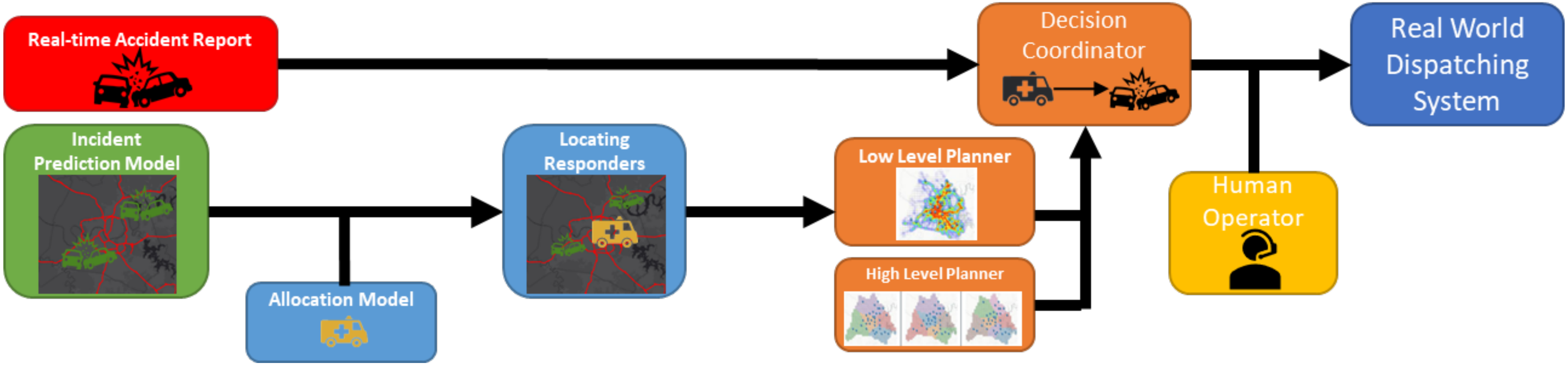
Learning Incident Prediction Models Over Large Geographical Areas for Emergency Response

Sayyed Mohsen Vazirizade, Ayan Mukhopadhyay, Geoffrey Pettet, Said El Said, Hiba Baroud, Abhishek Dubey

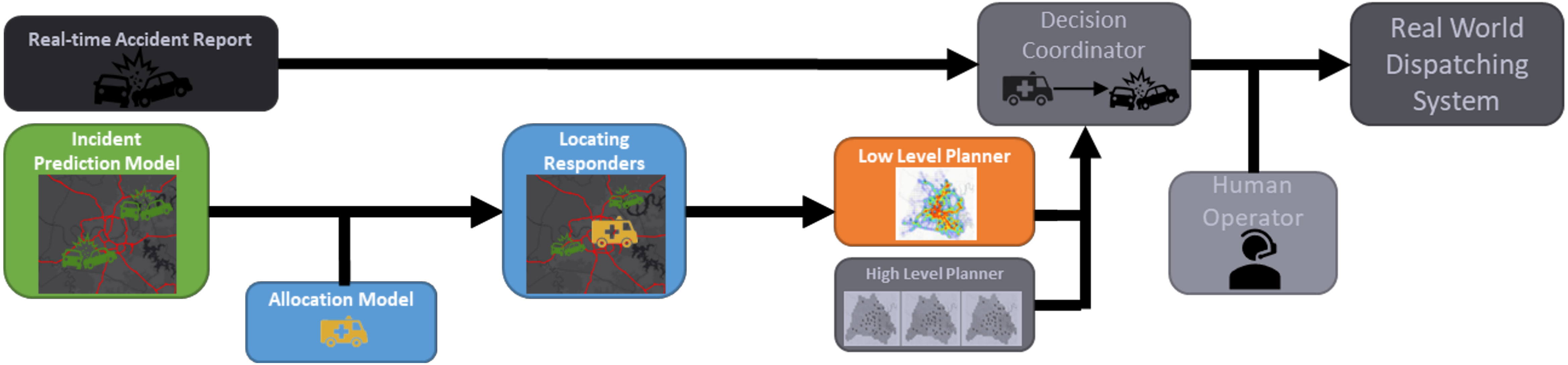
Vanderbilt University

Sponsored by Tennessee Department of Transportation (TDOT) and National Science Foundation

Emergency Response Management (ERM)

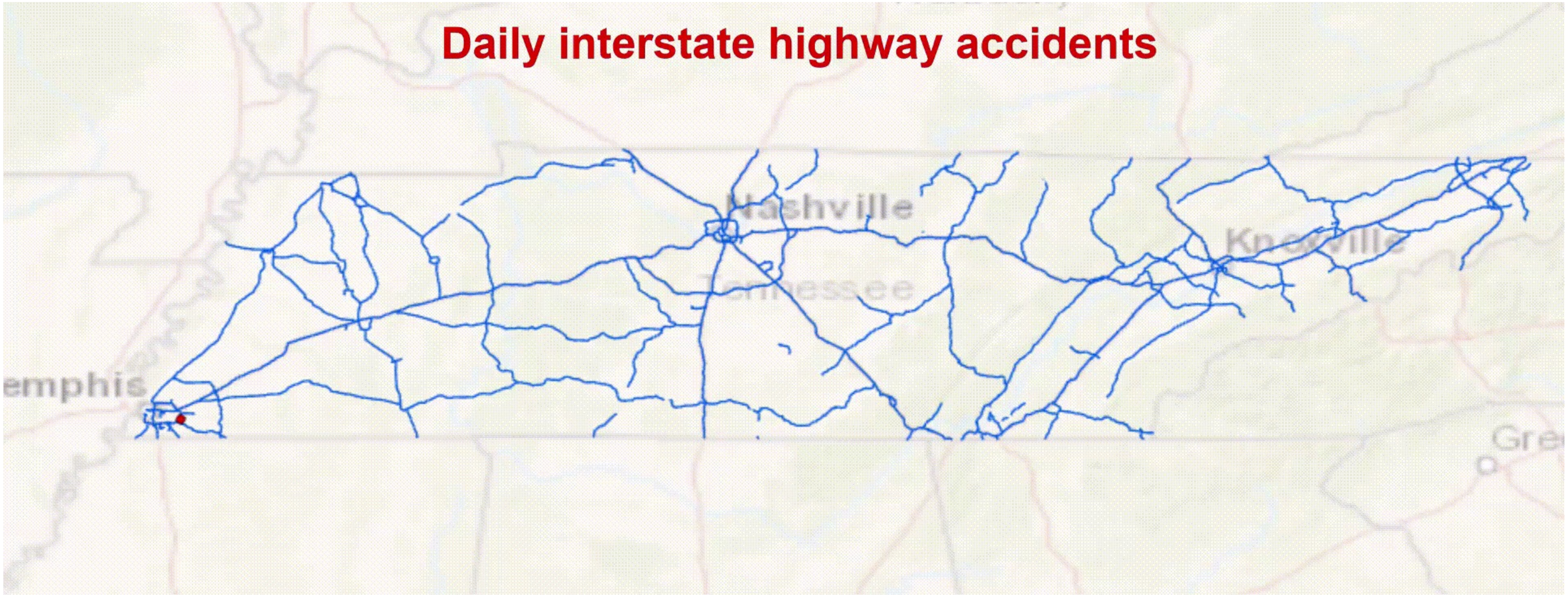


Emergency Response Management (ERM)



Problem Statement

- Respond efficiently to all incidents spread over a large geographical area with limited resources

