Using AI for spatial predictions of driver behavior

International Transport Forum
Roundtable on Artificial Intelligence in
Road Traffic Crash Prevention, 10-12 February 2021

Apostolos Ziakopoulos
& George Yannis

Department of Transportation
Planning and Engineering,
National Technical University of Athens

www.nrso.ntua.gr/apziak

apziak@central.ntua.gr

Research scope and questions

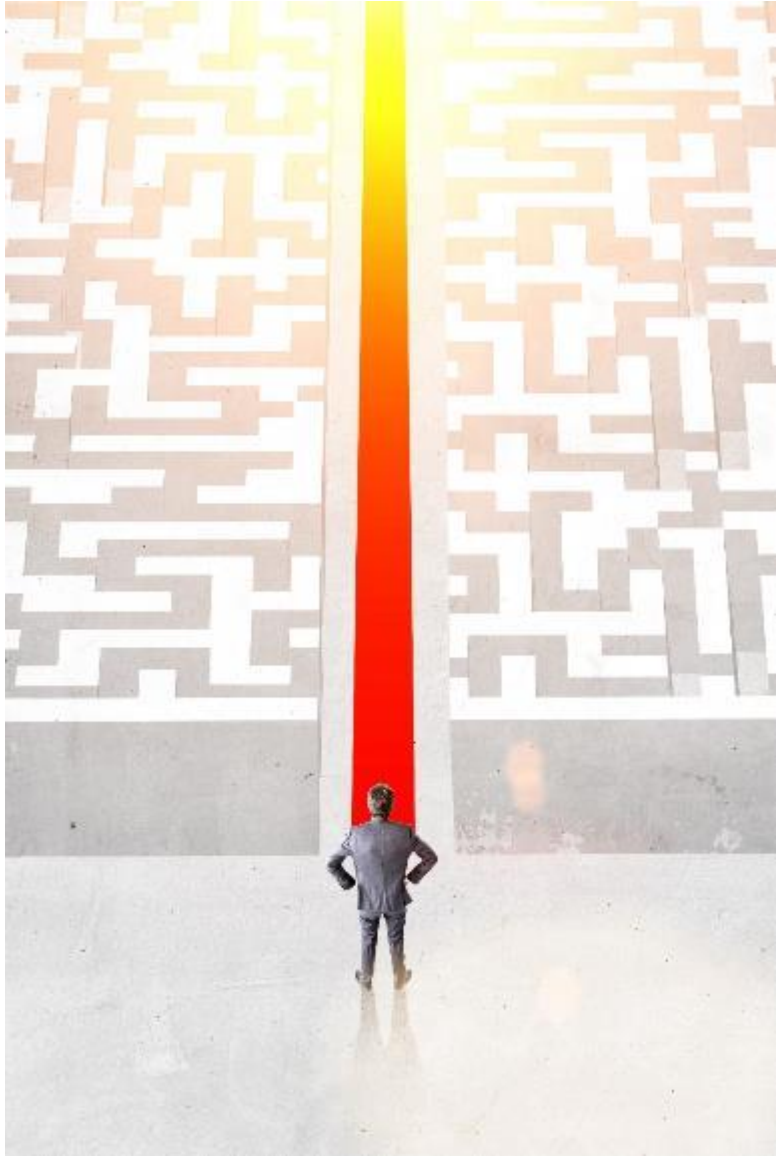
Spatial analysis of harsh event frequencies (harsh brakings/accelerations) in road segments

1. How can **high-resolution naturalistic driving** smartphone data and road segment **geometric** and **road network** characteristic data be **combined** (map-matched) and **examined** in road safety investigations?

2. How can **harsh event** frequencies be **analyzed spatially** in urban networks, and can AI methods be used for that purpose?

3. Which **road geometry** and **road network characteristics** affect harsh event frequencies in urban road network environments?

4. **How transferable** are the previous results in a different study area? Can reliable predictions be conducted?



Merits of harsh event examination

Harsh events: **harsh brakings** and **harsh accelerations** recorded by smartphone sensors for telematics-based vehicle insurance primes

- Parameters measuring **road safety levels** (correlations with spatial and temporal headways)
- Inherently linked with **driver risk** (Tselentis et al, 2017)
- **Different phenomena**, correlations with different variables (Ziakopoulos et al, 2020)

Considerable **comparative advantages** for their investigation:

1. Applications in driver **evaluation** and **classification** (Bonsall et al., 2005; Gündüz et al., 2018).

2. **Proactive** road safety indicators anticipating safety-critical events (Zohar et al., 2014; Jansen & Wesseling, 2018); evaluations **before** crashes occur

3. Non-aggressive driving reduces **emissions** by up to 40% (Alessandrini, 2012)

4. Investigated by the **insurance** industry (Paefgen et al., 2012; 2014)

5. Apparent **research gaps** in the investigation of harsh event frequencies

Data collection (1/2): Digital map road geometry data

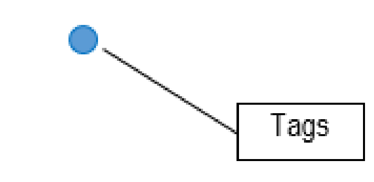
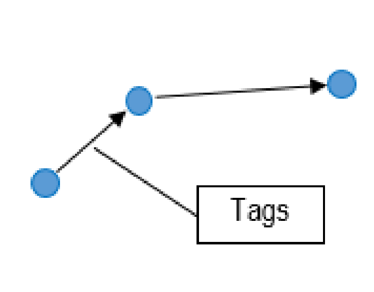
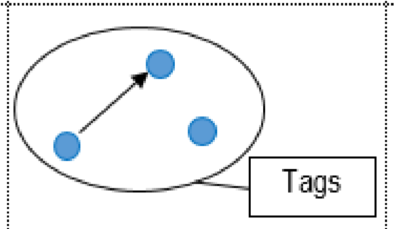
Data of **road segment geometry** and **road network** characteristics on a **microscopic level** from digital maps

OpenStreetMap: Open source digital map platform
Hierarchical elements:
1. **Nodes**
2. **Ways** from node groups
3. **Relations** from node and way groups

Obtaining a wealth of data in WGS84 through API queries (Overpass Turbo API through Overpass Query Language)

