Transportation Commission / FAST (e-mail: patels@rtcsnv.com).

the mentioned errors. A model for the relationship between the video ratings and the distribution of the human-errors taking into consideration relative ratings is developed through repetitive experimental studies of a single video. In addition, this paper presents the use of the developed method for identification of faulty detectors.

In Kang's paper [3] an anisotropic magnetoresistive (AMR) sensor was tested by measuring traffic volumes on the highway with the detector and comparing with the exact traffic volumes in a highly congested traffic. It was found that performance of this sensor highly depended on the rate at which vehicles were flowing. Verification done in this study did not define "exact traffic volumes". Even though defining it is not the main scope of Kang's paper, statistical analysis must take error distributions under consideration for accuracy and reliability of results.

Interestingly, Chen's paper [1] presented an $L\infty$ norm Path flow estimator (PFE) model in order to handle inconsistencies of traffic counts and the systematic bias of the total demand estimate encountered in the PFE model. This technique was shown to be capable of determining the maximal absolute error needed to define the set of inequality constraints, traffic counts and capacities, while estimating the path flows. This research work even though not directly related to the scope of this paper, provides a very interesting approach for handling errors in counts.

In Fathy's research work [2], traffic movements at junctions were measured using image processing techniques. The results of the operations of the proposed algorithms show an error rate of 9.5%. For the purpose of the study presented in this paper, 9.5% error rate is not acceptable. Error rates in that same range are usually associated with commercial or open source software for video processing. Moreover, in our setting the task is even more difficult because of the variations in viewing angles and zoom values of the cameras. Hence, manual vehicle counts had to be performed for the work presented in this paper.

In Zhuang's work on statistical methods to estimate vehicle count using traffic cameras [8], two methods were developed for constructing traffic models. One model used statistical machine learning based on Gaussian models and the other used analytical derivation from the origin-destination (OD) matrix. It was found that Gaussian-based statistical learning method outperformed correlation coefficient based method. Simulations showed that it reduced the average estimation error by up to 72%. Variance estimation can also be provided. This method is particularly useful for roads with more dynamic and uncertain traffic. When training data is missing, the developed OD matrix based method is superior to the statistical learning methods. This is true provided that traffic is somewhat stable.

Zimmerman [9] studied type 1 error probability of the student t-test which arises due to unequal variances combined with unequal sample sizes. The Welch student t-test is known to eliminate these effects. Zimmerman found conditional probabilities of rejecting the null hypothesis with various conditions imposed on the sample variances. It was found that the inspection of sample data solely does not suffice when it comes to deciding to use a pooled-variance or separate-variance.

The overall algorithm developed in this paper involves the following six steps:

- Step 1: Collect flow detector data and corresponding videos for the same timeframe.
- Step 2: Obtain the video data by manual observation.
- Step 3: Conduct a human bias experiment on a selected video of a mid-difficulty of observation.
- Step 4: Obtain a bias-model for the relationship between the video ratings and the distribution of the human-error.
- Step 5: Use the bias-model to perform a modified Welch's t-test on the paired detector data and video data.
- Step 6: Identify potentially faulty detectors using the developed algorithm.

The first four steps develop a model for incorporating human bias in processing of video data and obtain its parameters from data. The next two steps then utilize this model to identify faulty detectors.

The developed methodology is presented in Section II. The case study is given in Section III and the data analysis is given in Section IV . Finally, conclusions are provided in Section V.

This paper addresses the issue of validating and identifying a given sensor by using another sensor as a benchmark. However, the given benchmark might also have errors that must be accounted for in order to improve the result of the validation and faulty sensor identification. We address this issue specifically for the case where the sensors to be validated are the flow detector sensors and the benchmark sensors are video detectors. If we use automated video processing to obtain the flow values and compare with the flow detectors, the errors inherent in video processing make the process unreliable. Human processing of the videos and then counting can reduce this error. However, some videos can be difficult to judge and if those videos are watched by multiple people, one can get varied results. This paper develops a completely new algorithm to solve this problem. Figure 1 shows the comparison between the three methods, the first method comparing the two systems, the second method with human processing which includes bias, and the third method using the new proposed algorithm.

Our algorithm first develops the model of the error which is created due to human error and then uses that in the statistical processing in a parametrized way to reduce its impact. Our paper describes the details of this new methodology by applying it to a specific example.

## II. Developed Methodology

In this section, we present a methodology to deal with the inherent bias of human-observed ground truth data, video data, giving rise to its subjective and dynamic erroneous nature. The method is based on Welch's t-test. It proposes a model for the relationship between human-errors and relative video ratings. It also uses these models to develop a methodology for

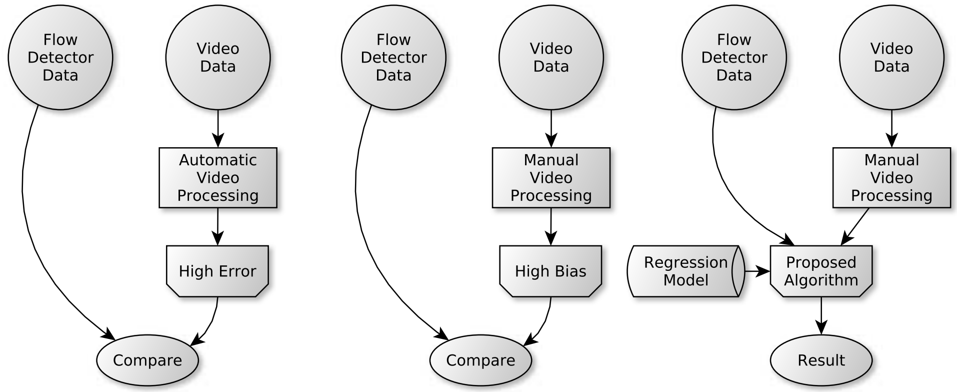



Fig. 1: Comparing the three Different Methods

identification of faulty detectors. The developed methodology will then be demonstrated through a case study in Section III. The next subsection II-A presents the fundamentals of testing two variables for the differences in their means in the case of equal and unequal variances. Subsection II-B develops the methodology when one of the variables is noisy, which is the case for the data studied in this paper.