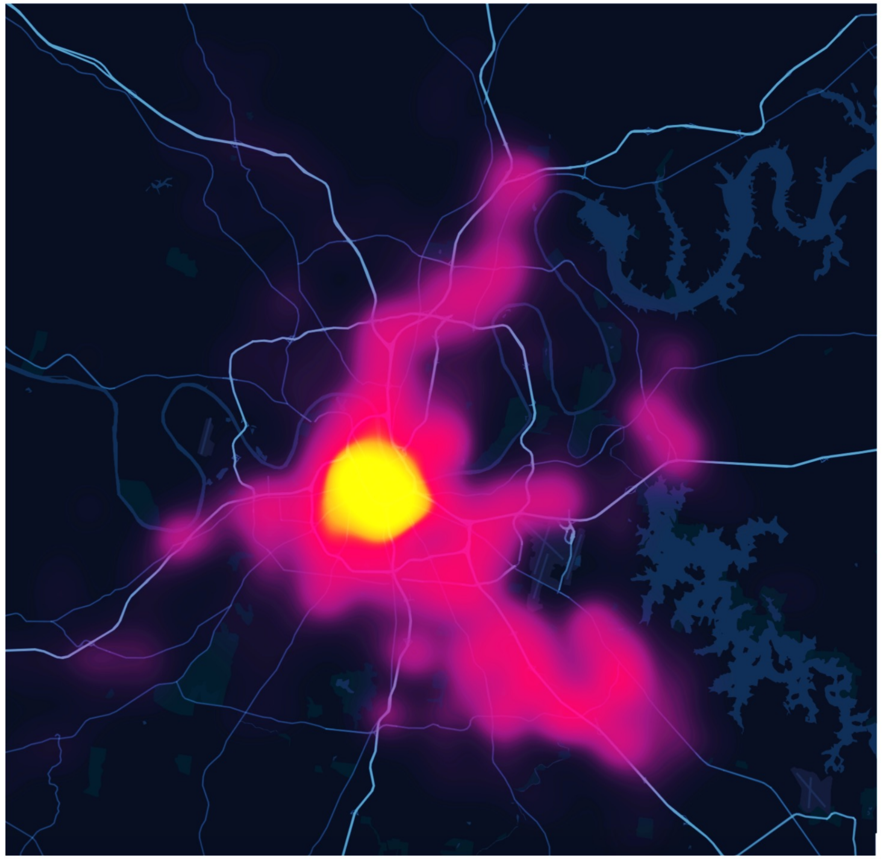


Our Approach: Proactive ERM Resource Allocation

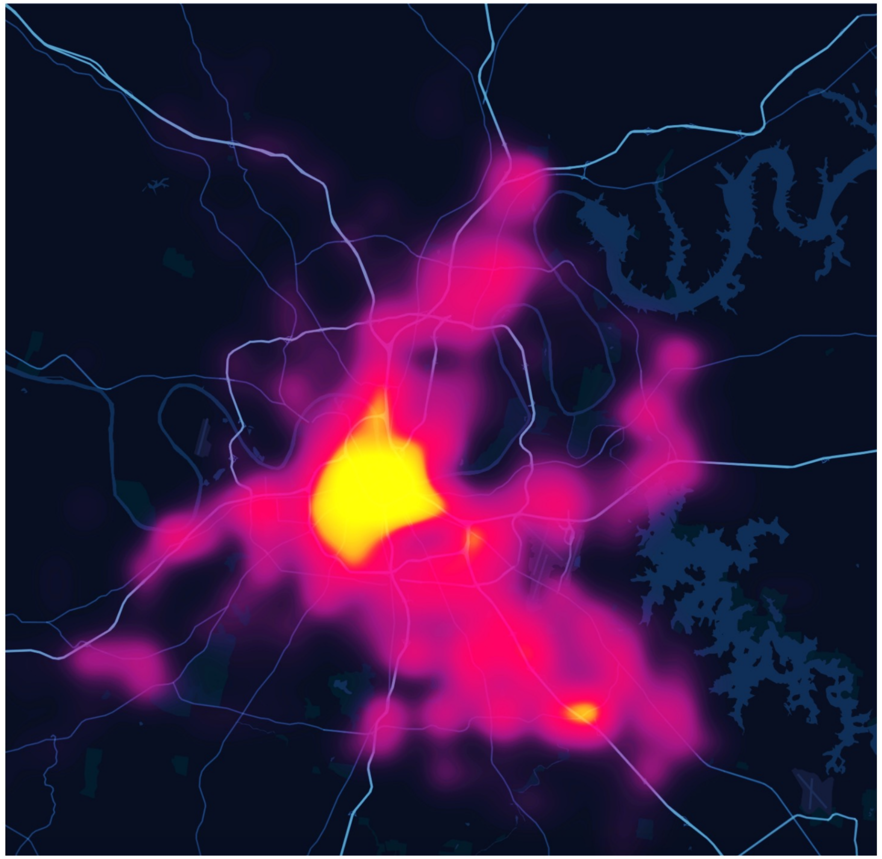
Sense	Observe the environment
Detect	Detect incidents in data streams before they are reported
Predict	Forecast *where* and *when* incidents will occur based on current environment
Allocation	Dynamically allocate (e.g., position) responders across the area based on predictions
Dispatch	When an incident is reported, quickly send a responder to service it

Background: Prediction Example

- We use **survival analysis** - a class of methods to find *inter-arrival* times.
- <u>**Inter-arrival time**</u>: $t_i = x_i - x_{i-1}$
- We use **Maximum Likelihood Estimation** to estimate parameters.
- However, even in this case we had to aggregate incidents within a grid and the prediction works over the grid. [sparsity is around 80 %].
- However, when we focus on key interstates:
    - Due high to spatial and temporal granularity the sparsity levels become extremely high.



Incidents predicted by model for January 2019



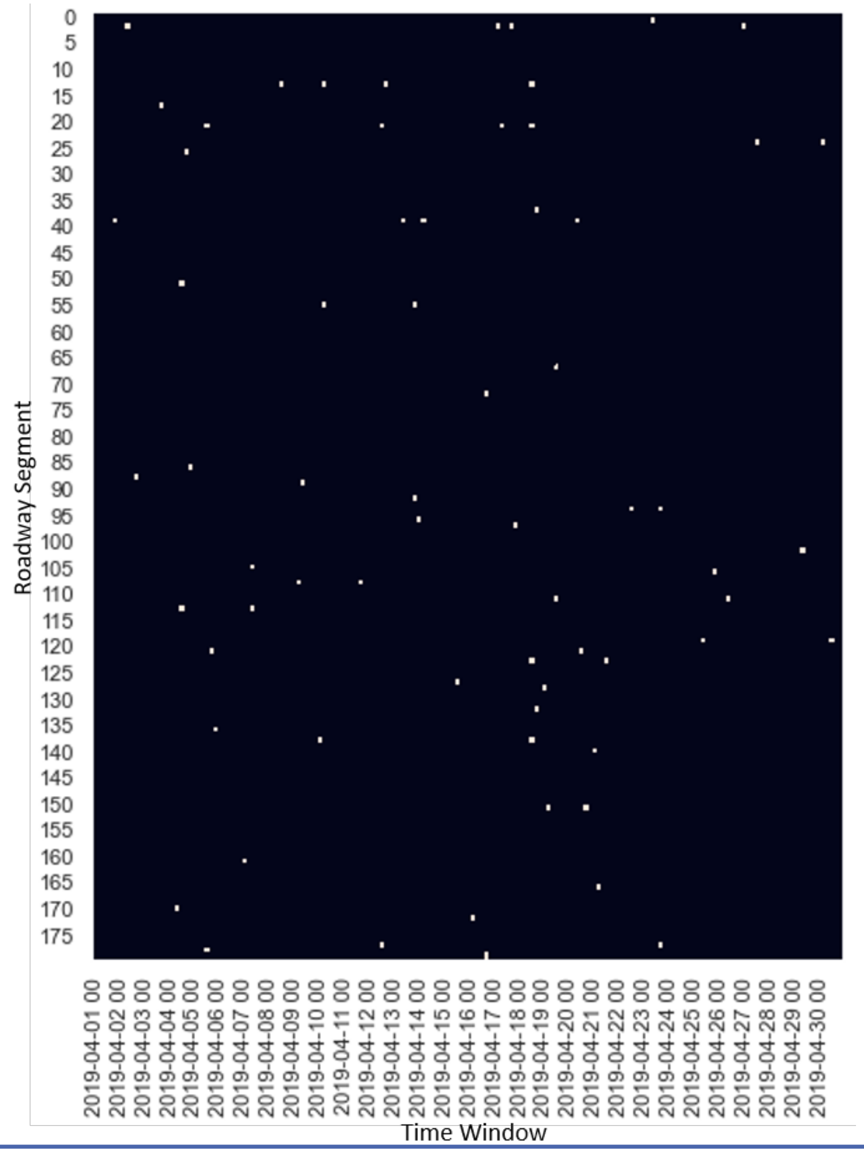
Real incident distribution January 2019.

Formalizing the Problem

- The goal is to design a function, $f\ (Y \mid X, w)$
- Where Y represents a measure of incident, such as a count, occurrence (presence of incidents during a specific time period,) severity, etc.
- X represents the parameters regarding the model such as rainfall and location of the segment.
- w paraments of the model we aim to adjust (usually using MLE.)

Challenges

- Data Collection: Many factors are involved in road accidents, which requires collecting various types of data from assorted resources with different resolution and quality.
- Data Integration: Due to the enormous size of data sets and their incompatibility, combining them in is not trivial.
- Sparsity: Although frequency of road accidents is high, when viewed from the perspective of total time and space, incidents are actually rare events.
- Irregular incident occurrence: Accidents are random in nature, in particular in high spatial-temporal resolution.
- Resolution: While archiving a very high-resolution prediction is virtually impossible, reducing the resolution may cause aggregation bias and unobserved heterogeneity. Therefore, resolution level should be selected strategically and based the application of the model.



Collected Data

Dataset	Range	Size	Rows	Features	Source	Frequency	Type	Description
-	-	-	-	Time of day	derived	-	Temporal	We divide each day into six 4-hour time windows.
-	-	-	-	Weekend	derived	-	temporal	A binary feature that denotes weekdays.
Incident	02/01/2017 to 05/01/2020	21MB	80,000	Past Incidents in the last window	derived	-	Spatio-temporal	Number of incidents on the segment in the last time window of 4 hours
				Past Incidents in a day	derived	-	Spatio-temporal	Number of incidents on the segment in the last day
				Past Incidents in a week	derived	-	Spatio-temporal	Number of incidents on the segment in the last week
				Past Incidents in a month	derived	-	Spatio-temporal	Number of incidents on the segment in the last month
Weather	02/01/2017 to 06/01/2020	300MB	1,400,000	Visibility	Weatherbit	1 hour	Spatio-temporal	A measure of the distance at which an object or light can be clearly discerned.
				Wind Speed	Weatherbit	1 hour	Spatio-temporal	Speed of wind.
				Precipitation	Weatherbit	1 hour	Spatio-temporal	Amount of precipitation.
				Temperature	Weatherbit	1 hour	Spatio-temporal	It is the reported temperature.
Traffic	04/01/2017 to 12/01/2020	1.2TB	30,000,000,000	Congestion	derived	5 minutes	Spatio-temporal	Congestion is the ratio of the difference between free flow speed and the current speed to free flow speed
				Free Flow Speed	INRIX	5 minutes	spatial	The speed at which drivers feel comfortable if there is no traffic and adverse weather condition.
				Traffic Confidence	INRIX	5 minutes	Spatio-temporal	A confidence score regarding the accuracy of the traffic data (we collect this directly from INRIX).
Roadways	Static	81MB	80,000	Lanes	INRIX	static	Spatial	Number of lanes for a roadway segment.
				Miles	derived	static	Spatial	Length of a roadway segment.
				iSF	derived	static	Spatial	Inverse scale factor which represents the the curvature of a roadway segment.

Training

We use rolling windows for training.
Different models are evaluated based on the average of these 12 windows.