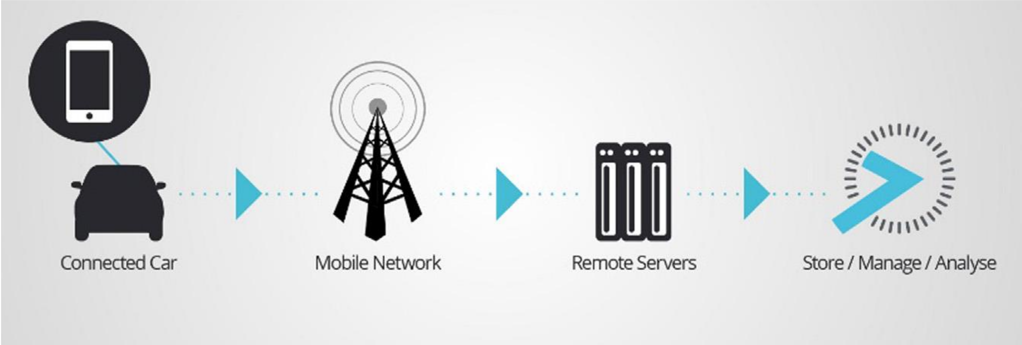
NASA SRTM topography

Altitude data provided by NASA:
- **Freely** available
- **Altitude** resolution per 10 cm – compared with OSM altitudes for verification – some accuracy **issues**
- Majority of **populated areas** available

Hierarchy level	Core Element	Schematic representation	OSM symbol	
1	Node		Node	
			Tag	
2	Way		Open polyline	
			Closed polyline	
			Area	
3	Relation		Relation	

Data collection (2/2): Naturalistic driving data from smartphones

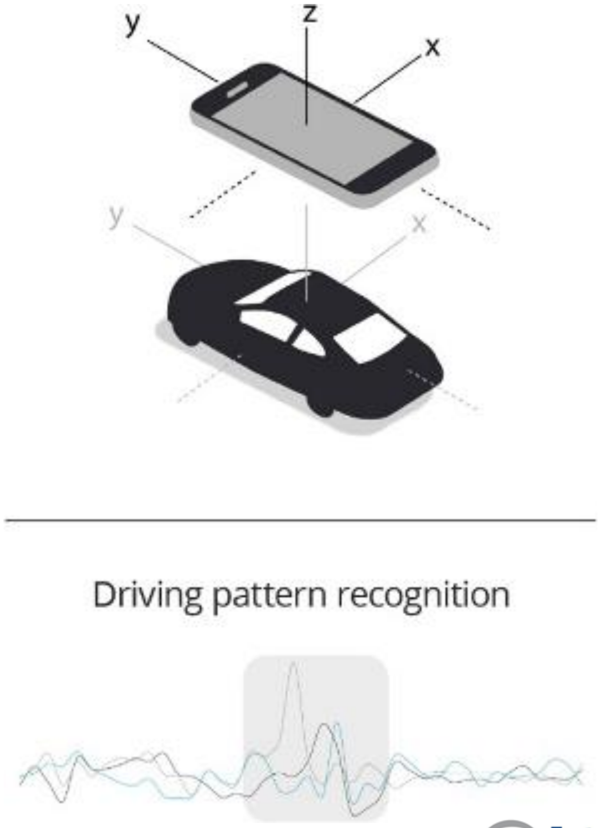
Naturalistic driving data from real-world conditions obtained from smartphones (per trip-second), primarily recorded for telematics-based vehicle insurance primes



Source: OSeven Telematics, (2020)

Utilization of the application/platform of OSeven Telematics

- APIs utilization for data reading from **smartphone sensors**
- **Exploited sensors**: GPS, accelerometer, gyroscope, device orientation
- **Transmission** from smartphone to central storage database
- **Data cleaning** and **processing** via a series of filtering, signal processing, Machine Learning (ML) and scoring algorithms
- Several data are provided, **indicatively**: trip position, speed, acceleration, harsh brakings/accelerations, event intensity, speeding, mobile phone use
- **Total anonymity** during all data handling phases (GDPR)



High resolution big data from driver trips including behaviour indicators

Data processing: Geometric characteristics (1/2)

Calculation of geometric characteristics based on **OSM node coordinates**

Roadway segment length

- Calculation based on modern geoids/ellipsoid models through available libraries
- Sum of elementary lengths (2 nodes each)

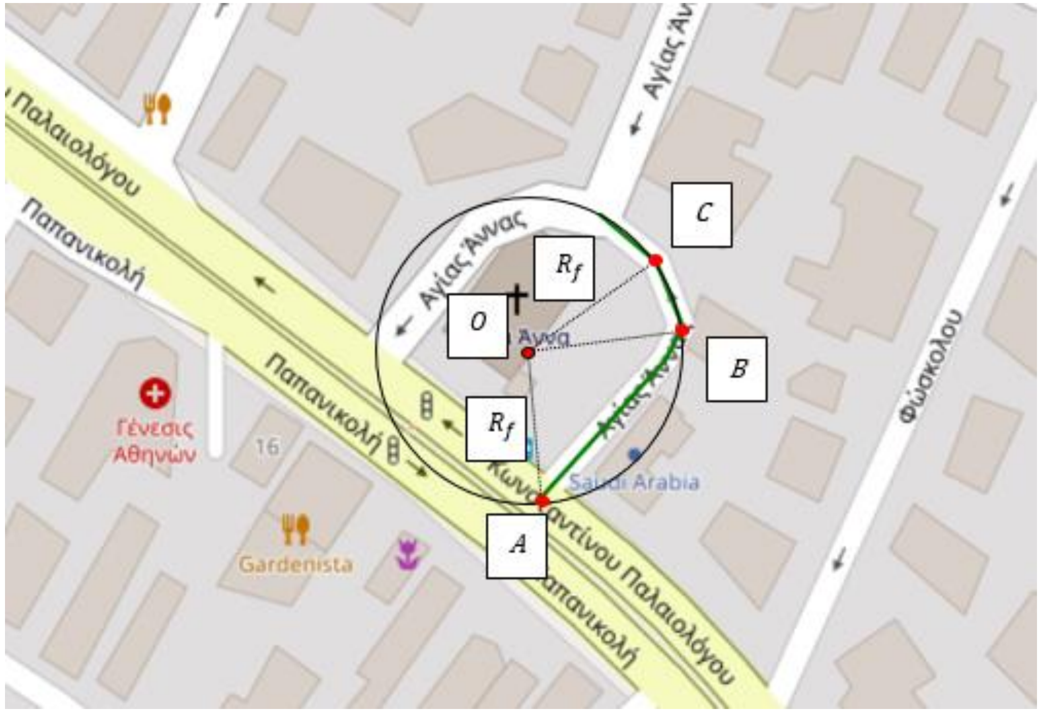
Determination of road segment centroids

Gradient

- Sum of elementary gradients (2 nodes each)
- Road segment average, weighted by elementary lengths