### A. Fundamentals

If the actual number of vehicles detected is done accurately by the flow detectors and the manual video counting, then we expect the correlation between the two variables $X$ and $Y$ to be very high. Moreover, we expect that these two variables should not be independent. We will explore these two ideas further in Section IV.

The assumptions and the basic implementation of a paired t-test are standard statistical tools and can be found in many sources on statistics, such as [4]. In this section, we present the fundamentals on which the developed paired testing technique is developed. This technique deals particularly with handling biased data exhibiting subjective errors when used for validation. Given two independent random variables $X$ and $Y$, where $X$ has a normal distribution with $0$ mean and variance $1$ , and $Y$ has a chi-square distribution with $n$ degrees of freedom, then the random variable $T$ given by 1 has a t-distribution.

$$T = \frac{X}{\sqrt{Y/n}} \tag{1}$$

For the sample data, we use the random variable $X$ to correspond to the detector data to be verified $x_i$, and the random variable $Y$ to correspond to the sample data of the ground-truth, or video data, $y_i$. If we assume equal variances for $X$ and $Y$, letting $N$ be the number of observations, then we can use t-statistic as given by

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2 + s_y^2}{N}}} \tag{2}$$

Here, the sample means are given by:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N},\ \overline{y} = \frac{\sum_{i=1}^{N} y_i}{N} \tag{3}$$

and sample variances are:

$${s_x}^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1},\ {s_y}^2 = \frac{\sum_{i=1}^{N} (y_i - \overline{y})^2}{N-1} \tag{4}$$

Assuming unequal variance, we can use the the Welch's t-test (see [7]) based on the Welch-Satterthwaite equation (see [6]). For Welch's t-test, we use the t-statistic given by Equation 2. The degrees of freedom obtained by the approximation based on the Welch-Satterthwaite equation is

$$\nu = \frac{(\frac{s_x^2 + s_y^2}{N})^2}{\frac{s_x^4 + s_y^4}{N^2(N-1)}} \tag{5}$$

As the sample size increases the t-distribution tends to exhibit normal distribution characteristics with mean $0$ and variance $1$ and hence we can use the Z-test to test for the null hypothesis of the two random variables $X$ and $Y$ given that they have the same mean.

### B. Noisy Ground-truth $Y$

The problem statement for this paper deals with comparing flow detector data with data obtained from human estimation of the flow data watching the video of the same region. The video recordings are made with varying levels of zoom and also varying perspective angles. Hence estimating flow from the video detectors is not error free. The human processing of video data introduces human bias error, which we model to minimize its effect in comparison analysis for the two methods.

Let $r_i$ be a discrete value representing the video clarity score assigned by the various human counters. This essentially converts the problem into the analysis of the difference of the two random variables $X_i$ and $(Y_i + \eta_i(r_i))$ as provided in Equation 6. The noise term $\eta_i(r_i)$ is a stochastic random variable whose distribution depends on the rating parameter $r_i$.

$$E_i = X_i - (Y_i + \eta_i(r_i)) \tag{6}$$

Let $r_i \in [0, 10]$ be the discrete ranking for every lane where 0 corresponds to very clear and 10 corresponds to the highest difficulty with regards to clarity and counting ability. We let $d_{ij_i}$ be the counts of detector $D_i$ where $i \in \{1, 2, \cdots, N_d\}$ and $N_d = 146$ which is the total number of detectors; $j_i \in \{1, 2, \cdots, N_i\}$, where $N_i$ is the number of lanes for detector $D_i$. In the examined detector data, the number of lanes does not exceed five, that is, $N_i \in \{1, 2, \cdots, 5\}$. Corresponding to each $d_{ij_i}$, we have $v_{ij_i}$, which is the manual vehicle counts from the video camera $C_i$ corresponding to the detector $D_i$ for lane $L_{j_i}$. Similarly, We also have scores $w_{ij_i}$ given by the human observer for the difficulty level associated with obtaining the vehicle counts $v_{ij_i}$.

Now, let us denote the total number of comparisons by $N$ which is shown in Equation 7.

$$N = \sum_{i=1}^{N_d} N_i \tag{7}$$

We list each detector data in a single vector whose elements are given by $x_i$ where $i \in \{1, 2, \cdots, N\}$; similarly, the manual video data is given by $y_i, i \in \{1, 2, \cdots, N\}$, and the corresponding ratings of the video clarity measurements are given by $r_i, i \in \{1, 2, \cdots, N\}$. The percentage difference between $d_{ij}$ and $v_{ij}$ for location $i$ and lane $j$ is given by:

$$pd_{ij} = \frac{d_{i_{j_i}} - v_{i_{j_i}}}{h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})} 100 \tag{8}$$

where $h$ is given by Equation 9.

$$\begin{array}{llll} \text{if} & r_{i_{j_i}} \geq \rho & \text{then} & h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}}) = d_{i_{j_i}} \\ \text{otherwise} & & & h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}}) = v_{i_{j_i}} \end{array} \tag{9}$$

In Equation 9, $\rho$ is a threshold value of rating quality that is used to determine which of the two data points, taken from video or from detectors, is more reliable. In this study, $\rho$ is taken to be 5. The average percentage difference between the two sets of vehicle counts is given by:

$$\overline{pd} = \frac{\sum_{i=1}^{N_d} \sum_{j_i=1}^{N_i} \frac{d_{i_{j_i}} - v_{i_{j_i}}}{h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})}}{N} 100 \tag{10}$$

Equation 10 must be modified to account for the varying reliability of each data point. As mentioned previously, weighing each data point must consider the video ratings, subjectivity of the given ratings, and human-error. The weighted percentage difference is given by Equations 11 and 12.

$$wpd_{ij} = \frac{d_{i_{j_i}} - v_{i_{j_i}}}{h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})} \frac{1}{s(r_{i_{j_i}})} 100 \tag{11}$$

$$\overline{wpd} = \frac{\sum_{i=1}^{N_d} \sum_{j_i=1}^{N_i} \frac{d_{i_{j_i}} - v_{i_{j_i}}}{s(r_{i_{j_i}}) h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})}}{\sum_{i=1}^{N_d} \sum_{j_i=1}^{N_i} (1/s(r_{i_{j_i}}))} \tag{12}$$

Quantification of $s(r)$ will be demonstrated in the following section.

