Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data

Apostolos Ziakopoulos
Civil – Transportation Engineer
PhD Candidate – Researcher
National Technical University of Athens
www.nrso.ntua.gr/apziak
apziak@central.ntua.gr
July 2020
Spatial analysis of harsh event frequencies (harsh brakings/accelerations) in road segments

Exploitation of multi-parametric high-resolution data:

1. Road segment geometric and road network characteristic data from digital maps

2. Naturalistic driving data from smartphone sensors

3. High resolution traffic data

Scope of the dissertation

Apostolos Ziakopoulos | Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data
Thorough review of 132 international scientific studies of spatial analysis applications in road safety

Available methodologies:
1. Geographically Weighted Regression (GWR)
2. Bayesian Conditional Autoregression (CAR)
3. Full/Empirical Bayesian Analyses
4. Machine learning approaches
5. Kernel density approaches etc.

Wide array of parameters related to:
1. Road traffic (speed, traffic volume, vehicle-kilometers)
2. Road environment (gradient, curvature, lane number/width, intersection number/density etc.)
3. Demographic characteristics (population, road user age)
4. Socio-economic characteristics (income, employment)
5. Land use (commercial, industrial, residential)

Several available unit scales for spatial analysis (road segment, TAZ, region, grid structures)
The majority of studies analyze crash frequency specially with count-data models (GWPR/CAR Poisson)

Additional issues:
1. **Boundary** problem
2. **Modifiable areal unit** problem
3. Lack of common working **framework**
4. Most research done in **modernized countries**
5. Harder examination of certain **parameters** due to lack of data or means of calculation (e.g. geometric characteristics)

All variables – parameters are examined and analyzed on a **spatial unit basis** (AADT/zone, average speed/road section)

Methodological advantages and disadvantages:
1. **Frequentist models** (e.g. GWPR): Intuitive interpretation, reduced fit capabilities
2. **Bayesian models** (e.g. CAR): Wide applications & adaptation to new data trends, lack of informative priors for initialization
3. **Machine learning** (e.g. SVM/CNN): Flexibility & handling of big data, harder interpretation – occasional ‘black box’ effect

Apostolos Ziakopoulos | Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data
Spatial analysis objectives are dictated by data availability:
• No research was found in urban road networks due to lack of data

Dependent variables:
• Limited analyses regarding crash injury severity
• No research pertinent with spatial analysis of harsh events was found

Despite precise hotspot location capabilities, there is a lack of transferability of spatial analysis results:
• No predictions are conducted for different study areas

Large margins for exploitation of new technological advancements for spatial analyses:
• Enhancement of existing data – production of new datasets
Parameters of exposure to danger

- Serve for the creation of a **common baseline** between models and results
- Most prevalent parameters: roadway length, vehicle-miles/kms, AADT

Meta-regressions: Original research

- **Quantitative investigation** of factors which systematically influence exposure parameters
- A means of investigating **heterogeneity** of scientific study results
- Conducted with the **inverse variance technique**

Results for road safety spatial analyses

1. AADT coefficients are positively correlated with taking **speed limit and road user age** into consideration
2. Roadway length coefficients are positively correlated with analyzing **only fatal crashes** compared to total crashes
3. AADT coefficients are positively correlated with analyzing crashes on a **county level** compared to TAZ level
## Literature review: Driver recording tools

A significant number of **driver recording tools** was identified

<table>
<thead>
<tr>
<th>Recording tools</th>
<th>Main advantages</th>
<th>Main disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surveys on opinion and stated behaviour</strong></td>
<td>low cost, flexibility</td>
<td>hypothetical questions, lower detail/reliability, biased data</td>
</tr>
<tr>
<td><strong>Past police or hospital record investigation</strong></td>
<td>low cost, official records</td>
<td>missing data/variables, underreporting, maintenance requirements, time delays</td>
</tr>
<tr>
<td><strong>Direct observer method</strong></td>
<td>observer specialization, removal of intermediaries</td>
<td>high person-hours, lack of randomization, observer bias</td>
</tr>
<tr>
<td><strong>Driving simulator</strong></td>
<td>safe environment, greater experimental control, precise reaction time</td>
<td>learning effects, nausea, high costs, maintenance requirements</td>
</tr>
<tr>
<td><strong>Naturalistic driving - Vehicle instrumentation (&amp; On-road driving)</strong></td>
<td>examination of real traffic conditions and events, uses in driver training – evaluation, interdisciplinary extensions</td>
<td>rare traffic incidents, high costs, driver screening requirements, time-consuming process</td>
</tr>
<tr>
<td><strong>OBD/IVDRs</strong></td>
<td>Real-time recording, accurate indications of crash involvement probability</td>
<td>unclear sampling frame</td>
</tr>
<tr>
<td><strong>In-depth incident investigation</strong></td>
<td>identification and reconstruction of crash factors, research of injury prevention</td>
<td>insufficient reconstruction evidence, long analysis time, demanding data analysis</td>
</tr>
<tr>
<td><strong>Smartphone data exploitation</strong></td>
<td>seamless and rapid data recording and storage, programmable means, increased flexibility, phone distraction measurements</td>
<td>demanding in data storage/analysis, upfront costs during development, lower cost during collection</td>
</tr>
</tbody>
</table>

