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?

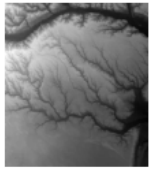
Future question: Where/which road could be more dangerous?

Two prerequisite knowledge:

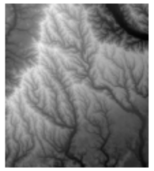
1. **DEM (Digital Elevation Models) Data**
2. **(Linear) Regression Model**

DEM Data

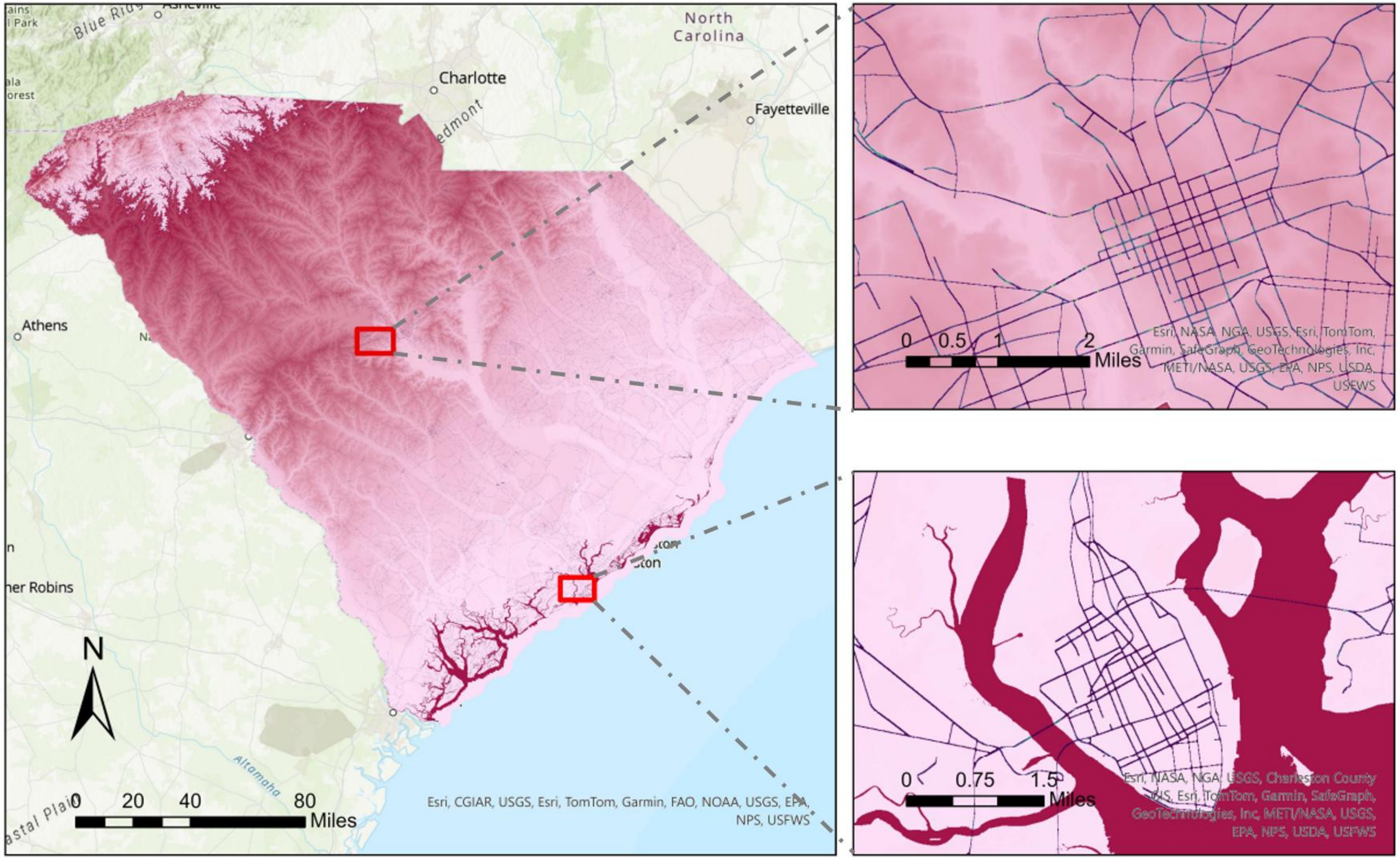
Digital Elevation Models, or DEMs, are three-dimensional models of the "bare earth" surface, showing **elevation** and **topography** without trees, buildings, or other objects. DEMs are created from a variety of sources. USGS DEMs used to be derived primarily from topographic maps. Those are being systematically 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



VIRGINIA TECH.

