Microscopic Analysis of Work Zone Delay using Big Data and Logistic Regression

Yeganeh Morshedzadeh, B.Sc.
Master's Student
The University of British Columbia

Suliman Gargoum, Ph.D.
Assistant Professor
The University of British Columbia

Paper prepared for presentation at the Transportation Data and Analytics session of the 2023 TAC Conference & Exhibition, Ottawa, ON

Abstract

Work zone traffic management poses a significant challenge due to the need to minimize traffic delays while maintaining a safe environment for workers. In order to effectively manage traffic through work zones, it is necessary to understand how delays are affected by both internal and external factors, such as the work zone's design characteristics, road geometry, weather conditions, and traffic conditions. Despite the literature's attention to modeling work zone delays, most studies are limited to analyzing data from a small number of work zones and consider limited travel time observations. Macroscopic travel time prediction is also a key focus of existing literature. To address those gaps and provide a better understanding of factors affecting delay in work zones, this study uses a big dataset consisting of 15 million travel time observations collected every 2-3 minutes at 624 work zones located between the western borders of Alberta and Vancouver, BC. A microscopic delay analysis was performed whereby the impact of spatial, temporal, environmental, and segment-related variables on work zone delays was assessed using ordinal logistic regression. Based on the model statistics, high delays were associated with hourly traffic volume, peak hours, weekend travel, length of the work zone, and summertime, whereas precipitation affected delays in the opposite direction.

Introduction

Highway networks are the backbone of transportation infrastructure, facilitating the movement of people and goods across the world. The construction and maintenance of highways, roads, and bridges are essential for ensuring the safe and efficient movement of people and goods. However, work zones, where construction activities such as paving, patching, and utility works are carried out, can cause significant traffic delays, putting workers and drivers at risk. For instance, sudden lane closures, narrowed roads, and reduced speed limits can cause congestion, leading to longer travel times and increased travel costs. Additionally, construction sites can expose workers to hazards such as moving vehicles and heavy machinery, which can result in serious injuries or fatalities. According to a report by the United States Department of Transportation Federal Highway Administration (FHWA) [1], work zones result in a loss of 60 million vehicles per hours of capacity. Therefore, it is essential to effectively manage traffic delays through work zones to ensure the safety of workers and minimize the impact on drivers. Effective management requires accurate measurement and estimation of anticipated delays and proactive monitoring of their causes. A better understanding of the factors that impact travel time through work zones can help in developing effective traffic management strategies that improve safety and reduce the impact on drivers.

The aim of this paper is to model the relationship between travel time through work zones and several internal and external factors, including work zone length, temporal factors, environmental factors, and traffic-related factors. This study uses a big dataset, which captures travel time observations every 2-3 minutes from 624 work zones between 2021 and 2022. This comprehensive dataset enables a microscopic analysis of variation in delay within the same day, across different days during a week, and on different days during a month. Unlike previous work, this study employs an ordinal logistic regression model and provides insights into the impacts of season and weather conditions on work zone delay, which has not been explored in most previous studies. The primary goal of this study is to provide a more comprehensive understanding of the factors impacting travel time through work zones, ultimately leading to more effective traffic management and improved safety for workers and drivers.

Previous Work

Various factors have been investigated in connection with the impact of delays in work zones. Meng and Weng's [2] model showed that the length of the transition area has a far more significant effect on traffic delay than the length of the activity area in a work zone, particularly in light traffic conditions. Hou et al. [3] studied traffic flow and developed a model based on work zone data from freeways and signalized arterials in St. Louis, Missouri, USA. The most influential factors affecting traffic flow varied based on the type of road, with the latest interval's look-back traffic flows at upstream, downstream, and current sites, or traffic flows from the three preceding intervals at the current site being the most significant factors. Du et al. [4], [5] used data obtained from probe vehicles to forecast work zone delays. Their research involved 430 work zones across multiple freeways in New Jersey during 2014. Factors examined in this research included the count of closed lanes, total lanes, work zone length, average upstream traffic speed, upstream traffic volume, heavy vehicle percentage, work duration, and starting and ending times. The second study involved 181 work zones on various New Jersey freeways. The model's effectiveness was assessed in three instances with lengths of 3.6, 0.5, and 0.3 km. The gathered data was limited to short segments with observations over a brief period. Abdelmohsen and El-Rayes [6] employed an innovative approach to minimize crash probabilities and delay by determining the ideal values for work zone characteristics. The model sought to identify the best values for work zone speed limit, initiation time, shoulder use, lateral clearance, length, access and egress methods, and temporary traffic control measures. To achieve minimal traffic delays, the multi-objective optimization model established the maximum value for work zone speed limit, initiation time, shoulder use, and minimum segment length. Conversely, minimizing crash indexes required the opposite approach. Similarly, they presented an optimization model for reducing work zone delays and lowering construction costs [7].

