Journal of Transportation Technologies, 2023, 13, 615-631
https://www.scirp.org/journal/jtts
ISSN Online: 2160-0481
ISSN Print: 2160-0473

# Traffic Flow Characteristics in Work Zone and Non-Work Zone Environment and Its Impact on Road Crashes at the Segment Level

**Dorcas O. Okaidjah, Christopher M. Day**

Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, USA
Email: okaidjah@iastate.edu, cmday@iastate.edu

**How to cite this paper:** Okaidjah, D.O. and Day, C.M. (2023) Traffic Flow Characteristics in Work Zone and Non-Work Zone Environment and Its Impact on Road Crashes at the Segment Level. *Journal of Transportation Technologies*, **13**, 615-631.
https://doi.org/10.4236/jtts.2023.134027

**Received:** July 21, 2023
**Accepted:** September 17, 2023
**Published:** September 20, 2023



## Abstract

This study investigates relationships between congestion and travel time performance metrics and crashes on road segments. The study focuses on work zone routes in Iowa, utilizing 2021 commercially-available probe vehicle data and crash data. Travel time performance metrics were derived from the probe vehicle data, and crash counts were obtained from the crash data. Additional variables included road characteristics (traffic volume, road type, segment length) and a categorical variable for the presence of a work zone. A mixed effect linear regression model was employed to identify relationships between road segment crash counts and the selected performance metrics. This was accomplished for two sets of models that include congestion performance measures at different defining threshold values, along with travel time performance measures. The study results indicate that the congestion indicators, certain travel time performance measures, and traffic counts were statistically significant and positively correlated with crash counts. Indicator variables for rural interstate locations and non-active work zones have a stronger influence on crash count than those for municipal interstate locations and active work zones. These findings can inform decision-makers on work zone safety strategies and crash mitigation planning, especially in high traffic volume areas prone to congestion and queues.

## Keywords

Probe Data, Congestion Mile Hours, Queue Mile Hours, Speed Deficit, Crashes, Work Zones

## 1. Introduction

There is considerable literature on the relationship between crashes and congestion but relatively little on the influence of probe data-derived performance metrics on road segment crashes. A few prior studies have sought to explore such relationships. Dimitrijevic *et al.* developed multiple models that employed performance metrics to access short-term crash risks on highway segments using reactive data (driver and vehicle age) and proactive data (speed, volume to capacity ratio, etc.) [1]. The results show the integration of reactive and performance metrics related data resulted in better prediction of crashes.

Several previous studies have shown that the presence of work zones influences crash frequency [2] [3] [4]. Work zone environments pose a danger to drivers and account for approximately 700 fatalities, 24,000 injury crashes, and 52,000 non-injury crashes annually in the United States [2]. In 2020, work zones accounted for approximately 102,000 crashes and 857 fatalities in the United States [5]. Work zones tend to have varying speeds, and differences in speed have been shown to influence crashes, especially rear-end crashes in work zone locations [6]. Though less severe, higher congestion levels are associated with more crashes [7].

To date, limited studies have examined the influence of travel performance metrics on crash frequency using probe data and statistical methods. The present study seeks to do so at the road segment level. Several performance measures are defined in the present study that quantifies the amount of congestion occurring on the segment as well as travel time characteristics. Additionally, few studies have also included work zone status integrated with probe data performance metrics. The present study also includes this variable to determine how congestion and travel time characteristics of segments, as revealed by probe data co-influence the safety performance along with work zone activity. The study gives insight into how congestion and queueing conditions influence road segment crashes in different road environment conditions assisting transportation safety engineers in enforcing strategies to reduce road segment crashes. Therefore, the study aims to evaluate performance metrics in work zone and non-work zone environments and how such metrics influence road segment crashes.

## 2. Literature Review

In recent years, probe data has been used to analyze performance metrics on roadways [8] [9]. Such performance measures are able to provide detailed information about how roadways are operating. Some studies have employed traffic flow data to analyze crashes [1] [10] [11].

Mekker *et al.* [12] employed probe data to characterize crash rates on interstate segments by classifying crashes as related to queued or free-flowing conditions. Three years of probe vehicle data and crash data were used. The study revealed that more crashes occur on congested interstate segments than on uncongested ones. Also, commercial vehicles accounted for over 87% of fatal crashes during congestion and 39% in free flow conditions.

Probe data has also been used to assess delays resulting from work zone [13] [14]. Performance measures quantifying congestion, queueing, and travel time reliability have been widely used by agencies to monitor mobility performance. However, sparse literature exists where these metrics have been assessed in relation to crash occurrences in work zones. A recent study utilized probe data to conclude that historical speed variation is essential in predicting work zone mobility [15]. Although some research has been conducted on the impact of work zone presence on crashes, most prior studies were not examined on the road segment level, as in the present study.

Other studies have utilized multiple data sources to access work zone safety. Speed data has been employed to examine speed limit compliance in work zones using a Tobit regression model. Study findings concluded a higher probability of speeding in the presence of other speeding vehicles and when traffic volumes are high [16]. Tobit model has also been employed to examine crashes at the intersection level [17]. Ozturk *et al.* [4] examined relationships between work zones and highway safety. Preliminary findings indicated a 24.4 percent increase in crash rate under the work zone condition. Negative binomial regression models employed in the study showed that work zones significantly increase the risk of crashes on roads. Another study utilized meta-analysis and meta-regression models to explore the influence of work zones on safety performance. Study findings revealed work zone length significantly affects crash count, but the work zone duration does not. Both work zone duration and length have a positive influence on crash count [3]. Additionally, another study, although not on work zones, integrated real time traffic data from loop detectors: average flow per lane, average occupancy, etc., and weather data to understand the severity and probability of crashes in urban arterials [11].

Researchers have examined the influence of work zone presence on crash types. Using negative binomial models, Khattak *et al.* [2] analyzed the effects of work zone presence on injury and non-injury crashes. Data for the study included crash, road inventory, and work zone activity data. The results indicated that the work zone crash rate was 21.5% higher than the preceding non-work zone period for freeway segments. Crash frequency increased with the work zone duration, segment length, and AADT. Work zone duration significantly increased injury and non-injury crash frequency. Aside from the influence of work zones on crash types, crash severity and collision type have also been investigated. An Ohio study examined fatal and injury crashes in interstate work zones using spatial analysis over three years. The results showed significant differences in fatal and injury crash proportions and rear-end crashes at various work zone locations. The study concluded that work zone activity areas were danger prone, and varying speeds contributed to rear-end crash occurrences [6]. Another study investigated fatal crashes in work zone and non-work zone locations in Georgia. The authors concluded that work zone fatal crashes mostly involved more than one vehicle and are rarely induced by vertical and horizontal alignment compared to non-work zone crashes. The results indicated that work zones influence the collision type of fatal crashes [18]. A Missouri study employed statistical data analysis to identify factors influencing work zone crash severity. Several factors were examined, including road conditions, weather, light condition, traffic conditions, segment geometry, and type of crash, among others. A multinomial logistic regression model was employed to identify significant factors and compare different crash severities: property damage, minor injuries, and fatal injuries. The results indicated that the majority of the crashes were property damage crashes and rear end collision type crashes [19].