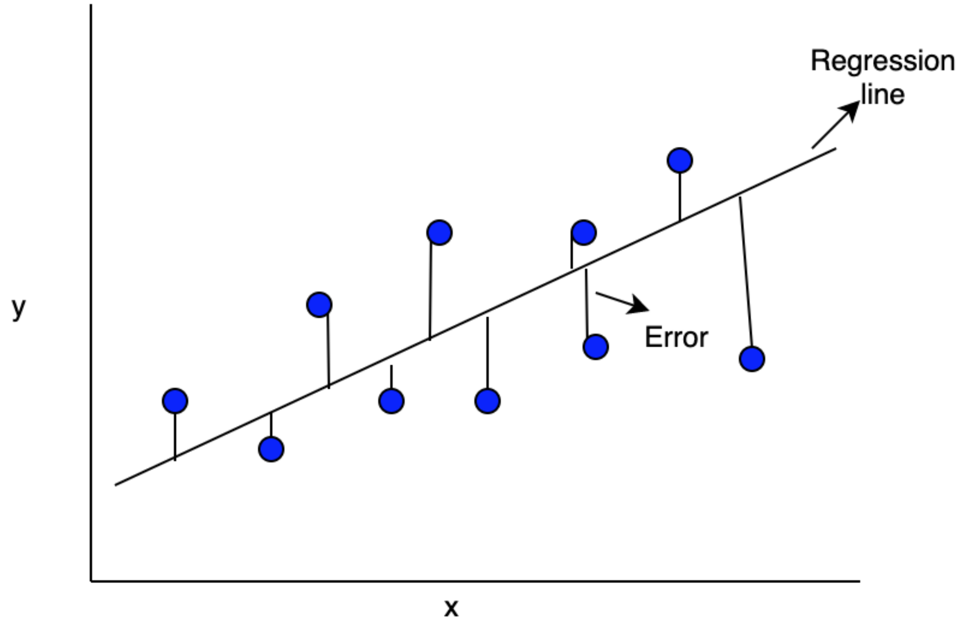
https://www.earthdata.nasa.gov/topics/land-surface/digital-elevation-terrain-model-dem

https://apps.nationalmap.gov/downloader/

Regression model

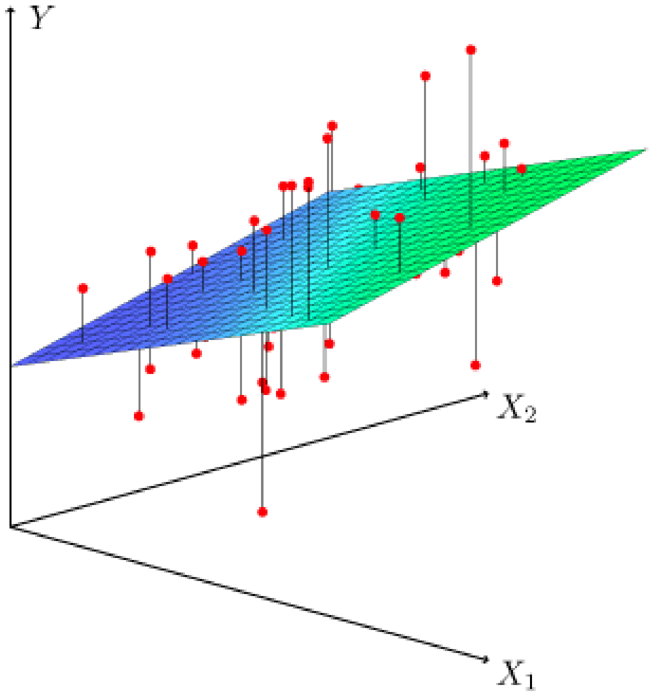
People use Regression model for finding relationship, **Inference**, or prediction

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*



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

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

VIRGINIA TECH

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

$$\log L(\beta) = \sum_{i=1}^{n} \log \lambda(s_i) - \int_W \lambda(s)\, ds$$

$$= \sum_{i=1}^{n} X(s_i)^T \beta - \int_W exp\left(X(s)^T \beta\right) ds$$

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

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 10trees or car crashes in an area, but never -3or 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 ($0,1,2,\ldots$) 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.