Curvature

- Menger's formula per elementary triangle (3 nodes each)
- Road segment average, weighted by elementary lengths



Data processing: Geometric characteristics (2/2)

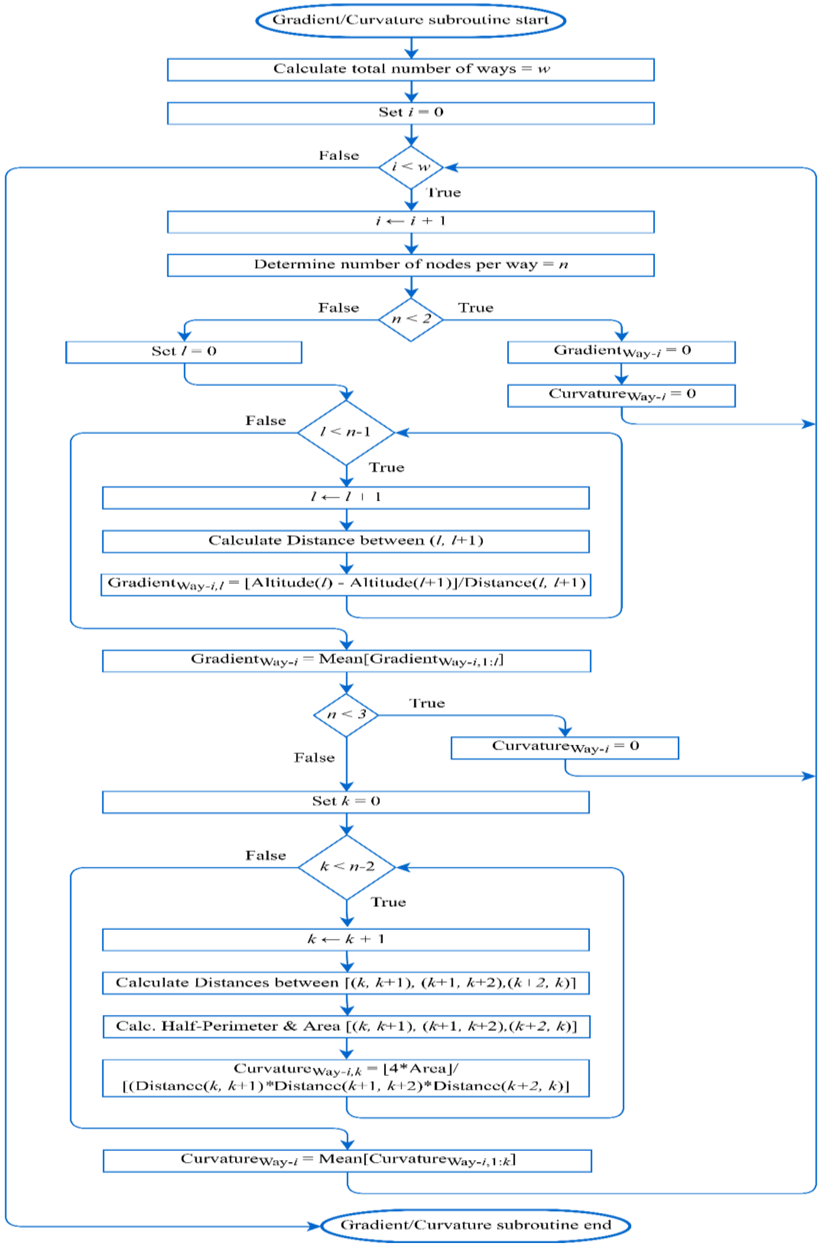
Neighborhood complexity calculation

- Measurement of density and complexity of immediate road segment environment: (i) in reality (ii) on the digital maps
- Logarithm of nodes within a window of 470m * 470m from each road segment centroid

Obtaining of additional **road segment characteristics** from OSM:

1. Presence of **pedestrian crossing**
2. Presence of **traffic lights**
3. **Lane number**
4. Road type
   (exclusion of walkways/footpaths/surfaces without vehicles)
5. Direction **number** (one-way or two-way)

Calculation with original purpose-made algorithms and sub-routines created in R-studio, iteratively for each road segment



Data processing: Map-matching (1/2)

Map-matching: Plotting of naturalistic driving data on maps after determination of the corresponding segment

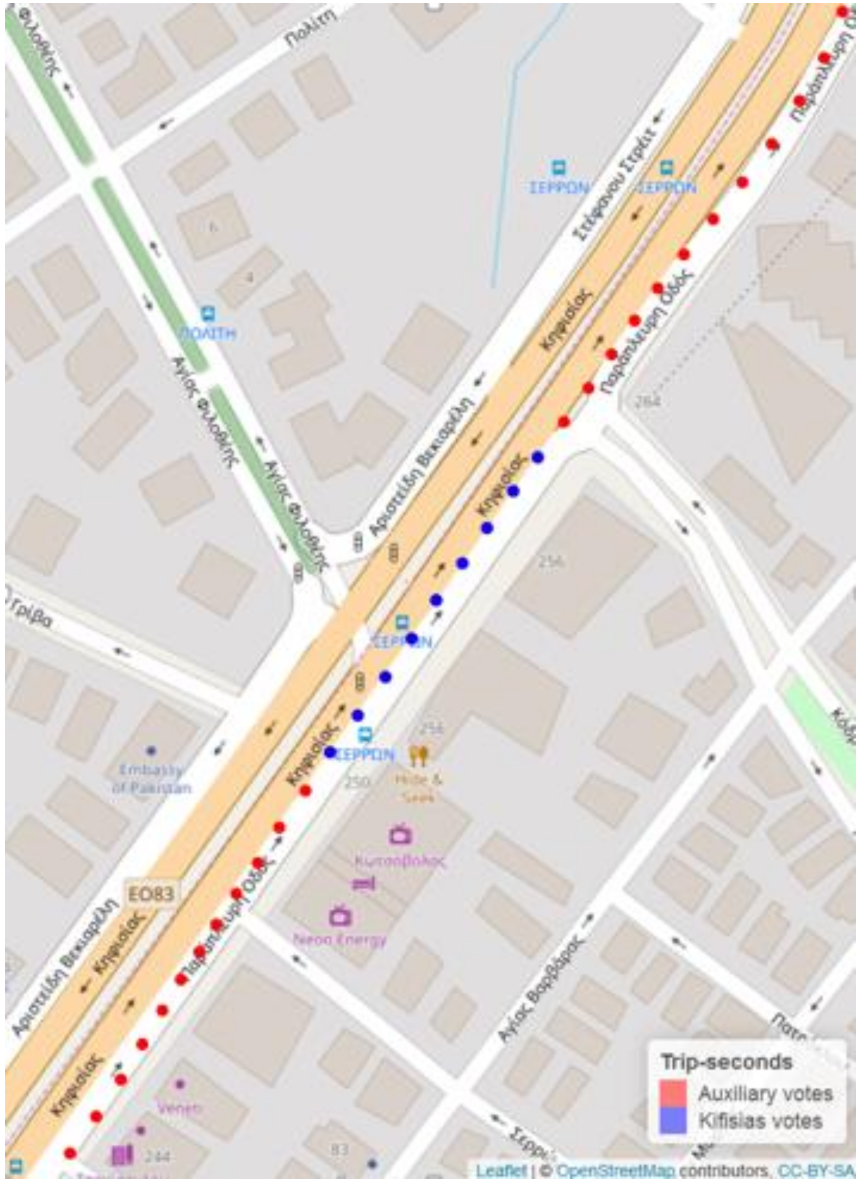
Matching of GPS trace to each road segment **per second**