## 3. Data and Methods

This study considered road segments where work zones were present (active) in some months of the year 2021. Active work zones represent road segments with some roadwork executed, while non-active work zones are those with no roadwork. This study examined 201 road segments with active work zones. The proprietary road segmentation scheme used by the vendor was employed for this study. Road segments in this scheme ranged between 0.068 - 0.971 miles. Information on work zone duration was also obtained from the Iowa Department of Transportation (DOT). **Table 1** shows the time periods when work zones were active on the selected routes.

Crash data were obtained from the Iowa DOT for the entire year of 2021. The crash data included crash key, date and time, crash severity, latitude, longitude, Route ID, and nearest milepost. A linear referencing scheme data containing information on road attributes such as length, number of lanes, etc., was also gathered from the Iowa DOT. Probe vehicle data was obtained from the vendor for all work zones listed in **Table 1**. The vehicle speeds were of interest from this dataset. This data consists of 1-minute average speeds for the segments previously mentioned. Along with the average speed, a "confidence" score is also included, which indicates whether the speed is based on real traffic observations or is inferred from historical data (because few probes are available). A confidence score of 30 which is the highest confidence score INRIX reports, was employed for this study. The score is the measure of confidence for the speed estimated for the segment. A score of 30 is based on real time data, unlike other

**Table 1.** 2021 Work zone information.

| Work zone name | General Location | Start Date | End Date |
|---|---|---|---|
| Group Ick | I-80, 0.5 mi east of Iowa Route 224 | 9/1/2021 | 11/4/2021 |
| Group Ibq | I-80, 2.5 mi east of Iowa Route 224 | 5/1/2021 | 11/4/2021 |
| Group 1cm | I-80, 1 mi west of Iowa Route 146 | 4/5/2021 | 8/31/2021 |
| Group 1cn | I-80, 2.5 mi west of Iowa Route 117 | 4/5/2021 | 10/14/2021 |
| Group 1co | I-35, bridges over US 30 | 4/5/2021 | 7/2/2021 |
| Group 3q | I-29, 1.4 mi north of Iowa Route 127 | 4/5/2021 | 8/19/2021 |
| Group 4ce | I-80, near US 6/169 interchange | 6/21/2021 | 12/8/2021 |
| Group 5an | I-35, 3.8 mi north of Iowa Route 92 | 4/1/2021 | 12/31/2021 |
| Group 5ao | I-80, near Exit 211 | 4/5/2021 | 8/19/2021 |

scores: 10 or 20 based on historical or a combination of expected and real time data. The present study includes only observations based on real data. The speed data was used to calculate average speed, free flow travel time, actual travel time, speed deficit, travel time index, congestion mile hours, and queue mile hours, as described in the next section. The next section also details the methodology for the study.

## Definition of Performance Measures

Travel Time Index is one measure for quantifying the congestion level on a road network. It is defined as the ratio of the actual travel time to the travel time under free-flow conditions on the same road segment:

$$\mathrm{TTI} = \frac{\text{Actual Travel Time}}{\text{Free Flow Travel Time}}$$

The free-flow travel time is often regarded as the baseline or "ideal" scenario, with no traffic-induced delays. The TTI provides a relative measure of the additional time needed to complete a journey during congested conditions compared to free-flow situations. The actual travel time was estimated for each road segment using the probe data average speed combined with the road segment length from Iowa DOT's linear referencing scheme. The free flow travel time was calculated using the probe date average speed and the posted speed limit of the road segment. Subsequently, the travel time index was obtained by dividing the actual travel time over the free flow travel time.

We define another performance measure, speed deficit, as the difference between the free-flow speed and the actual speed for a specified location or during a specific time period. This definition is slightly different than what was used in some previous mobility reports [20], which used a 45 mph reference speed rather than the free flow speed. Speed deficit quantifies the reductions in speed occurring on the link. A large speed deficit value can result from shorter periods of significant speed reductions or more extended periods with more moderate speed reductions. Overall, it tends to reflect the overall separation between ideal flow and actual flow. **Figure 1** shows an example heat map view of speed deficit for eight days for 14 segments in a work zone. The chart shows that the greatest amounts of congestion occurred on days 24, 27, and 28 in the month of September and were mostly limited to the first 8 hours of the day. The variation from day to day could be caused by varying work zone activities, fluctuating traffic demand, or a combination of these and other factors.

This paper uses Congestion Mile Hours (CMH) and Queue Mile Hours (QMH) to quantify congestion at different degrees of severity. Some previous mobility reports used CMH to total the amount of congestion along different routes for comparison purposes [20]. To calculate CMH, the number of 1-minute observations where speed is less than or equal to 45 mph are tabulated, divided by 60 to convert into hours, and then multiplied by the segment length to convert into mile-hours. QMH is the product of the duration of the queue in hours and the