VIRGINIA TECH

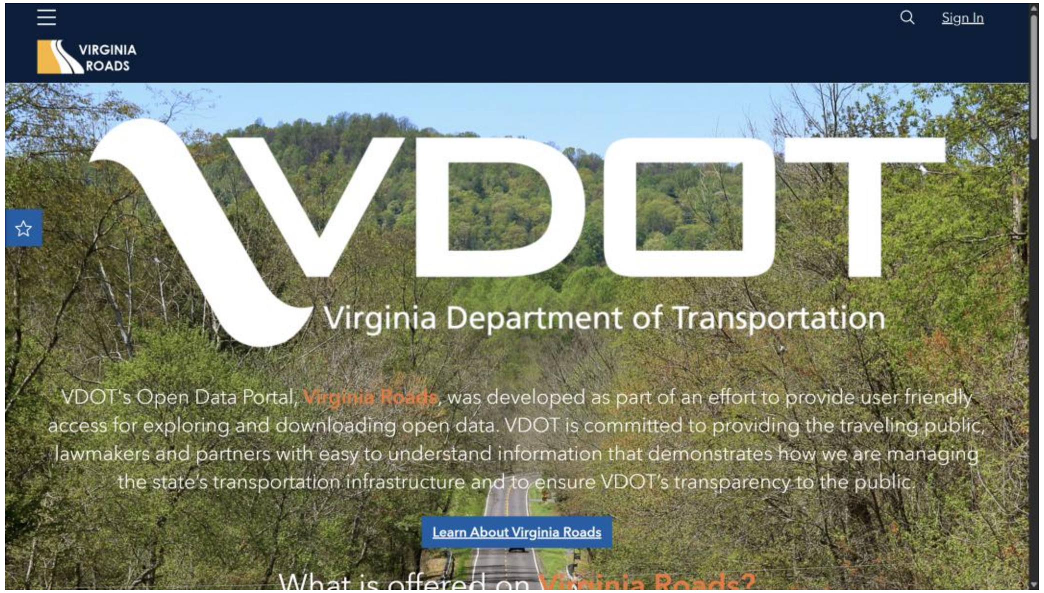
Data & Data processing

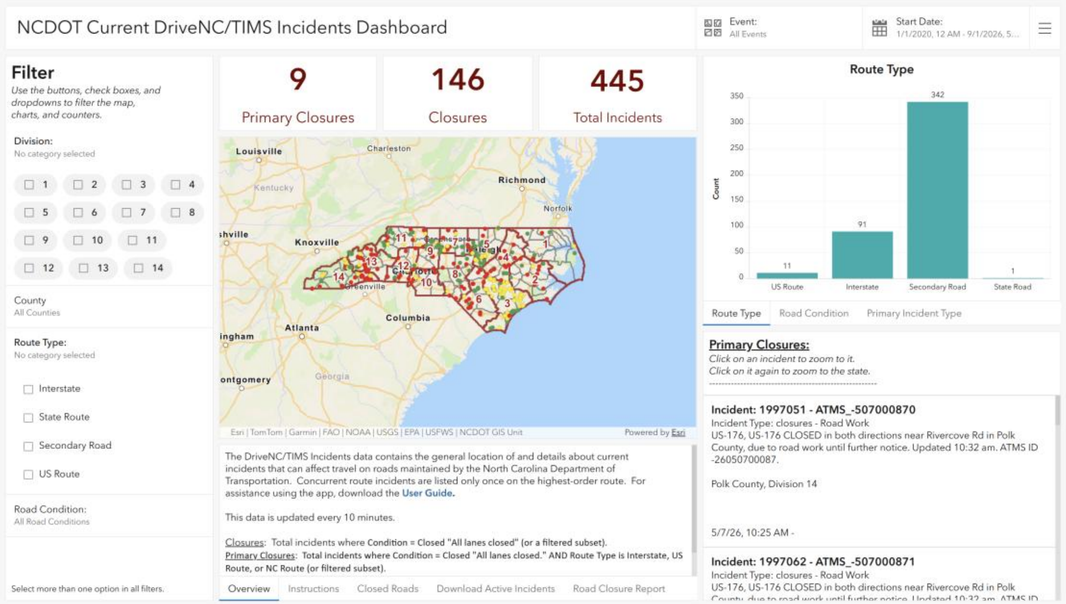
DOT DATA

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

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)



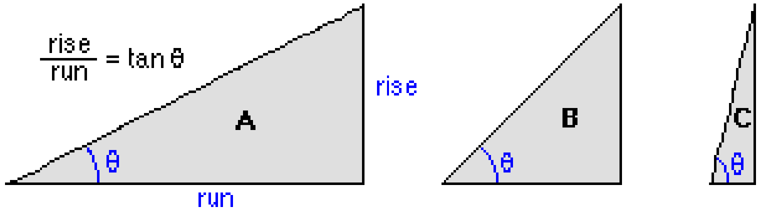


VIRGINIA TECH

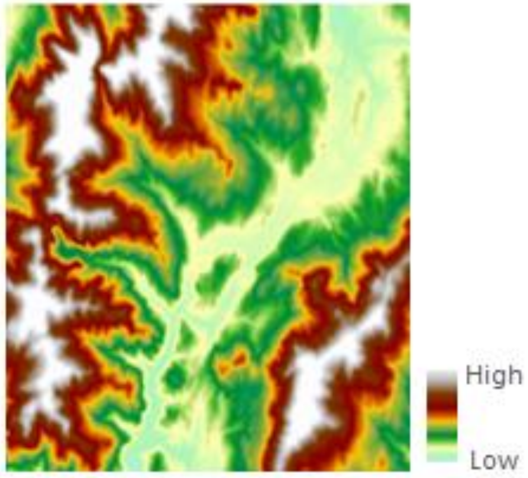
Road Slope

Degree of slope = θ

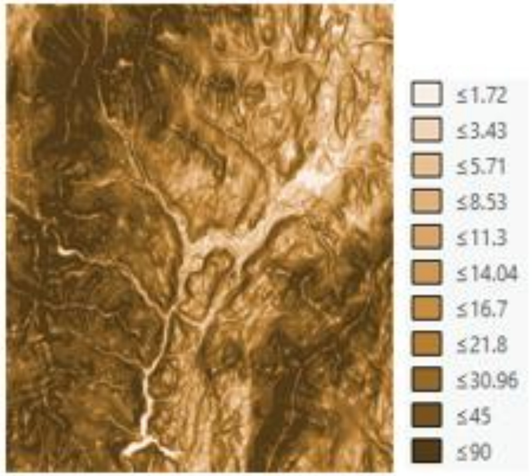
Percent of slope = $\frac{\text{rise}}{\text{run}} * 100$



