



# Modeling the lagged impacts of hourly weather and speed variation factors on the segment crash risk of rural interstate freeways: Applying a space–time-stratified case-crossover design

Zihang Wei <sup>a,\*</sup>, Subasish Das <sup>b</sup>, Yue Wu <sup>a</sup>, Zihao Li <sup>a</sup>, Yunlong Zhang <sup>a</sup>

<sup>a</sup> Zachry Department of Civil & Environmental Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

<sup>b</sup> Ingram School of Engineering, Texas State University, 601 University Dr, San Marcos, TX 78666, United States

## ARTICLE INFO

### Keywords:

Weather factors  
Hourly crash data  
Exposure-lag response association  
Time-series modeling  
Distributed lag nonlinear model  
Rural interstate

## ABSTRACT

In the realm of traditional roadway crash studies, cross-sectional modeling methods have been commonly employed to investigate the intricate relationship between the crash risk of roadway segments and variables including roadway geometrics, weather conditions, and speed distribution. However, these methodologies assume that the explanatory variables and target variable are only associated within the same time period. Although this assumption is well-founded for static factors like roadway geometrics, it proves inadequate when dealing with highly time-varying variables related to weather conditions and speed variation. Recent investigations have unveiled that these time-varying variables may exhibit lagged impacts on segment crash risk, necessitating the adoption of more comprehensive time-series modeling methods. This study employs two interpretable statistical methods, namely the distributed lag model (DLM) and the distributed lag nonlinear model (DLNM), to elucidate meaningful and interpretable patterns of the lagged impacts of weather and speed variation factors on segment crash risk. Empirical evidence based on crash data collected from rural interstate freeways in the state of Texas demonstrates coherent and interpretable lagged impact patterns of these variables. This study's results serve as strong support for the existence of lagged impacts on roadway segment-level crash risk, emphasizing the need for considering time-series effects in future crash modeling research. Furthermore, these findings could offer practical implications for the design of real-time crash warning systems and the effective implementation of variable speed limits to enhance road safety.

## 1. Introduction

Roadway traffic crashes have emerged as a critical global concern, representing a significant cause of both human fatalities and substantial financial losses. The United States, in particular, experienced a troubling escalation in this regard, with an estimated 42,915 individuals losing their lives in motor vehicle traffic crashes in 2021. This marked a distressing 10.5 % increase from the recorded 38,824 fatalities in 2020. This tragic figure reflects the highest recorded number of fatalities since 2005, underscoring the urgent need for comprehensive investigation and intervention (National Center for Statistics and Analysis, 2022). Academic research has diligently delved into the complex realm of roadway traffic crashes, seeking to elucidate the factors that contribute to this pervasive problem. Studies have meticulously explored various

aspects of crashes, revealing crucial relationships between crash risk and roadway geometric characteristics (Anderson et al., 1999; Haghghi et al., 2018; Miaou and Lum, 1993), speed distribution patterns (Garber and Gadipati, 1989; Imprialou et al., 2016; Pei et al., 2012) and weather-related factors (Brijs et al., 2008; Eisenberg, 2004; Scott, 1986; Yu and Abdel-Aty, 2014).

Traditionally, studies investigating the association between various factors and segment crash risk have predominantly relied on cross-sectional modeling methodologies. In such approaches, both explanatory and target variables are examined within a single time period, typically spanning yearly, monthly, daily, or hourly intervals (Eisenberg, 2004; Huang et al., 2010; Li et al., 2023; Peng et al., 2020; Wei et al., 2023, 2022; Yu et al., 2015). This technique has proven suitable for examining roadway geometry variables. Due to their static nature

\* Corresponding author.

E-mail addresses: [wzh96@tamu.edu](mailto:wzh96@tamu.edu) (Z. Wei), [subasish@txstate.edu](mailto:subasish@txstate.edu) (S. Das), [wywy@tamu.edu](mailto:wywy@tamu.edu) (Y. Wu), [scottlzh@tamu.edu](mailto:scottlzh@tamu.edu) (Z. Li), [yzhang@civil.tamu.edu](mailto:yzhang@civil.tamu.edu) (Y. Zhang).

and limited variation at specific locations, it is feasible to analyze their relationship with segment crash risk within fixed time frames. However, a distinct scenario arises concerning variables related to speed distribution patterns and weather-related factors, as these variables exhibit significant temporal variability (Lord and Mannering, 2010; Pei et al., 2012; Wang et al., 2018). Prior empirical investigations have elucidated that certain factors, such as precipitation, could exert lasting impacts on segment crash risk. This is primarily due to the time required for accumulated rainfall to be mixed with engine oil and gasoline residues present on the roadway surface during dry conditions (Eisenberg, 2004). Furthermore, other weather-related time-varying variables, such as visibility and temperature, may also engender lasting effects on segment crash risk, primarily arising from the “memory of driving behaviors.” For instance, drivers tend to reduce speed and acceleration and increase headway during low visibility conditions (Hoogendoorn et al., 2010). Even after the dissipation of low visibility, drivers may continue to adopt this cautious driving behavior for a prolonged period. Regarding the temperature factor, research has indicated that low temperature conditions, particularly freezing temperatures, are associated with an increased segment crash risk. This relationship may be attributed to the heightened challenges in driving maneuverability experienced under such adverse conditions (Yu and Abdel-Aty, 2014). Consequently, it is plausible to hypothesize that drivers might persist with the same cautious driving behavior even after the freezing temperature condition ends. This could contribute to the lagged impact of temperature on segment crash risk. Given the potential for time-varying variables to engender lagged impacts on segment crash risk, reliance solely on cross-sectional modeling approaches may hinder comprehensive investigations into these intricate relationships. Consequently, the integration of time-series modeling techniques becomes imperative to unravel the lagged effects of precipitation, visibility, and temperature on segment crash risk. Moreover, as for the speed variation factor, no related studies have investigated whether there is a lasting impact on segment crash risk. Due to the fact that speed variation is also highly time-varying and closely related to segment crash risk, this study also includes it into the time-series modeling process and investigates whether it exhibits lagged effects.

Recently, a growing number of studies have started to consider the lagged impacts of time-varying variables and have harnessed time-series modeling approaches to predict crash risk and other safety-related events. Notably, studies including Yuan et al. (2019), Li et al. (2020), and Zhang and Abdel-Aty (2022) have employed recurrent neural network (RNN) or long short-term memory (LSTM) models to predict real-time crash risk of roadway segments. In another study by Liu et al. (2023), a transformer encoder structure was applied to predict hard-brake events in real-time. When making predictions, these methods effectively account for temporal variation patterns by incorporating variables from both the current and preceding timesteps. The results of these studies have demonstrated that this can enhance the prediction performance of traffic crash risk, affirming that time-varying variables from the past exert significant effects on crash risk at the current timestep. Despite their commendable prediction performance, it is important to note that these methods are inherently data-driven and considered as black-box models. Consequently, they lack the ability to provide interpretable representations that elucidate the lagged impact patterns of these factors on segment crash risk. To address the aforementioned limitation on interpretability, researchers have turned to interpretable statistical time-series modeling techniques to explore the lagged impact patterns of weather-related factors on crash risk. Two related studies have investigated this topic within expansive spatial areas encompassing entire cities. Specifically, Xing et al. (2019) examined the lag effects of hourly weather factors on crash incidents, utilizing an integrated dataset comprising meteorological, traffic, and crash information in Hong Kong. Similarly, Zhan et al. (2020) investigated the lagged effects of hourly temperature and precipitation on road traffic casualty risk in the city of Shenzhen, China.

The studies conducted by Xing et al. (2019) and Zhan et al. (2020) solely focused on larger spatial resolutions. However, it is noteworthy that the majority of crash modeling research emphasizes investigating the relationship between crash risk and its contributing factors at the roadway segment level, with limited attention given to exploring the lagged impacts of these factors on segment-level crash risk. This study seeks to fill this research gap by employing two interpretable statistical time-series modeling methods, namely the distributed lag model (DLM) and distributed lag nonlinear model (DLNM), to explore the lagged impacts of precipitation, visibility, temperature, and speed variation on segment level crash risk. The data utilized for this investigation were collected from rural interstate highways in the state of Texas. Given the highly time-varying nature of these variables, the data were aggregated into 1-hour temporal resolution intervals, as this approach allows for the retention of the important time variation information. While larger aggregation intervals (e.g., yearly, monthly, and daily) adopted by traditional crash analysis models cannot capture this information effectively, it is also not recommended to use excessively small intervals like 10 or 5 min, as this may result in a diminishment of temporal variation between consecutive time steps. Such an approach could lead to a high temporal correlation issue, adversely impacting the modeling results. Thus, a 1-hour aggregation interval was adopted in order to maintain a balance between retaining essential time-variation information and reducing temporal correlations among consecutive time steps. Finally, due to the rare nature of crash observations after data was aggregated into 1 h intervals, a space-time-stratified case-crossover design was applied to rebalance the dataset.

## 2. Data preparation

The data preparation process contained three steps which are illustrated in the following flow chart (Fig. 1). Details on the data are provided after the flowchart.

### 2.1. Base roadway segment

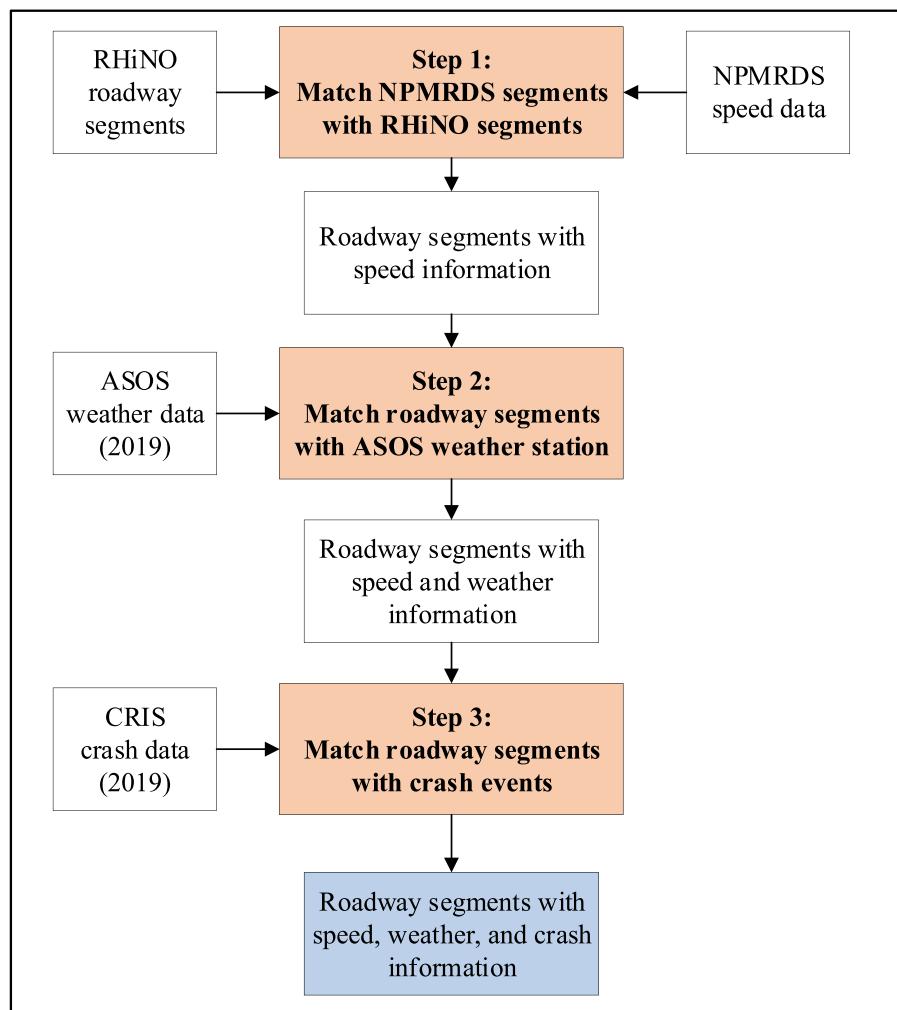
The base roadway segment network consists of rural interstate highways in the state of Texas. The data was collected from the Texas Department of Transportation (TxDOT) Road-Highway Inventory Network Offload (RHINO), which contains roadway GIS linework and roadway inventory attributes, including geometric features and traffic information. TxDOT submitted this dataset to the Federal Highway Administration (FHWA) as part of the Highway Performance Monitoring System (HPMS) program (Texas Department of Transportation, 2020).