#	2017									2018	2019												2020
1	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
2	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
3	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
4	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
5	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
6	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	...	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
7	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
8	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
9	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
10	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
11	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
12	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan

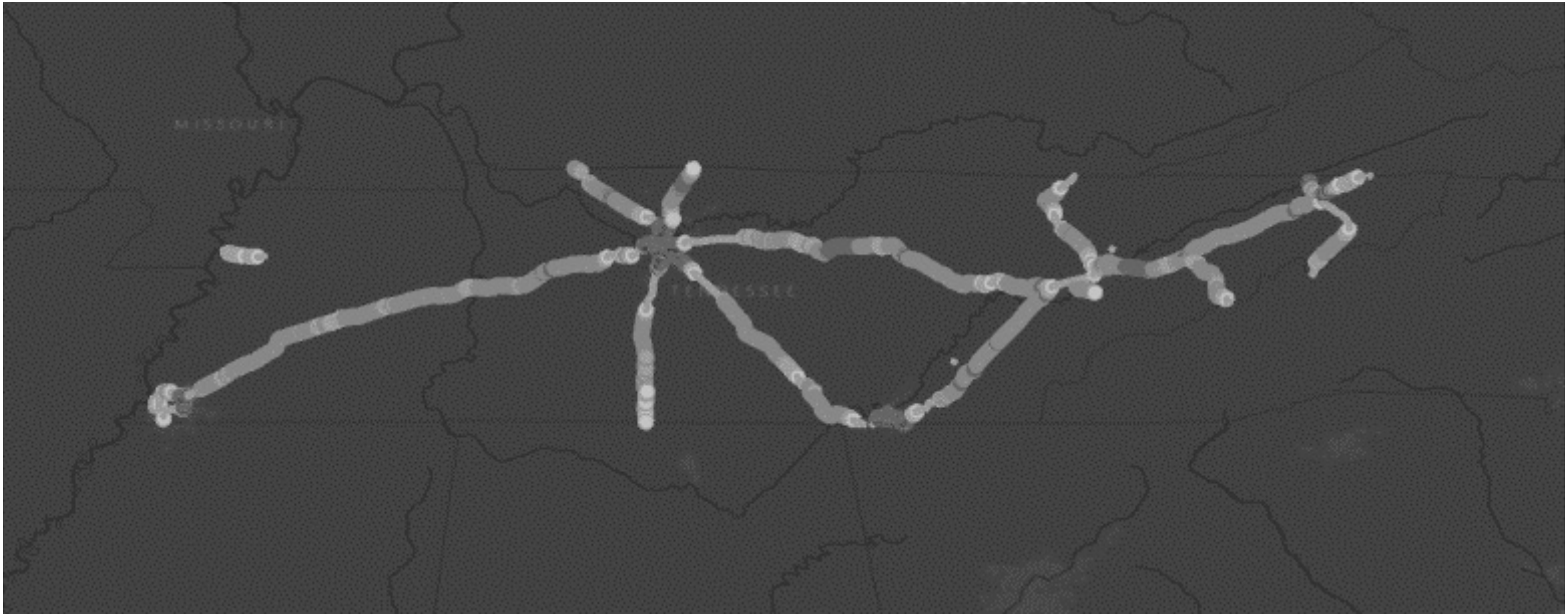
Evaluation

Model	Clustering	Resampling	Name	Accuracy	Precision	Recall	F1-Score	Pearson	Spearman
Naive			Naïve	**95.5**	3.8	4.2	4.0	**82.1**	60.8
LR	No cluster	No resampling	LR+NoR+NoC1	94.0	13.8	27.4	18.2	70.4	55.2
		RUS	LR+RUS+NoC1	93.0	12.8	32.3	18.3	63.1	54.7
		ROS	LR+ROS+NoC1	93.0	12.8	32.3	18.3	63.2	54.7
	clustering	No sample	LR+NoR+KM2	93.0	12.5	30.9	17.7	76.6	58.4
		RUS	LR+RUS+KM2	92.3	12.1	**34.4**	17.8	74.2	58.1
		ROS	LR+ROS+KM2	92.4	12.2	34.2	17.9	74.2	58.1
NN	[illegible]	[illegible]	NN+NoR+NoC1	94.9	19.2	32.8	[illegible]	[illegible]	58.5
		[illegible]	NN+RUS+NoC1	95.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	clustering	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		RUS	NN+RUS+KM2	94.7	18.4	32.7	23.3	73.1	54.6
		ROS	NN+ROS+KM2	94.7	18.3	33.1	23.3	74.5	55.4
Tree	No cluster	No resampling	RF+NoR+NoC1	95.0	19.0	31.8	23.6	78.7	63.4
		RUS	[illegible]	95.2	19.3	30.5	23.5	67.4	56.9
		ROS	[illegible]	[illegible]	[illegible]	[illegible]	22.1	79.2	**64.6**
		Class weights	[illegible]	[illegible]	[illegible]	[illegible]	**24.4**	77.1	62.5
	clustering	No resampling	[illegible]	95.1	[illegible]	[illegible]	23.2	79.8	62.3
		RUS	RF+RUS+KM2	95.0	19.4	32.5	24.2	73.8	57.6
		ROS	RF+ROS+KM2	95.1	18.3	28.7	22.2	80.1	63.6
		Class weights	RF+CW+NoC1	95.4	20.6	30.4	24.4	77.1	62.5
ZIP	No cluster	No resampling	ZIP+NoR+NoC1	94.4	14.6	26.8	18.9	74.0	58.0
		RUS	ZIP+RUS+NoC1	94.2	13.9	26.1	18.1	61.1	50.6
		ROS	ZIP+ROS+NoC1	94.2	13.9	26.7	18.2	61.2	50.6
	clustering	No resampling	ZIP+NoR+KM2	93.1	13.1	31.9	18.5	77.6	61.8
		RUS	ZIP+RUS+KM2	93.0	12.7	30.8	17.8	74.2	57.1
		ROS	ZIP+ROS+KM2	93.0	12.8	30.9	18.0	74.3	57.0

Which model is better?

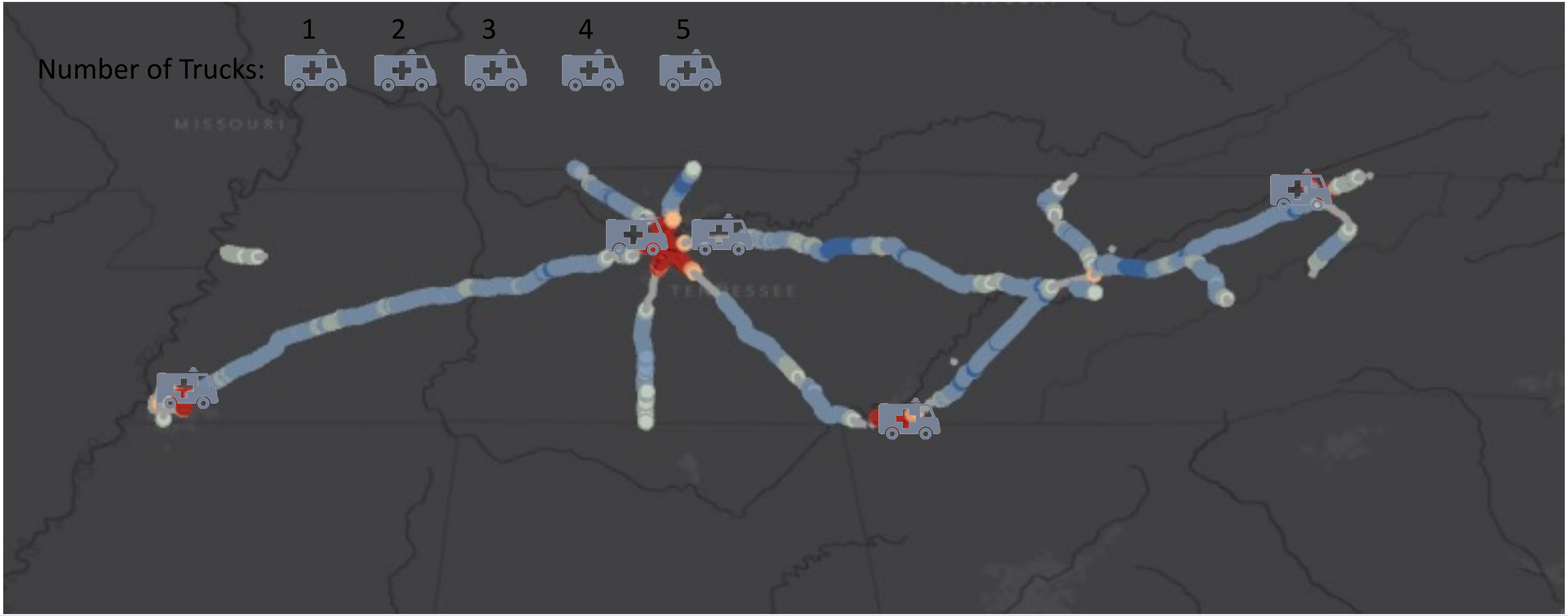
Summary of the Problem

The available resources (responders) are spatially distributed according to historical events or accident prediction rate.

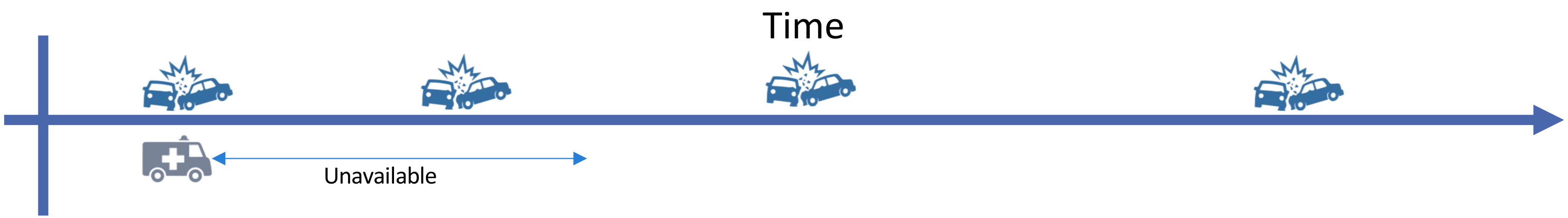


Hotspot Model

The available resources (responders) are spatially distributed according to historical events or accident prediction rate.

