

From Crash Data to Risk Fields

How Traffic Research in Geography Helps Us Understand Road Safety

Pengyu Chen, Ph.D. Student in Geospatial and Environmental Analysis
Department of Geography

Original research question:

Which factor influence crash fatality most, road slope?
Road curvature? Or other variables?

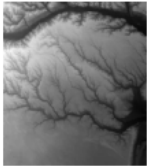
Future question: Where/which road could be more dangerous?

Two prerequisite knowledge:

1. DEM (Digital Elevation Models) Data
2. (Linear) Regression Model

DEM Data

Digital Elevation Models, or DEMs, are three-dimensional models of the “bare earth” surface, showing **elevation** and **topography** without trees, buildings, or other objects. DEMs are created from a variety of sources. USGS DEMs used to be derived primarily from topographic maps. Those are being systematically replaced with DEMs derived from high-resolution lidar and IfSAR (Alaska only) data.



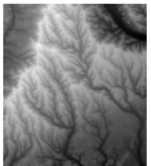
USGS 1/3 Arc Second n32w083 20220725

Published Date: 2022-07-25

Metadata Updated: 2022-07-25

Format: GeoTIFF

Extent: 1 x 1 degree



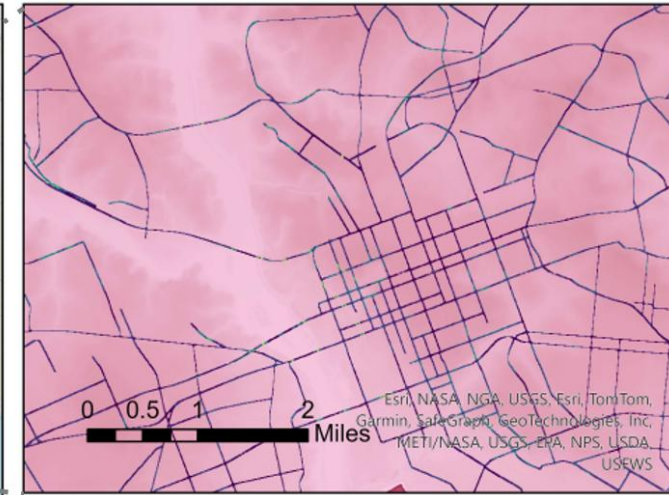
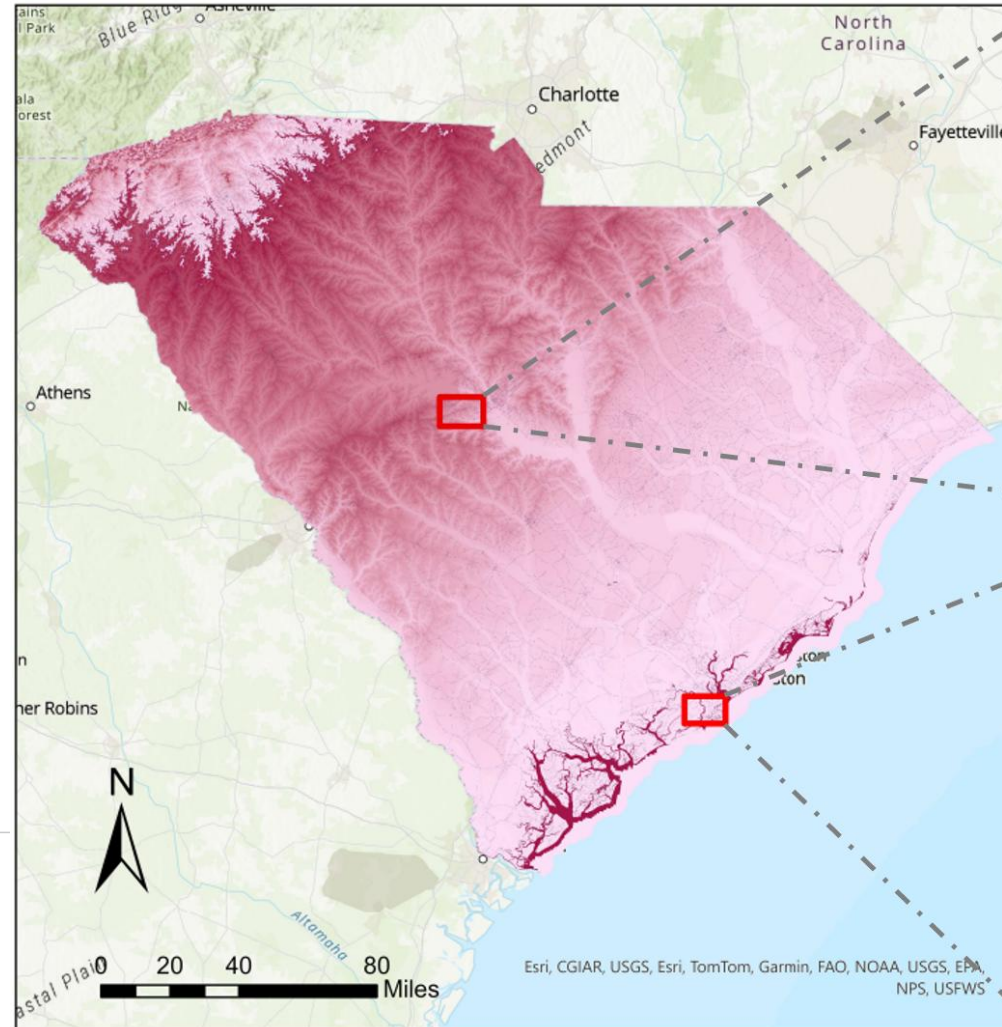
USGS 1/3 Arc Second n32w084 20230215

Published Date: 2023-02-15

Metadata Updated: 2023-02-15

Format: GeoTIFF

Extent: 1 x 1 degree



Regression model

People use Regression model for finding relationship, **Inference**, or prediction