In a study that considered both work zone features and non-work zone-related variables, Kamyab et al. [8] utilized historical lane closure data and to predict spatiotemporal mobility for upcoming lane closures. The research initially focused on cases where hourly traffic volume was unavailable, using the distribution of historical traffic speed as an alternative baseline for mobility. Modeling data was collected from 1160 work zones on Michigan interstates between 2014 and 2017. The authors employed probe vehicle data to obtain a mobility profile for historical observations. The three models' inputs included work zone configuration and features such as the number of open lanes, closed lanes, segment length, distance to lane closure location, roadway geometry, Annual Average Daily Traffic (AADT), and spatial and temporal aspects like month, day of the week, and hour of the day. The most effective model achieved an accuracy of 85%.

To determine the impact of work zones on non-recurrent delay separately, Chung [9], [10] used one year (2009) archive data from Korea’s freeways. The model used a dataset of 20,078 records gathered for 14 major freeways in Korea. It included work zone duration, traffic volume, work zone length, number of closed lanes, total number of lanes, type of work (e.g., weeding/planting, patching, lane marking, and facility maintenance), work zone location (e.g., tunnel, shoulder, bridge, and main lane). The results showed that the number of closed lanes, average daily delay, work zone length, and work zone duration directly impact the delay caused by work zones.

Mashhadi et al. [11] reviewed 40 studies from 1981-2019 on estimating work zone capacity rather than travel time and delay. They reported heavy vehicle percentage, work zone grade, work zone intensity, rural or urban roads, number of open lanes, number of closed lanes, work zone duration, work time, lane width, lane closure location (left/right), work zone length, weather condition, driver composition and population, existence of ramp, work zone speed limit, normal speed, weekday, and lateral clearance as the most common predictors of work zone capacity in previous studies. Another review by Weng and Meng [12] also overviews the factors impacting work zone capacity. More specifically, Bian and Ozbay [13] attempted to predict work zone capacity using multiple variables. They considered five factors in their study: total number of lanes, number of open lanes, heavy vehicle percentage, work intensity, and work duration. They used data from 225 work zones on freeways in Maryland, Oregon, Texas, Indiana, South Carolina, Kansas, and Missouri. In a similar study, Weng et al. [14] used road type, number of closed lanes, number of open lanes, lane closure location, heavy vehicle percentage, work intensity, work duration, work time (daytime or night-time), capacity measurement method, and year of data collection to study capacity of work zones.

While several studies have attempted to model work zone delays, but they often fall short in understanding the relationship between explanatory variables and delay. Many of these studies are limited by the number of observations and small datasets, hindering the examination of variables' impacts on microscopic work zone delay. Furthermore, some studies have used estimated capacity information instead of observed data when developing travel time prediction models. Additionally, data from previous studies were collected across different corridors, introducing multiple confounding factors that could obscure the relationship between variables and delay if not properly accounted for. Moreover, many studies have not controlled for external factors unrelated to work zone characteristics, such as weather and temporal variables. Meaning that the majority of research has focused on work zone-related variables without considering weather conditions or environmental factors. In this study, we aim to address these limitations by examining a more comprehensive set of variables, including weather and environmental factors, and using a larger dataset to better understand the relationship between explanatory variables and work zone delay.