Apostolos Ziakopoulos | Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data
Literature review: Harsh events

Harsh events: harsh brakings and harsh accelerations occurring during naturalistic driving

- Parameters measuring road safety levels (correlations with spatial and temporal headways)
- Different phenomena, correlations with different variables
- Correlation with driver risk

Considerable comparative advantages for investigation:

1. Applications in driver evaluation and classification

2. Proactive road safety indicators – evaluations before crashes occur

3. Considerable research gaps regarding the investigation of harsh event frequencies
Research questions

1. How can smartphone data and map data be combined (map-matched) and examined in road safety investigations?

2. How can harsh event frequencies be analyzed spatially in these environments, and which methods are appropriate for that purpose?

3. Is there spatial autocorrelation present in harsh event frequencies for road segments in urban road environments?

4. Which road geometry and road network characteristics affect harsh event frequencies in urban road network environments? How transferable are the previous results in a different study area?

5. Do traffic and driver behavioural parameters have a statistical impact on harsh event frequencies?
Methodological approach

Spatial analysis of harsh event frequencies along two pillars:

1. **Urban road networks**
   Predictive modelling – Measurement of result transferability

2. **Urban arterials**
   Explanatory modelling – Examination of additional traffic and driver behaviour characteristics (without result transferability)

**Exploratory** spatial analyses (Moran’s I coefficients and variograms)

Selection of **four spatial analysis** methods
1. Geographically Weighted Poisson Regression (GWPR)
2. Conditional Autoregressive Prior Regression (CAR)
3. Extreme Gradient Boosting (**XGBoost**) with Random Cross-Validation (**RCV**)
4. Extreme Gradient Boosting (**XGBoost**) with Spatial Cross-Validation (**SPCV**)

**Model averaging** for harsh event predictions in urban road networks
Theory of exploratory spatial analyses

“Everything is related to everything else, but near things are more related than distant things.”
– Tobler, 1970

1. Moran’s I
   • Measurement of spatial autocorrelation
   • Can be calculated globally [-1, 1] and locally (∈ R)
   • Several available geographical weighting options

2. Variograms
   • Measurement of distance of spatial autocorrelation
   • Graphs of (semi)variance of the measured quantity by distance
   • Theoretical (mathematical curves) and Empirical (real observation cluster points)
   • Additional information: geographical anisotropy and cyclicity
Theory of spatial statistical models

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)
   • 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)
Evaluation of model performance

Model performance metrics:
Difference between true and predicted values

1. Root Mean Squared Error (RMSE)
2. Mean Absolute Error (MAE)
3. Root Mean Squared Log Error (RMSLE)

Additional indicator:
4. Custom accuracy (CA – percentage):
   Percentage of accurate predictions within a ± 1 margin over total number of predictions

Performance metric selection is also affected by input and output data

Frequencies: Natural numbers (positive integer or zero values)
Data collection (1/3): 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**
- Majority of **populated areas** available

Source: SRTM website, (2020)
Naturalistic driving data from real-world conditions obtained from smartphones (per trip-second)

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)

Obtained high resolution big data from driver trips including behaviour indicators
Data collection (3/3): Traffic data

Traffic data in urban arterials provided by the Traffic Management Centre of Attica Region

Instrumentation in urban arterial corridors in Attica
• 550 inductive loop detectors
• 217 computer vision traffic cameras
• 24 variable message signs (VMS)

Regulation of ~ 1500 traffic lights in 850 intersections

Through vehicle time occupancy as a percentage, the TMC collects:
• Occupancy [% of time]
• Traffic volume [vehicle number / temporal unit]

Secondarily, traffic speed [km/h] is calculated as well

Several temporal resolutions for data: 1 h, 5 min or 90 s (high resolution)

Source: TMC, (2020)
Study areas

1. **Urban road networks (URNs)**
   (i) Chalandri (spatial model calibration for URNs)
   (ii) Omonoia (accuracy evaluation – transferability assessment for URNs)

Road geometry and naturalistic driving high resolution data collection

2. **Urban arterials (UA)**
   Kifisias Avenue (increased modelling depth with added characteristics in UA)

Road geometry, naturalistic driving and traffic high resolution data collection
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)

First trip-second on this MDW:
Number of total passes: +1
Number of total seconds driven: +1
If ha/hb: total ha/hb number: +1
If speeding: total speeding seconds: +1
If using mobile: total mobile seconds: -1