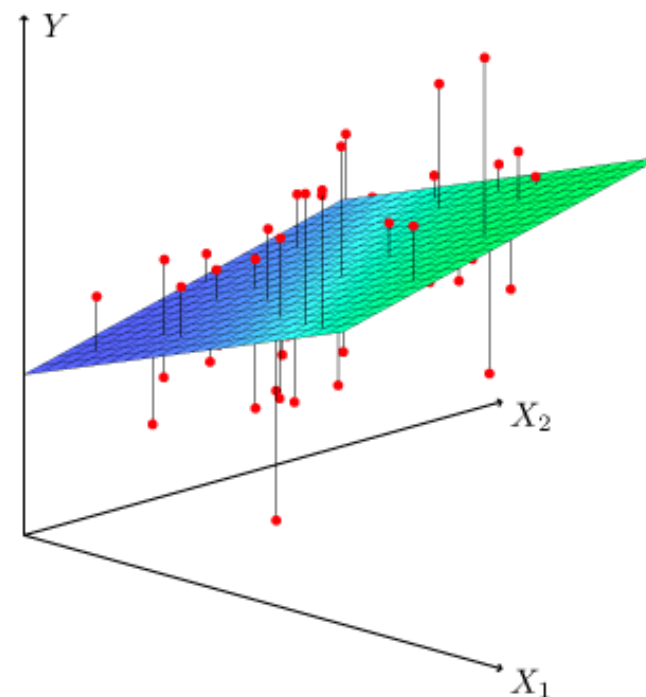
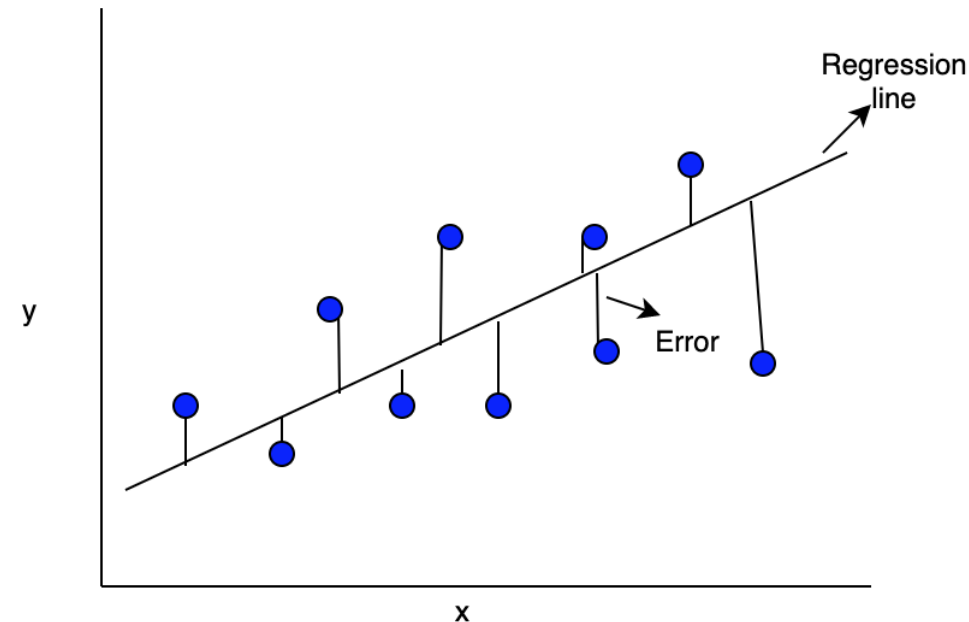
- 1) See how variables connect to each other.
- 2) **See if a relationship is real or just luck.**
- 3) Guess future numbers.

The model we use: Inhomogeneous Poisson point process

For today: Think it as a *Linear regression*

$$Y_i = \beta_0 + \beta_1 X_1 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$



Inhomogeneous Poisson point process (IPPP)

$$\log \lambda(s) = \beta_0 + \beta_1 X_1(s) + \cdots + \beta_p X_p(s) = X(s)^T \beta$$

$\lambda(s)$: intensity function, Represents the density or rate of event occurrence at position s .

Log-Likelihood Function of IPPP

$$\begin{aligned} \log L(\beta) &= \sum_{i=1}^n \log \lambda(s_i) - \int_W \lambda(s) ds \\ &= \sum_{i=1}^n X(s_i)^T \beta - \int_W \exp(X(s)^T \beta) ds \end{aligned}$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

1. Why use Poisson?

Because we are dealing with **counting rare, independent events**.

- **No negatives, whole numbers only:** You can observe 0,1,2, or 10 trees or car crashes in an area, but never -3 or 2.4.
- **Natural fit for counts:** The standard bell curve (Normal distribution) assumes numbers can be negative and continuous. The Poisson distribution is built specifically for non-negative integers (0,1,2,...) where occurrences happen independently at a certain average rate.

2. Why use a Point Process?

Because the **exact location** (x, y) of each event is the data itself.

- In classic statistics, the locations are fixed first (like choosing 50 pre-set weather stations) and then measuring a value at each one (like temperature).
- In spatial point patterns (trees in a forest, disease cases in a city, lightning strikes), the **locations where things appear are random and informative**. A point process models both *how many* events happen and *where* they choose to appear across continuous space.

Data & Data processing

DOT DATA

Traffic & Road Conditions

Emergency Response

Report Potholes

GIS Mapping

Traffic Counts

Public Transit

Motorist Info

GIS/MAPPING

Here you can find GIS data; interactive and digital maps for viewing and download; you can browse and order hard copies of our maps. Our KML web services are available for download; the Street Finder application identifies SCDOT maintained roads.

For GIS or Mapping products not present on this site please contact GIS/Mapping at 803-737-1677 or [send us an email](#).

Maps

PDF maps are available for download.

Hard copies of other map types (county, city, etc.) may be purchased for \$1.00 a sheet plus shipping &

ArcGIS Online

Go to ArcGIS Site

Downloadable GIS Data

ESRI Shapefile data is available from our GIS site.

GIS Data

Street Finder

SCDOT has partnered with local governments to produce a database of road names statewide. We currently estimate our database to have road names for over 90% of state maintained roads and

Table 1

Description of variables used in the analysis.