### 2.2. Speed distribution data

Raw speed data were collected from FHWA's National Performance Management Research Dataset (NPMRDS). The NPMRDS contains travel time and speed data collected from a fleet of probe vehicles (e.g., cars and trucks). The NPMRDS can generate speed and travel time data using probe vehicle location information. Speed data were collected based on traffic message channels (TMC). TMC segments and base roadway network segments can be matched with each other using GIS software. This study calculates the hourly speed standard deviation using the raw speed data, which is collected in a 5-minute interval.

### 2.3. Weather condition data

Weather condition data were collected from the Automated Surface Observing System (ASOS) of the National Centers for Environmental Information. This study utilized all ASOS stations in Texas, and each base roadway segment was assigned to the ASOS station closest to it. A 30 miles buffer was applied when assigning ASOS weather stations to roadway segments. This ensured that the ASOS station assigned to each roadway segment was less than 30 miles away, which means the



**Fig. 1.** Flow chart of the data preparation process.

weather information collected at the assigned ASOS station can almost completely accurately reflect the weather condition of the roadway segment. Weather information, including precipitation, visibility, and temperature, was collected at each weather station, and it was averaged into one-hour intervals.

#### 2.4. Crash data

Crash data were collected within the state of Texas in 2019 through the Crash Record Information System (CRIS). Each crash record includes spatial and temporal information. Through GIS software, all crashes were assigned to the base roadway segments on which they occurred. The hourly crash number was summarized for each roadway segment. This study used the total number of crashes that occurred on a roadway segment in 1 h to represent the hourly crash risk. The hourly crash number of a roadway segment was used as the target variable in the modeling process.

#### 2.5. Space-time-stratified case-crossover design

Due to the rare nature of crash events, there are a significantly larger number of non-crash hours than crash hours after the data is aggregated at an hourly level. Thus, a space-time-stratified case-crossover design was applied to restructure the data in order to achieve more reliable results. The design compares an hourly observation where crashes happened (case hour) with hourly observations where no crash

happened (control hours) while controlling the spatial (e.g., roadway segment) and temporal (e.g., hour) factors. The control hours serve as the counterfactual exposure experience of each case hour. For each case hour, a stratum ID combining space and time dimensions was introduced and each case hour was matched with all control hours with the same stratum ID. Stratum ID in this study combines road ID, year, month, day of week, and hour (see Table 1). Specifically, a case hour is matched with 3 or 4 control hours on the same day of the week in the same month (Gao et al., 2022; Nyadanu et al., 2022; Wu et al., 2021).

Furthermore, the descriptive statistics of hourly crash occurrence numbers in the prepared space-stratified case-crossover dataset are summarized in Table 2.

#### 2.6. Descriptive statistic of Time-Series variables

Hourly precipitation, hourly visibility, hourly temperature, and hourly speed standard deviation are considered time-series variables in this study to investigate the lagged impact of these factors on segment crash risk. Since the exposures that happened in the preceding hours have lagged impact on segment crash risk in the current hour, the modeling process considers the exposure at the current hour when crashes occur and the exposure at the preceding hours before crashes occur. Based on the modeling results, the analysis in this study considers hourly precipitation for 8 preceding hours and considers hourly visibility, hourly temperature, and hourly speed standard deviation for 10 preceding hours. Table 3 presents the descriptive statistics of all time-

**Table 1**  
Space-time-stratified case-crossover design.

Road ID	year	month	day	hour	Day of Week	Total Crash	Stratum ID	Case / Control
1000	2019	1	2	14	4	1	1000 2019 1 4 14	Case
1000	2019	1	9	14	4	0	1000 2019 1 4 14	Control
1000	2019	1	16	14	4	0	1000 2019 1 4 14	Control
1000	2019	1	23	14	4	0	1000 2019 1 4 14	Control
1000	2019	1	30	14	4	0	1000 2019 1 4 14	Control
1001	2019	12	24	22	3	2	1001 2019 12 3 22	Case
1001	2019	12	3	22	3	0	1001 2019 12 3 22	Control
1001	2019	12	10	22	3	0	1001 2019 12 3 22	Control
1001	2019	12	17	22	3	0	1001 2019 12 3 22	Control

Note: lines in bold text are case hours.

**Table 2**  
Descriptive statistics of hourly crash number in the prepared dataset.

Hourly crash number	Count	Mean	Variance
0	27,400	0.238	0.197
1	8022		
2	175		
3	23		
4	7		

series variables at the current hour and each preceding hour separately for case hours and control hours. As shown by the table, the descriptive statistics show that time-series variables may have lagged impact on segment crash risk. For example, for hourly precipitation, the average hourly precipitation at the immediate preceding hour of the case hour is 0.0111 in., which is more than 2 times higher than the average hourly precipitation at the current hour (i.e., 0.0045 in.). Moreover, the averages of hourly precipitation at the 3rd to 6th preceding hours are also larger than that of the current hour, while as for the control hours, averages of hourly precipitation at the current hour and each preceding hour are all around 0.003 in.. This indicates that crashes that happen at the case hours are linked to the lagged impact brought by high precipitation exposures in the preceding hours. For hourly speed standard deviation, the average hourly speed standard deviation at the case hours is the highest (e.g., 5.5 mph), and the value decreases at each preceding hour (e.g., 4.1 mph at the immediate preceding hour, 3.8 mph at the 2nd preceding hour), while for the control hours, the averages of hourly speed standard deviation at the current hour and each preceding hour are all around 3.5 mph.

Note that this study does not consider roadway geometrics information in the modeling process. Since roadway geometric-related factors may also affect the segment crash risk, this study only included data from rural interstate highways in order to ensure all roadway segments in the prepared dataset have similar geometrics and minimize the effects of roadway geometrics on the modeling results.

### 3. Methodology

#### 3.1. Unconstrained distributed lag model

The introduction of an unconstrained distributed lag model is due to the realization that an exposure event may not just impact a target event within the same timestep. For example, an increase in segment crash risk at the current hour is linked with the precipitation level at the current hour, as well as the precipitation level at the 1st preceding hour, etc. Traditionally, when modeling the time-series relationship between the target event and the exposure event across multiple timesteps, the regression model of the following form has been widely applied.

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_q X_{t-q} \quad (1)$$

This model is termed the unconstrained distributed lag model, where  $X_t, X_{t-1}, \dots, X_{t-q}$  corresponds to the time-series exposure variables at  $q$

consecutive timesteps before the current timestep  $t$ . The goal of this regression modeling is to estimate the values of parameters  $\beta_0, \beta_1, \dots, \beta_q$ . For each parameter  $\beta_i, i \in [0, q]$ , it can be interpreted as: for a 1 unit increase in the exposure variable  $X_{t-i}$  at timestep  $t-i$  (i.e., the  $i^{th}$  preceding hour of the current timestep  $t$ ), there will be a  $\beta_i$  unit increase in the target variable  $Y_t$  at the current timestep  $t$ . Alternatively,  $\beta_i$  can also be viewed as: if there is a 1 unit increase in exposure variable at the current timestep  $t$ , there will be a  $\beta_i$  unit increase in target variable at the  $i^{th}$  lagged timestep. However, due to the high temporal correlated nature of most timeseries variables (e.g., the 4 variables applied in this study), many empirical studies have found that the high autocorrelation of the timeseries variables among consecutive timesteps can result in a high degree of collinearity of the model and consequently lead to unstable and poor estimations of the parameters  $\beta_0, \beta_1, \dots, \beta_q$ .

In two studies by Schwartz (2000) and Zanobetti et al. (2000), it was effectively proven that by applying polynomial functions or spline functions to constraint the parameters  $\beta_0, \beta_1, \dots, \beta_q$  of the model described in Eq. (1), the autocorrelation of the time-series variables can be controlled and lead to stable and accurate estimations of the parameters  $\beta_0, \beta_1, \dots, \beta_q$ . By adding these constraints, the model becomes the DLM applied in this study and moreover, for DLNM, it further transforms the timeseries exposure variables to nonlinear terms in order to discover the nonlinear relationship between the target events and the exposure events (Gasparrini et al., 2010). Detailed and comprehensive descriptions of the design of DLM and DLNM are introduced in the remaining paragraphs of this section.

#### 3.2. Model descriptions

##### 3.2.1. Representation of the model design

The representation of the model design used in this study can be written as in Eq. (2) (Gasparrini et al., 2010; Gasparrini, 2014):

$$f(\mu_t) = \alpha + \sum_j^J \sum_{i=1}^{N_j} \eta_i^j s_i^j(x, t) \quad (2)$$

Where  $t = 1, \dots, n$  is the time index of the event,  $\mu_t \equiv E(Y_t)$  is the expectation of the time series outcome  $Y_t$ , and  $f$  is a monotonic link function that assumes that  $Y_t$  comes from a distribution belonging to the exponential family.  $\alpha$  is the model intercept.  $J$  is the set of time-series variables.  $s_i^j(\bullet)$  is called the basis function of time-series variable  $j$  and  $N_j$  is the total number of basis functions applied to the lag space and exposure space of time-series variable  $j$ .  $\eta_i^j$  is the parameter of the variable created by the transformation  $s_i^j(\bullet)$ .  $N_j$  equals to  $v_l$  for the DLM and equals to  $v_l v_x$  for the DLNM. Based on the choices of the basis function  $s_i^j(\bullet)$ , the model is a DLM if nonlinear transformations are applied to the lag space and a DLNM if nonlinear transformations are applied to both the lag space and the exposure space.

##### 3.2.2. The basis function of DLM

For DLM, one basis function  $s_i(x, t)$  of a time series variable can be

**Table 3**

Descriptive statistics of variables at the current hour and each preceding hour.