Data Description

Data used in this study was collected by ATS Traffic's Intelitraffik Division (a traffic technology company with headquarters in Edmonton, AB). The raw data included approximately 15 million travel time observations collected between 2021 and 2022 along the 700 km corridor between Alberta borders and Vancouver, British Columbia, Canada. ATS Traffic has 710 planned work zones along the corridor as part of the Trans Mountain Expansion Project (TMEP). Out of the 710 work zones, 624 had already been activated and, thus, have been considered for this research. These work zones were located along the same corridor, which included various road types. The corridor included Multi-Lane Divided Roadways, Multi-Lane Undivided Roadways Two-Lane Roadways, and Low Volume Roadways with no center line.

Dependent Variable

Travel time was collected from Google's probe vehicle data every 2-3 minutes for each active work zone. Baseline time is estimated based on a combination of factors including speed limits, historical data, actual driving conditions, and the delay related to specific operation zone. In addition, SMATS sensors which use Bluetooth and Wi-Fi to capture travel time between pre-set locations and vehicles. The dependent variable that was used in the models is the level of delay. This variable is derived from delay. First, the delay was computed as the difference between the current and baseline estimates of travel time. Then the delay value was classified into one of four categories to give an intuition/sense of the level of delay. The first level, "*None*", represents situations where there was no delay (=0); "*Low*" represents a delay between 0 to 1 minute; "*Moderate*" represents a delay of 1 to 5 minutes, and "*High*" represents a delay over 5 minutes.

Independent Variables

Weather information was also appended to travel time information to identify if weather conditions impacted the level of delay. To do so, one-hour weather data records were gathered from 40 Environment Canada weather stations in BC from the Government of Canada environment resources (18). Weather data includes hourly Temperature, Precipitation and Wind Speed information. This information was linked to the travel time observations based on temporal and location data. An automated script was written to perform the execution and data collection procedure. It is worth noting that Wind Speed was later dropped from the models due to a low impact on delay.

Since traffic volume and congestion typically differ among different days and different seasons, two variables were defined. The Season variable represented the season during which data was collected. Since construction activity is often limited to summer and spring, only two seasons existed in the dataset. For the Day of Week variable, three categories were defined Monday/Friday = 1 (Week Onset/Offset); Tuesday, Wednesday, and Thursday = 2 (Midweek); and Saturday and Sunday = 3 (Weekend). Similar to season and day, traffic congestion also varies during different hours of the day. Hour of Day is a categorical variable that is used to capture the differences between peak and off-peak hours in a day. From 6 a.m. to 9:30 a.m. and 3 p.m. to 6:30 p.m. are considered peak hours; off-peak hours are the remainder of the day.

The speed limit information along the roadways was also considered in the models. Unlike operating speed, this is also a variable that work zone operations can control. Speed limit information was obtained using OpenStreetMap data for British Columbia [16]. Considering traffic volume information is also crucial when modelling the travel time of traffic and delays. This information was obtained through the BC Ministry of Transportation and Infrastructure's (BCMoTI's) Traffic Data Program. As part of this program, BCMoTI installs multiple permanent and temporary traffic counters sites at various locations throughout British Columbia [17]. This study used 30 counter-sites to estimate the traffic around each work zone asset. A script was written to automatically locate traffic count sites that were within proximity of each of the 624 work zones. It is worth noting here that some of the traffic counter sites were permanent while others were temporary. For temporary traffic count stations, traffic volumes during the times when travel time data was collected were not always available. To account for this, the hourly traffic volume (HTV) information for those records was estimated using the following equation recommended in the Alberta Highway Design Guide [18] for AADT projections.

$$HTV' = HTV \times \Big(1 + \big(TGR \times (CY - RY)\big)\Big) \qquad \textbf{(1)}$$

Where HTV' = Estimated HTV during the current year (CY)

HTV = Observed HTV during the most recent year where data is available (RY)