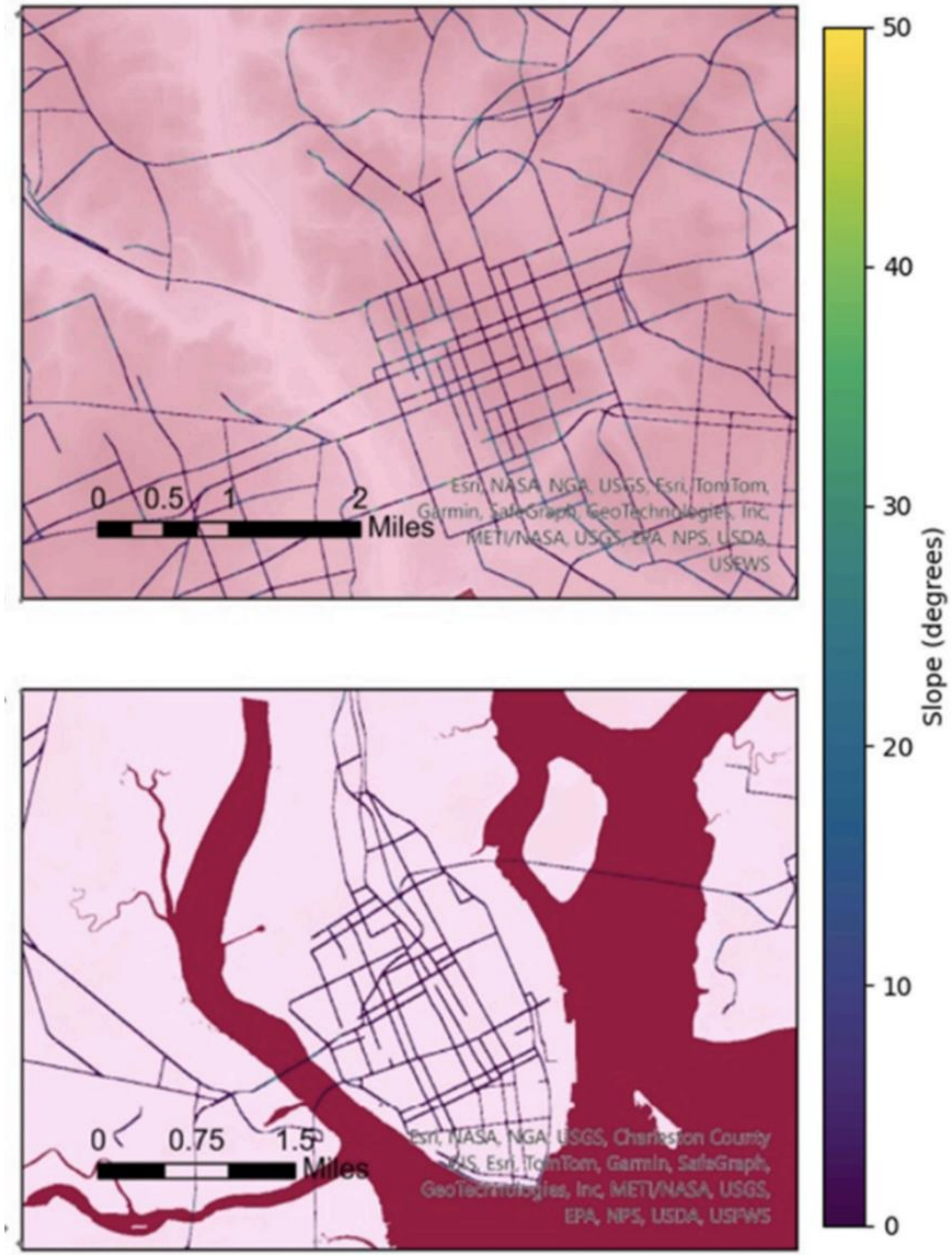
	A	B	C
Degree of slope =	30	45	76
Percent of slope =	58	100	373



Input elevation raster



Output slope raster
(in degrees)



https://pro.arcgis.com/en/pro-app/3.6/tool-reference/spatial-analyst/how-slope-works.htm

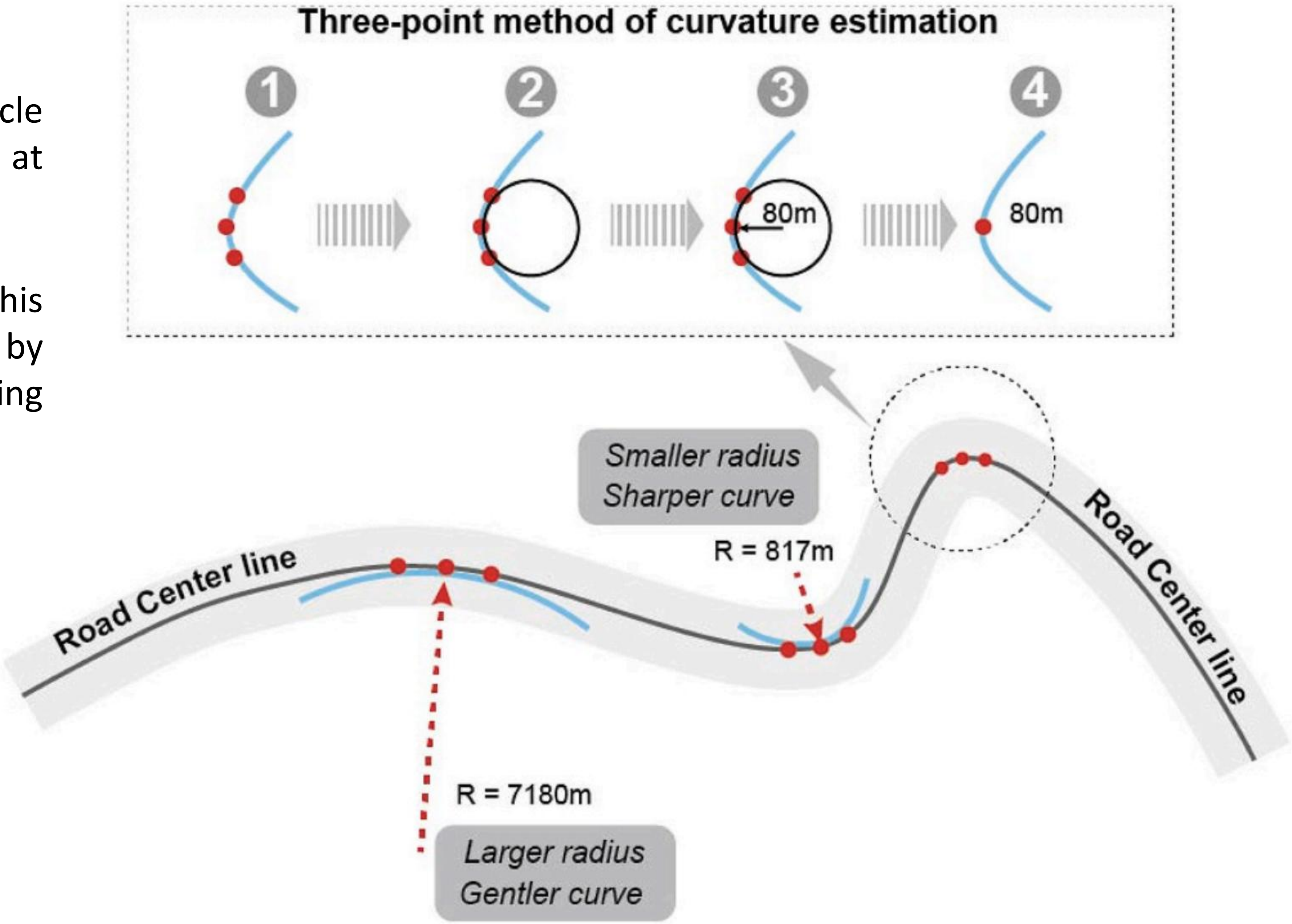
VIRGINIA TECH.