Variable	Description
Dependent Variable	
FatalCrash	Binary indicator for fatal crash (1 = fatal crash, 0 = background point)
Road Design	
Log (Slope)	Log-transformed road slope (in degrees), derived from DEM
Log(Curvature ⁻¹)	Log-transformed inverse road curvature (radius of fitted arc, in meters)
Dynamic Road Conditions	
Work Zone	Indicator for presence of a work zone (1 = yes, 0 = no)
Wet Surface	Indicator for wet or slippery surface condition (1 = wet, 0 = dry)
Log (AADT)	Log-transformed Average Annual Daily Traffic (normalized/interpolated)
Temporal and Environmental Factors	
Night	Indicator for nighttime crash (1 = night, 0 = day)
Weekend	Indicator for weekend crash (1 = Saturday/Sunday, 0 = weekday)
Bad Weather	Indicator of adverse weather (e.g., rain, snow)
Vehicle Characteristics	
Commercial Vehicle (CMV) Involved	Indicator for commercial motor vehicle involvement (1 = CMV, 0 = otherwise)

<https://www.scdot.org/travel/travel-mappinggis.html>

VIRGINIA
ROADS

Sign In

☆

VDOT

Virginia Department of Transportation

VDOT's Open Data Portal, **Virginia Roads**, was developed as part of an effort to provide user friendly access for exploring and downloading open data. VDOT is committed to providing the traveling public, lawmakers and partners with easy to understand information that demonstrates how we are managing the state's transportation infrastructure and to ensure VDOT's transparency to the public.

Learn About Virginia Roads

What is offered on **Virginia Roads**?

e-Stat

政府統計の総合窓口

統計で見る日本

e-Statは、日本の統計が閲覧できる政府統計ポータルサイトです

お問い合わせ | ヘルプ | English

ログイン | 新規登録

統計データを探す統計データの活用統計データの高度利用統計関連情報リンク集

トップページ / 統計データを探す / ファイル

選択条件:

ファイル

 /

道路の交通に関する統計

 /

交通事故統計月報

 /

月次

 /

2024年

 /

5月

政府統計一覧に戻る (すべて解除)

データセット

キーワードを入力

検索オプション

提供分類、表題を検索

データベース、ファイル内を検索

検索のしかた

データセット一覧

URLをコピー

一覧形式で表示

政府統計名	道路の交通に関する統計	詳細
提供統計名	交通事故統計月報	
提供周期	月次	
調査年月	2024年5月	

NCDOT Current DriveNC/TIMS Incidents Dashboard

Filter

Division: No category selected

County: All Counties

Route Type: No category selected

Road Condition: All Road Conditions

Select more than one option in all filters.

9 Primary Closures

146 Closures

445 Total Incidents

Route Type

Count

US Route	11
Interstate	91
Secondary Road	342
State Road	1

Primary Closures:

Click on an incident to zoom to it. Click on it again to zoom to the state.

Incident: 1997051 - ATMS_507000870

Incident Type: closures - Road Work

US-176, US-176 CLOSED in both directions near Rivercove Rd in Polk County, due to road work until further notice. Updated 10:32 am. ATMS ID -26050700087.

Polk County, Division 14

5/7/26, 10:25 AM -

Incident: 1997062 - ATMS_507000871

Incident Type: closures - Road Work

US-176, US-176 CLOSED in both directions near Rivercove Rd in Polk County, due to road work until further notice. Updated 10:32 am. ATMS ID -26050700087.

国家统计局

国家数据

查数ChaShu

登录 注册 English

统计热词

车流

搜索

首页 月度数据 季度数据 年度数据 普查数据 地区数据 部门数据 国际数据 出版物 我的收藏 帮助

首页 - 年度数据

简单查询 高级查询 经济图表

下载 打印

国家统计局 官方微博 官方微信 数据中国 帮助 返回首页

指标

2025年

2024年

2023年

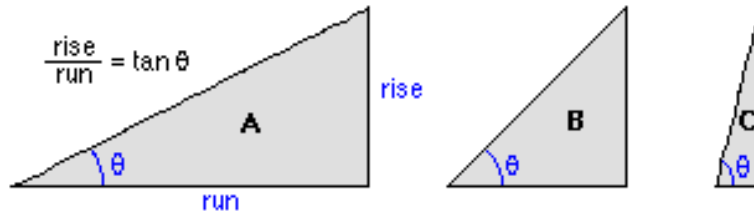
20

旅客运输量 (万人)	1712343.22	1708340.00	1574331.00	
铁路客运量 (万人)	460140.91	431240.00	385449.59	
公路客运量 (万人)	1149232.33	1178108.00	1101152.89	
水运客运量 (万人)	25955.29	25971.00	25770.74	
民用航空客运量 (万人)	77014.70	73021.00	61957.64	

Road Slope

Degree of slope = θ

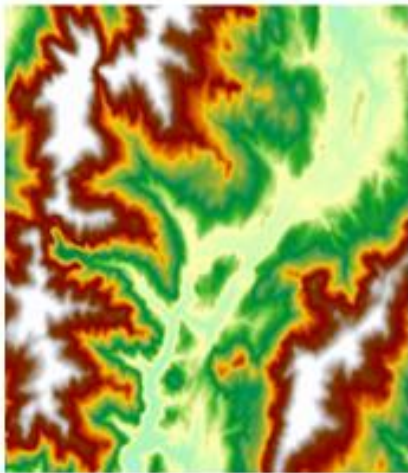
Percent of slope = $\frac{\text{rise}}{\text{run}} * 100$



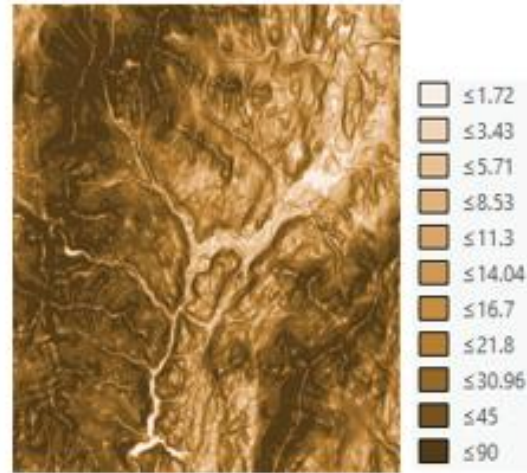
Degree of slope =

Percent of slope =

30	45	76
58	100	373



Input elevation raster



Output slope raster
(in degrees)



Road curvature

The radius of an equivalent circle represents the road curvature at a given point.