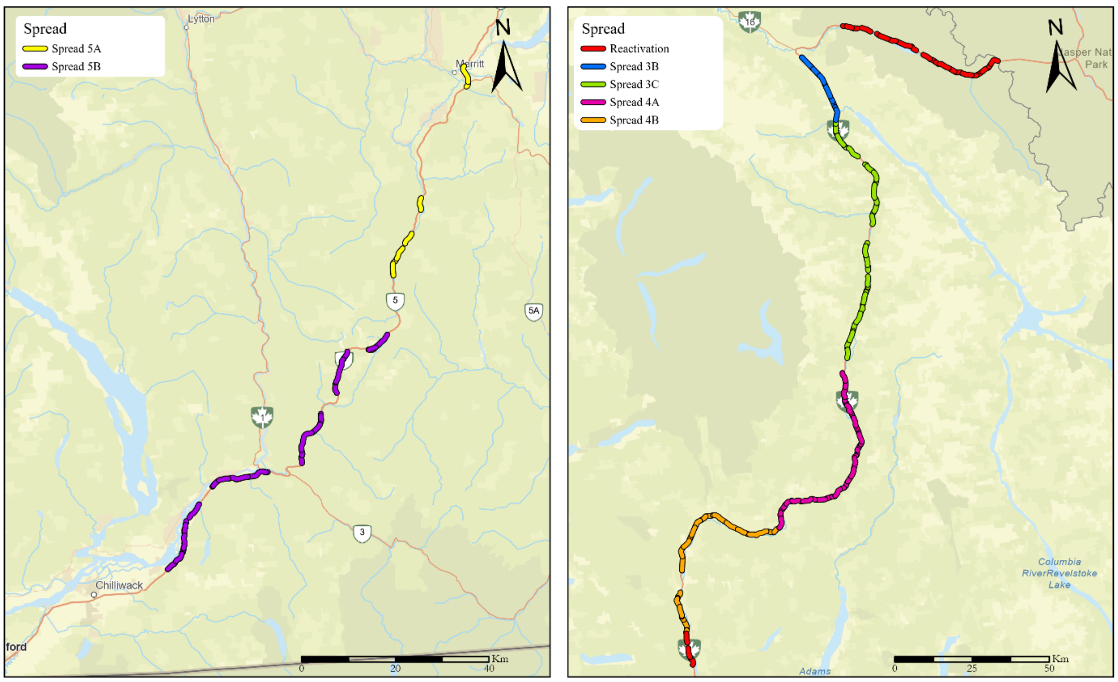
TGR = Traffic growth rate (assumed to be 1%)

CY = Current year

RY = Recent year — latest year that data is available.

Besides traffic volume, segment length was also considered in the models. This is defined as the distance between two points representing the start and end points of the traffic control area for a construction site as outlined in the approved traffic control plan. For this project, site access points are the places where traffic control is needed. Additionally, based on the road space and type of operation (e.g., to enter or exit site access point), traffic might be stopped for brief period and released from both ends at the same time; however, for an operation that involves equipment loading and offloading, traffic control needs are different. In other words, SLAT could be needed with complete stop for traffic when equipment is on move. This information was provided by ATS Traffic based on the work zone design drawings and the field layout.

Since construction activity for the Trans Mountain Expansion Project (TEMP) was split between four different contractors, the corridor was divided into seven different "*Operation Zones*", also known as "*Spreads*". Each operation zone covered a certain length of the corridor and ranged in length between 84 and 184km. Multiple work zones exist along the length of each operation zone, and **Figure 1** shows how different work zones are distributed among them. It is worth noting that construction activity and highway conditions varied significantly between different operation zones. This included variations in when the work took place, traffic control type, proximity to urban centers, traffic volume and composition, and highway type (e.g., two-lane two-way or multi-lane). Therefore, it was necessary to control for such a variable in the models. After augmenting, cleaning, and refining the dataset, the descriptive statistics of the nearly 2.9 million data records were estimated, as shown in **Table 1**.