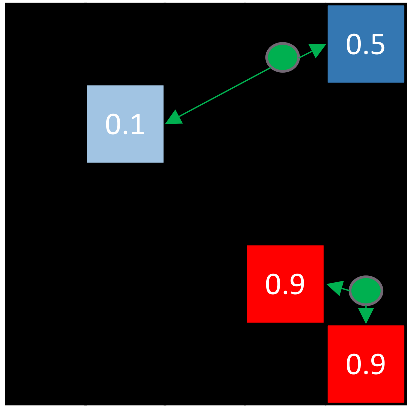



Clarksville
TENNESSEE


Time
Unavailable

Allocation

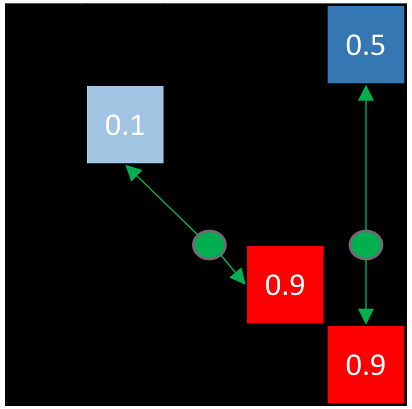
$\alpha = 0$



P-Median Problem

$$Performance = \sum d_{i,j} p_i$$

$\alpha > 0$



Modified P-Median Problem

$$Performance = \sum d_{i,j} p_i \, b_j$$

$$b_j = \left(\frac{\sum p'_i}{\sum p_i}\right)^{\alpha}$$

Algorithm 1: Greedy-Add Algorithm

input : Demand Edges E, Potential Responder Locations L, Segment Incident Likelihoods a_i $\forall e_i \in E$, Segment to Location Distances $d(i,j)$ $\forall e_i \in E, \forall l_j \in L$, Number of Responders p, Balance Factor α

output: Responder Locations X

Initialize $k := 0$, $X_k := \emptyset$;
while $k < p$ **do**
  $k := k + 1$;
  for location $l_{j'} \in L$, where $j' \notin X_{k-1}$ **do**
    $X'_k := X_{k-1} \cup l_{j'}$;
    Find nearest facilities y_i $\forall e_i \in E$, where $y_{e_i} \in X'_k$;
    Compute balance terms $b_j := \left(\frac{\sum_{e_i \in E} a_i \psi}{\sum_{e_i \in E} a_i}\right)^{\alpha}$ $\forall l_j \in L$ where $\psi := 1$ if $y_i = l_j$, $\psi := 0$ otherwise;
    Compute $Z^k_{j'} := \sum_{e_i \in E} a_e d(e_i, y_i) b_{y_{e_i}}$;
  end
  Best location $l^*_j :=$ argmin$_j$ Z^k_j;
  $X_k := X_{k-1} \cup j^*$;
  Return X_k
end

Evaluation

			Classification Metrics				Correl-ation	
Model	**Clustering**	**Resampling**	**Acc.**	**Prec.**	**Rec.**	**F1**	**Pear.**	**Spear.**
Naive			**95.5**	3.8	4.2	4.0	**82.1**	60.8
LR	No cluster	No resampling	94.0	13.8	27.4	18.2	70.4	55.2
		RUS	93.0	12.8	32.3	18.3	63.1	54.7
		ROS	93.0	12.8	32.3	18.3	63.2	54.7
	clustering	No sample	93.0	12.5	30.9	17.7	76.6	58.4
		RUS	92.3	12.1	**34.4**	17.8	74.2	58.1
		ROS	92.4	12.2	34.2	17.9	74.2	58.1
NN	No cluster	No resampling	94.9	19.2	32.8	24.0	71.7	58.5
		RUS	95.0	19.2	32.6	24.1	73.2	59.3
		ROS	94.9	19.1	32.8	23.9	69.3	54.7
	clustering	No sample	95.0	19.0	31.6	23.7	75.6	58.9
		RUS	94.7	18.4	32.7	23.3	73.1	54.6
		ROS	94.7	18.3	33.1	23.3	74.5	55.4
Tree	No cluster	No resampling	95.0	19.0	31.8	23.6	78.7	63.4
		RUS	95.2	19.3	30.5	23.5	67.4	56.9
		ROS	95.3	18.6	27.6	22.1	79.2	**64.6**
		Class weights	95.4	**20.6**	30.4	**24.4**	77.1	62.5
	clustering	No resampling	95.1	18.9	30.5	23.2	79.8	62.3
		RUS	95.0	19.4	32.5	24.2	73.8	57.6
		ROS	95.1	18.3	28.7	22.2	80.1	63.6
		Class weights	95.4	20.6	30.4	24.4	77.1	62.5
ZIP	No cluster	No resampling	94.4	14.6	26.8	18.9	74.0	58.0
		RUS	94.2	13.9	26.1	18.1	61.1	50.6
		ROS	94.2	13.9	26.7	18.2	61.2	50.6
	clustering	No resampling	93.1	13.1	31.9	18.5	77.6	61.8
		RUS	93.0	12.7	30.8	S	74.2	57.1
		ROS	93.0	12.8	30.9	18.0	74.3	57.0

Distance
p =10,15,20
α = 0, 0.5, 1, 2

Evaluation

Model	Clustering	Resampling	Classification Metrics				Correl-ation		Total travel distance of responders per accident (km)											
									p =10				p=15				p=20			
Model	Clustering	Resampling	Acc.	Prec.	Rec.	F1	Pear.	Spear.	α=0	α=0.5	α=1	α=2	α=0	α=0.5	α=1	α=2	α=0	α=0.5	α=1	α=2
Naive			**95.5**	3.8	4.2	4.0	**82.1**	60.8	39.48	38.44	43.21	45.35	26.29	25.78	27.34	26.78	19.29	19.43	20.36	23.12
LR	No cluster	No resampling	94.0	13.8	27.4	18.2	70.4	55.2	41.54	41.88	40.04	44.90	25.30	25.16	26.93	26.73	18.98	16.78	17.41	20.23
		RUS	93.0	12.8	32.3	18.3	63.1	54.7	42.90	43.41	39.97	44.36	25.07	25.38	26.01	26.94	19.07	18.05	17.00	20.41
		ROS	93.0	12.8	32.3	18.3	63.2	54.7	42.83	43.90	39.80	44.74	25.14	25.33	25.88	27.22	19.02	18.25	16.61	20.06
	clustering	No sample	93.0	12.5	30.9	17.7	76.6	58.4	40.79	39.44	42.57	44.81	24.44	25.14	26.21	27.79	18.55	19.39	18.95	21.45
		RUS	92.3	12.1	**34.4**	17.8	74.2	58.1	42.69	40.96	42.16	43.75	24.66	24.75	26.20	27.78	18.93	18.69	17.18	20.08
		ROS	92.4	12.2	34.2	17.9	74.2	58.1	42.78	40.89	42.71	44.22	24.58	24.84	26.18	28.29	18.87	18.66	17.04	19.90
NN	No cluster	No resampling	94.9	19.2	32.8	24.0	71.7	58.5	37.04	39.12	39.21	43.13	22.35	23.57	24.74	26.69	15.70	16.44	17.52	20.33
		RUS	95.0	19.2	32.6	24.1	73.2	59.3	37.44	39.07	37.83	43.84	22.24	23.85	24.97	27.64	16.40	16.21	17.05	20.27
		ROS	94.9	19.1	32.8	23.9	69.3	54.7	37.32	37.71	39.86	43.21	21.57	23.15	24.32	26.61	15.70	15.81	17.23	20.33
	clustering	No sample	95.0	19.0	31.6	23.7	75.6	58.9	39.32	39.88	39.61	43.09	23.18	23.96	24.58	27.34	17.46	17.15	17.00	20.16
		RUS	94.7	18.4	32.7	23.3	73.1	54.6	39.79	39.61	39.99	45.08	22.92	24.72	25.32	27.75	16.20	17.10	17.71	21.23
		ROS	94.7	18.3	33.1	23.3	74.5	55.4	38.60	38.24	40.66	45.50	22.23	23.78	25.04	27.40	16.31	16.89	18.00	20.81
Tree	No cluster	No resampling	95.0	19.0	31.8	23.6	78.7	63.4	40.81	38.28	39.62	44.46	23.21	22.99	24.30	26.22	16.88	16.36	16.49	19.97
		RUS	95.2	19.3	30.5	23.5	67.4	56.9	39.55	38.71	40.13	42.39	23.44	23.32	24.41	27.06	16.47	17.19	17.17	20.04
		ROS	95.3	18.6	27.6	22.1	79.2	**64.6**	41.14	39.86	40.37	45.29	23.72	23.78	25.12	26.82	17.89	16.53	16.68	20.14
		Class weights	95.4	**20.6**	30.4	**24.4**	77.1	62.5	39.79	39.46	39.91	44.58	23.14	23.14	24.09	26.56	16.24	16.51	17.68	20.04
	clustering	No resampling	95.1	18.9	30.5	23.2	79.8	62.3	41.40	38.81	39.88	43.16	22.98	23.02	24.56	26.75	16.88	16.25	16.89	19.90
		RUS	95.0	19.4	32.5	24.2	73.8	57.6	39.47	39.53	40.20	44.62	23.12	23.79	23.89	27.49	16.44	17.13	18.00	20.39
		ROS	95.1	18.3	28.7	22.2	80.1	63.6	40.94	39.70	40.82	44.21	23.36	23.48	24.45	26.60	16.92	16.08	16.80	20.38
		Class weights	95.4	20.6	30.4	24.4	77.1	62.5	39.53	38.50	40.95	45.12	23.54	23.54	23.94	27.58	16.29	17.34	18.12	20.37
ZIP	No cluster	No resampling	94.4	14.6	26.8	18.9	74.0	58.0	40.37	40.14	40.15	44.42	25.35	25.07	25.99	26.66	18.53	16.45	17.08	20.81
		RUS	94.2	13.9	26.1	18.1	61.1	50.6	44.57	45.68	40.89	44.23	25.72	25.43	26.86	27.54	18.93	19.26	16.93	19.91
		ROS	94.2	13.9	26.7	18.2	61.2	50.6	44.51	45.42	40.70	44.62	25.77	25.48	27.00	27.80	18.88	19.08	16.98	19.76
	clustering	No resampling	93.1	13.1	31.9	18.5	77.6	61.8	39.35	41.08	40.12	44.97	24.17	24.66	26.42	27.26	18.06	18.40	18.91	20.92
		RUS	93.0	12.7	30.8	S	74.2	57.1	43.46	41.76	42.50	45.17	24.67	26.08	26.56	28.36	19.36	19.85	17.47	21.08
		ROS	93.0	12.8	30.9	18.0	74.3	57.0	43.57	41.31	42.71	45.08	24.77	26.13	26.41	28.52	19.32	20.02	17.29	20.98