In practice, we approximate this radius using the circle defined by the three nearest neighboring points along the road.

$$\text{Curvature} = 1/\text{Radius}$$

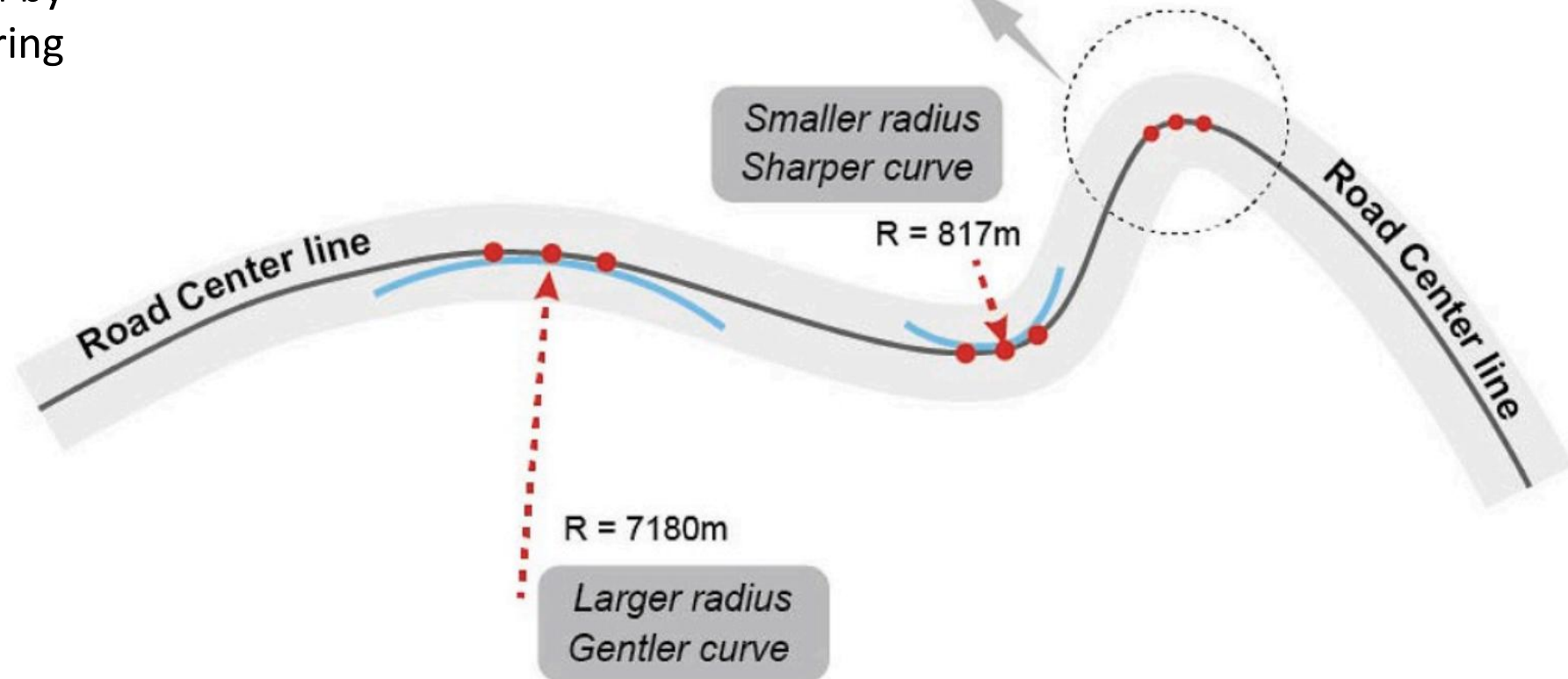
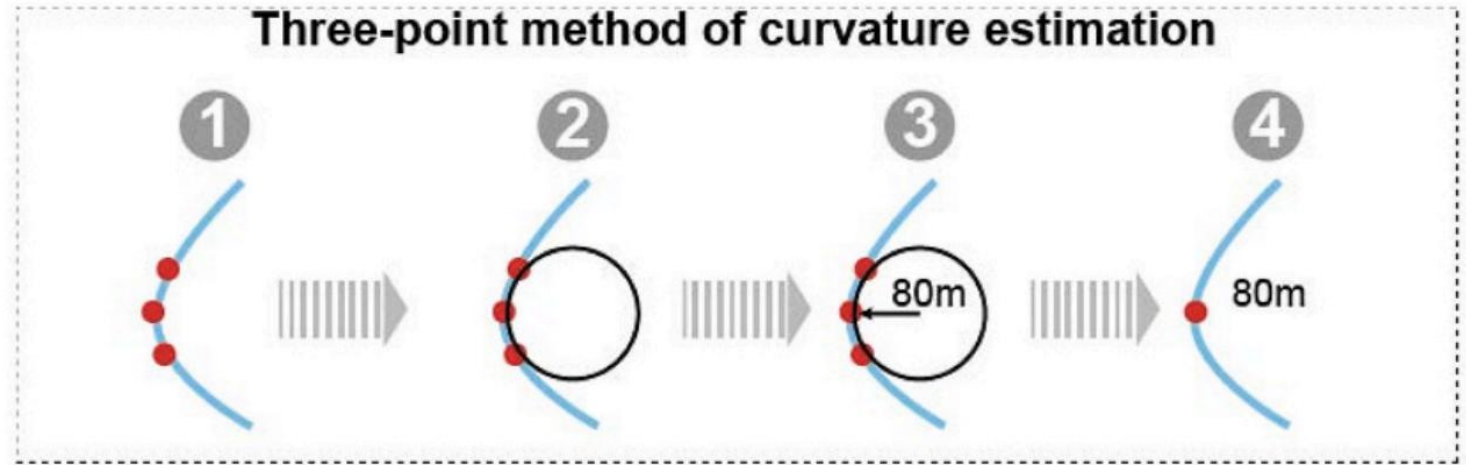