Variables	Mean	Std.	Min	Max	Variables	Mean	Std.	Min	Max
<i>Case hours (Number of crashes &gt; 0)</i>					<i>Control hours (Number of crashes = 0)</i>				
<i>Hourly precipitation (inches)</i>					<i>Hourly precipitation (inches)</i>				
Current hour	0.0045	0.0423	0	2	Current hour	0.0029	0.0383	0	2
Preceding hour 1	0.0111	0.3251	0	17	Preceding hour 1	0.003	0.0375	0	2
Preceding hour 2	0.0044	0.0388	0	1	Preceding hour 2	0.0034	0.1076	0	17
Preceding hour 3	0.0058	0.0648	0	2	Preceding hour 3	0.003	0.0376	0	2
Preceding hour 4	0.0074	0.2701	0	24	Preceding hour 4	0.0039	0.1484	0	17
Preceding hour 5	0.0162	0.5184	0	24	Preceding hour 5	0.0028	0.0366	0	2
Preceding hour 6	0.0094	0.2698	0	21	Preceding hour 6	0.0039	0.1515	0	24
Preceding hour 7	0.0039	0.0401	0	1	Preceding hour 7	0.0029	0.054	0	7
Preceding hour 8	0.0039	0.0426	0	1	Preceding hour 8	0.0043	0.2095	0	24
<i>Hourly visibility (miles)</i>					<i>Hourly visibility (miles)</i>				
Current hour	9.0020	2.2296	0	10	Current hour	9.1791	2.0383	0	10
Preceding hour 1	9.0029	2.2182	0	10	Preceding hour 1	9.1434	2.0833	0	10
Preceding hour 2	9.0330	2.1684	0	10	Preceding hour 2	9.1243	2.1127	0	10
Preceding hour 3	9.0328	2.1732	0	10	Preceding hour 3	9.1183	2.1112	0	10
Preceding hour 4	9.0396	2.1657	0	10	Preceding hour 4	9.1167	2.1164	0	10
Preceding hour 5	9.0629	2.1605	0	10	Preceding hour 5	9.1254	2.1026	0	10
Preceding hour 6	9.1048	2.1102	0	10	Preceding hour 6	9.1531	2.0673	0	10
Preceding hour 7	9.1318	2.0964	0	10	Preceding hour 7	9.1935	2.011	0	10
Preceding hour 8	9.1706	2.0503	0	10	Preceding hour 8	9.2399	1.9553	0	10
Preceding hour 9	9.2154	2.0088	0	10	Preceding hour 9	9.2764	1.9097	0	10
Preceding hour 10	9.2471	1.9649	0	10	Preceding hour 10	9.3157	1.8607	0	10
<i>Hourly temperature (°F)</i>					<i>Hourly temperature (°F)</i>				
Current hour	67.1679	17.2751	11	106	Current hour	67.279	17.2585	9	109
Preceding hour 1	66.6303	17.0922	9	104	Preceding hour 1	66.6909	17.1131	11	106
Preceding hour 2	66.1092	16.9440	8	103	Preceding hour 2	66.1792	16.9455	9	106
Preceding hour 3	65.6871	16.7986	9	104	Preceding hour 3	65.7378	16.7697	9	108
Preceding hour 4	65.4004	16.6734	9	105	Preceding hour 4	65.4493	16.6168	10	109
Preceding hour 5	65.2642	16.5773	10	106	Preceding hour 5	65.3159	16.5207	9	108
Preceding hour 6	65.3091	16.5472	10	105	Preceding hour 6	65.3399	16.4985	9	107
Preceding hour 7	65.5452	16.5650	10	105	Preceding hour 7	65.5709	16.523	9	108
Preceding hour 8	65.8580	16.6578	10	106	Preceding hour 8	65.9167	16.5632	9	108
Preceding hour 9	66.2151	16.7300	11	106	Preceding hour 9	66.3141	16.6109	10	108
Preceding hour 10	66.7053	16.8420	12	106	Preceding hour 10	66.8024	16.6606	10	108
<i>Hourly speed standard deviation (mph)</i>					<i>Hourly speed standard deviation (mph)</i>				
Current hour	5.5483	3.6371	0	28	Current hour	3.5568	1.8205	0	28
Preceding hour 1	4.0843	2.3973	0	23	Preceding hour 1	3.5554	1.7585	0	27
Preceding hour 2	3.7983	2.0703	0	24	Preceding hour 2	3.5397	1.7669	0	26
Preceding hour 3	3.7003	1.9014	0	21	Preceding hour 3	3.5369	1.7569	0	35
Preceding hour 4	3.6643	1.8909	0	24	Preceding hour 4	3.5293	1.7129	0	24
Preceding hour 5	3.6461	1.7913	0	19	Preceding hour 5	3.5265	1.7019	0	23
Preceding hour 6	3.6519	1.7534	0	19	Preceding hour 6	3.5349	1.6673	0	28
Preceding hour 7	3.6177	1.7486	0	23	Preceding hour 7	3.5573	1.6867	0	26
Preceding hour 8	3.6106	1.6678	0	23	Preceding hour 8	3.5704	1.6973	0	29
Preceding hour 9	3.6369	1.6641	0	19	Preceding hour 9	3.5829	1.667	0	29
Preceding hour 10	3.6348	1.6906	0	24	Preceding hour 10	3.6036	1.653	0	27

written as Eq. (3):

$$s_i(x, t) = \int_{t_0}^{t_1} x_u w_i(t-u) du \quad (3a)$$

$$= \int_{l_0}^L x_{t-l} w_i(l) dl \quad (3b)$$

$$\approx \sum_{l=l_0}^L x_{t-l} w_i(l), \text{ where } i = 1, \dots, v_l \quad (3c)$$

In Eq. (3a), the risk at time  $t$  can be defined as the integral of exposure magnitude  $x_u$  over the time period from  $t_0$  to  $t_1$ . Here,  $t_0$  and  $t_1$  are the first and last lag exposure times that are relevant to the risk at time  $t$ .  $w_i(\bullet)$  is one weight function of the lag space. There is a total of  $v_l$  weight functions of the lag space, where  $v_l$  is defined by the dimension of the transformation basis functions on the lag space.  $w_i(t-u)$  can be viewed as the relative contribution of lag exposure at time  $t-u$  on the risk at time  $t$ .  $w_i(\bullet)$  can also be called the lag-response function. Eq. (3a) can be rewritten as Eq. (3b), where  $L$  is the total length of the lag period that has contributed to the risk at time  $t$ .  $l_0$  is the start of the lag period, and

usually equals 0. Moreover, since real-life data are always discrete, Eq. (3b) can be estimated as Eq. (3c), where the lag interval is partitioned into discrete units and the integral is estimated as the summation of all the discrete units. This method allows the effects of an exposure event to be distributed over a specific lag period defined by  $L$ , and  $s_i(x, t)$  can be viewed as a part of the cumulative impact at time  $t$  of an exposure event within lag time period  $L$ .

By using matrix notation, it is possible to write Eq. (3) in a more compact form, as follows:

$$s(x, t, \eta) = q_{x,t}^T C \eta = w_{x,t}^T \eta \quad (4)$$

Where  $C$  is a matrix of shape  $(L-l_0+1) \times v_l$  which is obtained by applying  $v_l$  basis functions to the lag vector  $\delta = [l_0, l_1, \dots, L]^T$ .  $q_{x,t} = [x_{t-l_0}, x_{t-l_1}, \dots, x_{t-L}]^T$  denotes the exposure history vector containing exposure values at each time lag within the exposure period  $L$ . Note that in most cases  $l_0$  equals to 0.  $w_{x,t}$  is now a vector of length  $v_l$ , where each element is a weighted sum of all exposure history within lag period  $L$ . By stacking  $w_{x,t}^T$  row wise, we can get a matrix  $W$ , which can also be derived by Eq. (5):

$$W = QC \quad (5)$$

Where  $W$  is a matrix of the shape  $N \times v_l$  and  $Q$  is a matrix of shape  $N \times (L - l_0 + 1)$ , where  $N$  is the total number of observations. Note that each row of matrix  $Q$  is the vector  $q_{x,t}^T$  of a single observation, and each row of matrix  $W$  is  $w_{x,t}^T$  of a single observation. The parameters  $\eta$  corresponding to the  $v_l$  new variables can be estimated by including matrix  $W$  into the designed matrix. The original parameters  $\hat{\beta}$  representing the risk contribution at each time lag and the corresponding (co)variance matrix  $V(\hat{\beta})$  can be calculated after the estimated  $\hat{\eta}$  is obtained:

$$\begin{aligned}\hat{\beta} &= C\hat{\eta} \\ V(\hat{\beta}) &= CV(\hat{\eta})C^T\end{aligned}\quad (6)$$

After obtaining  $\hat{\beta}$ , which is a vector of length  $L - l_0 + 1$ , each element  $\hat{\beta}_l$  in  $\hat{\beta}$  can be viewed in two ways. One is forward, so that one unit change of the exposure variable in the current time step will result in a  $\hat{\beta}_l$  unit change of the risk after  $l$  time steps. The other is backward, which means that a  $\hat{\beta}_l$  unit change of risk at the current time step is the result of one unit change of the exposure at  $l$  time steps early. With  $\hat{\beta}$ , it is possible to unveil the impact of an exposure event at the current time step on the changes of risk level during a period after the exposure event happens.

### 3.2.3. The basis function of DLNM

For DLNM, one basis function  $s_{k,i}(x, t)$  of a time series variable can be written as Eq. (7):

$$s_{k,i}(x, t) = \int_{l_0}^L f_k(x_{t-l}) w_i(l) dl \quad (7a)$$

$$\approx \sum_{l=l_0}^L f_k(x_{t-l}) w_i(l), \text{ where } k = 1, \dots, v_x \text{ and } i = 1, \dots, v_l \quad (7b)$$

In Eq. (7a) and (7b), the definitions of  $v_l$ ,  $L$ , and  $l_0$  are the same as the definitions in Eq. (3). Unlike DLM, there is a new weight function  $f_k(\bullet)$  introduced in DLNM. This weight function is applied to the exposure space, and there is a total of  $v_x$  weight functions.  $v_x$  is defined by the dimension of the transformation basis function on the exposure space.  $f_k(\bullet)$  can also be called as the exposure-response function. The core ideas of DLM and DLNM are similar. However, DLM only applies transformation to the lag space, while DLNM not only applies transformations onto the lag space but also applies transformations onto the exposure space. The combination of the transformation basis functions of the lag space and exposure are called cross-basis functions, and the core concept of DLNM is to apply the cross-basis functions to transform the time-series variables and their corresponding lags. In this case, there are a total of  $v_x v_l$  cross-basis functions. The algebraic representations of the DLNM basis function are as follows. Let  $A_{x,t}$  equal to:

$$A_{x,t} = (1_{v_l}^T \otimes R_{x,t}) \odot (C \otimes 1_{v_x}^T) \quad (8)$$

$$s(x, t, \eta) = (1_{L-l_0+1}^T A_{x,t}) \eta = w_{x,t}^T \eta \quad (9)$$

Where  $w_{x,t}$  is the vector of  $v_x v_l$  new variables obtained by applying exposure-response function and lag response function to the exposure and lag space. Note that here  $\otimes$  represent the Kronecker product and  $\odot$  represent the Hadamard product.  $C$  is the same as Eq. (4).  $R_{x,t}$  is a matrix of shape  $(L - l_0 + 1) \times v_x$  which is obtained by applying  $v_x$  basis functions to the exposure history vector  $q_{x,t} = [x_{t-l_0}, x_{t-l_1}, \dots, x_{t-L}]^T$ .

By stacking  $w_{x,t}^T$  row wise, we can get a matrix  $W$  of the shape  $N \times v_l v_x$ . This matrix can be included into the design matrix to estimate parameters  $\eta$  corresponding to the  $v_x v_l$  new variables after obtaining the estimated variables  $\hat{\eta}$ . Parameters  $\hat{\beta}_{x_p}$  representing the risk contribution during the  $L - l_0 + 1$  lag period by exposure change and the corresponding (co)variance matrix  $V(\hat{\beta}_{x_p})$ , can be calculated using a similar

method as the DLM.

$$\begin{aligned}\hat{\beta}_{x_p} &= A_{x_p} \hat{\eta} \\ V(\hat{\beta}_{x_p}) &= A_{x_p} V(\hat{\eta}) A_{x_p}^T\end{aligned}\quad (10)$$

Note that  $A_{x_p}$  in Eq. (10) is obtained by doing the transformation in Eq. (8), where  $R_{x,t}$  in this case is a matrix obtained by applying  $v_x$  basis functions to an exposure history vector  $q_{x,t} = [x_p, x_p, x_p, \dots, x_p]^T$ , whose elements are all equal to the same exposure level  $x_p$ .  $\hat{\beta}_{x_p}$  can be interpreted as the risk level at each lagged timestep caused by level  $x_p$  exposure at the current time step.

In the discussion section, relative risk is used to represent the risk contribution at each time lag. For DLNM,  $\hat{\beta}_{x_p}$  in Eq. (10) are the parameters representing the risk level at each time step during the lag period. The relative risk is calculated as  $\exp(\hat{\beta}_{x_p})$ . For DLM, the relative risk is calculated as  $\exp(a\hat{\beta})$ , where  $a$  is the number of unit change in exposure intensity since  $\hat{\beta}$  in Eq. (6) represents the risk contribution at each time lag brought by one unit change of exposure. Thus, a relative risk greater than 1 means increasing segment crash risk, a relative risk less than 1 means decreasing segment crash risk, and a relative risk equal to 1 means no effect on the segment crash risk.