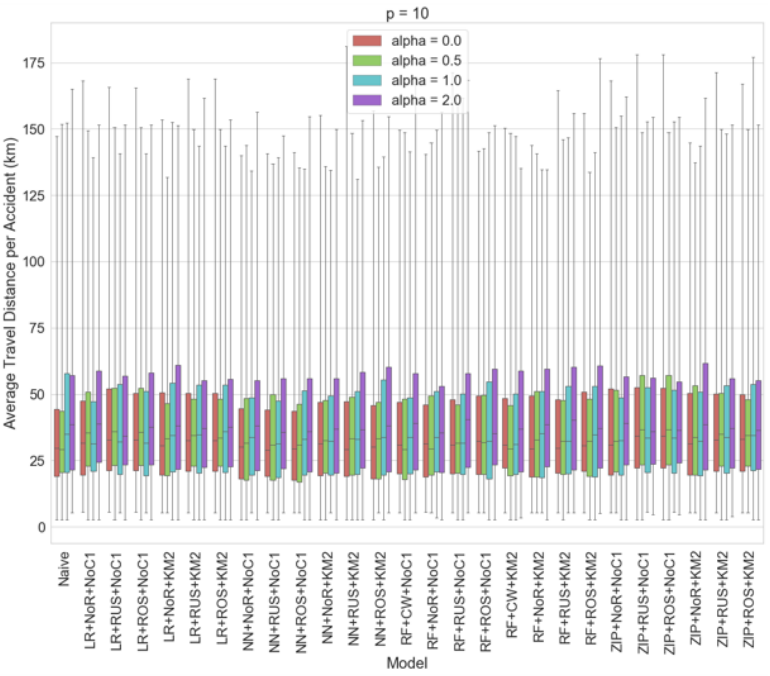
Evaluation

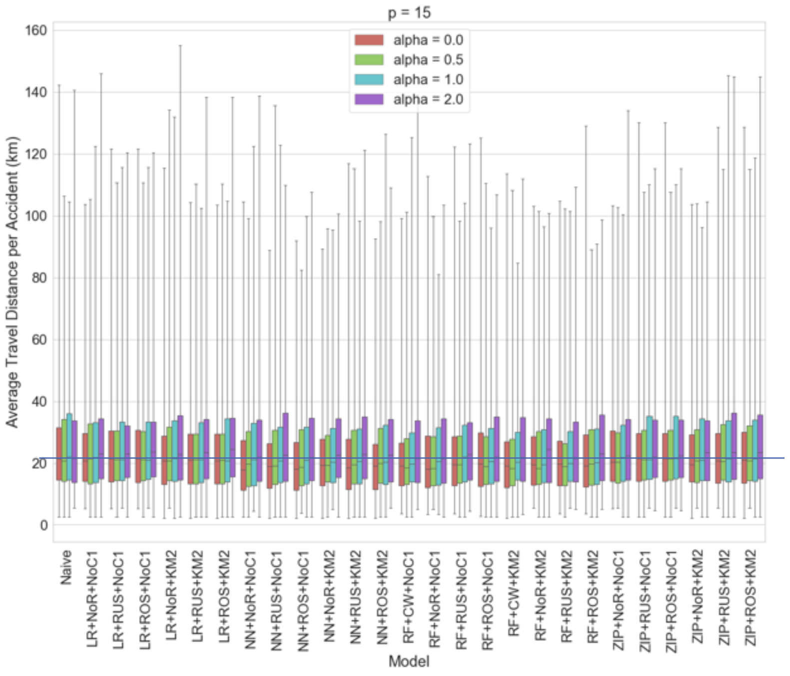
			Classification Metrics				Correl-ation		Total travel distance of responders per accident (km)												Average number of unattended accidents							
									p=10				p=15				p=20				10				15			
Model	**Clustering**	**Resampling**	**Acc.**	**Prec.**	**Rec.**	**F1**	**Pear.**	**Spear.**	**α=0**	**α=0.5**	**α=1**	**α=2**	**α=0**	**α=0.5**	**α=1**	**α=2**	**α=0**	**α=0.5**	**α=1**	**α=2**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**
Naive			95.5	3.8	4.2	4.0	82.1	60.8	39.48	38.44	43.21	45.35	26.29	25.78	27.34	26.78	19.29	19.43	20.36	23.12	0.54	0.49	0.48	0.46	0.02	0.01	0.01	0.01
LR	No cluster	No resampling	94.0	13.8	27.4	18.2	70.4	55.2	41.54	41.88	40.04	44.90	25.30	25.16	26.93	26.73	18.98	16.78	17.41	20.23	0.54	0.47	0.42	0.42	0.00	0.00	0.01	0.01
		RUS	93.0	12.8	32.3	18.3	63.1	54.7	42.90	43.41	39.97	44.36	25.07	25.38	26.01	26.94	19.07	18.05	17.00	20.41	0.56	0.52	0.46	0.46	0.00	0.00	0.01	0.00
		ROS	93.0	12.8	32.3	18.3	63.2	54.7	42.83	43.90	39.80	44.74	25.14	25.33	25.88	27.22	19.02	18.25	16.61	20.06	0.56	0.51	0.46	0.45	0.00	0.00	0.01	0.00
	clustering	No sample	93.0	12.5	30.9	17.7	76.6	58.4	40.79	39.44	42.57	44.81	24.44	25.14	26.21	27.79	18.55	19.39	18.95	21.45	0.53	0.41	0.43	0.44	0.02	0.01	0.01	0.01
		RUS	92.3	12.1	**34.4**	17.8	74.2	58.1	42.69	40.96	42.16	43.75	24.66	24.75	26.20	27.78	18.93	18.69	17.18	20.08	0.54	0.48	0.42	0.40	0.01	0.00	0.00	0.01
		ROS	92.4	12.2	34.2	17.9	74.2	58.1	42.78	40.89	42.71	44.22	24.58	24.84	26.18	28.29	18.87	18.66	17.04	19.90	0.54	0.48	0.42	0.41	0.01	0.00	0.00	0.01
NN	No cluster	No resampling	94.9	19.2	32.8	24.0	71.7	58.5	37.04	39.12	39.21	43.13	22.35	23.57	24.74	26.69	15.70	16.44	17.52	20.33	0.45	0.40	0.43	0.40	0.01	0.00	0.01	0.01
		RUS	95.0	19.2	32.6	24.1	73.2	59.3	37.44	39.07	37.83	43.84	22.24	23.85	24.97	27.64	16.40	16.21	17.05	20.27	0.47	0.41	0.43	0.45	0.00	0.01	0.01	0.01
		ROS	94.9	19.1	32.8	23.9	69.3	54.7	37.32	37.71	39.86	43.21	21.57	23.15	24.32	26.61	15.70	15.81	17.23	20.33	0.46	0.41	0.42	0.43	0.00	0.00	0.00	0.01
	clustering	No sample	95.0	19.0	31.6	23.7	75.6	58.9	39.32	39.88	39.61	43.09	23.18	23.96	24.58	27.34	17.46	17.15	17.00	20.16	0.44	0.40	0.42	0.42	0.00	0.00	0.00	0.00
		RUS	94.7	18.4	32.7	23.3	73.1	54.6	39.79	39.61	39.99	45.08	22.92	24.72	25.32	27.75	16.20	17.10	17.71	21.23	0.48	0.45	0.42	0.42	0.01	0.01	0.01	0.02
		ROS	94.7	18.3	33.1	23.3	74.5	55.4	38.60	38.24	40.66	45.50	22.23	23.78	25.04	27.40	16.31	16.89	18.00	20.81	0.48	0.41	0.44	0.41	0.00	0.00	0.00	0.01
Tree	No cluster	No resampling	95.0	19.0	31.8	23.6	78.7	63.4	40.81	38.28	39.62	44.46	23.21	22.99	24.30	26.22	16.88	16.36	16.49	19.97	0.51	0.44	0.42	0.42	0.00	0.00	0.00	0.02
		RUS	95.2	19.3	30.5	23.5	67.4	56.9	39.55	38.71	40.13	42.39	23.44	23.32	24.41	27.06	16.47	17.19	17.17	20.04	0.48	0.40	0.38	0.43	0.01	0.01	0.00	0.02
		ROS	95.3	18.6	27.6	22.1	79.2	**64.6**	41.14	39.86	40.37	45.29	23.72	23.78	25.12	26.82	17.89	16.53	16.68	20.14	0.53	0.46	0.44	0.42	0.01	0.01	0.00	0.00
		Class weights	95.4	**20.6**	30.4	**24.4**	77.1	62.5	39.79	39.46	39.91	44.58	23.14	23.14	24.09	26.56	16.24	16.51	17.68	20.04	0.46	0.41	0.40	0.41	0.01	0.01	0.00	0.01
	clustering	No resampling	95.1	18.9	30.5	23.2	79.8	62.3	41.40	38.81	39.88	43.16	22.98	23.02	24.56	26.75	16.88	16.25	16.89	19.90	0.49	0.42	0.42	0.43	0.01	0.00	0.00	0.01
		RUS	95.0	19.4	32.5	24.2	73.8	57.6	39.47	39.53	40.20	44.62	23.12	23.79	23.89	27.49	16.44	17.13	18.00	20.39	0.49	0.40	0.38	0.42	0.00	0.00	0.00	0.00
		ROS	95.1	18.3	28.7	22.2	80.1	63.6	40.94	39.70	40.82	44.21	23.36	23.48	24.45	26.60	16.92	16.08	16.80	20.38	0.51	0.45	0.43	0.40	0.01	0.00	0.01	0.01
		Class weights	95.4	20.6	30.4	24.4	77.1	62.5	39.53	38.50	40.95	45.12	23.54	23.54	23.94	27.58	16.29	17.34	18.12	20.37	0.48	0.38	0.38	0.41	0.01	0.00	0.00	0.01
ZIP	No cluster	No resampling	94.4	14.6	26.8	18.9	74.0	58.0	40.37	40.14	40.15	44.42	25.35	25.07	25.99	26.66	18.53	16.45	17.08	20.81	0.51	0.45	0.41	0.40	0.02	0.00	0.00	0.01
		RUS	94.2	13.9	26.1	18.1	61.1	50.6	44.57	45.68	40.89	44.23	25.72	25.43	26.86	27.54	18.93	19.26	16.93	19.91	0.59	0.53	0.51	0.48	0.01	0.02	0.01	0.00
		ROS	94.2	13.9	26.7	18.2	61.2	50.6	44.51	45.42	40.70	44.62	25.77	25.48	27.00	27.80	18.88	19.08	16.98	19.76	0.59	0.53	0.51	0.48	0.01	0.02	0.01	0.00
	clustering	No resampling	93.1	13.1	31.9	18.5	77.6	61.8	39.35	41.08	40.12	44.97	24.17	24.66	26.42	27.26	18.06	18.40	18.91	20.92	0.47	0.45	0.42	0.37	0.02	0.00	0.01	0.01
		RUS	93.0	12.7	30.8	S	74.2	57.1	43.46	41.76	42.50	45.17	24.67	26.08	26.56	28.36	19.36	19.85	17.47	21.08	0.55	0.48	0.49	0.46	0.00	0.01	0.00	0.01
		ROS	93.0	12.8	30.9	18.0	74.3	57.0	43.57	41.31	42.71	45.08	24.77	26.13	26.41	28.52	19.32	20.02	17.29	20.98	0.57	0.49	0.48	0.45	0.00	0.01	0.00	0.01