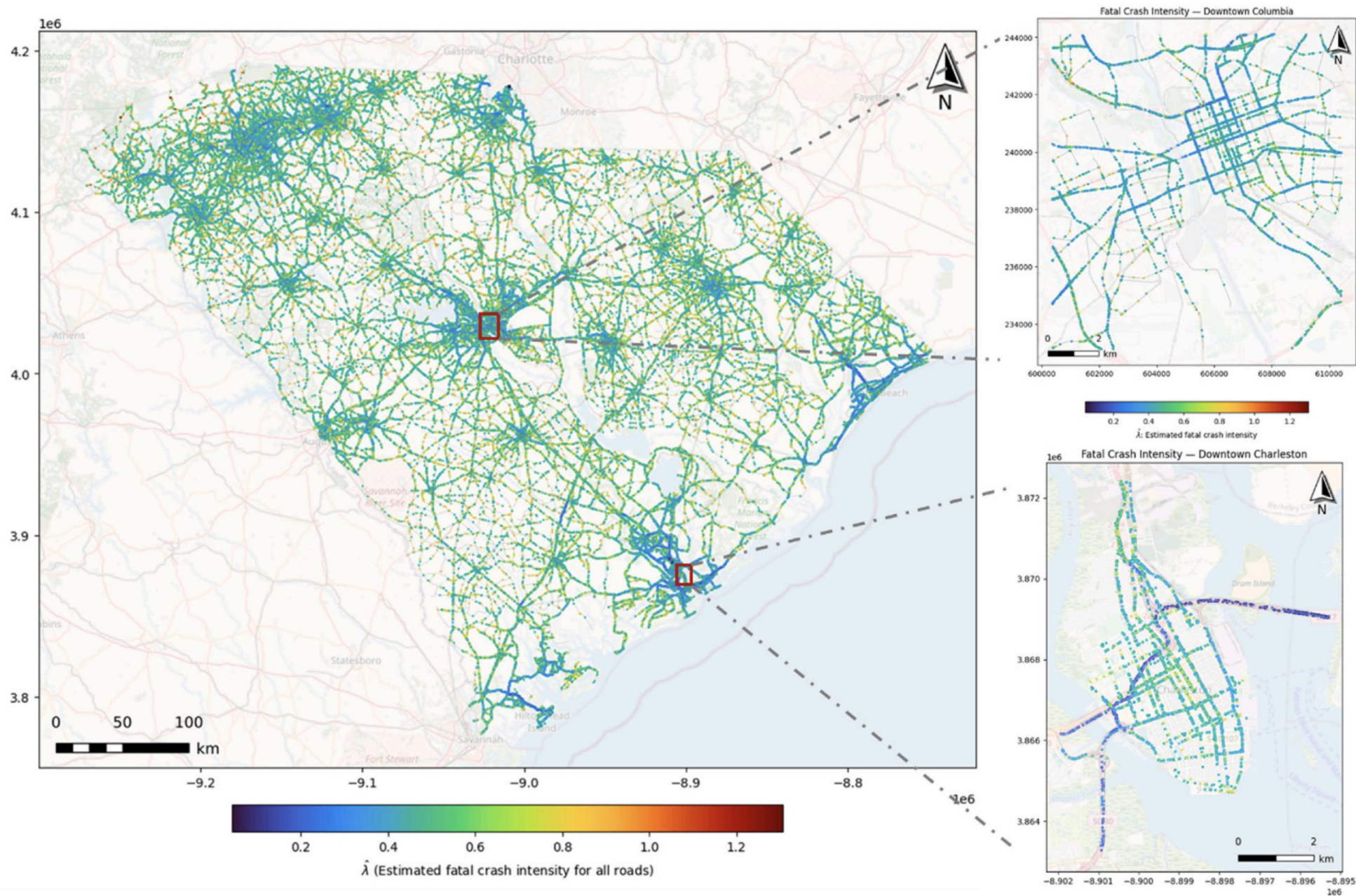
Table: Variable definitions in the analytical dataset

Variable	Description
Slope	Road slope (degrees), derived from DEM
Night	Indicator for nighttime crash (1 = yes, 0 = no)
Surface	Indicator for wet surface condition (1 = wet, 0 = dry)
Workzone	Indicator for work-zone presence (1 = yes, 0 = no)
AADT	Average Annual Daily Traffic (normalized)
Curvature	Road curvature (arc radius in meters)
Road	Road class (e.g., interstate, local)
death	Fatal crash indicator (1 = fatal, 0 = non-fatal)