(a) **(b)**

Figure 1: Work zones grouped by operation zone in TEMP: (a) Work zones in proximity to Kamloops and Valemount (b) Work zones in proximity to Abbotsford and Merritt

Table 1: Descriptive Statistics of Numerical Variables

	Minimum	Maximum	Mean	Std. Deviation
Under Construction Time (minute)	0.85	109.03	2.0551	0.7847
Baseline Time (minute)	1	189.9	1.9735	0.6473
Delay Time (minute)	0	106.16	0.1390	0.3055
Segment Length (kilometer)	1	7.92	3.2134	0.9829
Temperature (centigrade)	-15.4	29.9	8.8433	6.3374
Precipitation (millimeter)	0	7.7	0.1034	0.4349
Wind Speed (kilometer per hour)	0	41	7.6730	5.4389
Traffic Volume (vehicles per hour)	7.5	3926	302.4571	544.9376
Speed Limit (kilometer per hour)	70	120	101.6713	7.3441

Methodology

The method used in the analysis of the data was Ordinal Logistic Regression. Logistic regression makes no assumptions about the distributions of the independent variables, which can be binary, continuous, or categorical variables. Moreover, the ordinal nature of the response, in our case, the level of delay, also makes it challenging to meet the assumptions of the ordinary linear regression.

Model

Logistic regression is used to predict the odds of a certain event occurring based on the values of several independent variables. Let Y_i be a variable that denotes the level of delay observed at a particular location during a specific time, where $Y_i = 1, 2, 3, 4$ representing no delay, low, moderate, and high delay, respectively. In this case, $P\ (Y = k) = p,\ for\ k = 1\ to\ n$ represents the probability of a certain level of delay k being observed at instance i. This probability can be represented as log-odds, using a logit model of the following functional form:

$$logit\ (Y = k) = \ln\left(\frac{p_i}{1-(p_1+\cdots+p_n)}\right) = \ln\left(\frac{p_i}{p_{i+1}}\right),\ i = 1 \dots n \qquad \textbf{(2)}$$

In the case of ordinal logistic regression, the $P\ (Y \leq k)$ can be rewritten as:

$$P\ (Y \leq k) = p_1 + \cdots + p_n \qquad \textbf{(3)}$$

The odds of Y can then be expressed as:

$$odds\ (Y \leq k) = \frac{P(Y \leq k)}{1 - P(Y \leq k)} \qquad \textbf{(4)}$$

The logit can be expressed as:

$$logit\ (Y \leq k) = \ln\left(\frac{P(Y \leq k)}{1 - P(Y \leq k)}\right) \qquad \textbf{(5)}$$

The cumulative logistic model for ordinal response data is given by:

$$logit\ (Y \leq k) = \alpha_k + \beta_{k1}X_1 + \cdots + \beta_{kn}X_n, k = 1, \dots, n \qquad \textbf{(6)}$$

Where β_{km} are regression parameters of the model, and X_m are the covariates representing m factors that have alleged effects on the odds ratio of delay. The regression parameters are evaluated using an iterative maximum likelihood (ML) procedure. All analysis was done using IBM SPSS Statistics version 26.

Model Fit

Unlike linear regression, a well-accepted measure of goodness-of-fit (GOF), such as R^2, does not exist in logistic regression. Alternatively, the Chi-square method in the Omnibus test was used to assess model GOF. The Omnibus test assesses whether including explanatory variables in the model improves GOF. This is achieved by computing the difference between the deviance of the model before and after the variables are added, as in the following notation:

$$G = \chi^2 = -2LL_{Null} - (-2LL_k) \quad \textbf{(7)}$$

Where G is the goodness-of-fit estimate, and $-2LL$ is the deviance estimate for the null model and the model with k variable, respectively.

It is worth noting here that the Hosmer and Leshow (H-L) test cloud also be used to measure the difference between prediction with and without the model variables. However, the test has been found unreliable, especially when using a large sample size, since the test becomes too sensitive [19]. This means that even a small difference between observed and model-predicted values gets flagged as significant.

In addition to the goodness of fit, the Pearson correlation between independent variables was checked, and a very low correlation existed (<0.1).

Results and Discussion

Table 2 shows a summary of the ordinal logistic regression results. As evident in the table, all variables considered in the study had statistically significant effects on the level of delay observed at the work zones on the 5% confidence level (i.e. p-value <0.05).