*1) Identification of Faulty Detectors:* For any detector which has any instance where condition 13 is true, we apply the test given in Equation 9. We take the percentage threshold to be denoted by $p_T h$. This threshold is the 90th percentile value for the absolute value of the function $h$ applied to every data entry.

$$\frac{d_{i_{j_i}} - v_{i_{j_i}}}{h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})} \geq p_{Th} \tag{13}$$

Given detector $D_i$,

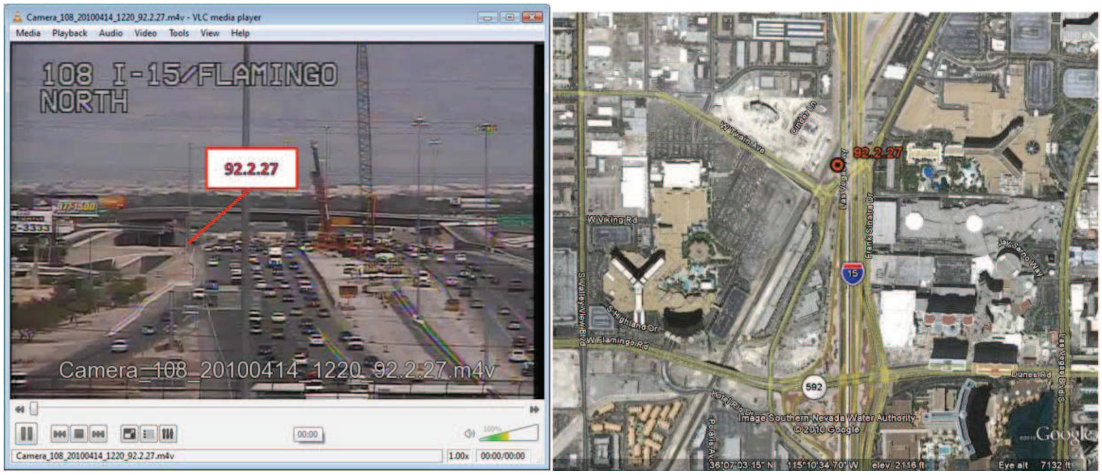
$$\begin{array}{ll} \text{if} & \text{then} \\ \dfrac{\sum_{j_i=1}^{N_i} \frac{d_{i_{j_i}} - v_{i_{j_i}}}{s(r_{i_{j_i}}) h(r_{i_{j_i}}, v_{i_{j_i}}, d_{i_{j_i}})}}{\sum_{j_i=1}^{N_i} (1/s(r_{i_{j_i}}))} > p_{Th} & D_i \text{ is defective} \\ \text{otherwise} & D_i \text{ is not defective} \end{array} \tag{14}$$

where $s(r_{i_{j_i}})$ is the relationship between the ratings and the distribution of the errors and is given by the curve fitting between the mean rating and the variance obtained from conducting an additional set of experiments targeted to quantify human bias in assigned ratings. The details of the process needed to obtain $s(r_{i_{j_i}})$ is given in Section IV.

## III. Data Description

Data from 146 detectors that are installed on Interstate-15 in the Las Vegas Metropolitan area in Nevada were provided, as shown in Figure 2. The detection technology is a combination of loop detectors and radar. Each device detects a number of lanes ranging from 1 to 5. Videos corresponding to the same time frame and location of a given detector were also provided. The time interval was 15 minutes for every detector. Manual counting was performed for every detector at each lane. It was noted that the quality of the videos was not only inconsistent among the different detectors but also among the various lanes within a video. To deal with this issue, each human counter assigned a discrete rating $r_i \in [0, 10]$ for each lane where 0 corresponded to the clearest view on the video and 10 corresponded to the highest difficulty with regards to clarity and counting ability.

A screenshot of a video from which the vehicle count was manually obtained is shown in Figure 3a. The location of the corresponding flow detector on Google Map is depicted in Figure 3b.




(a) Video for Traffic

(b) Location of the Flow Detector

Fig. 3: Traffic Video and Flow Detector Location


Fig. 2: Sensors, indicated by the circular markups, mounted on Interstate-15 North and South bounds

## A. Descriptive Statistics

The basic statistics for the variables $X$, $Y$, and $r$ are give in Listing 1.

Listing 1 indicates that the summary statistics for $X$ and $Y$ are, loosely speaking, close to each other. This is indicated by the fact, e.g. that the mean for $X$ is 275 and for $Y$ it is 270.7. Similarly, the values of the order statistics are close, such as for the first, second(median), and third quartiles. However, as the summary of the statistics for the difference of $X$ and $Y$ indicates that although their distributions have similar statistics, the percent difference shows more variation. We define the percent difference to be $D = 100(X - Y)/X$. The statistics for the percent difference are shown in Listing 2.