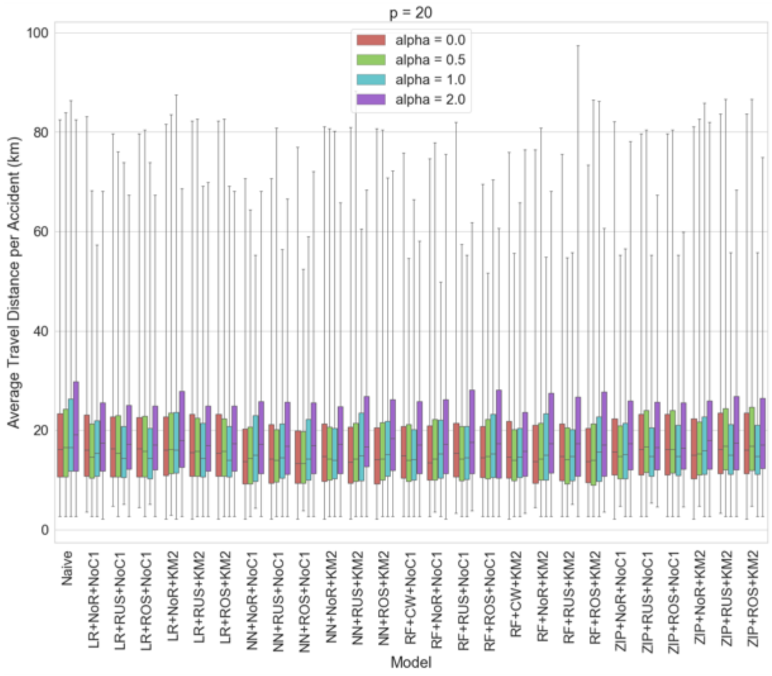
Evaluation

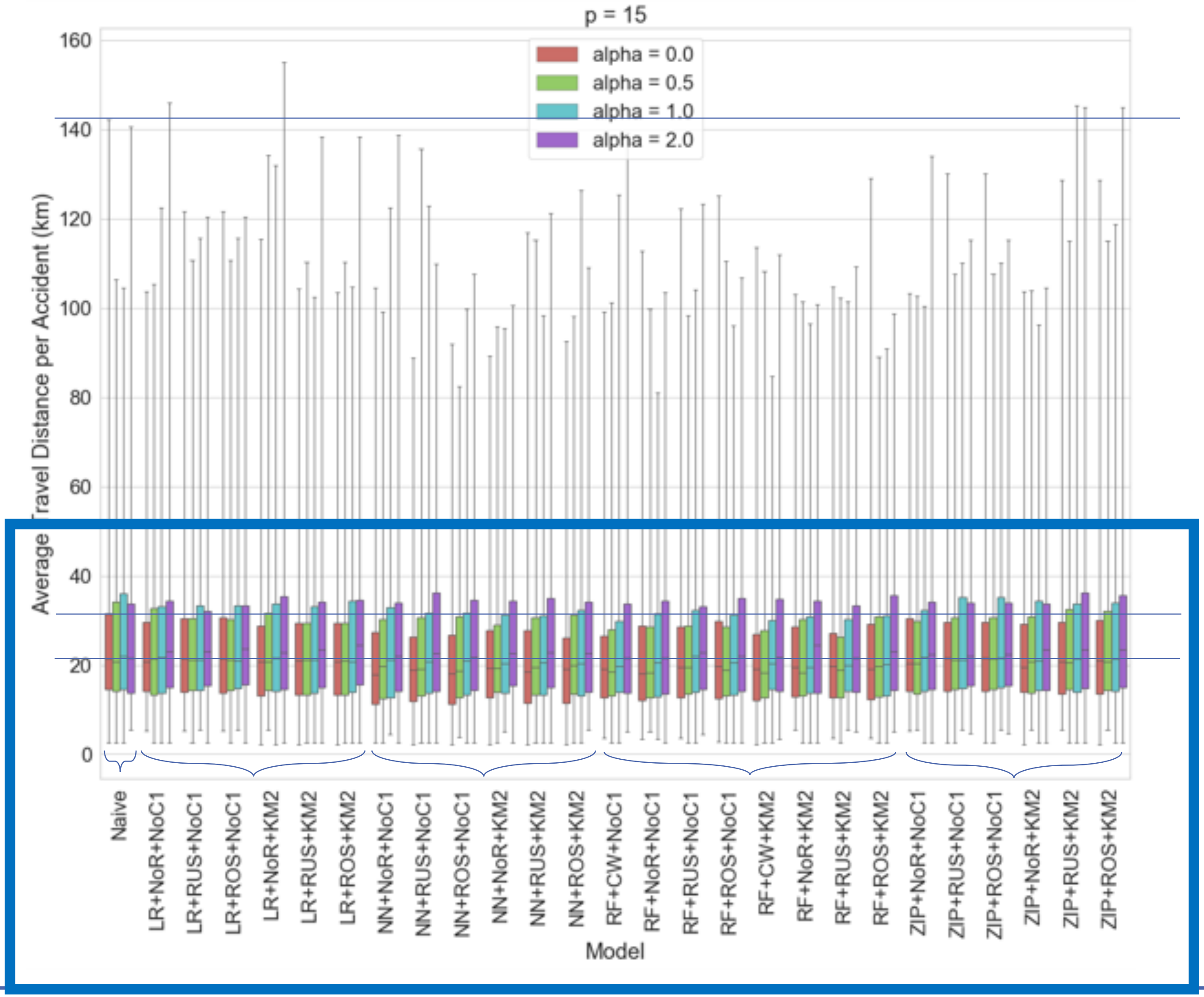
Model	Clustering	Resampling	Classification Metrics				Correl-ation		Total travel distance of responders per accident (km)												Average number of unattended accidents								Maximum number of unattended accidents							
									p =10				p=15				p=20				10				15				10				15			
Model	Clustering	Resampling	Acc.	Prec.	Rec.	F1	Pear.	Spear.	α=0	α=0.5	α=1	α=2	α=0	α=0.5	α=1	α=2	α=0	α=0.5	α=1	α=2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
Naive			95.5	3.8	4.2	4.0	82.1	60.8	39.48	38.44	43.21	45.35	26.29	25.78	27.34	26.78	19.29	19.43	20.36	23.12	0.54	0.49	0.48	0.46	0.02	0.01	0.01	0.01	15.00	14.00	14.00	16.00	2.00	1.00	1.00	2.00
LR	No cluster	No resampling	94.0	13.8	27.4	18.2	70.4	55.2	41.54	41.88	40.04	44.90	25.30	25.16	26.93	26.73	18.98	16.78	17.41	20.23	0.54	0.47	0.42	0.42	0.00	0.00	0.01	0.01	16.00	13.00	14.00	12.00	0.00	0.00	1.00	1.00
		RUS	93.0	12.8	32.3	18.3	63.1	54.7	42.90	43.41	39.97	44.36	25.07	25.38	26.01	26.94	19.07	18.05	17.00	20.41	0.56	0.52	0.46	0.46	0.00	0.00	0.01	0.00	17.00	17.00	15.00	15.00	0.00	0.00	1.00	0.00
		ROS	93.0	12.8	32.3	18.3	63.2	54.7	42.83	43.90	39.80	44.74	25.14	25.33	25.88	27.22	19.02	18.25	16.61	20.06	0.56	0.51	0.46	0.45	0.00	0.00	0.01	0.00	17.00	17.00	15.00	14.00	0.00	0.00	1.00	0.00
	clustering	No sample	93.0	12.5	30.9	17.7	76.6	58.4	40.79	39.44	42.57	44.81	24.44	25.14	26.21	27.79	18.55	19.39	18.95	21.45	0.53	0.41	0.43	0.44	0.02	0.01	0.01	0.01	17.00	12.00	13.00	15.00	3.00	1.00	1.00	2.00
		RUS	92.3	12.1	34.4	17.8	74.2	58.1	42.69	40.96	42.16	43.75	24.66	24.75	26.20	27.78	18.93	18.69	17.18	20.08	0.54	0.48	0.42	0.40	0.01	0.00	0.00	0.01	15.00	15.00	11.00	12.00	1.00	0.00	0.00	1.00
		ROS	92.4	12.2	34.2	17.9	74.2	58.1	42.78	40.89	42.71	44.22	24.58	24.84	26.18	28.29	18.87	18.66	17.04	19.90	0.54	0.48	0.42	0.41	0.01	0.00	0.00	0.01	15.00	15.00	11.00	15.00	1.00	0.00	0.00	1.00
NN	No cluster	No resampling	94.9	19.2	32.8	24.0	71.7	58.5	37.04	39.12	39.21	43.13	22.35	23.57	24.74	26.69	15.70	16.44	17.52	20.33	0.45	0.40	0.43	0.40	0.01	0.00	0.01	0.01	12.00	11.00	11.00	11.00	1.00	0.00	1.00	1.00
		RUS	95.0	19.2	32.6	24.1	73.2	59.3	37.44	39.07	37.83	43.84	22.24	23.85	24.97	27.64	16.40	16.21	17.05	20.27	0.47	0.41	0.43	0.45	0.00	0.01	0.01	0.01	12.00	11.00	11.00	12.00	0.00	1.00	1.00	1.00
		ROS	94.9	19.1	32.8	23.9	69.3	54.7	37.32	37.71	39.86	43.21	21.57	23.15	24.32	26.61	15.70	15.81	17.23	20.33	0.46	0.41	0.42	0.43	0.00	0.00	0.00	0.01	12.00	11.00	11.00	13.00	0.00	0.00	0.00	1.00
	clustering	No sample	95.0	19.0	31.6	23.7	75.6	58.9	39.32	39.88	39.61	43.09	23.18	23.96	24.58	27.34	17.46	17.15	17.00	20.16	0.44	0.40	0.42	0.42	0.00	0.00	0.00	0.00	15.00	12.00	12.00	14.00	0.00	0.00	0.00	0.00
		RUS	94.7	18.4	32.7	23.3	73.1	54.6	39.79	39.61	39.99	45.08	22.92	24.72	25.32	27.75	16.20	17.10	17.71	21.23	0.48	0.45	0.42	0.42	0.01	0.01	0.01	0.02	12.00	11.00	11.00	12.00	1.00	1.00	1.00	2.00