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.

Curvature = 1/Radius



VIRGINIA TECH

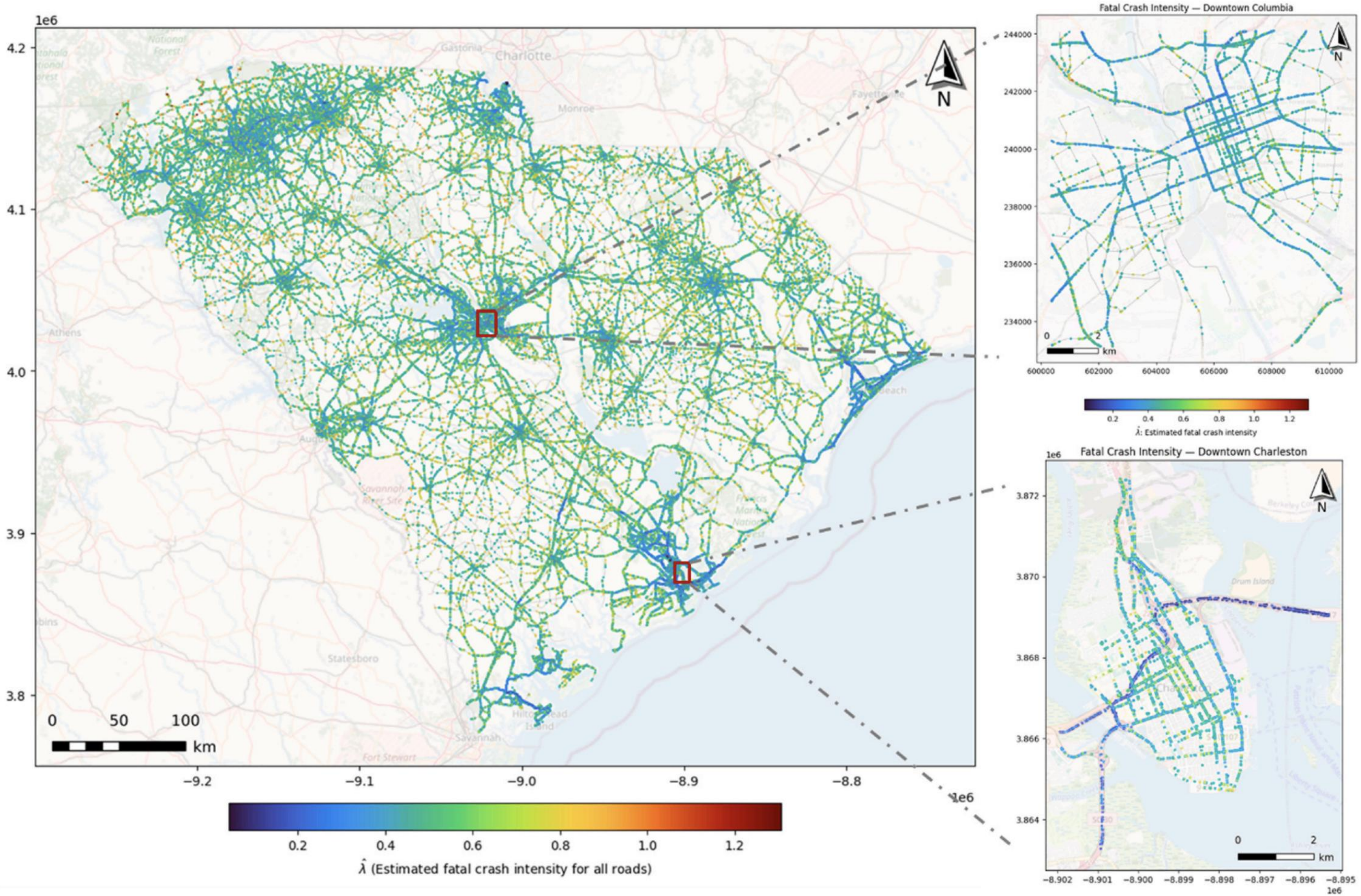
Table: Variable definitions in the analytical dataset