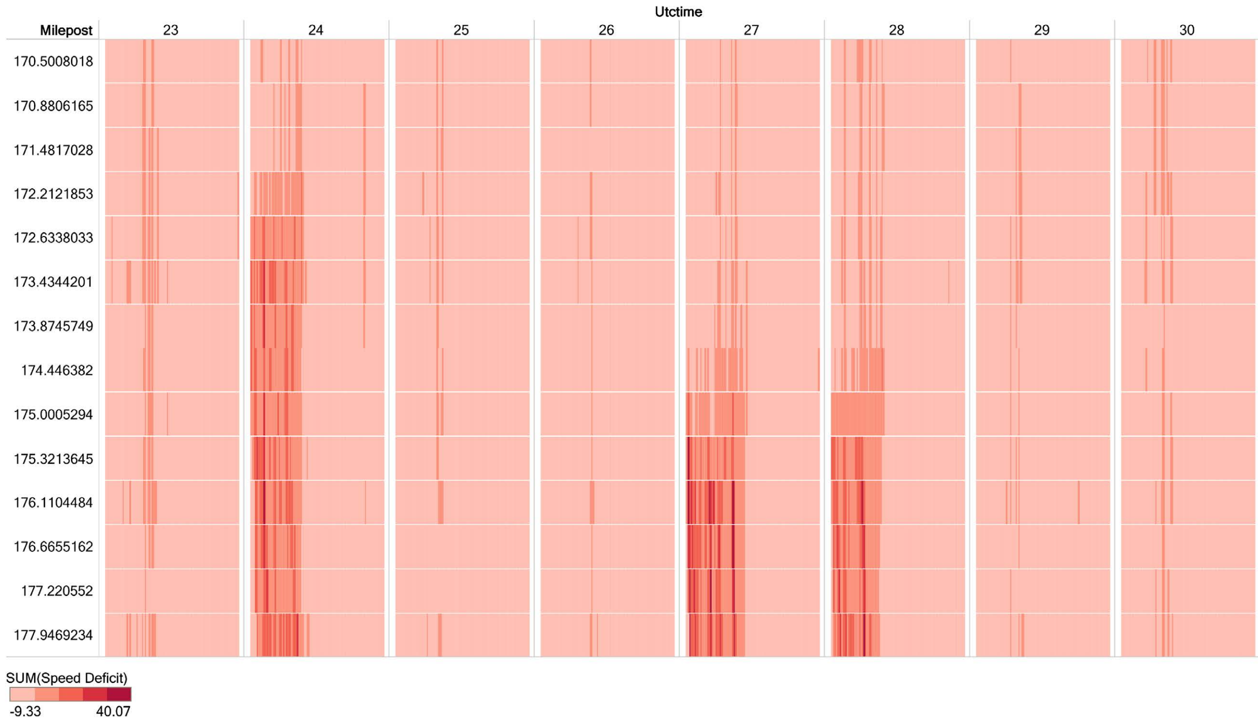


**Figure 1.** Speed deficit for a work zone for an 8-day period.

length of queue in miles. We set the threshold value of speed to 15 mph because it indicates a critical breakdown with traffic flow. Whereas CMH includes both moderate and severe congestion, QMH is intended to include only the most severe congestion, where traffic is experiencing stop-and-go conditions. CMH and QMH are useful performance measures for assessing road performance, identifying locations needing improvement, and evaluating strategies for congestion management.

Figure 2 displays example calculations of the number of CMH and QMH for an example work zone across the year 2021 for two directions of travel. As the chart shows, the number of CMH is greater than QMH (since QMH has a lower threshold speed), and there is considerable fluctuation in both performance measures over the year. For this work zone, the most severe congestion seems to occur around weeks 37 - 43 in the westbound direction.

The probe data were merged to the linear referencing scheme on the segment ID, so we obtained information on the road characteristics, such as segment length, traffic volume, Route ID, from measure (beginning milepost) and to measure (ending milepost). To obtain the count of segment crashes, we gathered road asset management crash data from IowaDOT, which had the following information: crash key, Route ID, milepost, and datetime. This data was merged with the crash key from the crash data, so we obtained a new crash dataset that had the crash key, Route ID, milepost and datetime, x coordinate, y coordinate, and crash severity. Since we had the milepost from the new crash dataset, we wrote a script to retrieve the count of crash if it falls within the from measure and to measure of a road segment.

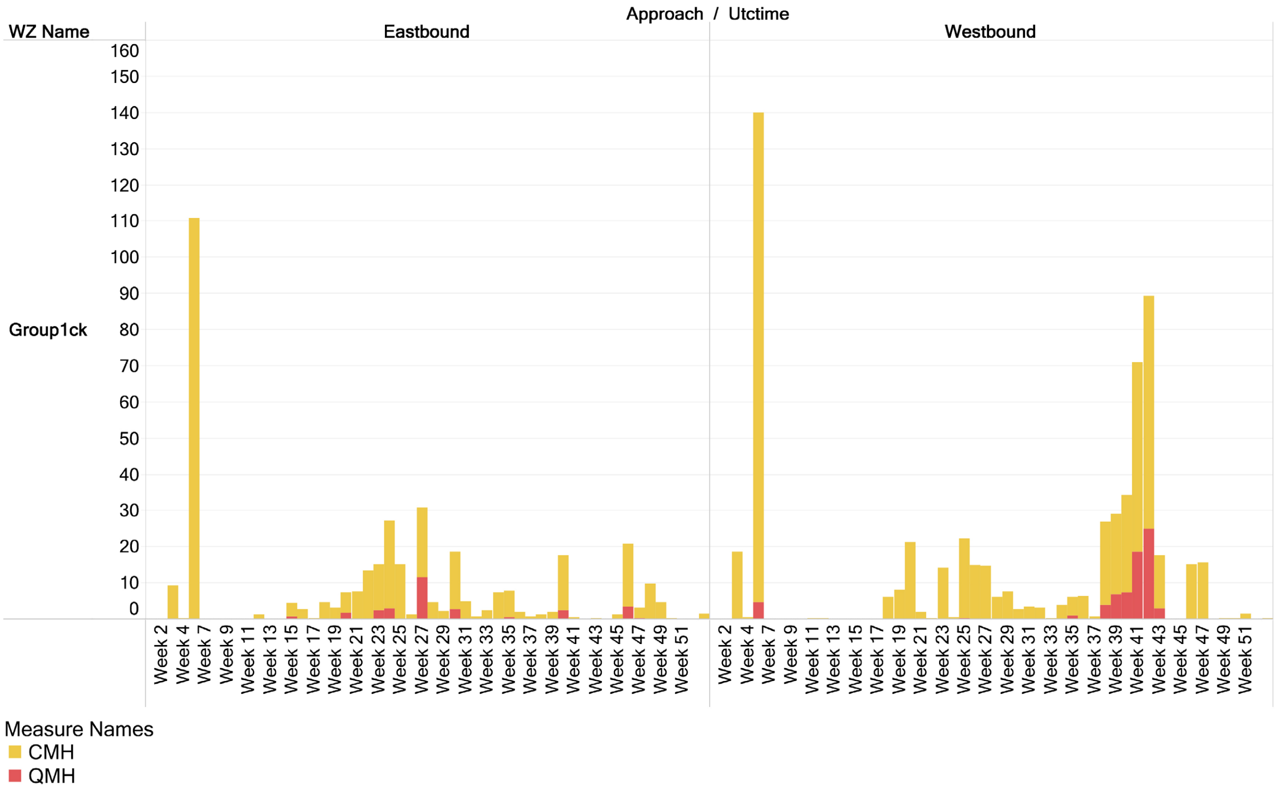


**Figure 2.** CMH and QMH for a work zone for the year 2021.

The descriptive statistics of the variables are provided in **Table 2**. **Table 2** shows that the average QMH is 10% higher for work zone months than for non-work zone months. However, there is a higher average CMH for non-work zone months than for work zone months. The average crash count in non-work zone months is much higher than in work zone months by a percentage of about 26%. The average travel time index has a similar value of about 1.0 for work zone and non-work zone months.