Identification of:

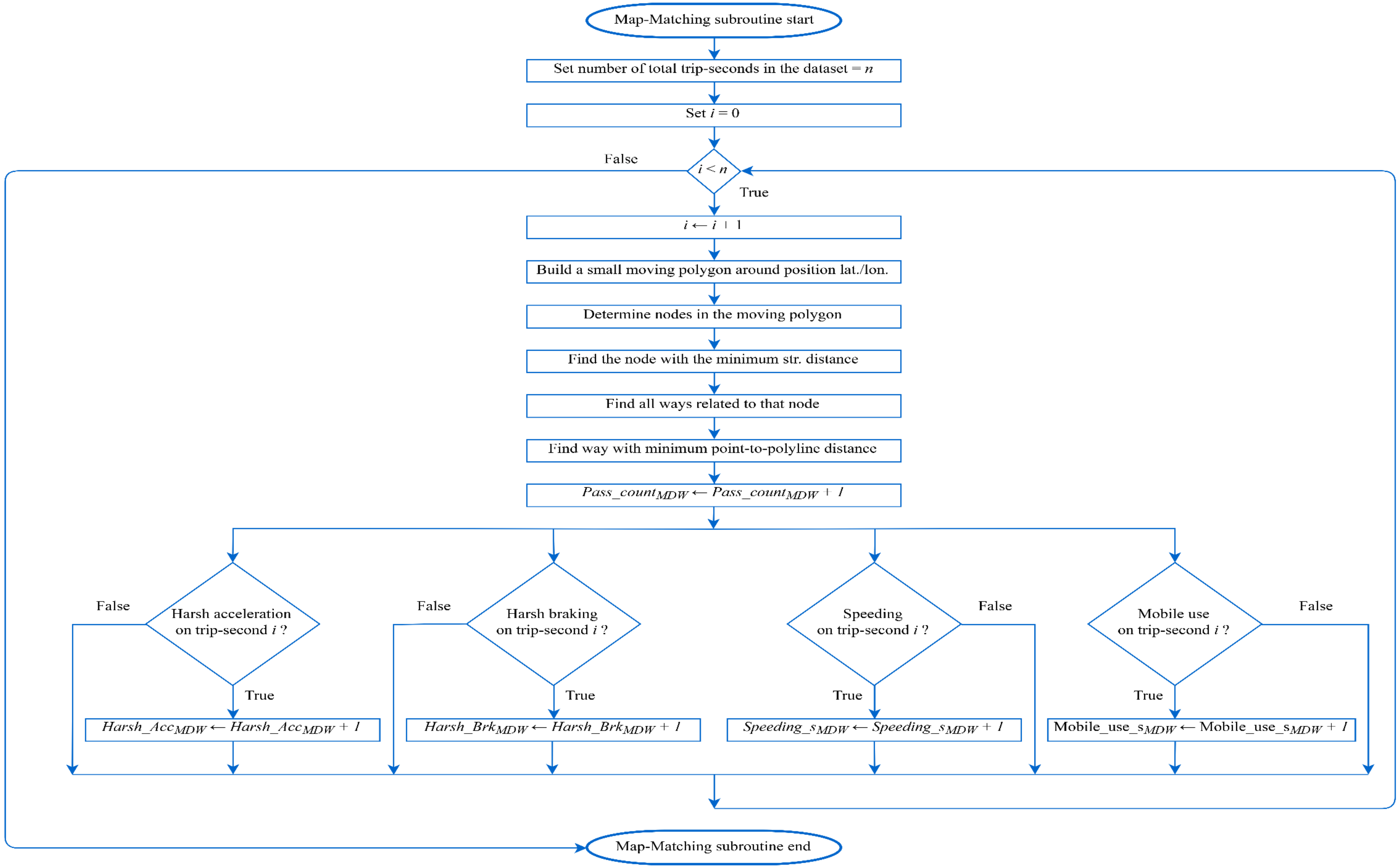
1. Nearest node (point-to-point distance)
2. Minimum distance way – MDW (point-to-polyline distance)
   - **Moving polygon** serving to reduce candidate ways
   - **Time-consuming** and **computationally demanding** process
   - **Corrections** are essential in dense road segments with parallel axes through a specialized vote-count algorithm

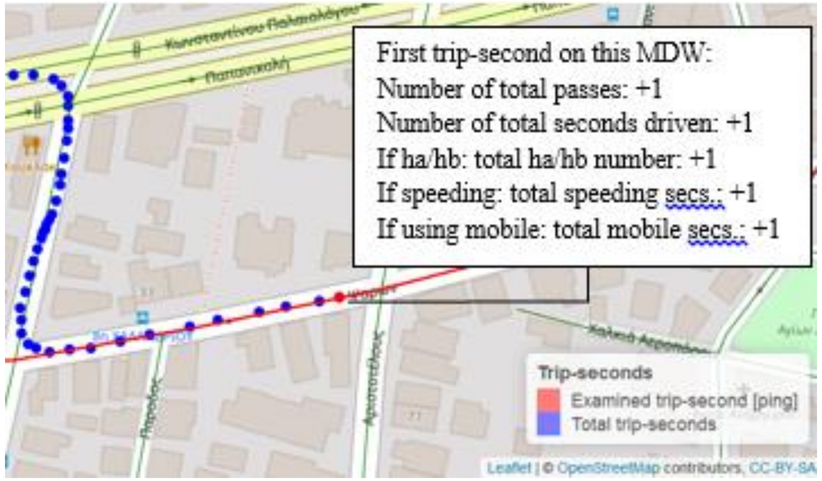
Recording and assignment per road segment**:**