**Listing 1** Summary Statistics

```
       x                y                r
 Min.  : 12.0    Min.  : 21.0    Min.  : 0.000
 1stQ.:192.5    1stQ.:196.0    1stQ.: 1.000
 Med.  :286.0    Med.  :278.0    Med.  : 2.000
 Mean  :275.0    Mean  :270.7    Mean  : 2.566
 3rdQ.:352.0    3rdQ.:343.0    3rdQ.: 4.000
 Max.  :946.0    Max.  :973.0    Max.  :10.000
```

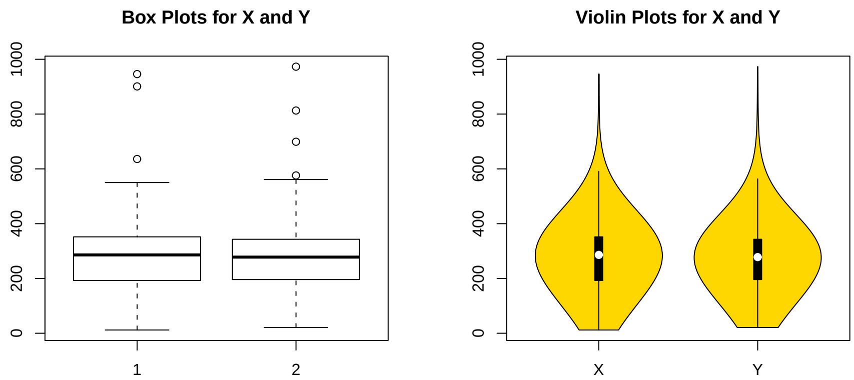
**Listing 2** Summary Statistics for the Difference

```
Min.     1stQ.   Median  Mean   3rdQ.  Max.
-1067.0 -4.822 0.495 -3.187 6.351 82.070
```

This analysis shows clearly that there are some outliers in the data based on the extreme values of the minimum and the maximum. To identify these outliers, we perform additional analysis. First, we present the boxplots and violin plots for the raw $X$ and $Y$ data. These are shown in Figure 4.

The boxplots show many outliers for $X$ and $Y$ data. The violin plots show the order statistical information and dispersion of these variables on the same plot. The same plots for the percent difference is shown in Figure 5.

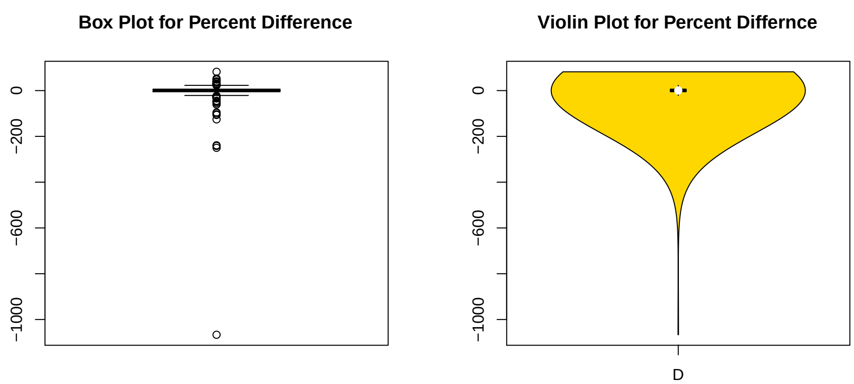
Figure 5 shows that there are some outliers in the data. The analysis of the outliers can help in identifying faulty flow detectors. We will remove the outliers for the current analysis, and then process the outliers after that to estimate faulty




(a) Box Plots

(b) Violin Plots

Fig. 4: Box and Violin Plots for X and Y




(a) Box Plot

(b) Violin Plot

Fig. 5: Box and Violin Plots for Percent Difference

detectors. Figure 6 shows the histogram plot for the percentage difference, $D = 100(X - Y)/X$ between the detector data and the video data, and its density plot.

## IV. Data Analysis

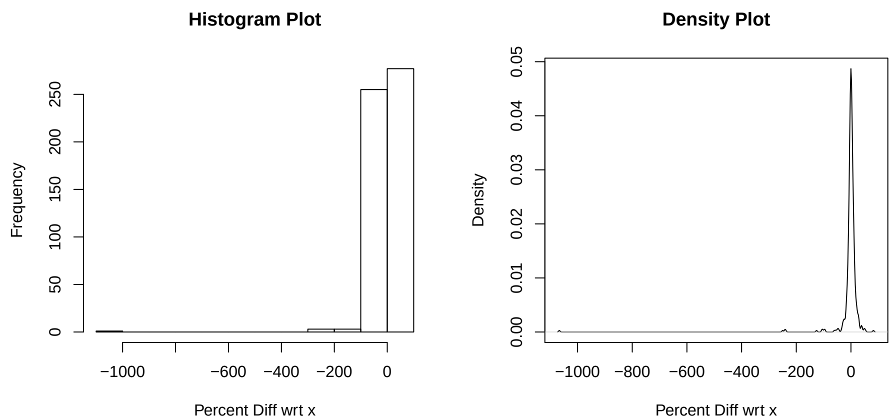
In this section, we present the inferential statistics based on the detector data, the manual video data, and the corresponding ratings. We also analyze the data obtained from the repeated manual video counting of a single site, to extract a model for the relationship between the ratings and the distribution of the errors. We use the open-source statistical package R for the descriptive and inferential data analysis.

### A. Determining $s(r)$

A methodology is needed in order to quantify the dynamic error that is inherently present in the video data due to the factors mentioned earlier. For this purpose, the ratings assigned by the human-counters are used. Subjective ratings were obtained experimentally by asking the human-counters to repeatedly rate and count the same video. The ratings for the five lanes of the video given by multiple raters is presented in Table I.

| Lane1 | r | Lane2 | r | Lane3 | r | Lane4 | r | Lane5 | r |
|---|---|---|---|---|---|---|---|---|---|
| 15.33 | 7 | 27.02 | 6 | -4.57 | 3 | -8.65 | 3 | 7.53 | 1 |
| 16.02 | 6 | 8.78 | 7 | 1.02 | 2 | -5.48 | 3 | 2.08 | 4 |
| 12.36 | 7 | 9.70 | 4 | 7.61 | 6 | -7.78 | 3 | 8.05 | 4 |
| 15.56 | 4 | 13.86 | 8 | -1.52 | 6 | -3.75 | 3 | 7.79 | 5 |
| 27.46 | 7 | 10.16 | 7 | -1.52 | 2 | -2.59 | 3 | 11.69 | 5 |
| 10.30 | 6 | 5.54 | 6 | 5.08 | 4 | -7.78 | 4 | 6.75 | 4 |
| 15.10 | 6 | 10.16 | 7 | -1.52 | 0 | -0.58 | 1 | 4.42 | 1 |