### 3.3. Quasi-Poisson regression

For the proposed analysis of DLM and DLNM, a quasi-Poisson regression method is applied to estimate the parameters  $\eta$  introduced in the previous sections. The reason for choosing the quasi-Poisson model is that the standard Poisson model assumes that the variance of count data (i.e., hourly crash numbers) is equal to the mean. However, as shown in Table 2, the mean and variance of hourly crash numbers are different. Thus, the quasi-Poisson model is chosen since it does not require variance equal to the mean, and instead the variance is modeled as a linear function of the mean (Ver Hoef and Boveng, 2007). Eq. (11) and (12) are the mathematical representations of the proposed quasi-Poisson model, where  $Y_t$  is the number of crashes of a specific roadway segment in a specific hour.  $precip_{t,l}$  is the cross-basis matrix of hourly precipitation,  $vsby_{t,l}$  is the cross-basis matrix of hourly visibility,  $spdstd_{t,l}$  is the cross-basis matrix of hourly speed standard deviation, and  $temp_{t,l}$  is the cross-basis matrix of hourly temperature.  $\eta_1, \eta_2, \eta_3, \eta_4$  are the corresponding parameters vectors with the parameters of the new variables created by the transformation.

$$E(Y_t) = \mu_t = \exp(\alpha + \eta_1 precip_{t,l} + \eta_2 vsby_{t,l} + \eta_3 spdstd_{t,l} + \eta_4 temp_{t,l}) \quad (11)$$

$$Y_t = poisson(\mu_t) \quad (12)$$

This study applied the 'dlm' R package version 2.4.7 and R version 4.2.1 to conduct the model analysis.

## 4. Results and discussions

A DLM and a DLNM (see Table 4) were fitted using the prepared dataset. The exposure-response function  $f(x)$  and lag-response function  $w(l)$  were determined by selecting the combination that can generate the most ideal model fitting after trying multiple options. For the DLM, the exposure-response functions  $f(x)$  for all four time-series variables were linear, indicating no transformations were performed to the exposure space. The lag-response functions  $w(l)$  for all four time-series variables were B-spline functions with 5 degrees of freedom. The total degrees of freedom ( $v_l$ ) of the cross-basis of all four time-series variables were 5. As for the DLNM, the exposure-response functions  $f(x)$  for all four time-series variables were polynomial.  $f(x)$  of hourly precipitation had a degree of freedom ( $v_x$ ) equal to 2, and  $f(x)$  for other time-series variables had a degree of freedom ( $v_x$ ) equal to 4. The lag-response functions  $w(l)$

**Table 4**

Detailed Information on the DLM and DLNM.

DLM	$f(x)$	$f(x)df$	$w(l)$	$w(l)df$	Total df	AIC
Precipitation	Linear	1	B-splines	5	5	36241.98
Visibility	Linear	1	B-splines	5	5	
Speed standard deviation	Linear	1	B-splines	5	5	
Temperature	Linear	1	B-splines	5	5	
<b>DLNM</b>	$f(x)$	$f(x)df$	$w(l)$	$w(l)df$	Total df	AIC
Precipitation	Polynomials	2	Polynomials	6	12	35928.73
Visibility	Polynomials	4	Polynomials	6	24	
Speed standard deviation	Polynomials	4	Natural cubic B-splines	5	20	
Temperature	Polynomials	4	Natural cubic B-splines	5	20	

\* df = degree of freedom.

for hourly precipitation and hourly visibility were polynomial with degrees of freedom ( $v_l$ ) equal to 6 (5 + Intercept), and the lag-response functions  $w(l)$  for hourly speed standard deviation and hourly temperature were Natural cubic B-splines with degrees of freedom ( $v_l$ ) equal to 5. The cross-basis of hourly precipitation had degrees of freedom ( $v_xv_l$ ) equal to 12, the cross-basis of hourly visibility had degrees of freedom ( $v_xv_l$ ) equal to 24, and the cross-basis of hourly speed standard deviation and hourly temperature had degrees of freedom ( $v_xv_l$ ) equal to 20.

This study used Akaike information criterion (AIC) scores to compare the performance of the DLM and the DLNM. AIC scores were calculated using Eq. (13).

$$AIC = -2\ln(L) + 2k \quad (13)$$

Where  $L$  is the log-likelihood estimation of the model and  $k$  is the degree of freedom of the model. Note that for the quasi-Poisson model, the AIC score is also referred to as quasi-AIC (Bolker, 2017). The models with lower AIC scores were considered to have better performance than the ones with higher AIC scores. Based on the results, the DLM's AIC score is 36241.98, and the DLNM's AIC score is 35928.73. The two models have similar performances, but the DLNM slightly outperforms the DLM.

#### 4.1. Hourly precipitation

The modeling results of the exposure-lag-response association of hourly precipitation are presented in this section. Table 5 presents the relative risk contribution at each lagged hour brought by hourly precipitation exposure; the numbers in the parentheses are the lower 95 % confidence interval and the upper 95 % confidence interval.

There is an obvious lagged impact on roadway crash risk brought by

precipitation exposure according to the modeling results of both the DLM and the DLNM (see Fig. 2 and Fig. 3). The results generated by DLM show that the instant impact of hourly precipitation on segment crash risk is minimal and the segment crash risk starts to increase and peaks at around 2 lagged hours after the precipitation event happened. After around 2 lagged hours, the adverse impact on roadway crash risk then disappears. The results from previous studies show that hourly precipitation has an immediate impact on crash risk (Xing et al., 2019; Zhan et al., 2020), but none of these studies used roadway segment-based data to analyze the crash risk of each individual roadway segment. In this study, the results show that the immediate impact of hourly precipitation on the crash risk of a roadway segment is small, and the impact is lagged.

Furthermore, since the DLNM also applies nonlinear transformation on the exposure space, the results not only present the delayed impact of precipitation exposure, but more detailed information is also revealed by the DLNM. Firstly, lower hourly precipitation levels do not have either a positive or negative impact on segment crash risk at any lagged hour. As shown in Fig. 3, the relative risk lines of hourly precipitation with levels of 0.1 in. and 0.5 in. are rather flat and very close to 1. This indicates that lower-level hourly precipitation exposure does not impact the segment crash risk too much at any lagged hour. Note that the increases in relative risk at lag hour 8 of 0.1 in. and 0.5 hourly precipitation are caused by large confidence intervals and do not properly reflect reality, and the relative risk at lag hour 8 should be minimal as well. Secondly, for higher hourly precipitation levels (e.g., 1, 1.5, 2, and 3 in. per hour), the crash risk of a roadway segment is smaller than normal during the first few hours after high hourly precipitation happened (see Fig. 3). This indicates that during the first few hours after

**Table 5**

Relative Risk at Each Lag Hour Resulted by Precipitation Exposure.

Lagged hours	Hourly Precipitation Intensity (Inches)					
	0.1	0.5	1	1.5	2	3
<b>DLM</b>	0	1(0.95,1.04)	0.98(0.77,1.24)	0.95(0.6,1.53)	0.93(0.46,1.89)	0.91(0.35,2.34)
	1	1.03(1.01,1.06)	1.19(1.07,1.31)	1.41(1.16,1.72)	1.67(1.24,2.25)	1.99(1.34,2.95)
	2	1.04(1.01,1.07)	1.21(1.06,1.37)	1.46(1.13,1.88)	1.77(1.21,2.59)	2.14(1.28,3.55)
	3	1.02(1.01,1.04)	1.12(1.04,1.21)	1.26(1.09,1.47)	1.42(1.13,1.78)	1.6(1.18,2.16)
	4	1.01(1,1.01)	1.04(1.01,1.07)	1.09(1.03,1.15)	1.13(1.04,1.23)	1.18(1.05,1.32)
	5	1(1,1.01)	1.02(1,1.05)	1.05(0.99,1.11)	1.08(0.99,1.17)	1.1(0.98,1.23)
	6	1.01(1,1.01)	1.04(1.01,1.07)	1.08(1.02,1.15)	1.13(1.02,1.24)	1.17(1.03,1.33)
	7	1.01(1,1.02)	1.04(0.99,1.08)	1.07(0.98,1.18)	1.11(0.97,1.28)	1.16(0.96,1.38)
<b>DLNM</b>	8	0.99(0.98,1.01)	0.96(0.9,1.03)	0.93(0.82,1.06)	0.9(0.74,1.09)	0.87(0.67,1.12)
	0	1.05(0.94,1.17)	1.06(0.75,1.5)	0.73(0.37,1.46)	0.33(0.05,2.06)	0.1(0.4,1.4)
	1	1.08(1.01,1.16)	1.28(1,1.64)	1.11(0.68,1.79)	0.65(0.21,2.02)	0.26(0.03,2.45)
	2	1.05(1,1.1)	1.18(0.99,1.4)	1.16(0.83,1.63)	0.96(0.42,2.22)	0.67(0.12,3.59)
	3	1.02(0.99,1.06)	1.11(0.96,1.28)	1.18(0.89,1.56)	1.21(0.75,1.94)	1.19(0.56,2.54)
	4	1.02(0.99,1.05)	1.11(0.97,1.26)	1.22(0.95,1.58)	1.35(0.93,1.96)	1.48(0.91,2.42)
	5	1.02(0.99,1.05)	1.11(0.95,1.29)	1.22(0.91,1.65)	1.35(0.87,2.09)	1.48(0.84,2.61)
	6	1.01(0.98,1.04)	1.05(0.91,1.21)	1.1(0.84,1.44)	1.15(0.77,1.71)	1.2(0.72,2.02)
	7	1.01(0.96,1.06)	1.01(0.81,1.27)	0.93(0.59,1.46)	0.76(0.37,1.59)	0.57(0.18,1.74)
	8	1.13(1.02,1.26)	1.41(1.03,1.93)	0.99(0.53,1.86)	0.35(0.06,2)	0.06(0,2.3)

\*Numbers in the parentheses are: (lower 95% confidence interval, upper 95% confidence interval).

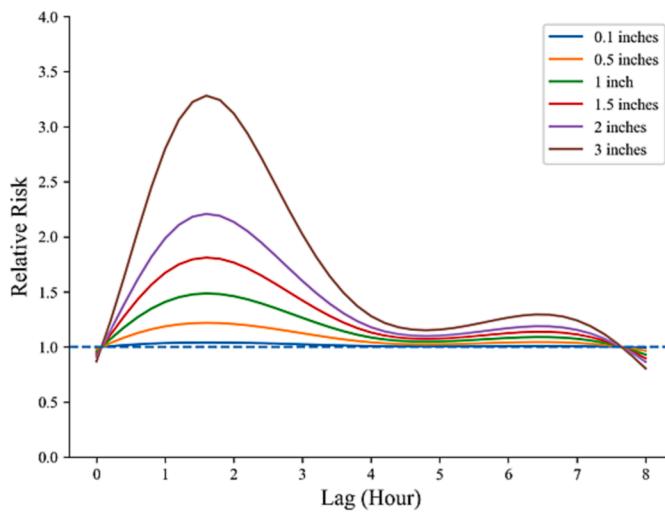


Fig. 2. DLM exposure-lag-response association for hourly precipitation.