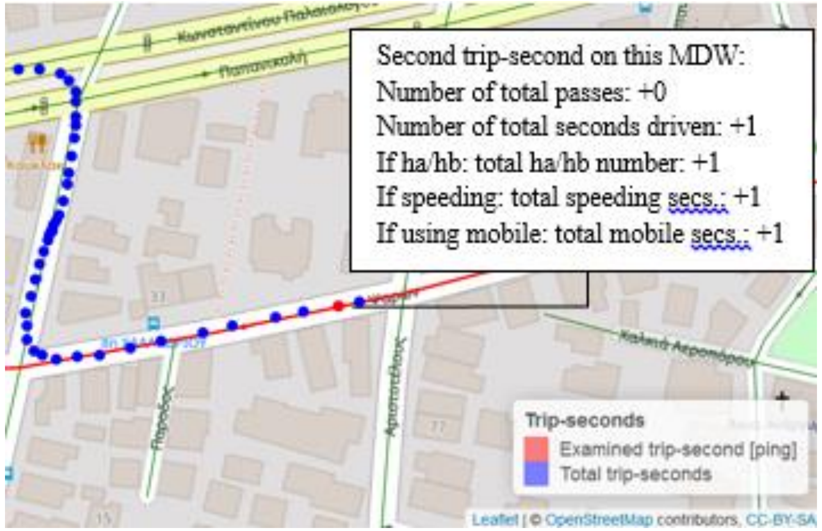
1. Pass count
2. Harsh brakings/accelerations
3. Speeding seconds
4. Mobile use seconds



Data processing: Map-matching (2/2)







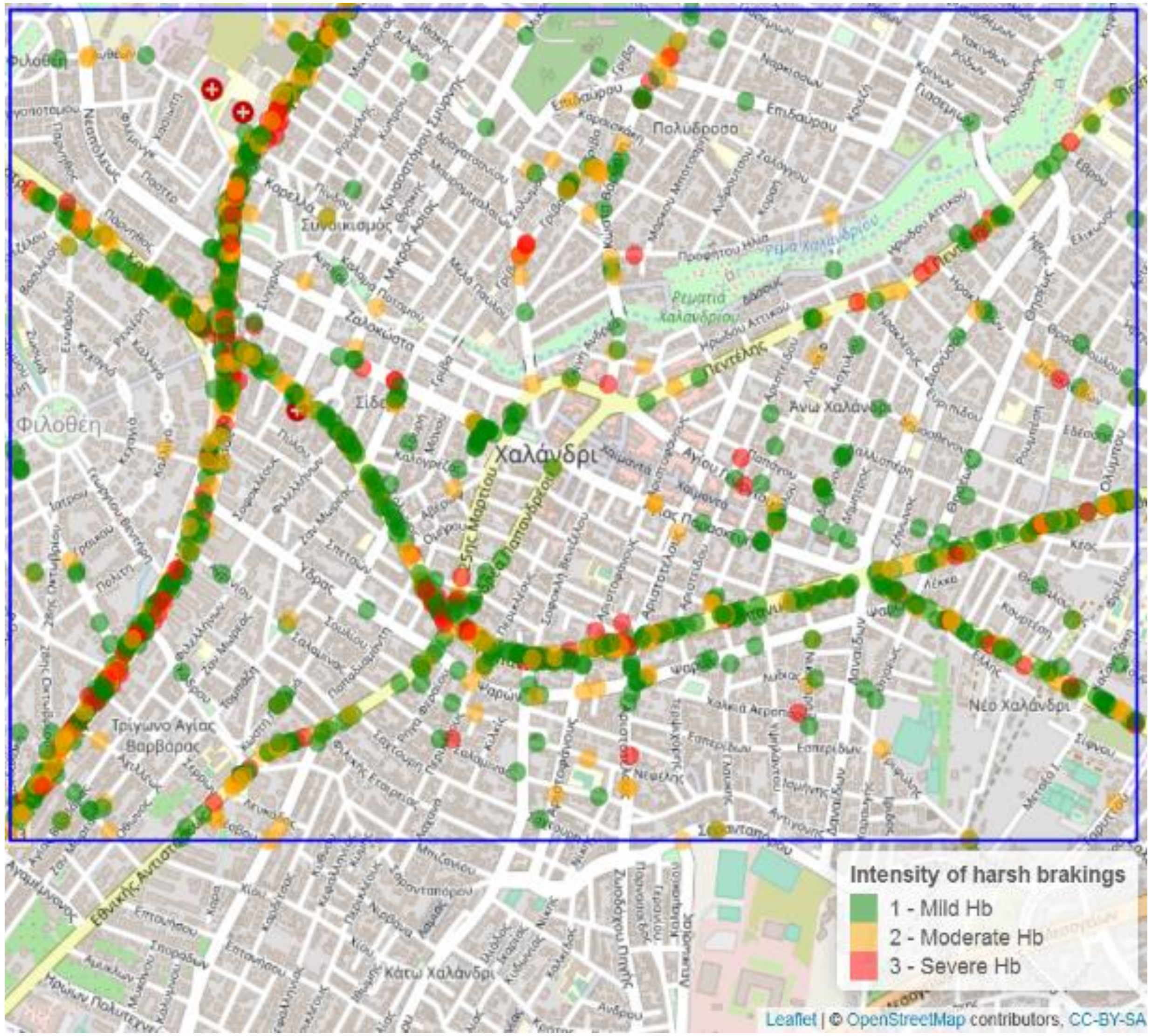
Sample description (1/2) – Chalandri urban road network

869 road segments (removal of 14 footways) with **4293** nodes
- **49** road segments with traffic lights
- **80** road segments with pedestrian crossings

Naturalistic driving data:
- Trips between 01-10-2019 & 29-11-2019 – **2 months**
- **A total of** 3294 trips from 230 drivers
- 1,000,273 **driving seconds**: average trip duration 304 s
- 1348 harsh **brakings**
- 921 harsh **accelerations**

90% of road segments feature **at least 1** trip

Variable distributions
- **Positive** skewness (larger right tails)
- **High** kurtosis (non-normal distributions)



Sample description (2/2) – Omonoia urban road network

1237 road segments (removal of 78 footways) with **6115** nodes

- **319** road segments with traffic lights
- **317** road segments with pedestrian crossings