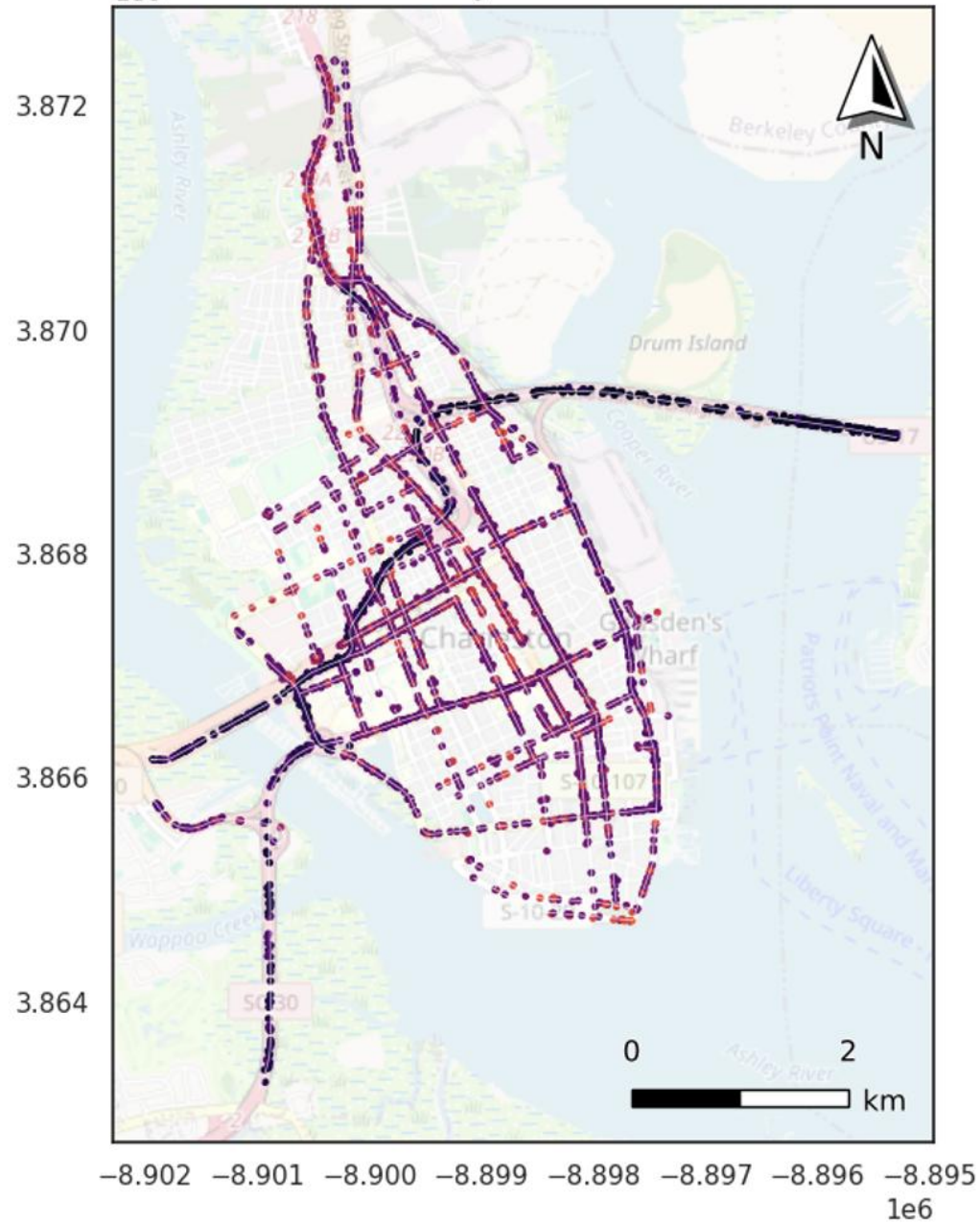
$$Y_i (\text{death}) = \beta_0 + \beta_1 X_1(\text{slope}) + \beta_2 X_2(\text{night}) + \cdots + \beta_n X_n + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Results

Result: Fatal Intensity (λ) map



1e6 Fatal Crash Intensity — Downtown Charleston



1e6 Fatal Crash Intensity — Downtown Columbia



**Result: Tables show
how variable influence
road fatal crash results**

Table 2
Poisson point process model using log-transformed variables (all road types).

	Estimate	Std. Error	z-value	p-value
DV: FatalCrash				
<i>Road Design</i>				
Log(Slope)	0.063	0.014	4.512	<0.001 ***
Log(Curvature ⁻¹)	-0.025	0.013	-1.941	0.052 ●
<i>Dynamic Road Condition</i>				
Work Zone1	0.205	0.095	2.159	0.031 *
Wet Surface	-0.049	0.045	-1.093	0.274
Log (AADT)	-0.865	0.018	-47.609	<0.001 ***
<i>Temporal and Environmental Factors</i>				
Night	0.570	0.027	21.394	<0.001 ***
Weekend	0.159	0.032	4.885	<0.001 ***
Bad Weather	0.049	0.042	1.167	0.243
<i>Vehicle Characteristics</i>				
CMV Involved	0.812	0.043	18.801	<0.001 ***
Intercept	1.618	0.136	11.914	<0.001 ***

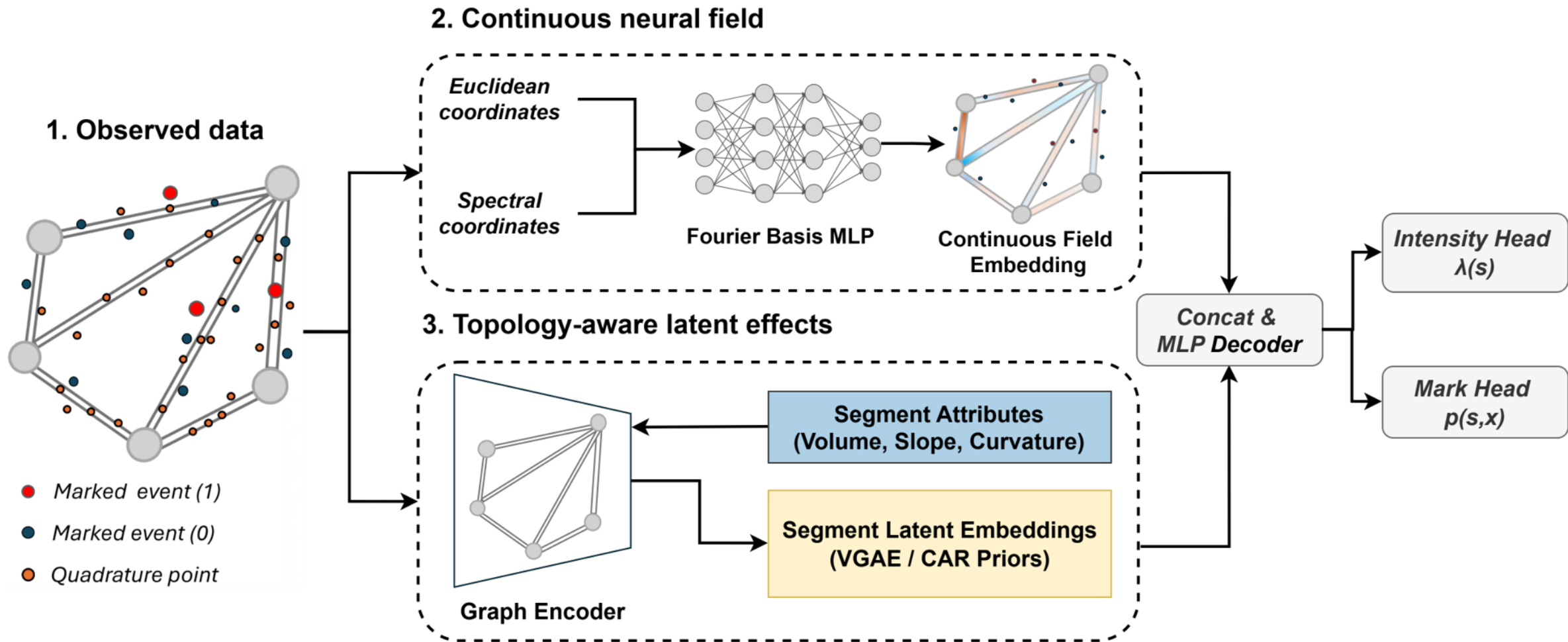
Sample Size (N): 602,955; signifiance codes: <0.001 ‘***’ <0.01 ‘**’ <0.05 ‘*’
<0.1 ‘●’