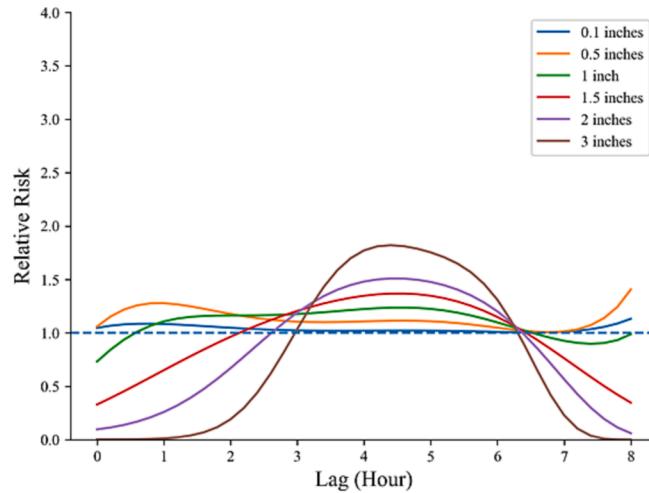
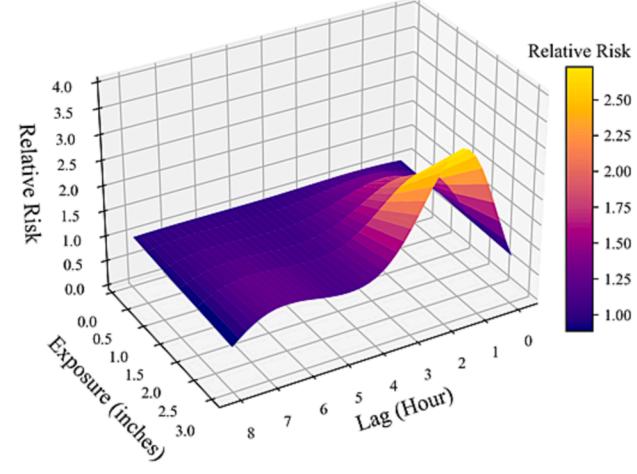


Fig. 3. DLNM exposure-lag-response association for hourly precipitation.

high hourly precipitation happened, crashes are less likely to happen. This is because during the first few lagged hours after high hourly precipitation occurs, drivers tend to drive more safely by slowing down due to heavy rainfall. Previous studies also concluded that higher precipitation leads to safer driving behavior (Eisenberg, 2004; Wei et al., 2023). Although it is not obvious in Fig. 2, this instant decrease in roadway crash risk brought by high-level hourly precipitation is also captured by the results of the DLM. As shown in Table 5, the relative risk at lag hour 0 generated by the DLM decreases as hourly precipitation intensity increases. Lastly, the lagged impact of higher levels of hourly precipitation on segment crash risk starts to increase after the first few lagged hours and peaks around lag hour 4 (see Fig. 3). This is because after heavy rain stops, drivers on the roadway start to increase their driving speed, while at the same time, the roadway surface becomes slippery as rainfall accumulated on the roadway surface starts to mix with engine oil and gasoline accumulated on the roadway surface under dry conditions (Eisenberg, 2004). Slippery roadway surfaces and high driving speeds together cause the segment crash risk to increase at lagged hours.

Both the DLM and the DLNM can reveal the lagged impact of hourly precipitation exposure on segment crash risk. Since the DLNM not only applies a nonlinear transformation to the lagged space but also applies a nonlinear transformation to the exposure space, it is capable of revealing more information on the exposure-lag-response association.

It is important to note, as mentioned in the methodology section, that the relative risk values shown in Fig. 2 and Fig. 3 represent the increase in segment crash risk at each lagged hour solely due to the hourly precipitation exposure in the current hour. These values are not associated with the hourly precipitation exposure in any other hours. For a cumulative increase in segment crash risk at a particular hour, the lagged impacts of exposure events from multiple preceding hours on the segment crash risk at the current hour must be aggregated. For instance, to determine the total increase in segment crash risk at 5 pm due to precipitation, one should sum up the 1st hour of lagged impact of precipitation exposure at 4 pm, the 2nd hour of lagged impact of precipitation exposure at 3 pm, and so forth. This also applies to the other timeseries variables investigated in this study.

#### 4.2. Hourly visibility

The modeling results of the exposure-lag-response association of hourly visibility are presented in this section. Table 6 presents the relative risk effects at each lagged hour brought by hourly visibility exposure.

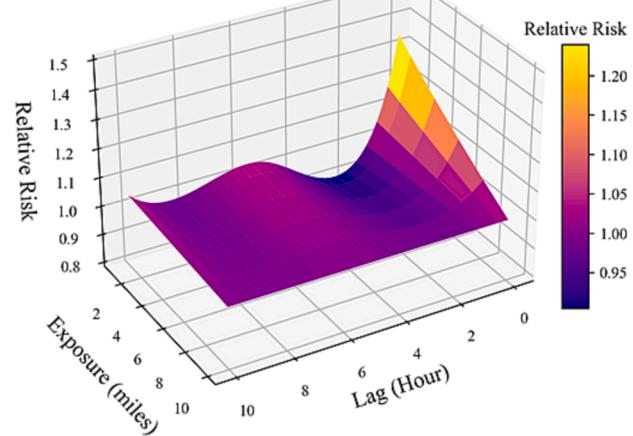
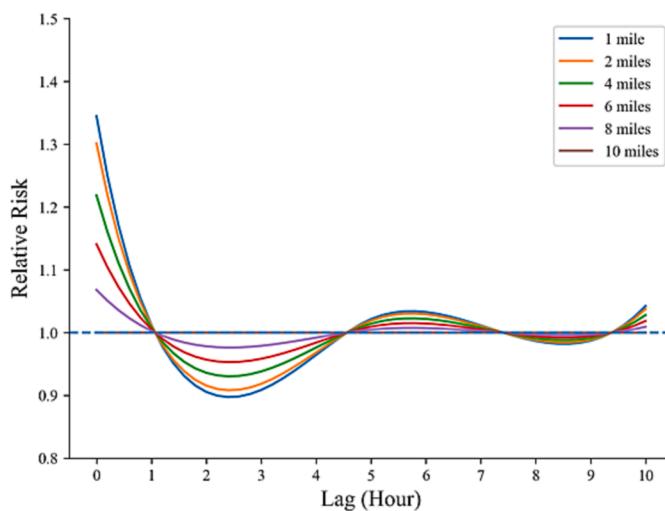
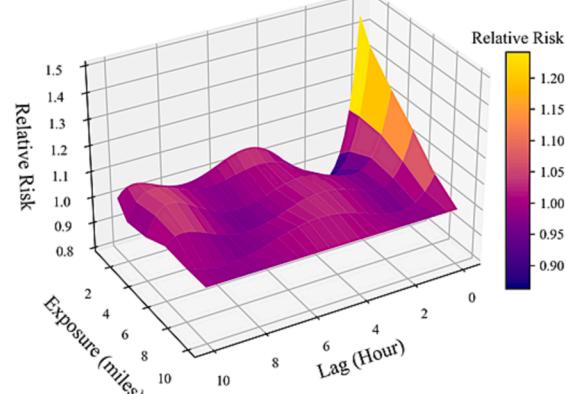
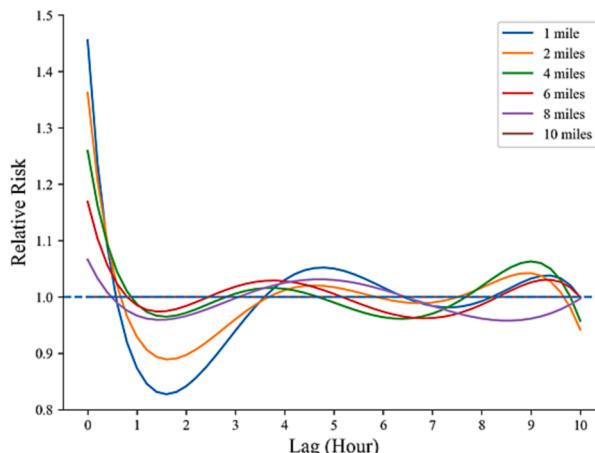
The patterns of the lagged impact of hourly visibility on segment crash risk obtained by the DLM and the DLNM are similar (see Fig. 4 and Fig. 5). The immediate impact on segment crash risk is very significant,

**Table 6**

Relative Risk at Each Lag Hour Resulted by Visibility Exposure.

	Lagged hours	Hourly Visibility Level (Miles)					
		1	2	4	6	8	10
<b>DLM</b>	0	1.34(1.19,1.52)	1.3(1.17,1.45)	1.22(1.12,1.32)	1.14(1.08,1.2)	1.07(1.04,1.1)	1(1,1)
	1	1.01(0.96,1.07)	1.01(0.96,1.06)	1.01(0.97,1.05)	1.01(0.98,1.03)	1(0.99,1.02)	1(1,1)
	2	0.91(0.84,0.98)	0.92(0.85,0.98)	0.94(0.89,0.99)	0.96(0.92,0.99)	0.98(0.96,1)	1(1,1)
	3	0.91(0.86,0.96)	0.92(0.87,0.96)	0.94(0.9,0.97)	0.96(0.94,0.98)	0.98(0.97,0.99)	1(1,1)
	4	0.96(0.92,1.02)	0.97(0.93,1.01)	0.98(0.94,1.01)	0.98(0.96,1.01)	0.99(0.98,1)	1(1,1)
	5	1.02(0.96,1.09)	1.02(0.96,1.08)	1.01(0.97,1.06)	1.01(0.98,1.04)	1(0.99,1.02)	1(1,1)
	6	1.03(0.98,1.09)	1.03(0.98,1.08)	1.02(0.99,1.06)	1.01(0.99,1.04)	1.01(1,1.02)	1(1,1)
	7	1.01(0.95,1.07)	1.01(0.96,1.06)	1.01(0.97,1.05)	1(0.98,1.03)	1(0.99,1.02)	1(1,1)
	8	0.99(0.91,1.07)	0.99(0.92,1.06)	0.99(0.94,1.05)	0.99(0.96,1.03)	1(0.98,1.02)	1(1,1)
	9	0.99(0.93,1.05)	0.99(0.94,1.04)	0.99(0.95,1.03)	0.99(0.97,1.02)	1(0.98,1.01)	1(1,1)
<b>DLNM</b>	10	1.04(0.91,1.19)	1.04(0.92,1.17)	1.03(0.94,1.12)	1.02(0.96,1.08)	1.01(0.98,1.04)	1(1,1)
	0	1.46(1.22,1.74)	1.36(1.17,1.59)	1.26(1.12,1.41)	1.17(1.06,1.29)	1.07(0.98,1.17)	1(1,1)
	1	0.87(0.76,1.01)	0.93(0.82,1.05)	0.99(0.9,1.08)	0.99(0.91,1.06)	0.97(0.9,1.03)	1(1,1)
	2	0.84(0.76,0.93)	0.9(0.82,0.98)	0.97(0.91,1.04)	0.98(0.93,1.04)	0.97(0.92,1.02)	1(1,1)
	3	0.94(0.86,1.03)	0.96(0.89,1.04)	1.01(0.95,1.07)	1.02(0.97,1.07)	1(0.95,1.04)	1(1,1)
	4	1.03(0.94,1.13)	1.01(0.93,1.1)	1.01(0.96,1.08)	1.03(0.98,1.08)	1.02(0.98,1.07)	1(1,1)
	5	1.05(0.97,1.13)	1.02(0.95,1.09)	0.99(0.94,1.04)	1.01(0.96,1.05)	1.03(0.99,1.07)	1(1,1)
	6	1.02(0.92,1.12)	1(0.92,1.08)	0.96(0.91,1.02)	0.97(0.92,1.03)	1.01(0.97,1.06)	1(1,1)
	7	0.98(0.91,1.08)	0.99(0.91,1.07)	0.97(0.92,1.03)	0.96(0.91,1.01)	0.98(0.94,1.03)	1(1,1)
	8	0.99(0.89,1.11)	1.02(0.92,1.12)	1.02(0.95,1.09)	0.99(0.93,1.05)	0.96(0.91,1.01)	1(1,1)
	9	1.03(0.89,1.2)	1.04(0.91,1.19)	1.06(0.97,1.17)	1.03(0.94,1.11)	0.96(0.91,1.03)	1(1,1)
	10	1(0.82,1.22)	0.94(0.79,1.12)	0.96(0.84,1.09)	1(0.89,1.12)	1(0.91,1.1)	1(1,1)

\*Numbers in the parentheses are: (lower 95% confidence interval, upper 95% confidence interval).