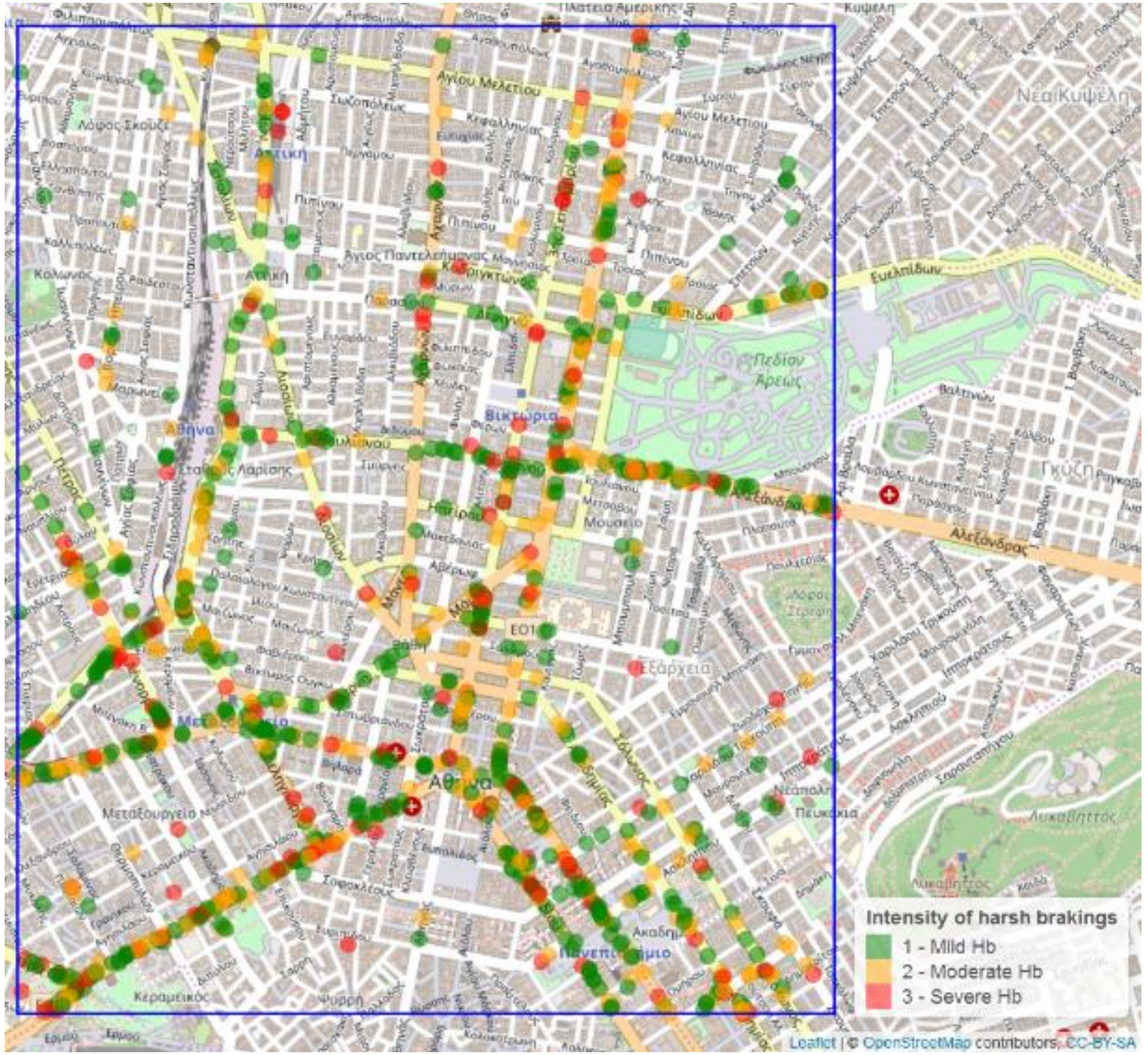
Naturalistic driving data:

- Trips between 01-10-2019 & 29-11-2019 – **2 months**
- **A total of** 2615 trips from 257 drivers
- 964,693 **driving seconds**: average trip duration 369 s
- 1036 harsh **brakings**
- 938 harsh **accelerations**

86% of road segments feature **at least 1** trip

Variable distributions

- **Positive** skewness (larger right tails)
- **High** kurtosis (non-normal distributions)



Arsenal of spatial statistical models & AI

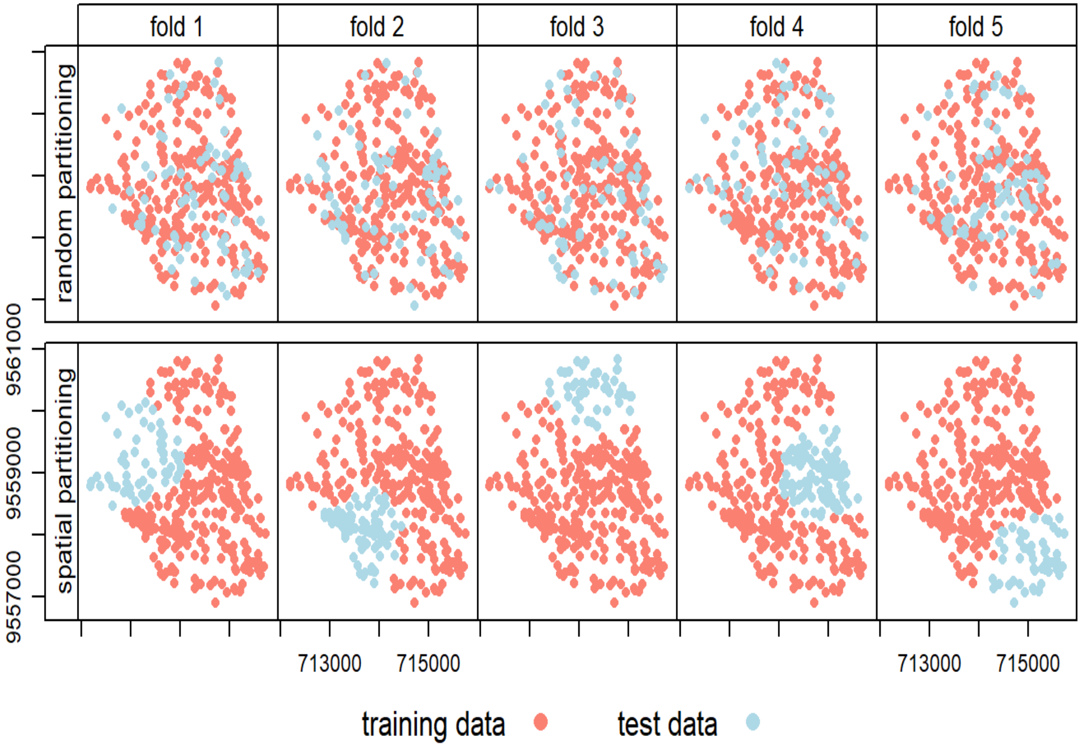
Integration of **spatial heterogeneity**