TABLE I: Ratings of the Video




(a) Histogram

(b) Density Plot

Fig. 6: Histogram and Density Plots for Percent Difference

The columns of the Table I show the percentage difference with respect to $X$ compared with the ratings for each lane of the video. For each lane, we obtain the mean rating as well as the variance of the percentages for that lane. These values are presented in Table II.

| | | | | | |
|---|---|---|---|---|---|
| Mean Rating | 6.14 | 6.42 | 3.29 | 2.86 | 3.43 |
| Variance Percent Difference | 29.75 | 48.83 | 18.20 | 9.25 | 9.14 |

TABLE II: Variance of the Percentage Difference in Counts vs Average Rating

The data in Table II is curve fitted using three curves, linear, power curve, and a log-linear curve. The plot of the data and the three curves are depicted in Figure 7. The code is presented in Appendix A.

The log-linear fit gives the best performance. Hence, we will use it for statistical inference. The formula for variance in terms of the rating score therefore, is given by Equation 15.

$$s(r) = e^{1.13+0.41r} \tag{15}$$

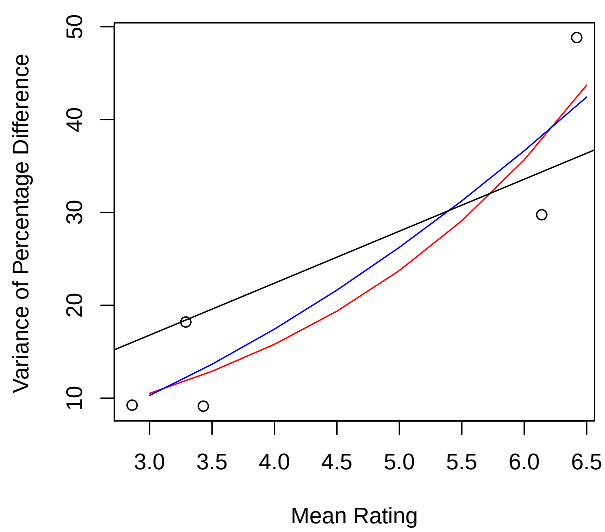



Fig. 7: Plot of Variance in Percent Difference to Mean Rating

### *B. Analysis of Non-faulty Detectors Using the Quantified Dynamical Error*

Now we perform paired t-test for the comparison of manual counting to the detector data from the non-faulty detectors. For this, we use weighted difference, where the weight is taken as the reciprocal of the curve fitting of the human error distribution and the mean ratings as we have developed in this paper. The data for the weighted differences has mean 0.03218569. The t-test results are shown in Listing 3.

The results show that the flow detectors work with a weighted accuracy of $3\%$. These results illustrate the advantage of using the proposed weighted approach in order to compare two data sets affected by subjective and dynamic measurement errors. That is, using the reciprocal of the variance to weight the data reduces artificial differences between the two data sets. Where the variance is obtained from an additional set of experiments with the goal of statistically quantifying the dynamic error. This reduction can be appreciated in Figure 8, where the density functions for the unweighted and weighted analysis are depicted, respectively. As depicted, There is significantly less variability between $d_{i_{j_i}}$ and $v_{i_{j_i}}$ when the weighted approach is used.

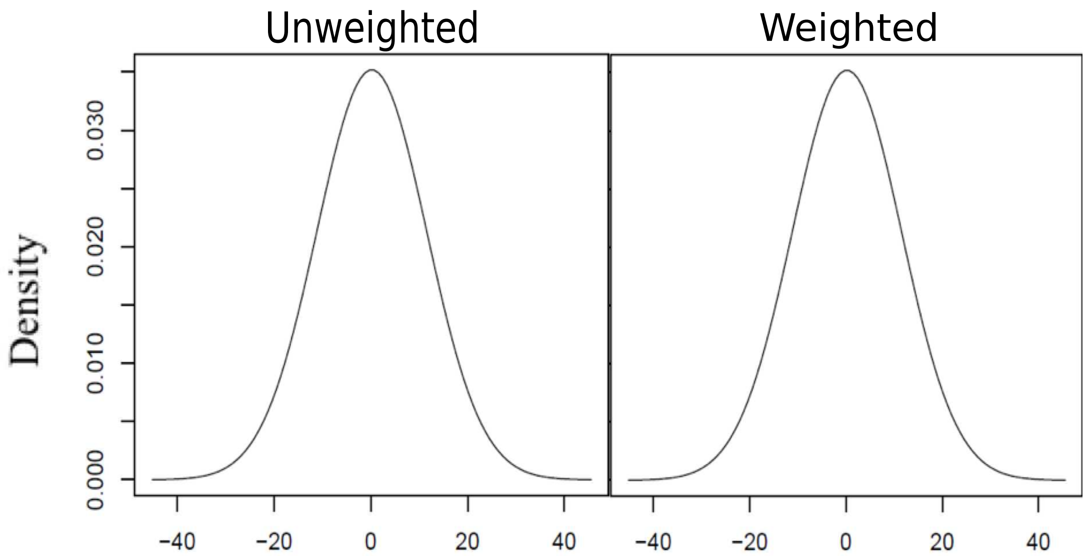



Fig. 8: Comparison results between unweighted and weighted t- statistics based on the developed subjective dynamic error

**Listing 3** Weighted t-test Results

```
> t.test(x)
        One Sample t-test
data:  x
t = 0.3742, df = 486, p-value = 0.7084
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.1368279  0.2011993
sample estimates:
 mean of x
0.03218569
```

### C. Identification of Potentially Faulty Detectors

The correlation between the flow detector values and the manually counted values is very high. The plot depicted in Figure 9 shows the linear relationship between the data and also the regression line obtained from the data.