**Fig. 4.** DLM exposure-lag-response association for hourly visibility.**Fig. 5.** DLNM exposure-lag-response association for hourly visibility.

and a lower visibility level results in a larger increase in segment crash risk. This finding echoes some previous studies (Das et al., 2018; Peng et al., 2017). Moreover, the impact of low visibility drastically decreases during the following lagged hours. This indicates that the increase in segment crash risk brought by low visibility is instantaneous, and since visibility does not have the same kind of lagged impact as precipitation does once the low visibility event ends, segment crash risk decreases quickly after the low visibility event ends.

However, one thing that is captured by the models is that during the first few hours after a low visibility event happened (e.g., lag hours 1, 2, and 3), the relative crash risk decreases below 1 (see Fig. 4 and Fig. 5). This indicates that a roadway segment is safer than normal during the first few hours after a low visibility event happened. One possible explanation of this is that during the first few lag hours (e.g., lag hours 1, 2, and 3) after a low visibility event happened, most of the drivers driving onto the roadway come from roadways in the vicinity areas where lag hour 0 (when low visibility happened) also has low visibility, and these drivers are more likely to continue driving cautiously (Hogendoorn et al., 2010).

#### 4.3. Hourly temperature

The modeling results of the exposure-lag-response association of hourly temperature are presented in this section. The relative risk effects at each lag hour brought by hourly temperature exposure are listed in Table 7.

The modeling results of the DLM show that there is neither an instant nor lagged impact on roadway crash risk by hourly temperature exposure (see Fig. 6). The relative risk lines of all temperature levels are flat and close to 1. However, the results from the DLNM provide more information on the impact of hourly temperature exposure on roadway crash risk (see Fig. 7). For higher temperature levels (e.g., 30°F, 50°F, 70°F, and 90°F), there is no obvious impact on roadway crash risk. This agrees with the results of the DLM. However, for lower temperature levels (e.g., 10°F and 20°F), the results from DLNM show that there is an immediate impact on roadway crash risk. Especially for 10°F and 20°F hourly temperatures, there is a significant increase in roadway crash relative risk at lag hour 0. During the few hours after low temperature exposure, the crash relative risk decreases to below 1 and then gradually returns to 1. This pattern is also seen in the modeling results of hourly visibility. The possible explanation for this pattern is that low

temperature exposure is highly possible to cause wet roadway surfaces to become icy, which can adversely increase segment crash risk. Moreover, during the first few hours right after low temperature exposure ends, most drivers on the roadway segment come from roadway segments in vicinity areas that also experienced low temperature, and thus the drivers have adopted more cautious driving behaviors (Brijs et al., 2008). Thus, during the first few hours after low temperature exposure happened, when they are driving on the current roadway segment they are still driving in a cautious manner, which makes the segment crash risk lower than normal.

A possible reason why the DLM does not capture the impact of low hourly temperature on segment crash risk is that unlike other time-series variables analyzed in this study, since the temperature is largely dependent on seasonality, it is extremely rare to see hourly temperature changes a lot in the dataset analyzed in this study. For example, it is nearly impossible to see temperature changes from 10°F in the current hour to 80°F in the next hour. As a result, the DLM is not able to generate any valuable results. The advantage of the DLNM is that since it applies a nonlinear transformation onto the exposure space (in this study polynomial transformation), a relatively small variation in hourly temperature can be magnified by the transformation. This potentially helps the model to find the instant and lagged impact of low hourly temperature on segment crash risk.

#### 4.4. Hourly speed standard deviation

The modeling results of the exposure-lag-response association of hourly speed standard deviation are presented in this section. Table 8 presents the relative risk effects at each lag hour brought by hourly speed standard deviation exposure.

The results from the DLM and DLNM both show that there are no obvious lagged impacts of high speed variation exposure on segment crash risk (see Fig. 8 and Fig. 9). The instant increase in segment crash risk at the current hour (i.e., lag hour 0) is significant and quickly disappears in the following lagged hours, and the relative segment crash risk stays around 1. The reason for this is that speed variation is not always consistent on roadway segments close to each other. During the lagged hours after a high speed variation event happened on one segment, drivers driving on this segment did not necessarily experience high speed variations at the current hour (i.e., lag hour 0) when they were driving on segments in vicinity areas. Thus, the assumption on the

**Table 7**

Relative Risk at Each Lag Hour Resulted by Temperature Exposure.

	Lagged hours	Hourly Temperature Level (°F)					
		10	20	30	50	70	
<b>DLM</b>	0	0.99(0.94,1.05)	0.98(0.88,1.11)	0.98(0.82,1.16)	0.96(0.72,1.29)	0.95(0.63,1.42)	0.93(0.55,1.57)
	1	0.99(0.95,1.02)	0.97(0.91,1.04)	0.96(0.87,1.05)	0.93(0.79,1.09)	0.9(0.72,1.13)	0.88(0.66,1.17)
	2	0.99(0.94,1.03)	0.98(0.89,1.07)	0.96(0.84,1.1)	0.94(0.75,1.18)	0.92(0.67,1.26)	0.89(0.59,1.35)
	3	0.99(0.97,1.02)	0.99(0.94,1.04)	0.98(0.9,1.07)	0.97(0.85,1.12)	0.96(0.79,1.17)	0.95(0.74,1.22)
	4	1(0.97,1.03)	1(0.95,1.06)	1(0.92,1.1)	1(0.88,1.17)	1(0.83,1.24)	1(0.79,1.32)
	5	1.01(0.97,1.05)	1.02(0.94,1.11)	1.03(0.91,1.16)	1.04(0.85,1.29)	1.06(0.79,1.42)	1.08(0.74,1.57)
	6	1.01(0.98,1.04)	1.02(0.96,1.08)	1.03(0.95,1.12)	1.05(0.91,1.21)	1.07(0.88,1.31)	1.1(0.85,1.41)
	7	1.01(0.98,1.04)	1.02(0.96,1.08)	1.02(0.94,1.12)	1.04(0.9,1.21)	1.06(0.86,1.3)	1.08(0.82,1.41)
	8	1(0.96,1.05)	1(0.92,1.11)	1(0.88,1.17)	1(0.81,1.3)	1.03(0.74,1.44)	1.04(0.68,1.6)
	9	1(0.97,1.03)	1(0.94,1.07)	1(0.91,1.11)	1(0.86,1.19)	1(0.81,1.27)	1.02(0.76,1.36)
<b>DLNM</b>	10	1(0.94,1.07)	1(0.88,1.14)	1(0.83,1.21)	1(0.74,1.38)	1(0.65,1.57)	1(0.58,1.78)
	0	3.47(1.08,11.17)	1.55(0.94,2.57)	1.08(0.84,1.38)	0.97(0.9,1.05)	1(0.94,1.08)	0.97(0.8,1.17)
	1	1.32(0.88,1.98)	1.17(0.98,1.39)	1.08(0.99,1.18)	1.01(0.98,1.03)	1(0.97,1.02)	0.96(0.9,1.03)
	2	0.67(0.25,1.82)	0.94(0.6,1.45)	1.06(0.84,1.33)	1.03(0.96,1.1)	0.99(0.93,1.05)	0.96(0.81,1.15)
	3	0.59(0.26,1.32)	0.85(0.61,1.22)	1(0.84,1.2)	1.03(0.97,1.09)	0.98(0.94,1.03)	0.97(0.84,1.11)
	4	0.83(0.42,1.65)	0.88(0.65,1.19)	0.94(0.8,1.09)	1(0.96,1.05)	0.98(0.94,1.02)	0.98(0.88,1.1)
	5	1.21(0.4,3.63)	0.95(0.58,1.55)	0.9(0.7,1.16)	0.98(0.91,1.05)	0.99(0.93,1.06)	1(0.82,1.21)
	6	1.27(0.65,2.49)	1.01(0.75,1.36)	0.93(0.81,1.08)	0.97(0.93,1.01)	1.02(0.98,1.06)	1.03(0.92,1.15)
	7	1.1(0.45,2.71)	1.05(0.7,1.56)	1(0.82,1.22)	0.97(0.92,1.03)	1.04(0.99,1.09)	1.06(0.91,1.22)
	8	0.97(0.33,2.84)	1.06(0.66,1.7)	1.05(0.82,1.34)	0.99(0.92,1.06)	1.04(0.97,1.1)	1.06(0.89,1.27)
	9	0.96(0.61,1.52)	1.04(0.86,1.27)	1.06(0.96,1.16)	1.01(0.99,1.04)	1(0.98,1.03)	1.04(0.97,1.12)
	10	1(0.25,3.96)	1.02(0.56,1.84)	1.04(0.78,1.39)	1.04(0.96,1.13)	0.96(0.89,1.03)	1.01(0.82,1.25)

\*Numbers in the parentheses are: (lower 95% confidence interval, upper 95% confidence interval).

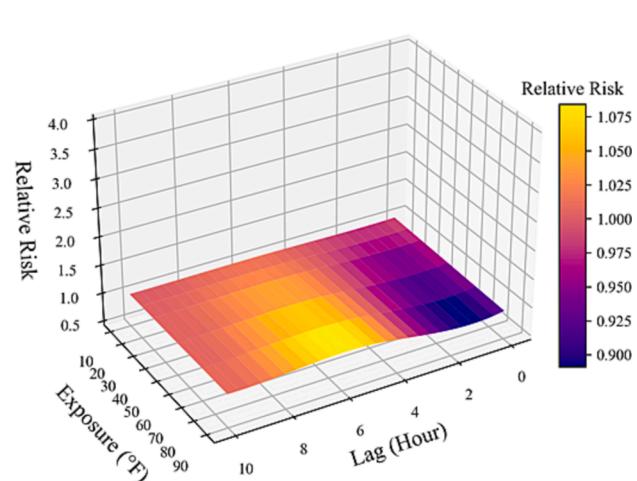
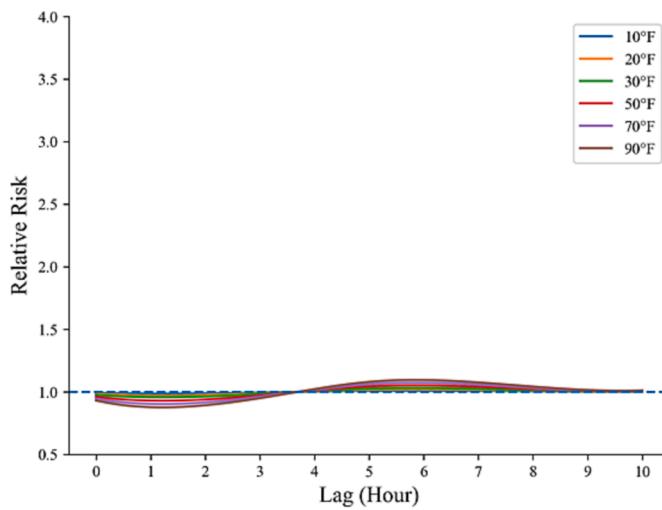


Fig. 6. DLM exposure-lag-response association for hourly temperature.