As expected, increases in traffic volume result in higher levels of congestion and, as a result, are associated with an increase in the level of delay. The model also reveals that segment length increases are associated with a slight increase in the level of delay. The odd ratio shows that a unit increase in segment length (1km) increases the likelihood of delay level increasing by 0.4%. The relationship between the speed limit on the roadways and the level of delay is inversely proportional. The model shows that a unit increase in the speed limit is associated with a 0.2% decrease in the likelihood that delay would increase from one level to the next. Since the speed limit often changes in units of 20km/hr, the findings indicate that the likelihood of observing a high level of delay drops by around 4% for roads with a speed limit of 100km/hr compared to roads where the speed limit is 80km/hr. It is important to note here that changes in the speed limit must not be carried out unless a comprehensive safety assessment is conducted. It is also worth noting that the speed limit considered here is the speed limit on the highway and not the speed limit through the construction zone, which typically drops down to 50km/hr.

Environmental factors were also found to have statistically significant impacts on the level of delay observed at a particular location during a certain time instance. The model shows that compared to summer, the likelihood of observing a high level of delay drops during the spring season by 36.8%. The Wald estimate (an estimate which measures the size of the impact each variable has) is also relatively high for the seasonal variable. One explanation why the summer experiences more delays through work zones is the fact that the highways provide access to many national parks and campsites. Therefore, more tourist activity is expected during the summer months.

Table 2: Results of Ordinal Logistic Regression Model

Variable	Level	Estimate	Std Error	Wald	df	p-value	Odd Ratio
Delay	Low	-0.2982	0.0696	18.363	1.0	<0.0001	0.7422
	Moderate	-0.2479	0.0696	12.687	1.0	0.0004	0.7805
	High	-8.6365	0.1048	6792.323	1.0	0.0000	0.0002
Season	Spring	-0.4588	0.0091	2528.1217	1.0	0.0000	0.6320
	Summer	0*	.	.	.	.	1.0
Day Type	Week Onset/Offset	-0.2172	0.0080	738.7481	1.0	0.0000	0.8047
	Midweek	-0.3622	0.0073	2455.5701	1.0	0.0000	0.6961
	Weekend	0*	.	.	.	.	1.0
Hour Type	Peak	0.3328	0.3611	2303.6811	1.0	0.0000	1.4148
	Off-Peak	0*	.	.	.	.	1.0
Operation Zone	Reactivation	1.7951	0.0079	52101.5633	1.0	0.0000	6.0199
	Spread 4B	1.4358	0.0230	3908.3220	1.0	0.0000	4.2030
	Spread 5B	0*	.	.	.	.	1.0
Segment Length		0.0037	0.0003	145.4509	1.0	0.0000	1.0037
Temperature		0.0126	0.0006	415.5969	1.0	0.0000	1.0126
Precipitation		-0.3570	0.0075	2292.0313	1.0	0.0000	0.6998
Traffic Volume		0.0006	<0.0001	1174.5342	1.0	0.0000	1.0006
Speed Limit		-0.0020	0.0007	8.8212	1.0	0.0030	0.9980

* Set to zero because this parameter is redundant.

Temperature increases are also associated with a higher likelihood that delay increases. This could be attributed to inefficient traffic management in work zones at times when the temperature is high. It could also be attributed to a higher volume of travel towards parks at times of high temperature. In contrast to temperature, an increase in precipitation is affiliated with lower levels of delay. A unit increase in precipitation decreases the likelihood of delay by 30%. This could be due to the fact that most construction activity is stopped, and no workers are present on-site during times of high precipitation, hence, smoother traffic flow. Although the speed limit is lower when it is precipitating, it is still higher when there is activity undergoing in work zones.

The impacts of temporal factors on delay are also intuitive. As expected, peak hours are associated with a higher delay level compared to off-peak. The odds ratio shows that the likelihood of observing a higher delay is 41.5%. It is worth noting that this was not the case for work zones where construction activity took place overnight; hence, the need to include a variable in the model that controls for operation zone (i.e., the difference between work zone that operates in different conditions under different contractors).