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)

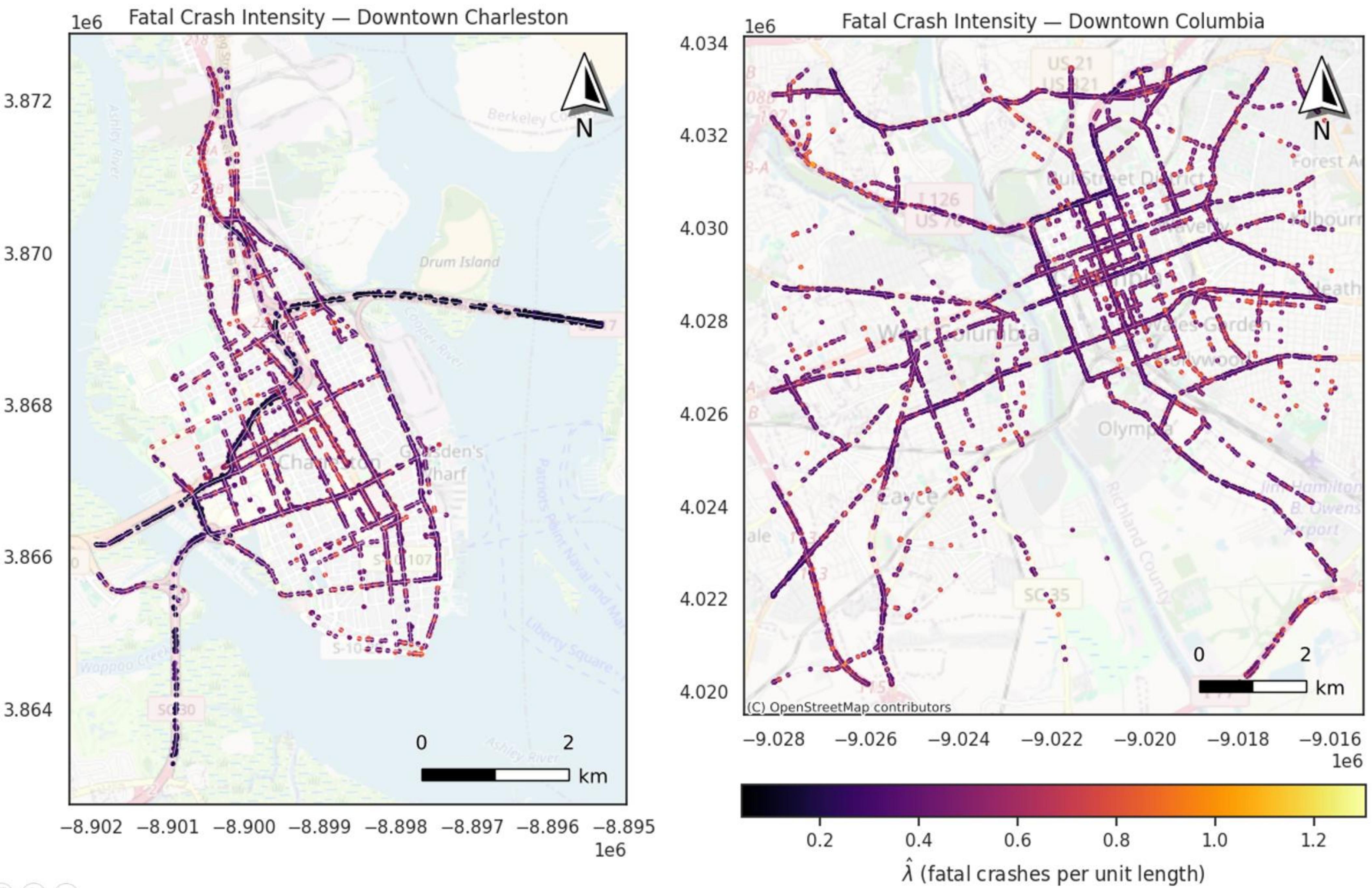
$$Y_i\,(\text{death}) = \beta_0 + \beta_1 X_1(slope) + \beta_2 X_2(night) + \cdots + \beta_n X_n + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$$

Results

Result: Fatal Intensity (λ) map



VIRG TECH


Fatal Crash Intensity — Downtown Charleston
1e6
3.872
3.870
3.868
3.866
3.864
−8.902
−8.901
−8.900
−8.899
−8.898
−8.897
−8.896
−8.895
1e6
N
Drum Island
0
2
km
Fatal Crash Intensity — Downtown Columbia
1e6
4.034
4.032
4.030
4.028
4.026
4.024
4.022
4.020
−9.028
−9.026
−9.024
−9.022
−9.020
−9.018
−9.016
1e6
N
0
2
km
(C) OpenStreetMap contributors
0.2
0.4
0.6
0.8
1.0
1.2
$\hat{\lambda}$ (fatal crashes per unit length)