Event **frequencies: Log-normal Poisson** framework

1. Geographically Weighted Poisson Regression **(GWPR)**
   - Frequentist functional models: local micro-regressions are conducted, b coefficients can vary locally

2. Conditional Autoregressive Prior Regression **(CAR)**
   - Bayesian functional models: Bayesian regressions are conducted with spatially structured and unstructured terms, b coefficient distributions are obtained

3. Extreme Gradient Boosting **(XGBoost) – AI methods**
   - Machine learning: Multiple additive regression trees (ensemble), obtained information regarding variable contribution (gain)
   - Random Cross-Validation – **RCV**
   - Spatial Cross-Validation – **SPCV**



Source: Lovelace et al. (2019)

Harsh braking spatial analyses in urban road networks

Positive correlation:
Segment length
Pass count

Negative correlation:
Gradient
Neighborhood complexity
Road type [Residential]

Marginally positive correlation:
Road type [Secondary]
Traffic lights
Pedestrian crossing

Marginally negative correlation:
Road type [Tertiary]

Independent variables	GWPR	CAR	RCV XGBoost	SPCV XGBoost
	Coefficients	Mean posterior values	Gain values	Gain values
Intercept	0.4636	**-1.4134**	N/A	N/A
Gradient	**-2.4864**	**-9.7538**	**0.0806**	**0.0860**
Curvature	—	—	**0.0444**	**0.0626**
Neighborhood complexity	**-0.2919**	**-0.1787**	**0.0344**	**0.0684**
Segment length	**0.0039**	**0.0075**	**0.1436**	**0.1400**
Pass count	**0.0040**	**0.0086**	**0.6788**	**0.6271**
Traffic lights: Yes [Ref.: Traffic lights: No]	**0.2563**	-0.0902	**0.0037**	**0.0010**
Pedestrian crossing: Yes [Ref.: Pedestrian crossing: No]	-0.1463	**0.3820**	**0.0024**	**0.0024**
Lanes: 2 [Ref.: Lanes: 1]	**-0.2435**	**-0.1713**		
Lanes: 3 [Ref.: Lanes: 1]	**0.3669**	**-0.5719**	**0.0072**	**0.0048**
Lanes: 4 [Ref.: Lanes: 1]	0.3578	**1.9169**		
Road type: secondary [Ref.: Road type: primary]	**1.0520**	-0.1094		
Road type: tertiary [Ref.: Road type: primary]	-0.0070	**-1.6389**	**0.0049**	**0.0078**
Road type: residential [Ref.: Road type: primary]	**-1.0084**	**-2.5578**		
Sigma-phi^2 [Spatially structured effects]	N/A	**700.3172**	N/A	N/A
Sigma-theta^2 [Spatially unstructured effects]	N/A	**2.3455**	N/A	N/A
Performance metrics				
RMSE	3.2954	1.2830	1.4215	1.8293
MAE	1.3048	0.4115	0.4971	0.4994
RMSLE	0.5569	0.1727	0.3140	0.2390
CA	80.90%	96.32%	90.56%	91.71%

Harsh braking prediction & transferability

Predictions on Omonoia test area

1. Geographically Weighted Poisson Regression **(GWPR)**
   - Local b-coefficient fluctuations are not transferable
   - Predictions using global Poisson regression

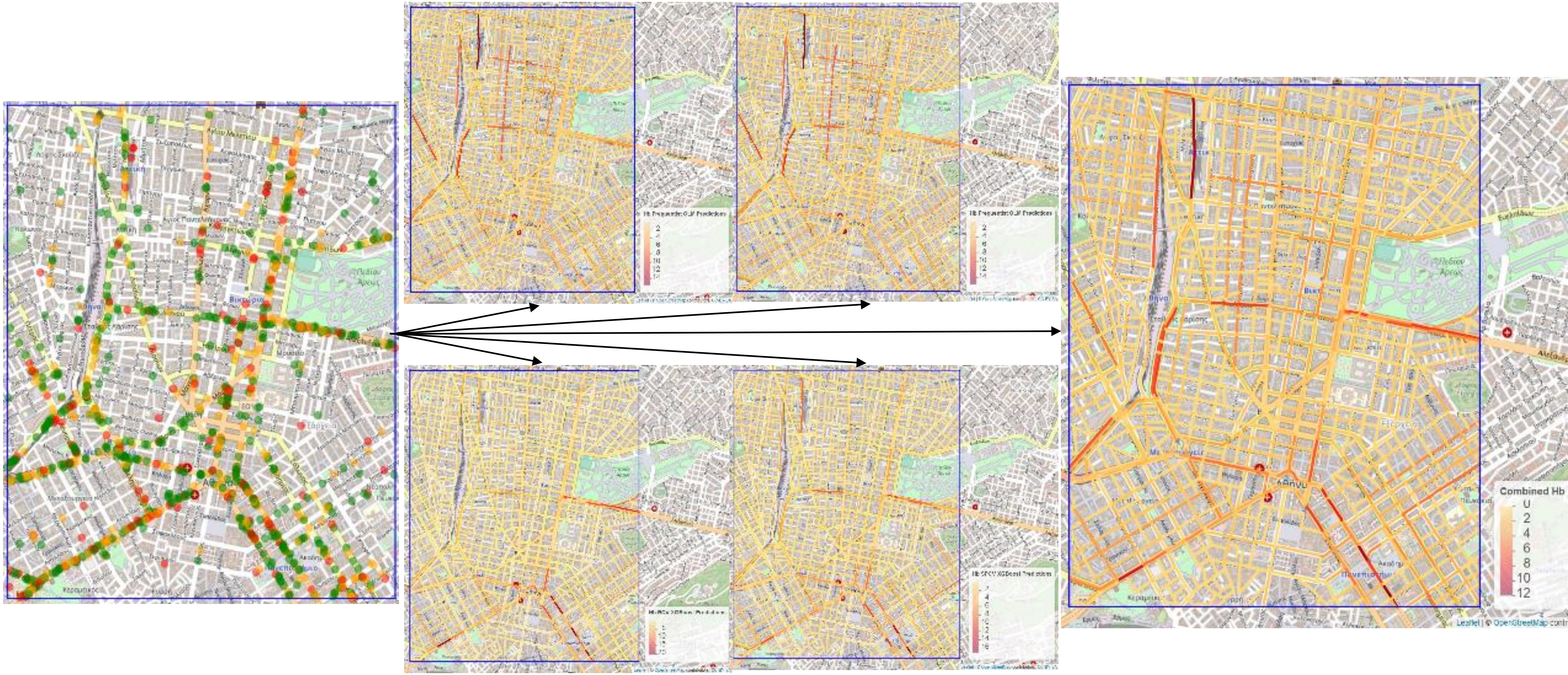
2. Bayesian Conditional Autoregressive Prior Regression **(CAR)**
   - Spatially structured and unstructured effects are not transferable
   - Predictions using new Bayesian Poisson regression