Compared to weekends, both midweek and week onset and offset are affiliated with a lower level of delay through work zones. The odds ratio shows that the likelihood of high delay drops by 30.4% for midweek. Similarly, for week onset and offset (i.e., Monday and Friday), the level of delay is expected to drop by 19.5%. These results indicate that delay through work zones increases due to intra-city travel, which increases on weekends when compared to other days of the week.

The results of this study provide insights on how traffic could be managed in work zones and how a lower delay could be achieved along corridors of similar characteristics. The findings reveal that lower delay at work zones could be achieved by (1) reducing work zone activities during weekends, (2) increasing activities during off-peak hours, and (3) considering more effective and productive traffic and work zone management strategies during hot summer months when the traffic volume is higher than other seasons.

Conclusions, Discussion, and Future Research

Identifying the factors that impact travel time through a work zone is a prerequisite to efficient work zone traffic management. This paper develops an ordinal logistic regression model using a big dataset collected along 624 work zones in British Columbia and Alberta. The model revealed meaningful relationships between several variables and work zone delays. This includes an increase in work zone segment length being linked to a slight rise in delay. Additionally, according to the model, the probability of noticing a significant delay during the spring decreases by 36.8% compared to the summer. Increases in temperature are also found to be affiliated with delay increases. As anticipated, peak hours are correlated with a higher delay when compared to off-peak. The results also showed that delay at work zones increases on weekends compared to other days of the week. This is an extremely interesting finding since increases in delay occur during weekends despite the contractors working on limiting activity during that period.

This research has several practical applications and could be usefully explored in further research. Firstly, the findings of this paper could be used to effectively manage traffic flow through a work zone. For instance, prioritizing off-peak hours and cooler temperatures for work zone operations. Secondly, the dataset used in this study consists of 2.9 million observations making it possible to run deep learning models to accurately predict travel time and delays. Machine learning algorithms demonstrated high accuracy in variable prediction and pattern recognition. Thirdly, the findings could also be used to build a robust model to predict work zone travel time at locations where construction activity is planned. In fact, the authors are currently working on building a deep learning model to predict travel time observations at work zones based on this study's findings. Another valuable application is to calculate the contractor penalty for cost overruns. In addition, the model can assist work zone designers and schedulers in determining the ideal start and end timings for work zones. In the case studied, it would be interesting to use the model to help transportation engineers better create and assess traffic mitigation and management plans by analyzing the performance of work zones. Although the findings of this paper are essential for the mentioned applications, the work could be improved by considering more work zone-related variables in the models, such as work zone configurations and traffic control strategy. Unfortunately, such information was not available to the authors at the time of the study.

Acknowledgements

The authors disclosed receipt of the following financial support for this research article: This project was sponsored by Mitacs Accelerate, ATS Traffic's (Intelitraffik), and UAE University's CBE Annual Research Program (CARP).

References

[1] T. Lomax, D. Schrank, and B. Eisele, "Urban Mobility Report 2021 — Mobility Division," Jun. 2019. Accessed: May 15, 2022. [Online]. Available: https://mobility.tamu.edu/umr/

[2] Q. Meng and J. Weng, "Impact Analysis of Work Zone Configuration, Traffic Flow and Heavy Vehicle Percentage on Traffic Delay at Work Zones," *Proc. East. Asia Soc. Transp. Stud.*, vol. 2011, pp. 289–289, 2011, doi: 10.11175/eastpro.2011.0.289.0.

[3] Y. Hou, P. Edara, and C. Sun, "Traffic Flow Forecasting for Urban Work Zones," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1761–1770, Aug. 2015, doi: 10.1109/TITS.2014.2371993.

[4] B. Du, S. Chien, J. Lee, L. Spasovic, and K. Mouskos, "Artificial Neural Network Model for Estimating Temporal and Spatial Freeway Work Zone Delay Using Probe-Vehicle Data," *Transp. Res. Rec.*, vol. 2573, no. 1, pp. 164–171, Jan. 2016, doi: 10.3141/2573-20.