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				
Road Design				
Log(Slope)	0.063	0.014	4.512	<0.001 ***
Log$\left(\text{Curvature}^{-1}\right)$	−0.025	0.013	−1.941	0.052 •
Dynamic Road Condition				
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 ***
Temporal and Environmental Factors				
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
Vehicle Characteristics				
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 '•'

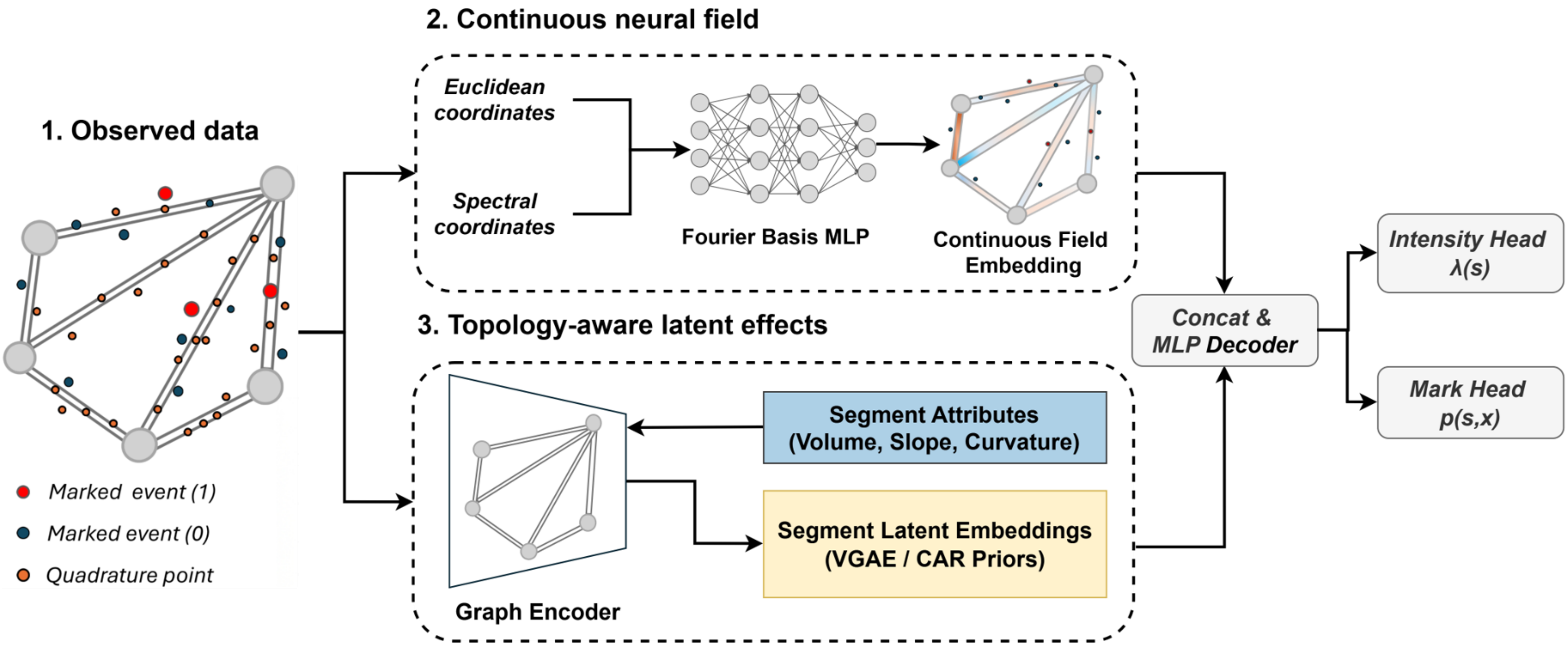
Poisson model estimates by road type with log-transformed variables.