The variables were tested for multicollinearity. A high degree of correlation between two or more predictor variables makes it unlikely that the individual effects can be separated, resulting in unreliable estimates of regression coefficients. The correlation matrix is shown in **Figure 3**. It provides pairwise correlation coefficients for all the predictor variables, with values ranging from −1 to 1, which signifies the degree and direction of the linear relationship between two predictors. A coefficient close to 1 represents a strong positive linear relationship, indicating that as one predictor increases, the other is also likely to increase. Conversely, a coefficient close to −1 indicates a strong negative linear relationship, meaning as one predictor increases, the other tends to decrease. A coefficient near 0 suggests little to no linear correlation between the predictors. Based on the results of the correlation matrix analysis, we eliminated TTI and average speed from our modeling process. TTI had a strong positive correlation with speed deficit and a strong negative correlation with average speed. Average speed had a strong negative correlation with speed deficit.

**Table 2.** Descriptive statistics of variables (Segment level).

| Analysis Periods | Variable | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| **All months** | Crash count | 1.544 | 1.899 | 0 | 11 |
| | CMH | 11.714 | 11.819 | 1.026 | 76.508 |
| | QMH | 1.332 | 2.576 | 0 | 16.538 |
| | AADT | 28,677 | 6740 | 1670 | 42,700 |
| | TTI | 1.005 | 0.021 | 0.979 | 1.189 |
| | Speed deficit | −0.043 | 1.425 | −1.966 | 11.034 |
| **Work zone months only** | Crash count | 0.6542 | 1.1720 | 0 | 11 |
| | CMH | 4.349 | 7.0149 | 0.0107 | 50.499 |
| | QMH | 1.295 | 2.241 | 0.0105 | 16.134 |
| | AADT | 28,815 | 6514 | 14,600 | 42,700 |
| | TTI | 1.005 | 0.0316 | 0.9743 | 1.208 |
| | Speed deficit | −0.146 | 1.764 | −1.966 | 11.034 |
| **Non-work zone months only** | Crash count | 0.879 | 1.384 | 0 | 10 |
| | CMH | 7.279 | 7.254 | 0.602 | 55.793 |
| | QMH | 1.159 | 2.713 | 0.008 | 16.134 |
| | AADT | 28,815 | 6514 | 14,600 | 42,700 |
| | TTI | 1.006 | 0.0169 | 0.989 | 1.185 |
| | Speed deficit | 0.060 | 0.969 | −1.204 | 10.232 |

A quantile-quantile plot with the variables chosen is indicated in Figure 4. A mixed effect linear model was suitable for analysis based on the approximate linearity of the data, as shown in Figure 4. Also, the road segment length and curve differ, although most of the roads are relatively straight. The mixed effect linear model is suitable to account for this randomness in road characteristics.

We considered two statistical models since CMH and QMH are slightly correlated, having a correlation coefficient of 0.61 as indicated by Figure 1. The two models included different congestion metrics as independent variables:

- CMH, AADT, speed deficit
- QMH, AADT, speed deficit

Road geometry characteristics such as traffic volume, number of lanes and road types were also considered for inclusion in the models. The number of lanes was eliminated since all the roads had two lanes in each direction of travel, hence did not present any variance that would be useful in the analysis. Finally, a categorical variable for work zone status was created from the time periods in which the work zones were active.

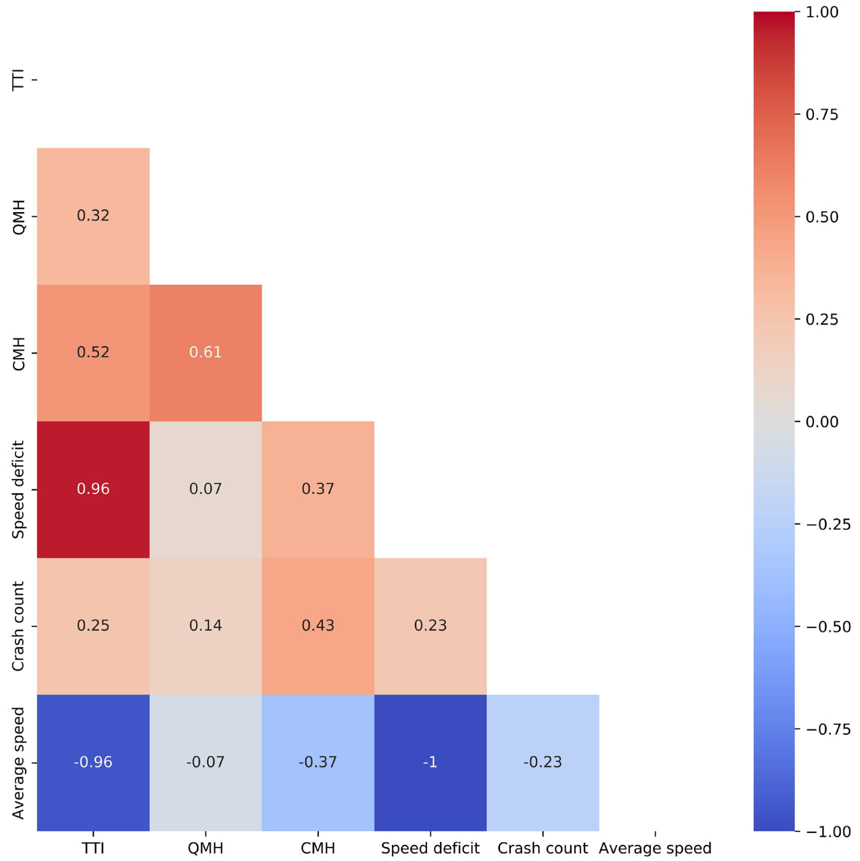


Figure 3. Correlation matrix of travel time performance metrics.