		ROS	94.7	18.3	33.1	23.3	74.5	55.4	38.60	38.24	40.66	45.50	22.23	23.78	25.04	27.40	16.31	16.89	18.00	20.81	0.48	0.41	0.44	0.41	0.00	0.00	0.00	0.01	13.00	11.00	14.00	11.00	0.00	0.00	0.00	1.00
Tree	No cluster	No resampling	95.0	19.0	31.8	23.6	78.7	63.4	40.81	38.28	39.62	44.46	23.21	22.99	24.30	26.22	16.88	16.36	16.49	19.97	0.51	0.44	0.42	0.42	0.00	0.00	0.00	0.02	13.00	12.00	12.00	11.00	0.00	0.00	0.00	1.00
		RUS	95.2	19.3	30.5	23.5	67.4	56.9	39.55	38.71	40.13	42.39	23.44	23.32	24.41	27.06	16.47	17.19	17.17	20.04	0.48	0.40	0.38	0.43	0.01	0.01	0.00	0.02	13.00	12.00	13.00	13.00	1.00	1.00	0.00	2.00
		ROS	95.3	18.6	27.6	22.1	79.2	64.6	41.14	39.86	40.37	45.29	23.72	23.78	25.12	26.82	17.89	16.53	16.68	20.14	0.53	0.46	0.44	0.42	0.01	0.01	0.00	0.00	16.00	13.00	11.00	14.00	1.00	1.00	0.00	0.00
		Class weights	95.4	20.6	30.4	24.4	77.1	62.5	39.79	39.46	39.91	44.58	23.14	23.14	24.09	26.56	16.24	16.51	17.68	20.04	0.46	0.41	0.40	0.41	0.01	0.01	0.00	0.01	12.00	11.00	12.00	12.00	1.00	1.00	0.00	1.00
	clustering	No resampling	95.1	18.9	30.5	23.2	79.8	62.3	41.40	38.81	39.88	43.16	22.98	23.02	24.56	26.75	16.88	16.25	16.89	19.90	0.49	0.42	0.42	0.43	0.01	0.00	0.00	0.01	12.00	13.00	11.00	12.00	1.00	0.00	0.00	1.00
		RUS	95.0	19.4	32.5	24.2	73.8	57.6	39.47	39.53	40.20	44.62	23.12	23.79	23.89	27.49	16.44	17.13	18.00	20.39	0.49	0.40	0.38	0.42	0.00	0.00	0.00	0.00	13.00	10.00	12.00	12.00	0.00	0.00	0.00	0.00
		ROS	95.1	18.3	28.7	22.2	80.1	63.6	40.94	39.70	40.82	44.21	23.36	23.48	24.45	26.60	16.92	16.08	16.80	20.38	0.51	0.45	0.43	0.40	0.01	0.00	0.01	0.01	13.00	13.00	11.00	12.00	2.00	0.00	1.00	1.00
		Class weights	95.4	20.6	30.4	24.4	77.1	62.5	39.53	38.50	40.95	45.12	23.54	23.54	23.94	27.58	16.29	17.34	18.12	20.37	0.48	0.38	0.38	0.41	0.01	0.00	0.00	0.01	12.00	10.00	10.00	12.00	1.00	0.00	0.00	1.00
ZIP	No cluster	No resampling	94.4	14.6	26.8	18.9	74.0	58.0	40.37	40.14	40.15	44.42	25.35	25.07	25.99	26.66	18.53	16.45	17.08	20.81	0.51	0.45	0.41	0.40	0.02	0.00	0.00	0.01	12.00	15.00	12.00	14.00	3.00	0.00	0.00	1.00
		RUS	94.2	13.9	26.1	18.1	61.1	50.6	44.57	45.68	40.89	44.23	25.72	25.43	26.86	27.54	18.93	19.26	16.93	19.91	0.59	0.53	0.51	0.48	0.01	0.02	0.01	0.00	16.00	17.00	15.00	15.00	2.00	3.00	2.00	0.00
		ROS	94.2	13.9	26.7	18.2	61.2	50.6	44.51	45.42	40.70	44.62	25.77	25.48	27.00	27.80	18.88	19.08	16.98	19.76	0.59	0.53	0.51	0.48	0.01	0.02	0.01	0.00	16.00	17.00	15.00	15.00	2.00	3.00	2.00	0.00
	clustering	No resampling	93.1	13.1	31.9	18.5	77.6	61.8	39.35	41.08	40.12	44.97	24.17	24.66	26.42	27.26	18.06	18.40	18.91	20.92	0.47	0.45	0.42	0.37	0.02	0.00	0.01	0.01	13.00	15.00	12.00	12.00	3.00	0.00	1.00	1.00
		RUS	93.0	12.7	30.8	S	74.2	57.1	43.46	41.76	42.50	45.17	24.67	26.08	26.56	28.36	19.36	19.85	17.47	21.08	0.55	0.48	0.49	0.46	0.00	0.01	0.00	0.01	16.00	13.00	15.00	12.00	0.00	1.00	0.00	1.00
		ROS	93.0	12.8	30.9	18.0	74.3	57.0	43.57	41.31	42.71	45.08	24.77	26.13	26.41	28.52	19.32	20.02	17.29	20.98	0.57	0.49	0.48	0.45	0.00	0.01	0.00	0.01	17.00	13.00	15.00	12.00	0.00	1.00	0.00	1.00