Poisson model estimates by road type with log-transformed variables.

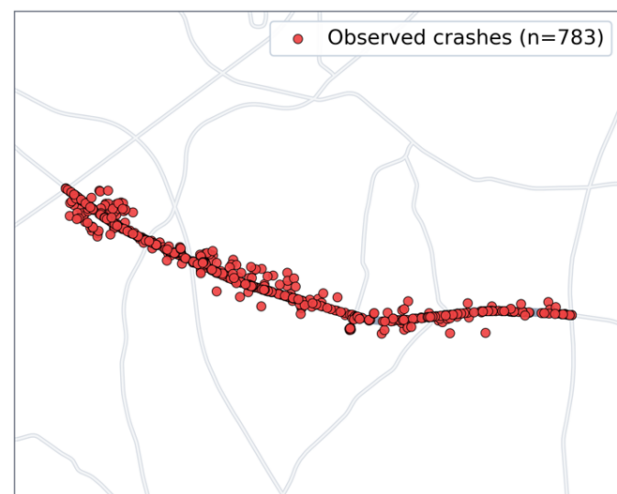
DV: FatalCrash	Stratified model of highways			Stratified model of local roads		
	Estimate	Std. Err.	p	Estimate	Std. Err.	p
Road Design						
Log (Slope)	0.0646	0.0185	<0.001***	0.0546	0.0247	0.027*
Log(Curvature ⁻¹)	-0.0065	0.0156	0.675	-0.0901	0.0197	<0.001***
Dynamic Road Condition						
Work Zone	0.1222	0.1108	0.270	0.3498	0.2985	0.241
Wet Surface	-0.0446	0.0582	0.444	-0.1503	0.0924	0.104
Log (AADT)	-0.9164	0.0289	<0.001***	-0.9904	0.0260	<0.001***
Temporal and Environmental Factors						
Night	0.7062	0.0339	<0.001***	0.6932	0.0494	<0.001***
Weekend	0.2109	0.0406	<0.001***	0.1745	0.0607	0.004**
Bad Weather	0.0267	0.0563	0.635	0.1113	0.0858	0.195
Vehicle Characteristics						
CMV Involved	0.8978	0.0522	<0.001***	0.7893	0.1033	<0.001***
Intercept	1.7431	0.2512	<0.001***	2.2809	0.1884	<0.001***

Sample Size (N): Highway = 438,200, Local = 291,810; Pseudo R²: Highway = 0.227, Local = 0.333; significance codes: <0.001 '***' <0.01 '**' <0.05 '*'.

Ongoing work



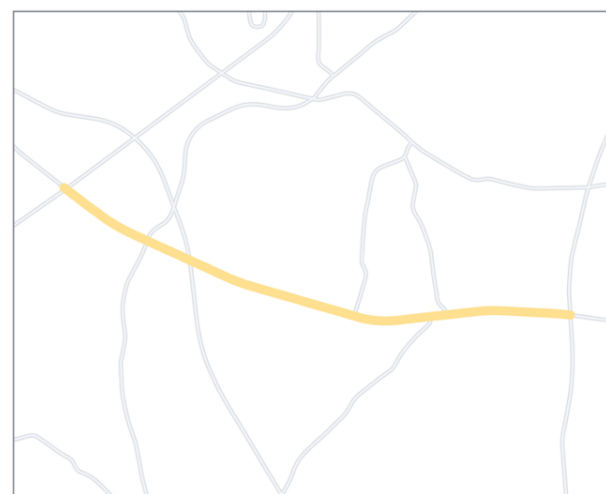
(a) Ground truth ($L = 4544$ m)



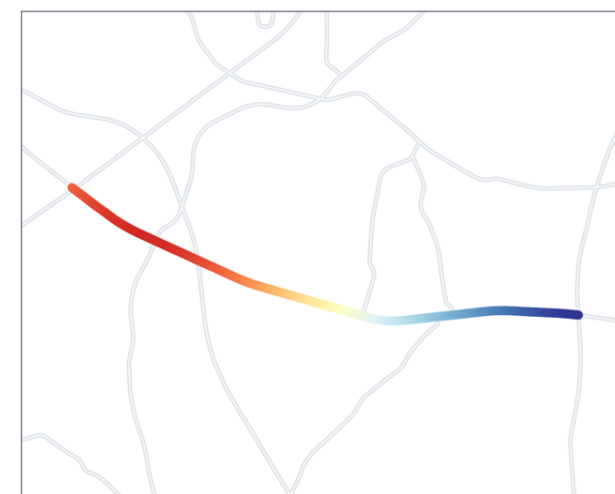
(b) GCN baseline ($\lambda = 46.0$, constant)



(c) GAT baseline ($\lambda = 122.4$, constant)



(d) Ours (continuous field)



Hazard rate λ (crashes / km)