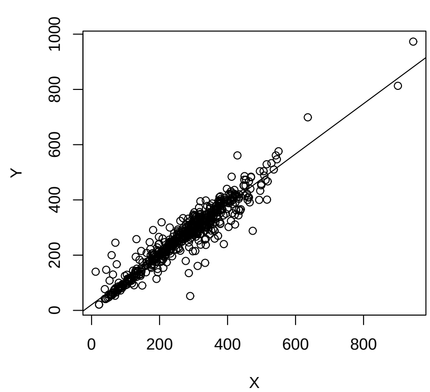



Fig. 9: Plot showing Linear Relationship

A call to the linear regression function in R produces the result shown in Listing 4.

**Listing 4** Regression Analysis Result

```
> fit <- lm(Y ~ X)
> summary(fit)

Call:
lm(formula = Y ~ X)

Residuals:
      Min        1Q    Median        3Q       Max
-232.4050  -15.5883   -0.8347   17.0857  160.7384

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.57962    3.92018    5.25  2.2e-07 ***
X            0.90974    0.01302   69.87  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 37.07 on 537 degrees of freedom
Multiple R-squared: 0.9009,     Adjusted R-squared: 0.9007
F-statistic:  4882 on 1 and 537 DF,  p-value: < 2.2e-16
```

The regression result shows strong linear relationship with the intercept $0.90974$. The values of Pearson, Spearman, and Kendall rank coefficients are given in Table III.

| **Coefficient Name** | **value** |
|---|---|
| Pearson Coefficient | 0.95 |
| Spearman Coefficient | 0.94 |
| Kendall Coefficient | 0.81 |

TABLE III: Correlation Coefficients

Applying the Empirical Cumulative Distribution Function (ECDF) on the percent difference data, we find that $90\%$ of the values are within $\pm 21\%$ difference. We use this percentage as the threshold for designing a decision system to identify faulty detectors. We use percentage of $X$ and also of $Y$ and identify all the detectors whose errors from either percentage is greater than $21\%$. Table IV shows some data that is analyzed to identify potentially faulty detectors. The columns of the data in Table IV are given in Table V.

| x | y | r | PDx | RoadID | SegID | Lane | DevID | Sign Diff | Pdy |
|---|---|---|---|---|---|---|---|---|---|
| 122 | 144 | 3 | -18.03 | 23 | 1 | 1 | 8 | -1 | -15.28 |
| 116 | 118 | 2 | -1.72 | 23 | 2 | 1 | 8 | -1 | -1.69 |
| 71 | 81 | 2 | -14.08 | 39 | 2 | 1 | 8 | -1 | -12.35 |
| 277 | 179 | 7 | 35.38 | 39 | 2 | 2 | 8 | 1 | 54.75 |
| 132 | 258 | 8 | -95.45 | 39 | 2 | 3 | 8 | -1 | -48.84 |
| 43 | 147 | 9 | -241.86 | 56 | 2 | 1 | 13 | -1 | -70.75 |
| 146 | 215 | 9 | -47.26 | 56 | 2 | 2 | 13 | -1 | -32.09 |
| 230 | 300 | 9 | -30.43 | 56 | 2 | 3 | 13 | -1 | -23.33 |
| 319 | 328 | 9 | -2.82 | 56 | 2 | 4 | 13 | -1 | -2.74 |
| 439 | 366 | 9 | 16.63 | 49 | 2 | 1 | 12 | 1 | 19.95 |
| 340 | 378 | 9 | -11.18 | 49 | 2 | 2 | 12 | -1 | -10.05 |
| 340 | 303 | 9 | 10.88 | 49 | 2 | 3 | 12 | 1 | 12.21 |
| 286 | 135 | 9 | 52.80 | 49 | 2 | 1E | 12 | 1 | 111.85 |
| 372 | 270 | 9 | 27.42 | 49 | 2 | 2E | 12 | 1 | 37.78 |
| 328 | 343 | 3 | -4.57 | 49 | 3 | 1 | 15 | -1 | -4.37 |
| 400 | 368 | 3 | 8.00 | 49 | 3 | 2 | 15 | 1 | 8.70 |
| 367 | 353 | 3 | 3.81 | 49 | 3 | 3 | 15 | 1 | 3.97 |

TABLE IV: Sample Data

| | |
|---|---|
| $x$ | Flow detector data |
| $y$ | Manually Counted Video data |
| $r$ | Rating |
| $PDx$ | Percent Difference $100(X-Y)/X$ |
| $RoadID$ | Road identification number |
| $SegID$ | Road segment identification number |
| $Lane$ | Lane number |
| $DevID$ | Flow detector identification number |
| $SignDiff$ | $(X-Y)/\|X-Y\|$ |
| $PDy$ | Percent Difference $100(X-Y)/Y$ |

TABLE V: Variables in the Data

The algorithm to find out if a given detector is possibly faulty is as follows. For any detector which has any instance where condition 13 is true, we apply the the test given in Equation 14. We take the percentage threshold $p_{Th}$ to be equal to $21\%$. This threshold is the $90^{th}$ percentile value for the absolute value of the function $h$ applied to every data entry.

Table VI shows the data analysis where the possible faulty detectors are identified by red cells. Unique detectors are identified by unique combinations of RoadID and SegID fields. The pink colored cells show the detectors that satisfy the condition given by Equation 13. The red colored cells show the detectors that satisfy the condition given by Equation 14.

This analysis reveals that 18 sensors require further investigation as they could be potentially faulty. It is recommended that further investigation involves actual vehicle counts so as to minimize any uncertainty.