Evaluation of the Prediction Models in Real Scenarios

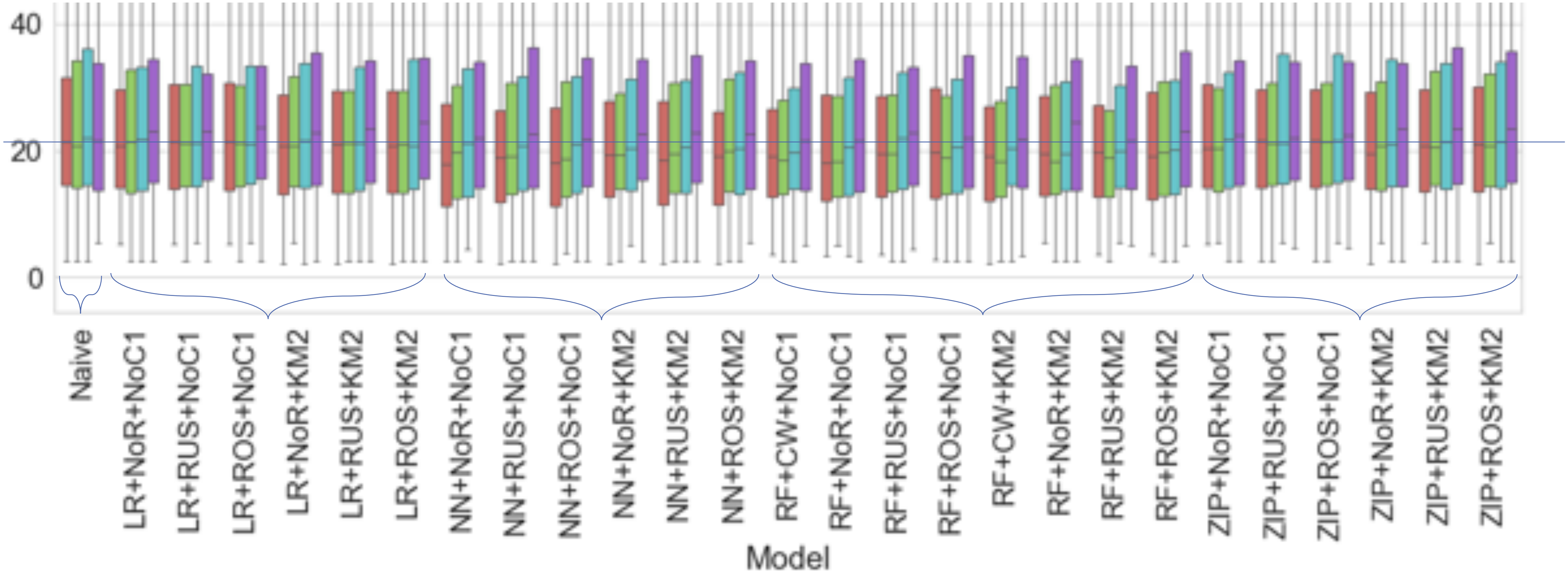


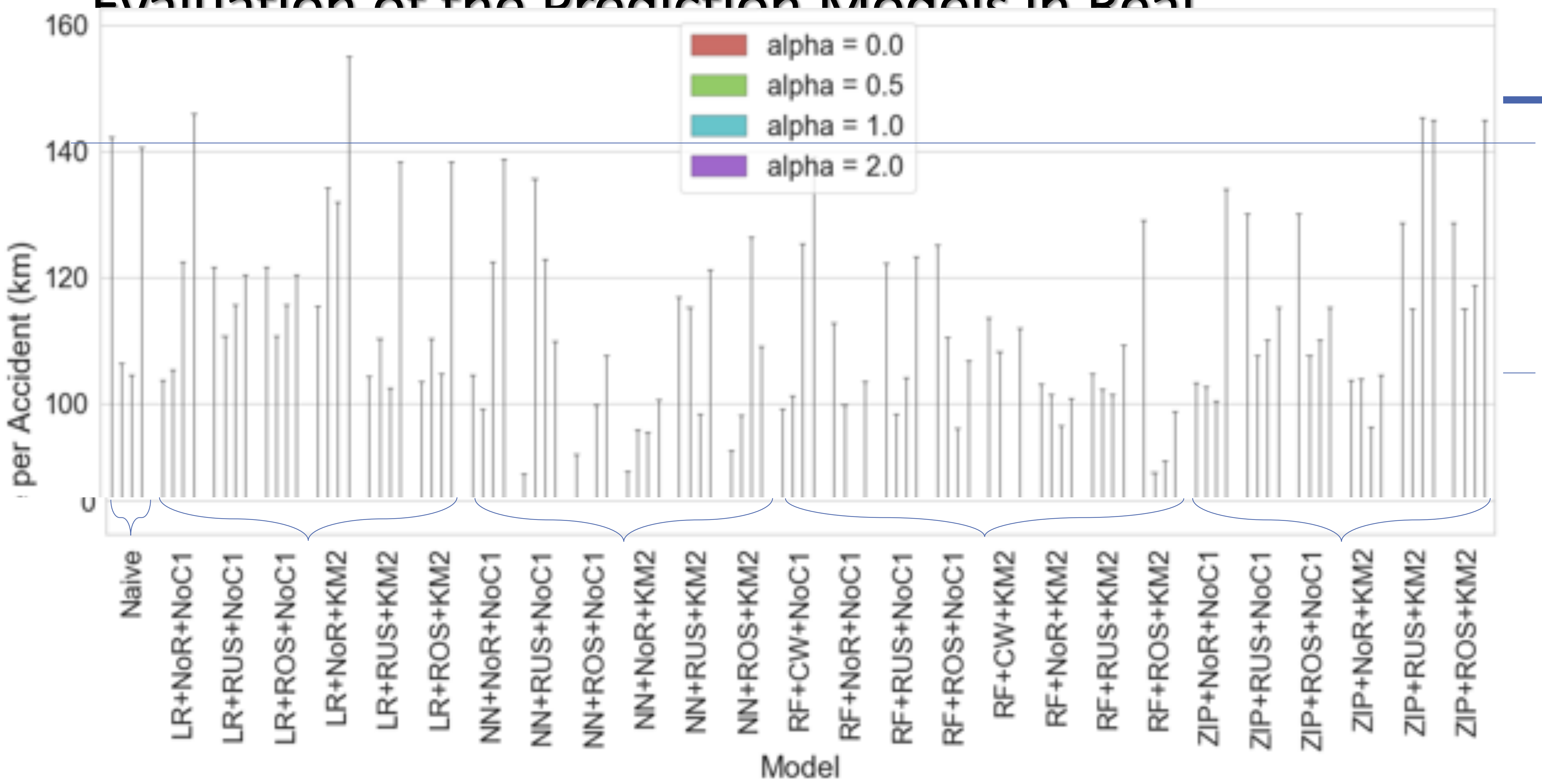





p = 15
alpha = 0.0
alpha = 0.5
alpha = 1.0
alpha = 2.0
Average Travel Distance per Accident (km)
160
140
120
100
80
60
40
20
0
Naive
LR+NoR+NoC1
LR+RUS+NoC1
LR+ROS+NoC1
LR+NoR+KM2
LR+RUS+KM2
LR+ROS+KM2
NN+NoR+NoC1
NN+RUS+NoC1
NN+ROS+NoC1
NN+NoR+KM2
NN+RUS+KM2
NN+ROS+KM2
RF+CW+NoC1
RF+NoR+NoC1
RF+RUS+NoC1
RF+ROS+NoC1
RF+CW+KM2
RF+NoR+KM2
RF+RUS+KM2
RF+ROS+KM2
ZIP+NoR+NoC1
ZIP+RUS+NoC1
ZIP+ROS+NoC1
ZIP+NoR+KM2
ZIP+RUS+KM2
ZIP+ROS+KM2
Model

Evaluation of the Prediction Models in Real Scenarios

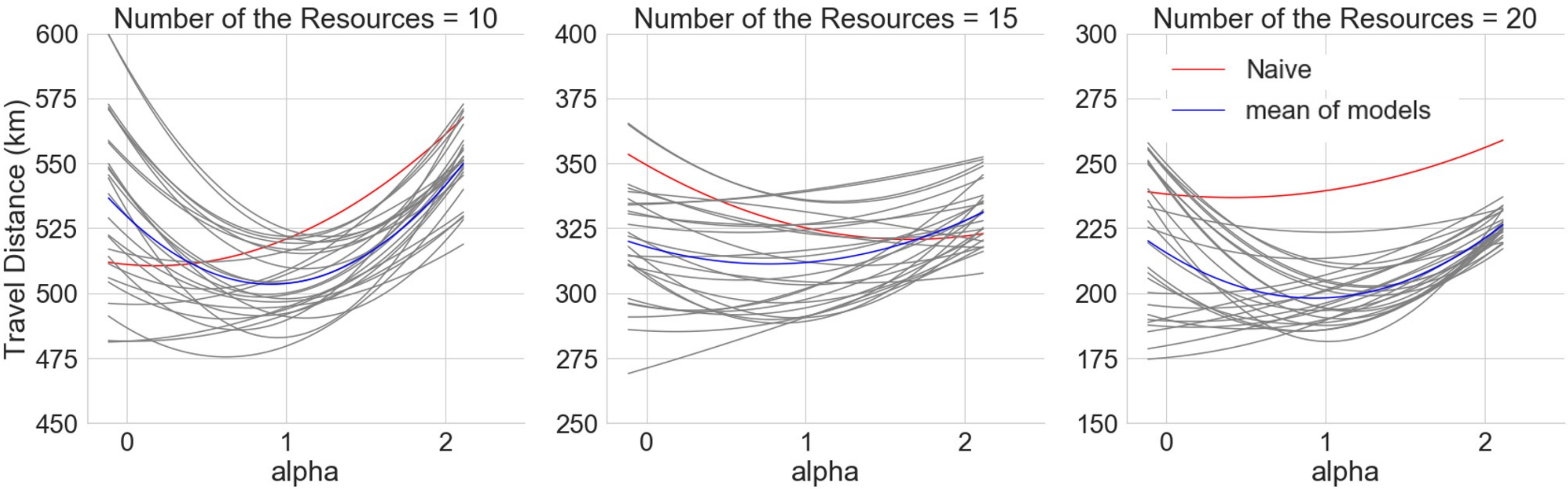




alpha = 0.0
alpha = 0.5
alpha = 1.0
alpha = 2.0
160
140
120
100
0
per Accident (km)
Naive
LR+NoR+NoC1
LR+RUS+NoC1
LR+ROS+NoC1
LR+NoR+KM2
LR+RUS+KM2
LR+ROS+KM2
NN+NoR+NoC1
NN+RUS+NoC1
NN+ROS+NoC1
NN+NoR+KM2
NN+RUS+KM2
NN+ROS+KM2
RF+CW+NoC1
RF+NoR+NoC1
RF+RUS+NoC1
RF+ROS+NoC1
RF+CW+KM2
RF+NoR+KM2
RF+RUS+KM2
RF+ROS+KM2
ZIP+NoR+NoC1
ZIP+RUS+NoC1
ZIP+ROS+NoC1
ZIP+NoR+KM2
ZIP+RUS+KM2
ZIP+ROS+KM2
Model

Effect of Hyperparameter α

- As the fleet size increases, the influence of α decreases.



Further Details

- https://arxiv.org/abs/2106.08307

Cornell University

arXiv.org > cs > arXiv:2106.08307

Search...

Help | Adv

Computer Science > Machine Learning

[Submitted on 15 Jun 2021]

Learning Incident Prediction Models Over Large Geographical Areas for Emergency Response Systems

Sayyed Mohsen Vazirizade, Ayan Mukhopadhyay, Geoffrey Pettet, Said El Said, Hiba Baroud, Abhishek Dubey

Principled decision making in emergency response management necessitates the use of statistical models that predict the spatial-temporal likelihood of incident occurrence. These statistical models are then used for proactive stationing which allocates first responders across the spatial area in order to reduce overall response time. Traditional methods that simply aggregate past incidents over space and time fail to make useful short-term predictions when the spatial region is large and focused on fine-grained spatial entities like interstate highway networks. This is partially due to the sparsity of incidents with respect to the area in consideration. Further, accidents are affected by several covariates, and collecting, cleaning, and managing multiple streams of data from various sources is challenging for large spatial areas. In this paper, we highlight how this problem is being solved for the state of Tennessee, a state in the USA with a total area of over 100,000 sq. km. Our pipeline, based on a combination of synthetic resampling, non-spatial clustering, and learning from data can efficiently forecast the spatial and temporal dynamics of accident occurrence, even under sparse conditions. In the paper, we describe our pipeline that uses data related to roadway geometry, weather, historical accidents, and real-time traffic congestion to aid accident forecasting. To understand how our forecasting model can affect allocation and dispatch, we improve upon a classical resource allocation approach. Experimental results show that our approach can significantly reduce response times in the field in comparison with current approaches followed by first responders.

Subjects: **Machine Learning (cs.LG)**
Cite as: arXiv:2106.08307 **[cs.LG]**
(or arXiv:2106.08307v1 **[cs.LG]** for this version)

Thank you for your attention.

NSF

VANDERBILT UNIVERSITY

TN TDOT Department of Transportation

ISIS