50

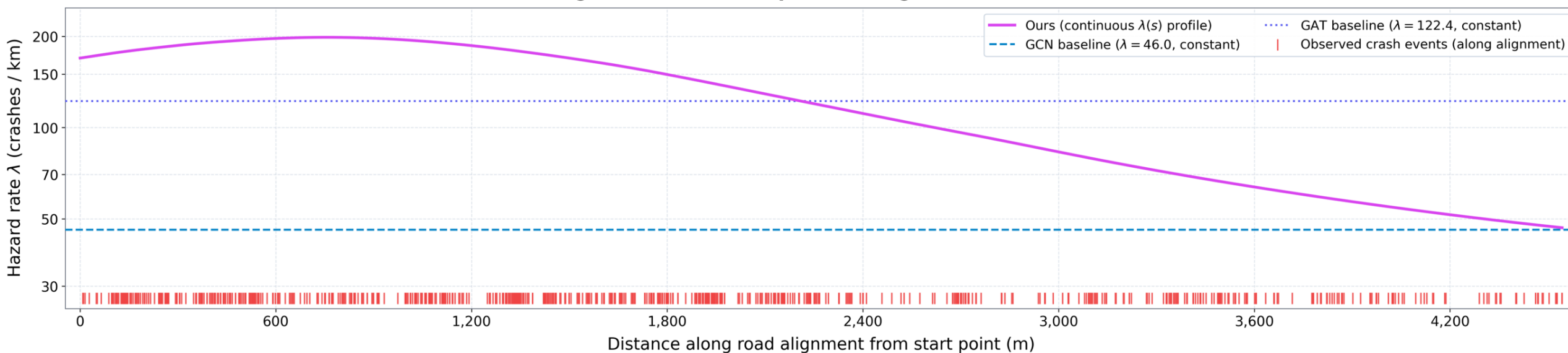
70

100

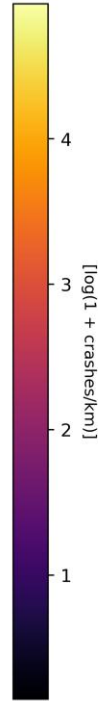
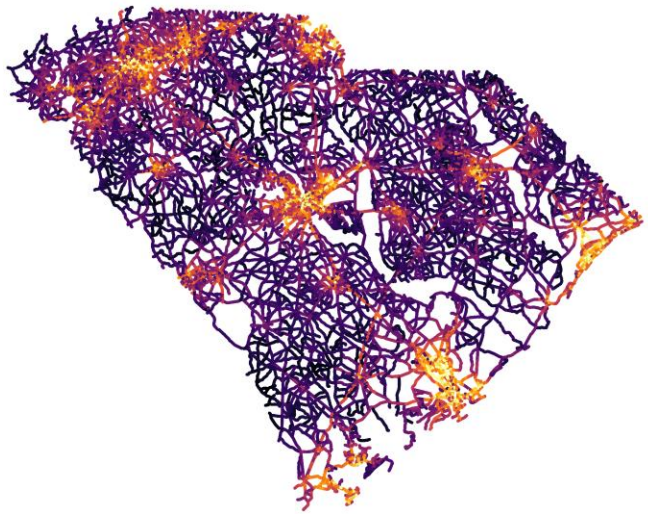
150

200

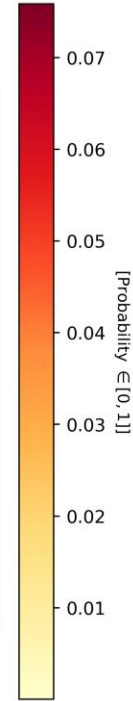
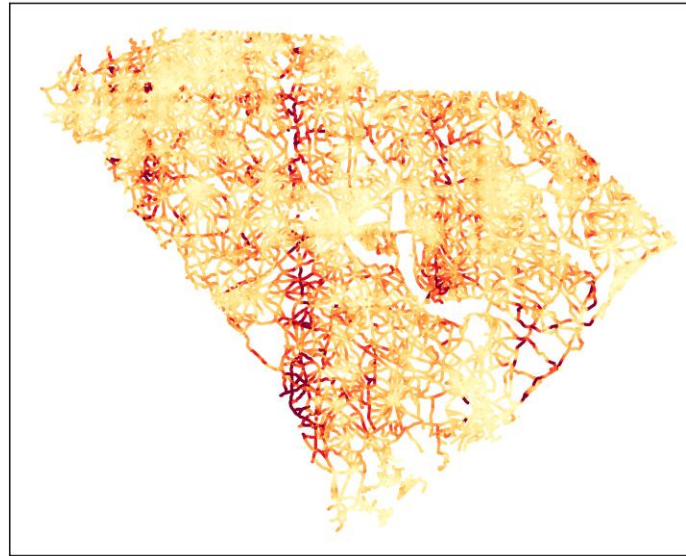
Sub-segment hazard-rate profile along the selected corridor



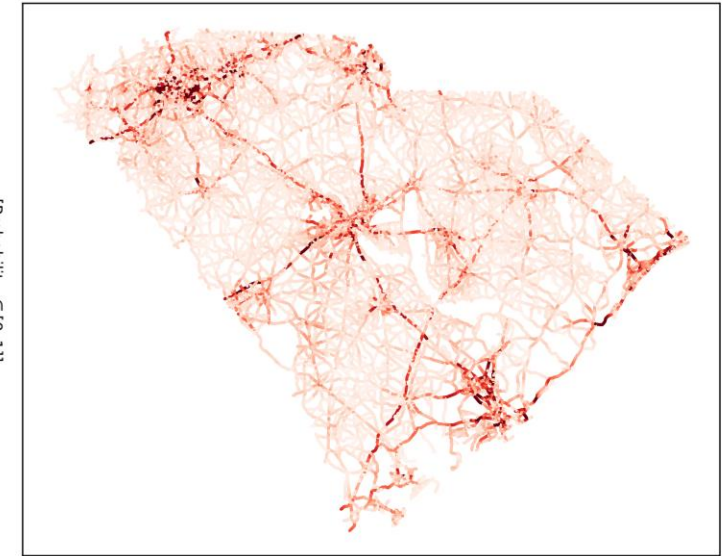
Crash Intensity $\log(1 + \hat{\lambda}(s))$



Fatality Probability $\mathbb{E}_x[p(s, x)]$



Fatal Crash Hazard $\log(1 + \hat{\lambda} \cdot \hat{p})$



Deep learning model provide more accurate & better results.

However, the results would be less explainable.

Questions?