### D. Complexity and Scalability

In this subsection we study the complexity of the algorithm and its implication issues. The algorithm requires the building of the bias model. However, that study is performed only once, and is used to build the regression model. Hence, this step has the same scalability as a standard regression model. Once, this step is complete, the main modification comes in the form of using a modified formula where we use the term $s(r_{i_{j_i}})$ for scaling. Hence, using this term does not add any complexity order to the processing of the data.

## V. Conclusions

In this study, model based approach is proposed and developed to accommodate subjective and dynamic errors inherently present in ground-truth video data from traffic videos used for verification of flow detector data . A model for the relationship between the video ratings and the distribution of the human-errors taking into consideration relative video clarity ratings is developed through repetitive experimental studies of a single video. Identification of faulty detectors was performed based on the developed methodology. Analysis was performed using flow detector data from a case study spanning Interstate-15 in the Las Vegas region from which we obtained a model relating the ratings to consistency variation. The results illustrated the advantage of the proposed approach over the standard Welch's t-test. The developed weighting technique provides a better estimate of the true performance of the detection system. The weighted data provides statistics for the hypothesis test and confidence interval that are closer to the corresponding values calculated using data with minimum errors. Comparing the two data sets without taking into account the reliability of the information may result in erroneous conclusions. For example, determining faulty sensors without weighting the information may erroneously increase the number of potentially faulty sensors.

## Appendix A
## Curve Fitting Listings

The analysis of this data for curve fitting is presented in the following code. The data for the mean ratings is stored in variable $rm$, and the variance of the percent differences in $rxv$, as seen in Listing 5.

**Listing 5** Variables for Mean rating and Variance

```
> rm
[1] 6.14 6.42 3.29 2.86 3.43
> rxv
[1] 29.75 48.83 18.20  9.25  9.14
```

The data is stored in a data-frame as shown in Listing 6.

**Listing 6** Data frame for the Variables

```
> ds <- data.frame(x = rm, y = rxv)
> str(ds)
'data.frame':   5 obs. of  2 variables:
 $ x: num  6.14 6.42 3.29 2.86 3.43
 $ y: num  29.75 48.83 18.2 9.25 9.14
```

| x | y | r | PDx | RoadID | SegID | Pdy | 1/s( r ) | pd | apd | v_pd | sumK | sumN | Def |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 122 | 144 | 3 | -18.03 | 23 | 1 | -15.28 | 0.09 | -15.28 | 15.28 | 1.44 | | | |
| 116 | 118 | 2 | -1.72 | 23 | 2 | -1.69 | 0.14 | -1.69 | 1.69 | 0.24 | | | |
| 71 | 81 | 2 | -14.08 | 39 | 2 | -12.35 | 0.14 | -12.35 | 12.35 | 1.76 | | | |
| 277 | 179 | 7 | 35.38 | 39 | 2 | 54.75 | 0.02 | 35.38 | 35.38 | 0.65 | | | |
| 132 | 258 | 8 | -95.45 | 39 | 2 | -48.84 | 0.01 | -95.45 | 95.45 | 1.16 | 0.17 | 3.56 | 20.64 |
| 43 | 147 | 9 | -241.86 | 56 | 2 | -70.75 | 0.01 | -241.86 | 241.86 | 1.95 | | | |
| 146 | 215 | 9 | -47.26 | 56 | 2 | -32.09 | 0.01 | -47.26 | 47.26 | 0.38 | | | |
| 230 | 300 | 9 | -30.43 | 56 | 2 | -23.33 | 0.01 | -30.43 | 30.43 | 0.25 | | | |
| 319 | 328 | 9 | -2.82 | 56 | 2 | -2.74 | 0.01 | -2.82 | 2.82 | 0.02 | 0.03 | 2.60 | 80.59 |
| 439 | 366 | 9 | 16.63 | 49 | 2 | 19.95 | 0.01 | 16.63 | 16.63 | 0.13 | | | |
| 340 | 378 | 9 | -11.18 | 49 | 2 | -10.05 | 0.01 | -11.18 | 11.18 | 0.09 | | | |
| 340 | 303 | 9 | 10.88 | 49 | 2 | 12.21 | 0.01 | 10.88 | 10.88 | 0.09 | | | |

TABLE VI: Faulty Detector Identification Worksheet Sample

The code and interaction for linear regression model with zero intercept (so as not to have non-positive variance) is given in Listing 7.

**Listing 7** Linear Regression

```
> rfit <- lm(rxv ~ rm + 0)
> abline(rfit)
> summary(rfit)

Call:
lm(formula = rxv ~ rm + 0)

Residuals:
       1        2        3        4        5
 -4.6118  12.9012  -0.2121  -6.7556 -10.0556

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
rm   5.5964     0.8732   6.409  0.00304 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.145 on 4 degrees of freedom
Multiple R-squared: 0.9113,     Adjusted R-squared: 0.8891
F-statistic: 41.08 on 1 and 4 DF,  p-value: 0.003045
```

The code and interaction for the power regression model is given in Listing 8.

The code and interaction for the log-linear regression model is given in Listing 9.

**Listing 8** Power Regression

```
> m.2 <- nls(y ~ rhs(x,intercept, power), data = ds, start =
list(intercept = 1, + power = 2),trace = T)
183.6028 :  1 2
180.8829 :  1.159547 1.916102
178.7074 :  1.355049 1.833367
178.1608 :  1.376444 1.831493
178.1608 :  1.376735 1.831399
> summary(m.2)

Formula: y ~ rhs(x, intercept, power)

Parameters:
          Estimate Std. Error t value Pr(>|t|)
intercept   1.3767     1.4787   0.931   0.4205
power       1.8314     0.6047   3.029   0.0564 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.706 on 3 degrees of freedom

Number of iterations to convergence: 4
Achieved convergence tolerance: 5.62e-06

> plot(rm,rxv)
> lines(s, predict(m.2, list(x = s)), lty = 1, col = "blue")
```

## References


[1] A. Chen, S. Ryu., and P. Chootinan. $l_\infty$-norm path flow estimator for handling traffic count inconsistencies: Formulation and solution algorithm. *Journal of Transportation Engineering*, 136:565575, 2010.