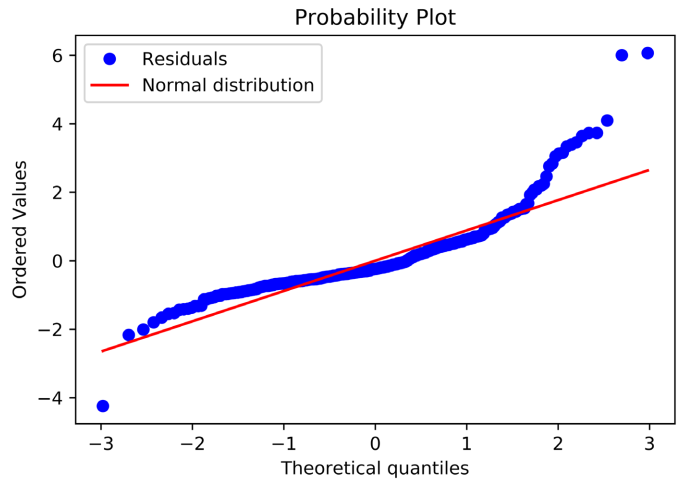


Figure 4. Quantile-quantile plot of predictor variables.

# 4. Results

## 4.1. Visualizations of Congestion Mile Hour (CMH) and Queue Mile Hour (QMH)

Figure 5 shows CMH and QMH along with AADT for work zone months, and Figure 6 shows the same for non-work zone months. The travel time index is

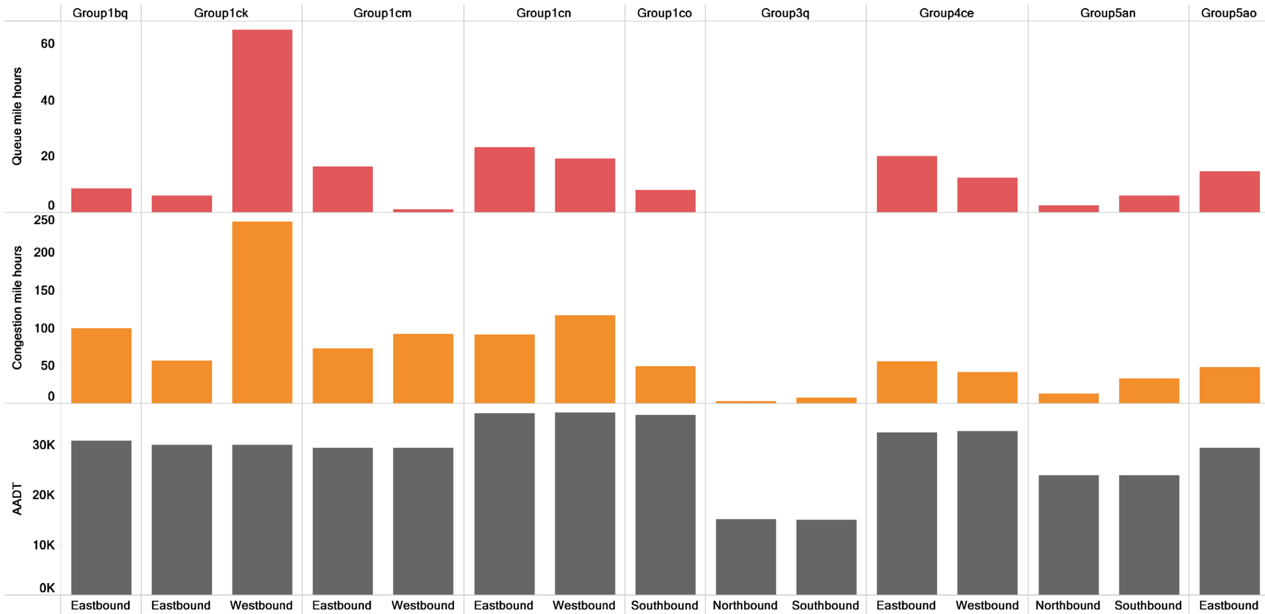


**Figure 5.** Travel time performance metrics for work zone months.

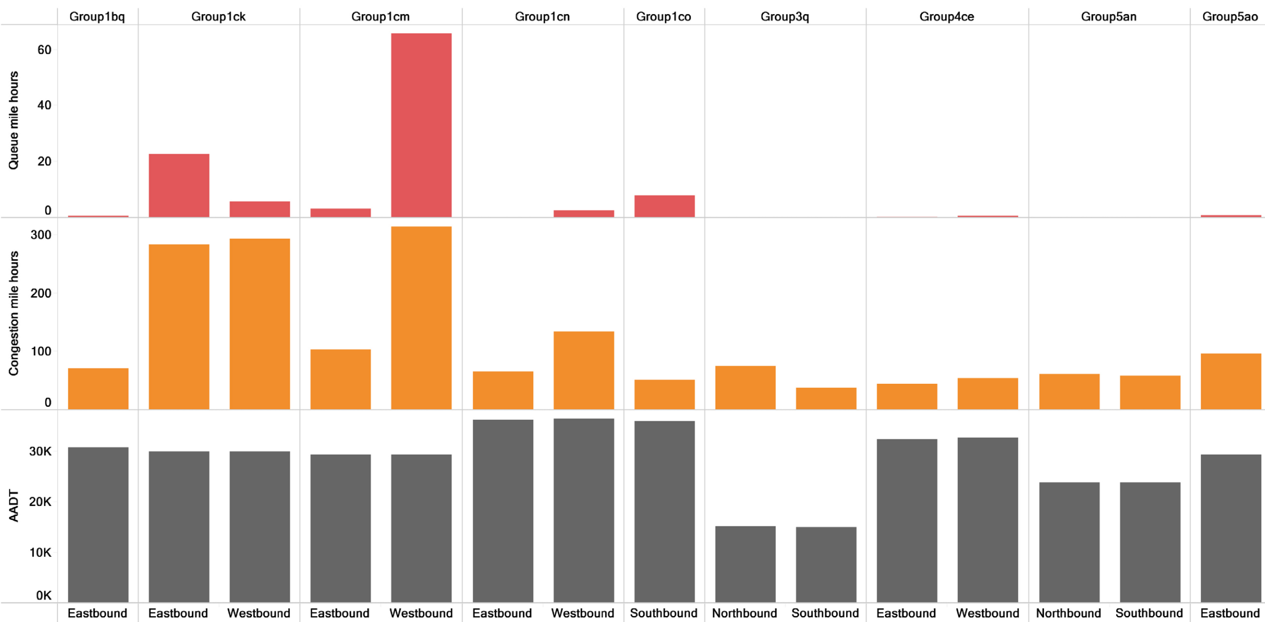


**Figure 6.** Travel time performance metrics non-work zone months.

overlaid on the CMH bar chart to allow comparison. The work zone names and travel directions are on the horizontal axes, and the vertical axes are AADT, CMH, and QMH. In Figure 5, work zone Group 1ck westbound experienced the largest amount of CMH and QMH. This work zone also has the highest traffic volume. This suggests, unsurprisingly, that higher traffic volumes correlate with increased CMH and QMH congestion.