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***

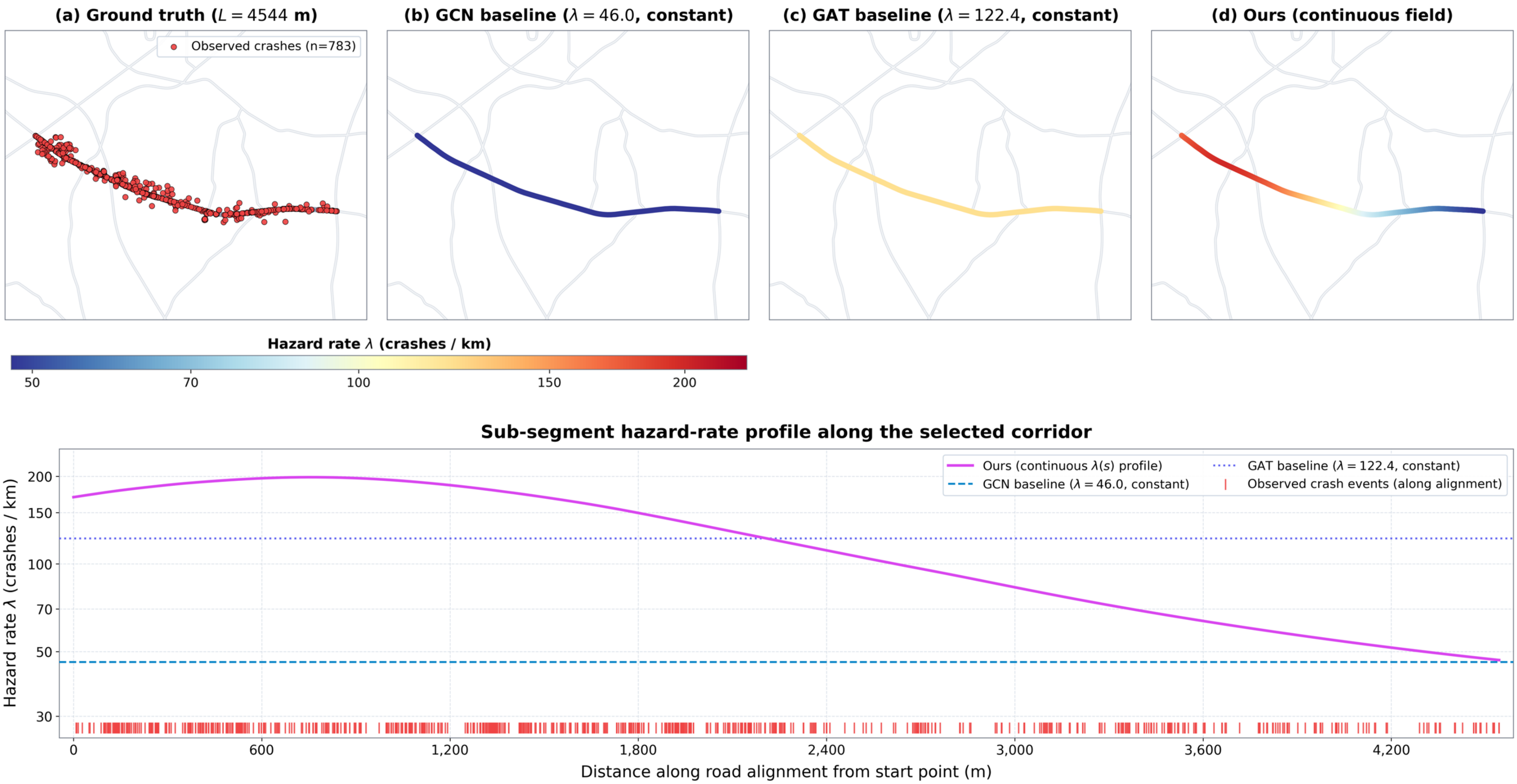
Sample Size (N): Highway = 438,200, Local = 291,810; Pseudo R^2: Highway = 0.227, Local = 0.333; significance codes: <0.001 '***' <0.01 '**' <0.05 '*'.

Paper link: https://www.sciencedirect.com/science/article/pii/S0143622826001633

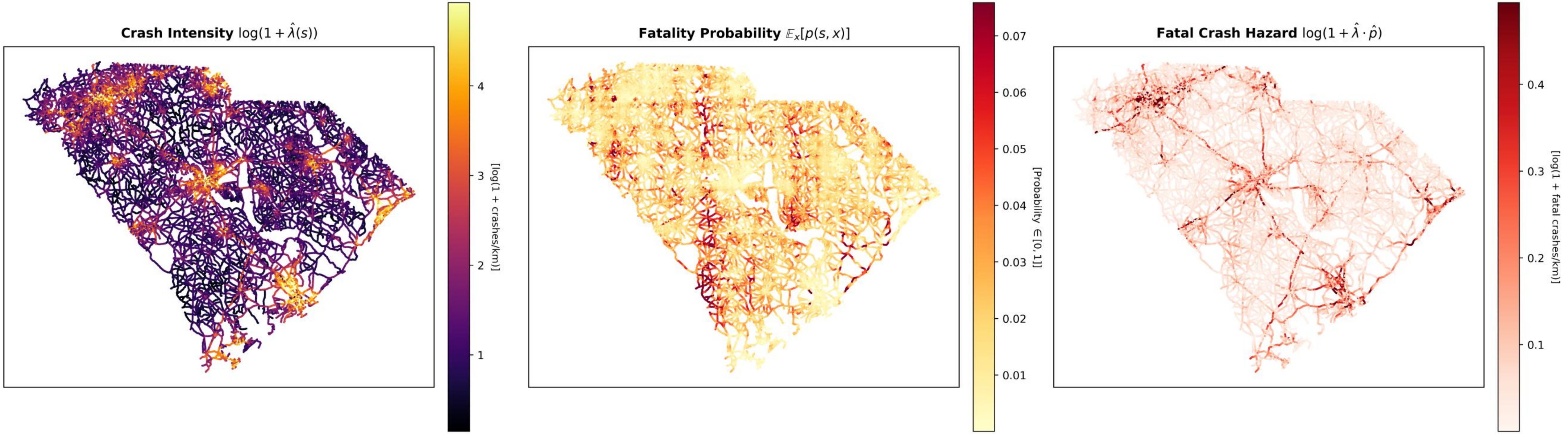
Ongoing work



VIRGINIA TECH



(a) Ground truth (L = 4544 m)
Observed crashes (n=783)
(b) GCN baseline (λ = 46.0, constant)
(c) GAT baseline (λ = 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
Ours (continuous λ(s) profile)
GCN baseline (λ = 46.0, constant)
GAT baseline (λ = 122.4, constant)
Observed crash events (along alignment)
Hazard rate λ (crashes / km)
200
150
100
70
50
30
0
600
1,200
1,800
2,400
3,000
3,600
4,200
Distance along road alignment from start point (m)



Deep learning model provide more accurate & better results.

However, the results would be less explainable.

VIRGINIA TECH

Questions?

VIRGINIA TECH