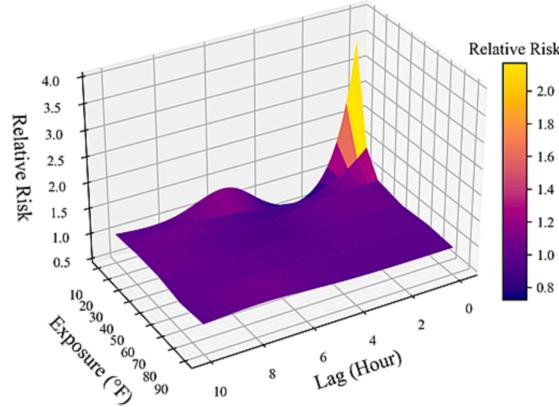
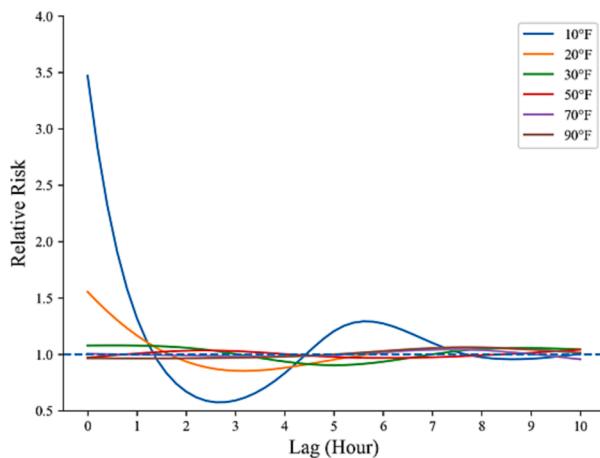


Fig. 7. DLNM exposure-lag-response association for hourly temperature.

**Table 8**

Relative Risk at Each Lag Hour Resulted by Speed Standard Deviation Exposure.

	Lagged hours	Hourly Speed Standard Deviation Level (mph)					
		5	6	7	8	9	10
<b>DLM</b>	0	2.03(1.98,2.07)	2.33(2.27,2.4)	2.69(2.6,2.77)	3.09(2.98,3.21)	3.56(3.42,3.71)	4.1(3.92,4.3)
	1	1.24(1.21,1.27)	1.3(1.26,1.34)	1.36(1.31,1.4)	1.42(1.36,1.47)	1.48(1.41,1.55)	1.54(1.47,1.62)
	2	0.98(0.95,1)	0.97(0.94,1)	0.97(0.93,1)	0.96(0.92,1.01)	0.96(0.91,1.01)	0.95(0.91,1.01)
	3	0.91(0.89,0.93)	0.89(0.87,0.92)	0.88(0.85,0.9)	0.86(0.83,0.89)	0.84(0.81,0.88)	0.83(0.79,0.87)
	4	0.93(0.91,0.96)	0.92(0.89,0.95)	0.91(0.87,0.95)	0.9(0.86,0.94)	0.88(0.84,0.93)	0.87(0.82,0.92)
	5	0.98(0.95,1.01)	0.98(0.94,1.01)	0.97(0.93,1.02)	0.97(0.92,1.02)	0.96(0.91,1.02)	0.96(0.91,1.02)
	6	0.99(0.96,1.02)	0.99(0.96,1.02)	0.99(0.95,1.02)	0.98(0.94,1.03)	0.98(0.93,1.03)	0.98(0.93,1.03)
	7	0.97(0.94,1)	0.96(0.93,1)	0.95(0.91,1)	0.95(0.91,1)	0.94(0.89,1)	0.94(0.88,1)
	8	0.94(0.9,0.97)	0.92(0.88,0.97)	0.91(0.86,0.96)	0.9(0.84,0.96)	0.89(0.83,0.95)	0.87(0.81,0.95)
	9	0.91(0.88,0.94)	0.9(0.86,0.93)	0.88(0.84,0.92)	0.86(0.82,0.91)	0.85(0.8,0.9)	0.83(0.78,0.89)
<b>DLNM</b>	10	0.92(0.86,0.98)	0.9(0.84,0.98)	0.89(0.81,0.97)	0.87(0.79,0.97)	0.86(0.77,0.97)	0.85(0.74,0.96)
	0	3.87(3.28,4.57)	4.81(4.07,5.68)	5.83(4.95,6.86)	6.9(5.9,8.07)	7.96(6.85,9.25)	8.96(7.76,10.35)
	1	1.61(1.46,1.77)	1.74(1.59,1.92)	1.88(1.71,2.05)	2(1.84,2.18)	2.12(1.95,2.31)	2.23(2.06,2.42)
	2	0.84(0.73,0.95)	0.82(0.72,0.93)	0.81(0.71,0.91)	0.8(0.71,0.9)	0.79(0.71,0.89)	0.79(0.71,0.89)
	3	0.67(0.6,0.75)	0.63(0.57,0.71)	0.61(0.54,0.68)	0.59(0.53,0.65)	0.57(0.52,0.64)	0.57(0.51,0.63)
	4	0.79(0.71,0.89)	0.77(0.69,0.85)	0.75(0.68,0.83)	0.74(0.67,0.82)	0.74(0.67,0.81)	0.74(0.67,0.82)
	5	1(0.86,1.15)	1(0.87,1.15)	1.01(0.88,1.15)	1.02(0.89,1.16)	1.04(0.91,1.18)	1.06(0.93,1.21)
	6	1.04(0.93,1.17)	1.05(0.94,1.17)	1.06(0.95,1.17)	1.07(0.97,1.18)	1.08(0.98,1.19)	1.1(0.99,1.21)
	7	0.96(0.84,1.1)	0.95(0.84,1.08)	0.94(0.84,1.06)	0.94(0.83,1.05)	0.93(0.83,1.04)	0.92(0.82,1.04)
	8	0.88(0.76,1.02)	0.86(0.75,0.99)	0.85(0.74,0.96)	0.83(0.73,0.94)	0.82(0.72,0.93)	0.8(0.7,0.92)
	9	0.85(0.76,0.94)	0.83(0.75,0.92)	0.82(0.74,0.9)	0.81(0.74,0.88)	0.8(0.73,0.87)	0.79(0.71,0.87)
	10	0.84(0.69,1.03)	0.83(0.68,1.01)	0.82(0.68,0.99)	0.82(0.69,0.98)	0.82(0.69,0.98)	0.83(0.69,0.99)

\*Numbers in the parentheses are: (lower 95% confidence interval, upper 95% confidence interval).

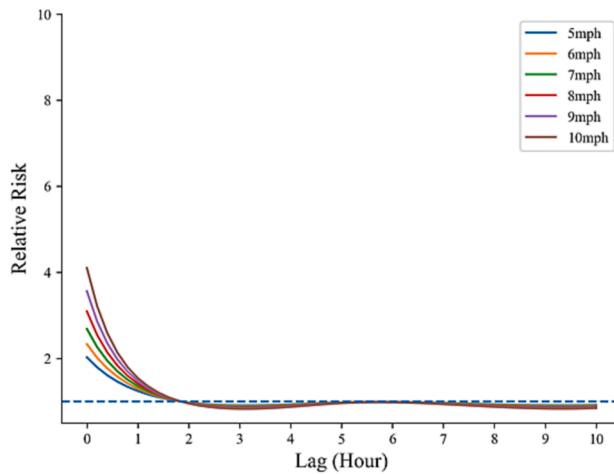


Fig. 8. DLM exposure-lag-response association for hourly speed standard deviation.

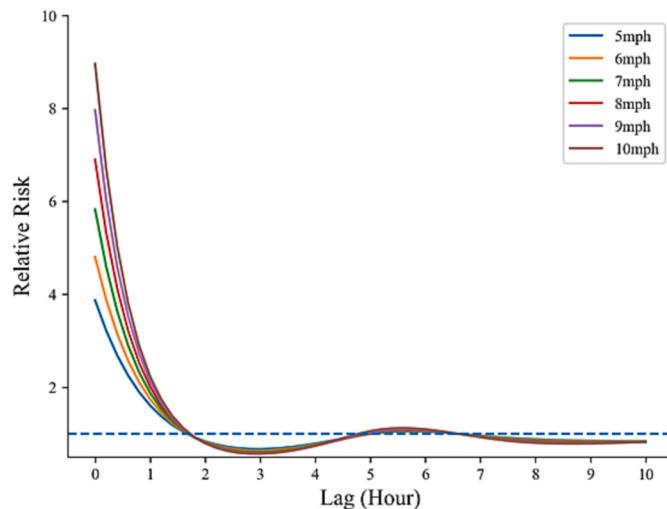
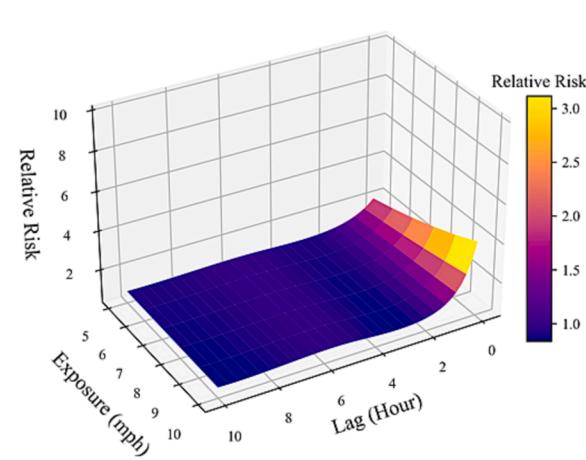
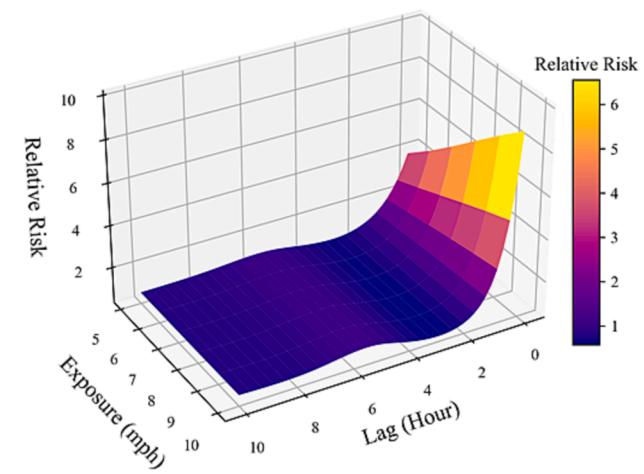


Fig. 9. DLNM exposure-lag-response association for hourly speed standard deviation.



memory of driving behaviors does not exist in this case.

Considering the presence of delayed impacts on the crash risk of roadway segments stemming from weather-related time-series variables, the conclusions drawn from this study hold significant implications for the implementation of various safety countermeasures. Several options are outlined below:

1. It is imperative to convey warning messages effectively and promptly through diverse channels, including in-vehicle driving assistance systems, variable message signs, and mobile phones.
2. The enhancement of road signage and signaling systems is pivotal to provide drivers with real-time weather event forecasts and timely hazard warnings. Moreover, to facilitate proactive safety management, it is crucial for displayed warning messages to adapt dynamically based on the lagged impact patterns on segment crash risk following weather-related exposure, thereby mitigating potential roadway hazards.
3. Implement a variable speed limit (VSL) system that adjusts speed limits based on real-time traffic conditions. VSL can help regulate traffic flow and encourage drivers to maintain safe speeds during and after adverse weather-related exposures. Specifically, VSL can lower speed limits when segment crash risk is high and increase them when the risk is low.

4. Launch a comprehensive traffic safety campaign to educate drivers about the potential lagged impact of adverse weather-related exposure. For example, drivers should be made aware of the increased crash risk even after precipitation has ceased, and they should be educated to exercise caution when driving in the aftermath of rain.