Figure 7 visualizes the safety performance of the work zones examined in this study as a line chart where the work zones are sorted by the number of crashes

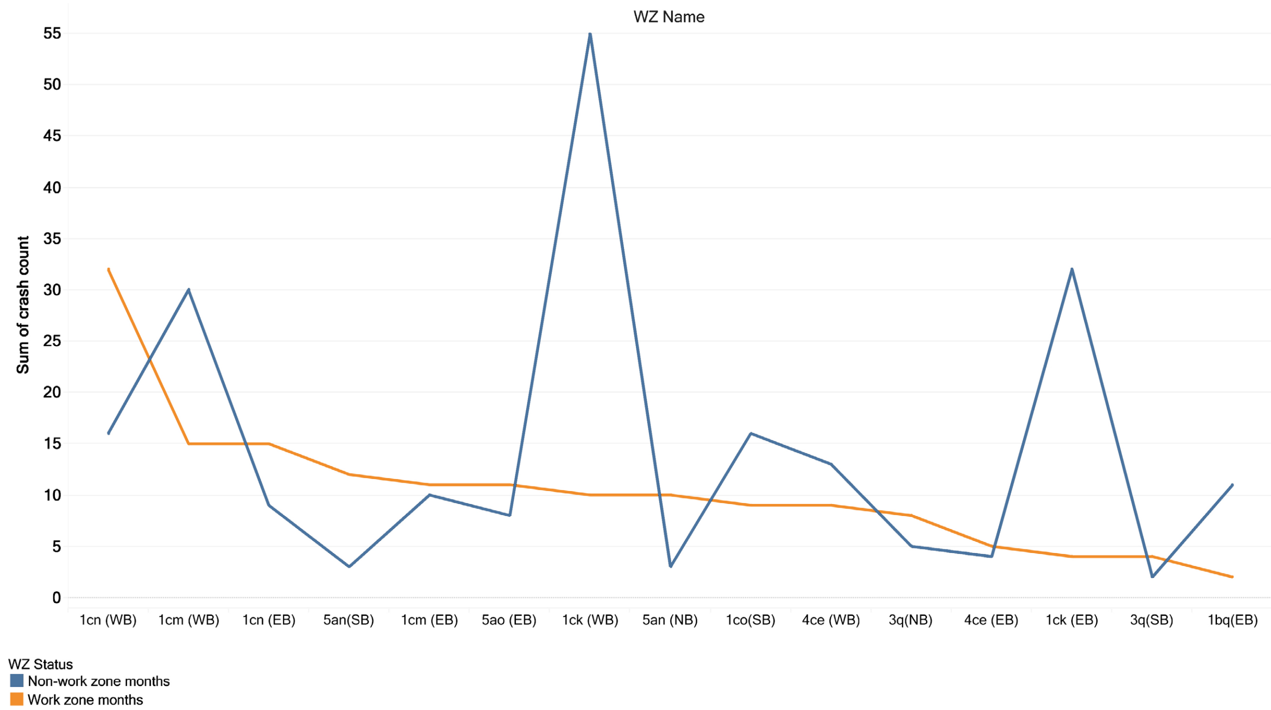


**Figure 7.** Crash count line chart for work zone and non-work zone months.

occurring in work zone months. The monotonically decreasing orange line shows this sorted value, while the blue line shows the number of crashes occurring in non-work zone months. The two lines do not share the same trend, demonstrating that non-work zone crash performance relative to other segments does not necessarily indicate which segments will perform better or worse when a work zone is in place. Variation in the nature of the road work (e.g., whether a lane needs to be closed or not) likely drives these differences.

### 4.2. Model Results and Discussion

The research utilized a mixed-effects linear model due to the linear nature of the data combined with the random variability inherent in the attributes of the road segments.

The mixed effect linear model results for the CMH and QMH models are shown in Table 3 and Table 4, respectively. Both models were trained on 480 observations grouped by 201 unique identifiers (road segments). The minimum group size is 2, and the maximum is 4, with an average group size of 2.4.

For the CMH model, traffic volume, speed deficit, and CMH are statistically significant variables positively correlated with the crash count. Although not statistically significant, the segment length is more positively correlated with crash count than the other variables considered, which are also positively correlated. For every one-unit increase in speed deficit, the crash count increases by an estimated 0.108 units. For every one-unit increase in CMH, traffic volume, and segment length, the approximate increase in crash count is 0.059, 0.534, and 0.667 units, respectively.

**Table 3.** Mixed linear model regression results (CMH model).

| Model | MixedLM | | Dependent Variable: | | Crash count | |
|---|---|---|---|---|---|---|
| No. observations | 480 | | Method: | | REML | |
| No. groups | 201 | | Scale | | 1.103 | |
| Min. group size | 2 | | Log-Likelihood | | −750.048 | |
| Max. group size | 4 | | Converged | | Yes | |
| Mean group size | 2.4 | | | | | |
| | Coef. | Std. Err | z | P > \|z\| | [0.025 | 0.975] |
| Intercept | −5.722 | 2.443 | −2.342 | 0.019 | −10.511 | −0.933 |
| C (Road type) (T. Rural intestate) | 0.295 | 0.341 | 0.865 | 0.387 | −0.373 | 0.963 |
| C (WZ status) (T. Non-active) | 0.024 | 0.099 | 0.245 | 0.807 | −0.170 | 0.219 |
| CMH | 0.059 | 0.009 | 6.397 | 0.000 | 0.041 | 0.077 |
| Speed deficit | 0.108 | 0.043 | 2.522 | 0.012 | 0.024 | 0.192 |
| Segment length | 0.667 | 0.424 | 1.563 | 0.116 | −0.164 | 1.498 |
| Log (Aadt) | 0.534 | 0.228 | 2.346 | 0.019 | 0.088 | 0.981 |
| Group Var | 0.221 | 0.072 | | | | |