3. Extreme Gradient Boosting **(XGBoost)**
   - **Seamless transferability** of machine learning ensemble trees/rules using both RCV and SPCV

SPCV XGBoost has the **best individual performance** from all implemented methods

Performance metrics	GWPR global Poisson	Bayesian Poisson	RCV XGBoost	SPCV XGBoost	Combined Average
RMSE	1.9792	1.9804	1.9834	1.8418	1.6114
MAE	1.0265	1.0290	0.8415	0.7542	0.6645
RMSLE	0.5508	0.5520	0.5484	0.5189	0.4514
CA	82.64%	82.74%	83.40%	85.27%	87.55%

Combined harsh braking predictions for the test urban network



Using **combined average**, spatial models **mitigate** their weaknesses and lead to a **balanced** predictive outcome for harsh brakings

Case study findings

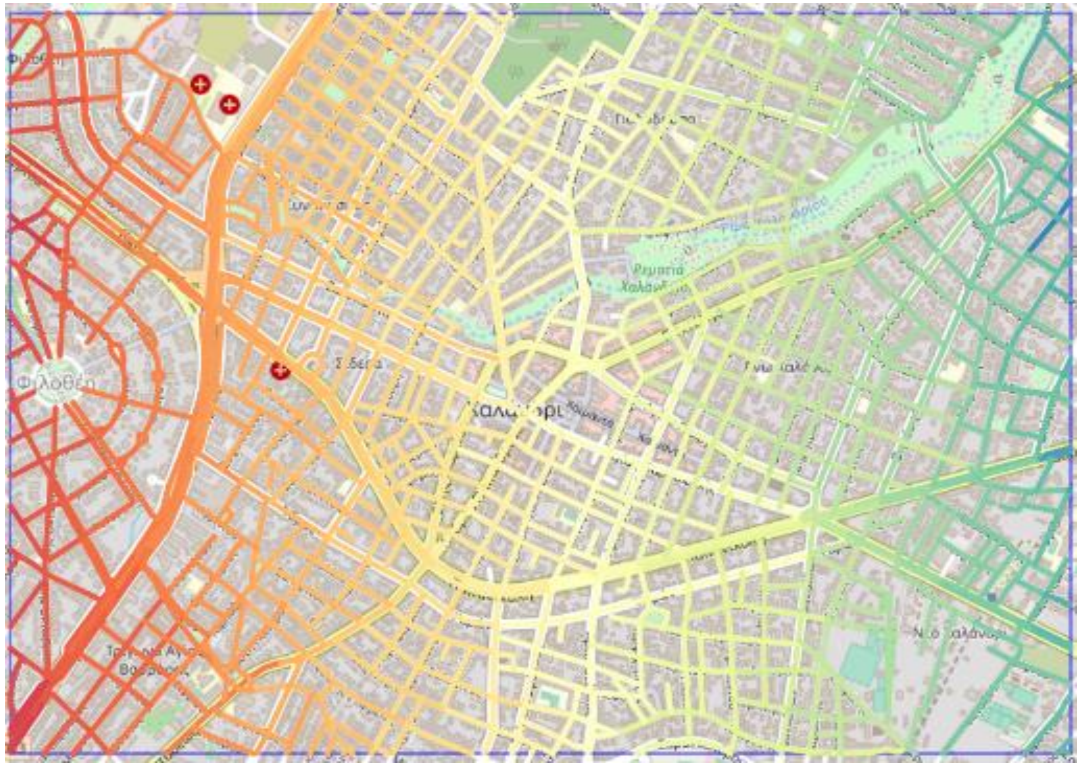
1. It is **possible to combine** high resolution **multi-parametric** naturalistic driving and geometric data that can be exploited to conduct meaningful spatial analyses on a road segment basis

2. The implementation of both **functional** spatial methods (GWPR, CAR, Moran's I and variograms) and **innovative AI-ML** methods (RCV & SPCV XGBoost) is feasible for spatial analyses of harsh braking frequencies on a road segment basis

3. **Precise predictions (87.6% accuracy)** of harsh braking frequencies can be **successfully conducted**. Several correlations were obtained.

4. Using **combined average**, spatial models **mitigate** their weaknesses and lead to a **balanced** predictive outcome for harsh brakings.

5. The analyses were mirrored for harsh accelerations (89% acc). A **more complete** image of harsh event hotspots is obtained.



Wider findings

1. **AI and ML Algorithms** can be easily and accurately transferable on different types of urban networks within a city

2. Highly useful diagnostic tools, like hotspot and critical segment **heatmaps** are created.

3. Smart and scientific evidence-based decision making of **Authorities** for road improvement, traffic management and good behaviour enforcement with great safety benefits

4. Targeted information and **feedback** (heatmaps) to the driver for significant behavioural change.

Future tasks – extension to industrial practices

1. **Correlation with crash data**
   Conducting spatial analyses including crash data per road segment – examination of possible hotspot overlap

2. **Investigation of further aspects**
   Temporal dimension, additional spatial/ML models, additional road environments, driver aggressiveness categories

3. **Creation of a seamless and constantly updating system**
   From smartphone data collection to heatmap rendition on a recurring basis using integrated AI algorithms

4. **Expanding benefits for road users and authorities**
   Road safety hotspot identification before crashes occur – Added information for pedestrians, professional drivers, mobility-impaired individuals

5. **Additional maps can be created for any indicator**
   E.g.: speeding, mobile phone use, emissions etc.

Using AI for spatial predictions of driver behavior

International Transport Forum

International Transport Forum
Roundtable on Artificial Intelligence in
Road Traffic Crash Prevention, 10-12 February 2021

Apostolos Ziakopoulos
& George Yannis

Department of Transportation
Planning and Engineering,
National Technical University of Athens

www.nrso.ntua.gr/apziak

apziak@central.ntua.gr