## 5. Conclusions

Weather factors and speed variation factors are closely related to the crash risk of roadway segments. Traditional methods often assume crashes only relate to weather and speed variation events that occurred during the same time period. In reality, weather and high speed variation events may also have impacts on segment crash risk during the lagged time periods after the events occurred. Previous studies rarely investigated the interpretable representations of lagged impacts of weather- and speed variation-related factors on segment crash risk. This study applied a DLM and a DLNM on a space-time-stratified case-crossover crash dataset to investigate the exposure-lag-response association between segment crash risk and four time-series variables, including hourly precipitation, hourly visibility, hourly temperature, and hourly speed standard deviation. The results show that hourly precipitation, hourly visibility, and hourly temperature have obvious lagged impacts on segment crash risk, while for hourly speed standard deviation, the modeling results indicate that there is no obvious lagged

impact on segment crash risk.

The instant impact of hourly precipitation exposure is very small at the current hour. The impact on segment crash risk increases and starts to peak during the first few lagged hours after an hourly precipitation event happened. Moreover, at the current hour, higher hourly precipitation intensity makes segment crash risk lower than that of lower hourly precipitation intensity because higher precipitation intensity makes drivers drive more cautiously. At the lagged hour when segment crash risk peaks, higher hourly precipitation intensity still brings higher peaked risk than lower hourly precipitation intensity. As for hourly visibility exposure, a low visibility event poses a significant impact on segment crash risk at the current hour. The results also show that during the first few lagged hours after low visibility occurred, the roadway is safer than normal. For hourly temperature exposure, the results from the DLM do not show obvious impacts of the temperature factor on segment crash risk. Since the DLNM applies nonlinear transformations to the exposure space, it revealed that although the higher hourly temperature does not have an obvious impact on segment crash risk, lower hourly temperature poses a significant impact on segment crash risk at the current hour, and it also has lagged impact.

The findings of this study provide compelling evidence for the lagged impact of crash-causing factors on roadway segment-level crash risk. This highlights the importance of incorporating time-series effects of these factors in future crash modeling studies, rather than relying solely on cross-sectional models. Moreover, the observed lagged impact patterns can offer valuable guidance for the development of real-time crash warning systems. For instance, following a precipitation exposure event, there is a delay before segment crash risk reaches its peak. During this period, the roadside warning displays can dynamically adjust their messages to inform drivers about the current crash risk level. Additionally, the implementation of a variable speed limit (VSL) system can be based on segment crash risk after exposure events. Specifically, the VSL system can advise higher speed limits when segment crash risk is lower, promoting smoother traffic flow, while recommending lower speed limits when segment crash risk is higher, prioritizing safety. However, it is important to note that this study applied 1-hour as the data aggregation interval. If applying smaller data aggregation interval, it may potentially affect the modeling outcomes and generate other useful insights to improve safety countermeasures. We acknowledge the challenges to acquire and process time-series data with smaller aggregation interval, however, it is necessary to conduct future studies to incorporate such data of shorter aggregation interval and model their lagged impact on segment crash risk.

Other future study directions include further exploring the exposure-lag-response association between these time-series variables and roadway crash risk on different roadway facility types other than rural interstate highways. Crash severity should also be considered in future studies. Moreover, the impact of non-time-series variables, such as geometric information on the exposure-lag-response association, can also be further investigated. Last but not least, the correlations between different independent variables such as precipitation and visibility; visibility and temperature should also be considered in the modeling process to enhance modeling results.

#### CRediT authorship contribution statement

**Zihang Wei:** Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing. **Subasish Das:** Data curation, Writing – original draft. **Yue Wu:** Writing – original draft, Visualization. **Zihao Li:** Writing – original draft. **Yunlong Zhang:** Supervision, Writing – review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

#### Data availability

Data will be made available on request.

#### References

- Anderson, I.B., Bauer, K.M., Harwood, D.W., Fitzpatrick, K., 1999. Relationship to Safety of Geometric Design Consistency Measures for Rural Two-Lane Highways. *Transportation Research Record* 1658, 43–51. <https://doi.org/10.3141/1658-06>.
- Bolker, B., 2017. Dealing with quasi- models in R.
- Brijs, T., Karlis, D., Wets, G., 2008. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis & Prevention* 40, 1180–1190. <https://doi.org/10.1016/j.aap.2008.01.001>.
- Das, S., Brimley, B.K., Lindheimer, T.E., Zupancich, M., 2018. Association of reduced visibility with crash outcomes. *IATSS Research* 42, 143–151. <https://doi.org/10.1016/j.iatssr.2017.10.003>.
- Eisenberg, D., 2004. The mixed effects of precipitation on traffic crashes. *Accident Analysis & Prevention* 36, 637–647. [https://doi.org/10.1016/S0001-4575\(03\)00085-X](https://doi.org/10.1016/S0001-4575(03)00085-X).
- Gao, X., Jiang, W., Liao, J., Li, J., Yang, L., 2022. Attributable risk and economic cost of hospital admissions for depression due to short-exposure to ambient air pollution: A multi-city time-stratified case-crossover study. *Journal of Affective Disorders* 304, 150–158. <https://doi.org/10.1016/j.jad.2022.02.064>.
- Garber, N.J., Gadireja, R., 1989. Factors affecting speed variance and its influence on accidents. *Transportation Research Record* 1213, 64–71.
- Gasparrini, A., 2014. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med* 33, 881–899. <https://doi.org/10.1002/sim.5963>.
- Gasparrini, A., Armstrong, B., Kenward, M.G., 2010. Distributed lag non-linear models. *Stat Med* 29, 2224–2234. <https://doi.org/10.1002/sim.3940>.
- Haghghi, N., Liu, X.C., Zhang, G., Porter, R.J., 2018. Impact of roadway geometric features on crash severity on rural two-lane highways. *Accident Analysis & Prevention* 111, 34–42. <https://doi.org/10.1016/j.aap.2017.11.014>.
- Hoogendoorn, R.G., Tamminga, G., Hoogendoorn, S.P., Daamen, W., 2010. Longitudinal driving behavior under adverse weather conditions: adaptation effects, model performance and freeway capacity in case of fog, in: 13th International IEEE Conference on Intelligent Transportation Systems. Presented at the 13th International IEEE Conference on Intelligent Transportation Systems, pp. 450–455. <https://doi.org/10.1109/ITSC.2010.5625046>.
- Huang, H., Abdel-Aty, M.A., Darwiche, A.L., 2010. County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record* 2148, 27–37. <https://doi.org/10.3141/2148-04>.
- Imprailou, M.-I.-M., Quddus, M., Pittfield, D.E., Lord, D., 2016. Re-visiting crash-speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention* 86, 173–185. <https://doi.org/10.1016/j.aap.2015.10.001>.
- Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention* 135, 105371. <https://doi.org/10.1016/j.aap.2019.105371>.
- Li, Z., Kong, X., Zhang, Y., 2023. Exploring Factors Associated With Crossing Assertiveness of Pedestrians at Unsignalized Intersections. *Transportation Research Record* 03611981221145140. <https://doi.org/10.1177/03611981221145140>.
- Liu, L., Racz, D., Vaillancourt, K., Michelman, J., Barnes, M., Mellem, S., Eastham, P., Green, B., Armstrong, C., Bal, R., O'Banion, S., Guo, F., 2023. Smartphone-based hard-braking event detection at scale for road safety services. *Transportation Research Part c: Emerging Technologies* 146, 103949. <https://doi.org/10.1016/j.trc.2022.103949>.
- Lord, D., Manning, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part a: Policy and Practice* 44, 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>.
- Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention* 25, 689–709. [https://doi.org/10.1016/0001-4575\(93\)90034-T](https://doi.org/10.1016/0001-4575(93)90034-T).
- National Center for Statistics and Analysis, 2022. Early Estimates of Motor Vehicle Traffic Fatalities and Fatality Rate by Sub-Categories 2021 (CrashStats Brief Statistical Summary. Report No. DOT HS 813 298). National Highway Traffic Safety Administration.
- Nyadanu, S.D., Tessema, G.A., Mullins, B., Pereira, G., 2022. Maternal acute thermophysiological stress and stillbirth in Western Australia, 2000–2015: A space-time-stratified case-crossover analysis. *Science of the Total Environment* 836, 155750. <https://doi.org/10.1016/j.scitotenv.2022.155750>.
- Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. *Accident Analysis & Prevention*, Intelligent Speed Adaptation + Construction Projects 48, 464–471. <https://doi.org/10.1016/j.aap.2012.03.005>.
- Peng, Y., Abdel-Aty, M., Shi, Q., Yu, R., 2017. Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. *Transportation Research Part c: Emerging Technologies* 74, 295–305. <https://doi.org/10.1016/j.trc.2016.11.022>.
- Peng, Y., Li, C., Wang, K., Gao, Z., Yu, R., 2020. Examining imbalanced classification algorithms in predicting real-time traffic crash risk. *Accident Analysis & Prevention* 144, 105610. <https://doi.org/10.1016/j.aap.2020.105610>.
- Schwartz, J., 2000. The Distributed Lag between Air Pollution and Daily Deaths. *Epidemiology* 11, 320–326.

- Scott, P.P., 1986. Modelling time-series of British road accident data. *Accident Analysis & Prevention*, Special Issue Accident Modelling 18, 109–117. [https://doi.org/10.1016/0001-4575\(86\)90055-2](https://doi.org/10.1016/0001-4575(86)90055-2).
- Texas Department of Transportation, 2020. Roadway Inventory [WWW Document]. accessed 2.23.21. <https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-inventory.html>.
- Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson Vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data? *Ecology* 88, 2766–2772. <https://doi.org/10.1890/07-0043.1>.
- Wang, X., Zhou, Q., Quddus, M., Fan, T., Fang, S., 2018. Speed, speed variation and crash relationships for urban arterials. *Accident Analysis & Prevention* 113, 236–243. <https://doi.org/10.1016/j.aap.2018.01.032>.
- Wei, Z., Das, S., Zhang, Y., 2022. Short Duration Crash Prediction for Rural Two-Lane Roadways: Applying Explainable Artificial Intelligence. *Transportation Research Record* 03611981221096113. <https://doi.org/10.1177/03611981221096113>.
- Wei, Z., Zhang, Y., Das, S., 2023. Applying Explainable Machine Learning Techniques in Daily Crash Occurrence and Severity Modeling for Rural Interstates. *Transportation Research Record* 2677, 611–628. <https://doi.org/10.1177/03611981221134629>.
- Wu, Y., Li, S., Guo, Y., 2021. 2021. Space-Time-Stratified Case-Crossover Design in Environmental Epidemiology Study, *Health Data Science* <https://doi.org/10.34133/2021/9870798>.
- Xing, F., Huang, H., Zhan, Z., Zhai, X., Ou, C., Sze, N.N., Hon, K.K., 2019. Hourly associations between weather factors and traffic crashes: Non-linear and lag effects. *Analytic Methods in Accident Research* 24, 100109.
- Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science* 63, 50–56. <https://doi.org/10.1016/j.ssci.2013.10.012>.
- Yu, R., Xiong, Y., Abdel-Aty, M., 2015. A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. *Transportation Research Part c: Emerging Technologies*, Special Issue on Road Safety and Simulation 50, 68–77. <https://doi.org/10.1016/j.trc.2014.09.016>.
- Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transportation Research Record* 2673, 314–326. <https://doi.org/10.1177/0361198119840611>.
- Zanobetti, A., Wand, M.P., Schwartz, J., Ryan, L.M., 2000. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 1, 279–292. <https://doi.org/10.1093/biostatistics/1.3.279>.
- Zhan, Z.-Y., Yu, Y.-M., Chen, T.-T., Xu, L.-J., Ou, C.-Q., 2020. Effects of hourly precipitation and temperature on road traffic casualties in Shenzhen, China (2010–2016): A time-stratified case-crossover study. *Science of the Total Environment* 720, 137482. <https://doi.org/10.1016/j.scitotenv.2020.137482>.
- Zhang, S., Abdel-Aty, M., 2022. Real-time crash potential prediction on freeways using connected vehicle data. *Analytic Methods in Accident Research* 36, 100239. <https://doi.org/10.1016/j.amar.2022.100239>.