**Table 4.** Mixed linear model regression results (QMH model).

| Model | MixedLM | | Dependent Variable: | | Crash count | |
|---|---|---|---|---|---|---|
| No. observations | 480 | | Method: | | REML | |
| No. groups | 201 | | Scale | | 1.127 | |
| Min. group size | 2 | | Log-Likelihood | | −766.263 | |
| Max. group size | 4 | | Converged | | Yes | |
| Mean group size | 2.4 | | | | | |
| | Coef. | Std. Err | z | P > \|z\| | [0.025 | 0.975] |
| Intercept | −9.307 | 2.541 | −3.662 | 0.000 | −14.288 | −4.327 |
| C (Road type) (T. Rural intestate) | 0.492 | 0.363 | 1.355 | 0.175 | −0.220 | 1.204 |
| C (WZ status) (T. Non-active) | 0.202 | 0.098 | 2.062 | 0.039 | 0.010 | 0.395 |
| QMH | 0.057 | 0.031 | 1.860 | 0.063 | −0.003 | 0.117 |
| Speed deficit | 0.217 | 0.041 | 5.297 | 0.000 | 0.137 | 0.298 |
| Segment length | 1.313 | 0.443 | 2.966 | 0.003 | 0.445 | 2.181 |
| Log (Aadt) | 0.850 | 0.238 | 3.573 | 0.000 | 0.384 | 1.316 |
| Group Var | 0.321 | 0.085 | | | | |

For the QMH model, the road type coefficient for rural interstate is 0.492, implying that when a segment is a rural interstate (compared to a municipal interstate), the crash count is expected to increase by 0.492 units, holding all explanatory variables constant. It's an absolute increase in the speed difference, not a relative change. The P-value (0.175) suggests this result isn't statistically significant at the 5% level. The coefficient for non-active work zone status is 0.202; this means the crash count increases by 0.202 compared to when it's active when all the independent variables remain the same.

The results of both models show that all the included variables have positive correlations with crash counts. Speed deficit, road segment length, and traffic volume are positively correlated with crash count with coefficient values of 0.217, 1.313, and 0.850, respectively. We observe that the most positively correlated variable for CMH was segment length, while that of the QMH model was traffic volume. We also observe that the speed deficit coefficient is higher in the QMH than in the CMH model. We attribute this to the fact that the travel time is much higher in what we have defined as queued conditions (speed > 15 mph) than in congested but not queued conditions (speed > 45 mph but <15 mph).

Both models indicate that there are more crashes in non-work zone months than in work zone months. This may be because, during work zone conditions, there are usually traffic control devices on the road to warn drivers of potential upcoming congestion, whereas, during non-work zone conditions, there may be little to no warning. In addition, non-work zone months had a larger positive coefficient in the QMH model when compared to the CMH model. This suggests that queued conditions may experience few crashes in work zone months because, when QMHs are accrued, the speed reductions have become so severe that the queuing is likely to have cascaded backward to other segments outside the defined limits of the work zone limits. Crashes are less likely to occur when the speeds are very low.

In contrast, more moderately congested conditions (with high CMH) are more likely to indicate the back of a queue, which is a much more hazardous condition with high-speed traffic approaching slow traffic. This finding agrees with a prior study which concluded that congested conditions were associated with more crashes than uncongested conditions [12]. Another study on expressways in Japan reported that congested conditions were associated with 70% of work zone crashes [21]. An additional study found that high variability in speeds was associated with more crashes [22]. It is likely that there is greater speed variation in a congested state than in a queued state, where most vehicles are stopped or traveling at low speeds. Our finding that traffic volume positively influences crashes concurs with another study that employed big data to examine safety performance [23] but contradicts Guo *et al.* [24], who did not find such a relationship.

Lastly, **Table 3** and **Table 4** show that the absence of a work zone (*i.e.*, when the "non-work zone" indicator = 1) is associated with a higher crash count. This contrasts with previous observations that work zone duration significantly increased injury and non-injury crash frequency [2].

The variance inflation factor (VIF) for the variables was also computed. VIF is a measure used in regression analysis to indicate the extent of multicollinearity within the explanatory variables. The general rule of thumb is that a VIF larger than 5 or 10 indicates a problematic amount of collinearity. VIF values for the CMH, QMH, speed deficit, segment length, and AADT were 2.15, 1.68, 1.30, 1.17, and 1.12, respectively. Since our VIF values are less than 5, we can confirm the acceptability of the mixed effect linear model for the analysis. The model included both fixed and random effects, accounting for variability within road segments. **Table 3** and **Table 4** indicate the group variance of the random intercept, *i.e.*, segments to be 0.221 and 0.321 for the CMH and QMH models, respectively.

## 5. Conclusions

This study used a mixed effect linear model to investigate the relationship between probe data-derived performance measures (CMH, QMH, and speed deficit) on crash counts at the road segment level. Road characteristics such as traffic volume and road type were also included in the modeling. The mixed effect linear models enabled us to factor in the differences in the road segment geometry. The analysis also factored in the periods for which work zones were active versus when they were not by including a work zone status variable. Two different model sets were developed for the analysis periods to eliminate multicollinearity. One model set used CMH, and the other used QMH.

The results indicate that CMH, QMH, speed deficit, segment length, and traffic volume correlated positively with crash count for the work zones examined in this study. Speed deficit, segment length, and traffic volume are more positively correlated with segment crashes in the QMH model than in the CMH model for non-active work zones and rural interstates.

A few interesting findings were also discussed, which help reveal the nature of work zone and non-work zone crashes. Non-active work zones are more statistically significant in the QMH model than the CMH model, implying for work zone conditions, CMH might be more significant than QMH. This may reflect the greater likelihood for locations with intermediate speeds (between 15 and 45 mph) to experience crashes since this is where the back of the queue is likely to be, whereas locations with very low speeds (under 15 mph) are less likely to experience crashes. Non-active work zones in the QMH model were more positively correlated with segment crashes than in the CMH model. This finding may reflect the effect of having traffic control devices during work zone conditions that may give drivers some awareness of the potential for crashes at a downstream location, while such devices are not present during non-work zone conditions.

The findings of this study can help provide transportation engineers with insights into correlations between the operational performance of a road segment and its impact on safety performance. The methods described here could be used to monitor performance in real time. This would enable more effective decision-making processes in the area of work zone planning and prioritization of road segments for countermeasures to improve safety. In particular, the differing results for the CMH and QMH models suggest that the locations with the most severe congestion (highest number of QMHs) may have less of a need for such countermeasures than those experiencing more CMHs but fewer QMHs. That is, the lower speeds associated with more severe congestion may itself tend to mitigate crashes within the work zone. In such conditions, however, it is important to provide drivers approaching but not yet within the work zone that they are approaching slow or stopped traffic.

One limitation of the study was weather data was not considered due to the manner in which the data was processed and aggregated. Also, the study was limited to one year of both probe and crash data. In future research, weather variables could be included in the models, additional years of data could be incorporated, and work zones on additional facility types could be used to expand on the initial models presented here. A time-of-day analysis could be factored into the study as well.

## Acknowledgements

We want to extend our gratitude to the Iowa Department of Transportation for acquiring the probe data used for this research.