[2] M. Fathy and M. Y. Siyal. Measuring traffic movements at junctions using image processing techniques. *Pattern Recognition Letters*, 18:493–500, 1997.

[3] M. H. Kang, B. W. Choi, K C. Koh, J. H. Lee, and G. T. Park. Experimental study of a vehicle detector with an amr sensor. *Sensors and Actuators A*, 118:278–284, 2005.

[4] Erwin Kreyszig. *Introductory Mathematical Statistics: Principle and Methods*. John Wiley & Sons Inc., 1970.

[5] Siyuan Liu, Yunhuai Liu, Lionel Ni, Jianping Fan, and Minglu Li. Towards mobility-based clustering. *The 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington DC:USA, 2010.

[6] F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2:4–27, 1946.

[7] B. L. Welch. The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34:28–35, 1947.

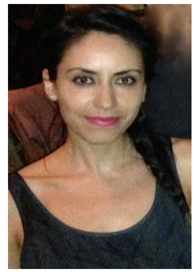
[8] P. Zhuang, Y Shang, and B. Hua. Statistical methods to estimate vehicle count using traffic cameras. *Multidim System Sign Process*, 20(7):121–133, 2009.

[9] D. W. Zimmerman. Conditional probabilities of rejecting h0 by pooled and separate-variances t tests given heterogeneity of sample variances. *Communications in Statistics*, 33:69–81, 2004.

**Listing 9** Loglinear Regression

```
> m.e <- nls(y ~ I(exp(1)^(a + b * x)), data = ds, start =
list(a = 0,+ b = 1), trace = T)
508654.9 :  0 1
61724.38 :  -0.5294191  0.9373290
6359.583 :  -0.4243993  0.7921467
542.5241 :  0.4408596 0.5634763
165.3981 :  1.0201739 0.4315203
159.0372 :  1.1207667 0.4090216
159.0220 :  1.1312375 0.4071317
159.0220 :  1.1320508 0.4069949
159.0220 :  1.1321105 0.4069849
> summary(m.e)

Formula: y ~ I(exp(1)^(a + b * x))

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a   1.1321     0.7566   1.496   0.2315
b   0.4070     0.1253   3.248   0.0475 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.281 on 3 degrees of freedom

Number of iterations to convergence: 8
Achieved convergence tolerance: 3.461e-06

> lines(s, predict(m.e, list(x = s)), lty = 1, col = "red")
> title(xlab="mean rating")
> plot(rm,rxv, xlab="Mean Rating",ylab="Variance of Percentage
Difference")
> lines(s, predict(m.e, list(x = s)), lty = 1, col = "red")
> lines(s, predict(m.2, list(x = s)), lty = 1, col = "blue")
```

**Pushkin Kachroo** is currently a visiting Professor at the University of California at Berkeley working with Professor Shankar Sastry. He is also the Lincy Professor of Transportation in the department of Electrical and Computer Engineering at the University of Nevada Las Vegas (UNLV). He is the director of the Transportation Research Center at UNLV and also the Associate director of the Mendenhall Innovation Program at the Thomas a Hughes College of Engineering at UNLV. He was an Associate Professor at Virginia Tech before he joined UNLV in 2007. He obtained his Ph.D. in Mechanical Engineering from University of California at Berkeley performing research in Vehicle Control in 1993 under Professor Masayoshi Tomizuka, and obtained another Ph.D. in Mathematics from Virginia Tech in Mathematics in the area of hyperbolic system of partial differential equations with applications to Traffic Control and Evacuation under Professor Joseph A. Ball. He has authored 10 books on traffic and vehicle control. He has authored more than 120 publications that include books, research papers, and edited volumes. He has taught about 30 different courses in Virginia Tech in the areas of electrical and computer engineering, and mathematics. Similarly, he also taught 30 different courses at UNLV since 2007. He has graduated more than 35 graduate students, and has been P.I. or Co P.I. on projects worth more than 4 Million Dollars. He was awarded the most outstanding new professor at Virginia Tech, and also has received many teaching awards and certificates, both at Virginia Tech and UNLV. He also received the faculty excellence award from CSUN UNLV in 2011.



**Neveen Shlayan** is currently an Assistant Professor at the State University of New York, Maritime. Previously, she worked at Philips Research on large scale lighting control for transportation systems. She was a research fellow at the Singapore-MIT Alliance for Research and Technology working on mesoscopic traffic simulations. She obtained her Ph.D. in Electrical Engineering from University of Nevada, Las Vegas in 2011 performing research in Cyber Physical Systems under the direction of Professor Pushkin Kachroo. She has also obtained a Masters in Mathematics in 2012 in the area of reconstruction of neutron density distributions. She has authored a book and a number of journal papers.

**Alexander Paz** Assistant Professor, P.E. Department of Civil and Environmental Engineering and Construction Transportation systems analysis, traffic operations and control, demand modeling and forecasting, intelligent transportation systems, human behavior and learning, traffic flow theory, and statistical econometric methods.

**Shankar Sastry** is the Dean of the College of Engineering at the University of California, Berkeley. He currently holds the Nippon Electronics Corporation (NEC) Distinguished Professorship in the College of Engineering and the Walter A. Haas School of Business. He was elected to the National Academy of Engineering in 2001 and is a recipient of the Donald P. Eckman Award in 1990. Sastry obtained his bachelor degree from Indian Institute of Technology, Bombay (1977), his master and PhD degrees from University of California, Berkeley (1979, 1980, 1981). His PhD advisor was Professor Charles Desoer. He became an Assistant Professor at Massachusetts Institute of Technology (1980-2), Assistant Professor, Associate Professor, and Professor at University of California, Berkeley (1983-present).

**Shital Patel** is a Principle ITD/Traffic Engineer, Regional Transportation Commission at the Freeway Arterial System of Transportation.