[5] B. Du, S. Chien, J. Lee, and L. Spasovic, "Predicting Freeway Work Zone Delays and Costs with a Hybrid Machine-Learning Model," *J. Adv. Transp.*, vol. 2017, p. e6937385, Aug. 2017, doi: 10.1155/2017/6937385.

[6] A. Z. Abdelmohsen and K. El-Rayes, "Optimizing the Planning of Highway Work Zones to Maximize Safety and Mobility," *J. Manag. Eng.*, vol. 34, no. 1, p. 04017048, Jan. 2018, doi: 10.1061/(ASCE)ME.1943-5479.0000570.

[7] A. Z. Abdelmohsen and K. El-Rayes, "Optimal Trade-Offs between Construction Cost and Traffic Delay for Highway Work Zones," *J. Constr. Eng. Manag.*, vol. 142, no. 7, p. 05016004, Jul. 2016, doi: 10.1061/(ASCE)CO.1943-7862.0001132.

[8] M. Kamyab, S. Remias, E. Najmi, S. Rabinia, and J. M. Waddell, "Machine Learning Approach to Forecast Work Zone Mobility using Probe Vehicle Data," *Transp. Res. Rec.*, vol. 2674, no. 9, pp. 157–167, Sep. 2020, doi: 10.1177/0361198120927401.

[9] Y. Chung, "Assessment of non-recurrent traffic congestion caused by freeway work zones and its statistical analysis with unobserved heterogeneity," *Transp. Policy*, vol. 18, no. 4, pp. 587–594, Aug. 2011, doi: 10.1016/j.tranpol.2011.02.003.

[10] Y. Chung, H. Kim, and M. Park, "Quantifying non-recurrent traffic congestion caused by freeway work zones using archived work zone and ITS traffic data," *Transportmetrica*, vol. 8, no. 4, pp. 307–320, Jul. 2012, doi: 10.1080/18128602.2010.522609.

[11] A. H. Mashhadi, M. Farhadmanesh, A. Rashidi, and N. Marković, "Review of Methods for Estimating Construction Work Zone Capacity," *Transp. Res. Rec.*, vol. 2675, no. 9, pp. 382–397, Sep. 2021, doi: 10.1177/03611981211002202.

[12] J. Weng and Q. Meng, "Estimating capacity and traffic delay in work zones: An overview," *Transp. Res. Part C Emerg. Technol.*, vol. 35, pp. 34–45, Oct. 2013, doi: 10.1016/j.trc.2013.06.005.

[13] Z. Bian and K. Ozbay, "Estimating Uncertainty of Work Zone Capacity using Neural Network Models," *Transp. Res. Rec.*, vol. 2673, no. 2, pp. 49–59, Feb. 2019, doi: 10.1177/0361198118825136.

[14] J. Weng, Y. Yu, and L. Ma, "Uncertainty-based prediction of work zone capacity using a Bayesian approach," *Proc. Inst. Civ. Eng. - Transp.*, vol. 172, no. 1, pp. 24–35, Feb. 2019, doi: 10.1680/jtran.17.00004.

[15] E. and C. C. Canada, "Historical Data - Climate - Environment and Climate Change Canada." https://climate.weather.gc.ca/historical_data/search_historic_data_e.html (accessed Jun. 22, 2022).

[16] "Download OpenStreetMap data for this region: British Columbia - Geofabrik Download Server." https://download.geofabrik.de/north-america/canada/british-columbia.html (accessed Jun. 30, 2022).

[17] "BC Traffic Data Program." https://www.th.gov.bc.ca/trafficData/ (accessed Jun. 20, 2022).

[18] "Alberta Transportation Highway Geometric Design Guide," *Alberta Government*, Apr. 2018. https://open.alberta.ca/dataset/highway-geometric-design-guide-1995-archived-versions/resource/6571aae5-09c5-4ce1-8d66-71661a529aad (accessed Jun. 28, 2022).

[19] P. Paul, M. L. Pennell, and S. Lemeshow, "Standardizing the power of the Hosmer–Lemeshow goodness of fit test in large data sets," *Stat. Med.*, vol. 32, no. 1, pp. 67–80, 2013, doi: 10.1002/sim.5525.