## Author Contribution Statement

D.O. Okaidjah conceived the study, performed data analysis, and drafted the paper. C.M. Day provided guidance on the study and helped draft the paper.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Dimitrijevic, B., Khales, S.D., Asadi, R. and Lee, J. (2022) Short-Term Segment-Level Crash Risk Prediction Using Advanced Data Modeling with Proactive and Reactive Crash Data. *Applied Sciences* (*Switzerland*), **12**, Article No. 856. https://doi.org/10.3390/app12020856

[2] Khattak, A.J., Khattak, A.J. and Council, F.M. (2002) Effects of Work Zone Presence on Injury and Non-Injury Crashes. *Accident Analysis & Prevention*, **34**, 19-29. https://doi.org/10.1016/S0001-4575(00)00099-3

[3] Theofilatos, A., Ziakopoulos, A., Papadimitriou, E., Yannis, G. and Diamandouros, K. (2017) Meta-Analysis of the Effect of Road Work Zones on Crash Occurrence. *Accident Analysis & Prevention*, **108**, 1-8. https://doi.org/10.1016/j.aap.2017.07.024

[4] Ozturk, O., Ozbay, K. and Yang, H. (2014) Estimating the Impact of Work Zones on Highway Safety.

[5] Work Zone Fatal Crashes and Fatalities—The National Work Zone Safety Informa-

tion Clearinghouse. https://www.workzonesafety.org/crash-information/work-zone-fatal-crashes-fatalities/#national

[6] Salem, O.M., Genaidy, A.M., Wei, H. and Deshpande, N. (2006) Spatial Distribution and Characteristics of Accident Crashes at Work Zones of Interstate Freeways in Ohio. 2006 *IEEE Intelligent Transportation Systems Conference*, Toronto, 17-20 September 2006, 1642-1647. https://doi.org/10.1109/ITSC.2006.1707460

[7] Chang, G.L. and Xiang, H. (2003) The Relationship between Congestion Levels and Accidents.

[8] Day, C.M., Remias, S.M., Li, H., Mekker, M.M., McNamara, M.L., Cox, E.D., *et al.* (2015) Performance Ranking of Arterial Corridors Using Travel Time and Travel Time Reliability Metrics. *Transportation Research Record*, **2487**, 44-54. https://doi.org/10.3141/2487-04

[9] Day, C.M., Haseman, R., Premachandra, H., Brennan, T.M., Wasson, J.S., Sturdevant, J.R., *et al.* (2010) Evaluation of Arterial Signal Coordination: Methodologies for Visualizing High-Resolution Event Data and Measuring Travel Time. *Transportation Research Record*, **2192**, 37-49. https://doi.org/10.3141/2192-04

[10] Nassiri, H., Mohammadpour, S.I. and Dahaghin, M. (2023) Forecasting Time Trend of Road Traffic Crashes in Iran Using the Macro-Scale Traffic Flow Characteristics. *Heliyon*, **9**, E14481. https://doi.org/10.1016/j.heliyon.2023.e14481

[11] Theofilatos, A. (2017) Incorporating Real-Time Traffic and Weather Data to Explore Road Accident Likelihood and Severity in Urban Arterials. *Journal of Safety Research*, **61**, 9-21. https://doi.org/10.1016/j.jsr.2017.02.003

[12] Mekker, M.M., Remias, S.M., Mcnamara, M.L. and Bullock, D.M. (2020) Characterizing Interstate Crash Rates Based on Traffic Congestion Using Probe Vehicle Data.

[13] Haseman, R.J., Wasson, J.S. and Bullock, D.M. (2010) Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking. *Transportation Research Record*, **2169**, 40-53. https://doi.org/10.3141/2169-05

[14] Du, B., Chien, S., Lee, J., Spasovic, L. and Mouskos, K. (2016) Artificial Neural Network Model for Estimating Temporal and Spatial Freeway Work Zone Delay Using Probe-Vehicle Data. *Transportation Research Record*, **2573**, 164-171. https://doi.org/10.3141/2573-20

[15] Kamyab, M., Remias, S., Najmi, E., Rabinia, S. and Waddell, J.M. (2020) Machine Learning Approach to Forecast Work Zone Mobility Using Probe Vehicle Data. *Transportation Research Record*, **2674**, 157-167. https://doi.org/10.1177/0361198120927401

[16] Debnath, A.K., Blackman, R. and Haworth, N. (2014) A Tobit Model for Analyzing Speed Limit Compliance in Work Zones. *Safety Science*, **70**, 367-377. https://doi.org/10.1016/j.ssci.2014.07.012

[17] Okaidjah, D., Haddad, M., Day, C. and Das, B. (2023) Spatial Analysis of Relationships between Intersection Safety, the Urban Built Environment, and Average Income Level: A Case Study of Des Moines. *Transportation Research Record*, **2677**, 803-817. https://doi.org/10.1177/03611981221108392

[18] Daniel, J., Dixon, K. and Jared, D. (2000) Analysis of Fatal Crashes in Georgia Work Zones. *Transportation Research Record*, **1715**, 18-23. https://doi.org/10.3141/1715-03

[19] Robin, P. (2014) Use of Multinomial Logistic Regression in Work Zone Crash

Analysis for Missouri Work Zones. Missouri University of Science and Technology, Rolla.

[20] Day, C.M., Mcnamara, M.L., Li, H., Sakhare, R.S., Desai, J., Cox, E.D., *et al.* (2015) Indiana Mobility Report and Performance Measure Dashboards.

[21] Xing, J., Takahashi, H. and Iida, K. (2010) Analysis of Bottleneck Capacity and Traffic Safety in Japanese Expressway Work Zones. In: *Transportation Research Board* 89*th Annual Meeting*, Transportation Research Board of the National Academies, Washington DC.

[22] Zhou, C. (2019) Exploring the Relationship between Interstate Crashes and Various Speed Metrics Based upon Probe Vehicle Data. Master's Thesis, Iowa State University, Ames.

[23] Shi, Q. and Abdel-Aty, M. (2015) Big Data Applications in Real-Time Traffic Operation and Safety Monitoring and Improvement on Urban Expressways. *Transportation Research Part C*: *Emerging Technologies*, **58**, 380-394. https://doi.org/10.1016/j.trc.2015.02.022

[24] Guo, M., Zhao, X., Yao, Y., Yan, P., Su, Y., Bi, C., *et al.* (2021) A Study of Freeway Crash Risk Prediction and Interpretation Based on Risky Driving Behavior and Traffic Flow Data. *Accident Analysis & Prevention*, **160**, Article ID: 106328. https://doi.org/10.1016/j.aap.2021.106328