Second trip-second on this MDW:
Number of total passes: +0
Number of total seconds driven: -1
If ha/hb: total ha/hb number: +1
If speeding: total speeding seconds: +1
If using mobile: total mobile seconds: -1
Data processing: Traffic parameter integration

Theory of three traffic states (indicatively Kerner, 2012)
1. Free flow
2. Synchronized flow
3. Congested flow

Matching of naturalistic driving data with traffic data spatio-temporally (closest measurement)

Classification of each trip-second per traffic state based on traffic data and on determined limits (Vlahogianni et al., 2008)

Map-matching of trips and maps: Creation of spatial data per traffic flow state for each road segment

Source: Vlahogianni et al. (2008)
Urban road networks: Sample description (1/2) – Chalandri

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)
Urban road networks: Sample description (2/2) – Omonoia

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)
Urban road networks: Exploratory spatial analyses (1/2)

**Global and local** Moran’s I coefficients (Chalandri area)
1. Distance-based weighting (DB)
2. k nearest-neighbor weighting (kNN)

Interpretation of k nearest-neighbors is more **reasonable:**
- Harsh event frequencies are influenced from the **more proximal** road environment
- **Positive spatial autocorrelation** manifests in harsh event frequencies

**Very few** outlier values appear for local Moran’s I (within $2\sigma$ per Anselin, 1995)

**Volatility of the coefficient:** appropriate for preliminary – exploratory analysis

<table>
<thead>
<tr>
<th>kNN Global Moran's I</th>
<th>Correlation threshold</th>
<th>k</th>
<th>Coefficient value</th>
<th>Expectation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harsh brakings</td>
<td>0.0</td>
<td>15</td>
<td>0.0806</td>
<td>-0.0012</td>
<td>0.000</td>
</tr>
<tr>
<td>Harsh accelerations</td>
<td></td>
<td>39</td>
<td>0.0945</td>
<td>-0.0012</td>
<td>0.000</td>
</tr>
<tr>
<td>Harsh brakings</td>
<td>0.1</td>
<td>5</td>
<td>0.1421</td>
<td>-0.0012</td>
<td>0.000</td>
</tr>
<tr>
<td>Harsh accelerations</td>
<td></td>
<td>5</td>
<td>0.2206</td>
<td>-0.0012</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Variograms of semivariance per directional axis [N-S, W-E]

Spherical theoretical variograms describe harsh event frequencies per road segment with a better fit

Spatial autocorrelation manifests mainly:
- Within 190 m from road segment centroids for harsh brakings
- Within 200 m from road segment centroids for harsh accelerations

In large theoretical road segment samples, harsh events are expected to have:
- Mean values of 4.83 and majority within [0.00, 9.65] for harsh brakings
- Mean values of 3.00 and majority within [0.00, 6.00] for harsh accelerations

- Geographical anisotropy along the N-S axis as opposed to the W-E axis
- Partial geographical cyclicity (wave patterns) along the N-S axis
Urban road networks: Harsh braking spatial analyses

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]

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Coefficients</th>
<th>CAR Mean posterior values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.4636</td>
<td>-1.4134</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>-2.4864</td>
<td>-9.7538</td>
<td>0.0806</td>
<td>0.0860</td>
</tr>
<tr>
<td>Curvature</td>
<td>—</td>
<td>—</td>
<td>0.0444</td>
<td>0.0626</td>
</tr>
<tr>
<td>Neighborhood complexity</td>
<td>-0.2919</td>
<td>-0.1787</td>
<td>0.0344</td>
<td>0.0684</td>
</tr>
<tr>
<td>Segment length</td>
<td>0.0039</td>
<td>0.0075</td>
<td>0.1436</td>
<td>0.1400</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0040</td>
<td>0.0086</td>
<td>0.6788</td>
<td>0.6271</td>
</tr>
<tr>
<td>Traffic lights: Yes [Ref.: Traffic lights: No]</td>
<td>0.2563</td>
<td>-0.0902</td>
<td>0.0037</td>
<td>0.0010</td>
</tr>
<tr>
<td>Pedestrian crossing: Yes [Ref.: Pedestrian crossing: No]</td>
<td>-0.1463</td>
<td>0.3820</td>
<td>0.0024</td>
<td>0.0024</td>
</tr>
<tr>
<td>Lanes: 2 [Ref.: Lanes: 1]</td>
<td>-0.2435</td>
<td>-0.1713</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lanes: 3 [Ref.: Lanes: 1]</td>
<td>0.3669</td>
<td>-0.5719</td>
<td>0.0072</td>
<td>0.0048</td>
</tr>
<tr>
<td>Lanes: 4 [Ref.: Lanes: 1]</td>
<td>0.3578</td>
<td>1.9169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road type: secondary [Ref.: Road type: primary]</td>
<td>1.0520</td>
<td>-0.1094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road type: tertiary [Ref.: Road type: primary]</td>
<td>-0.0070</td>
<td>-1.6389</td>
<td>0.0049</td>
<td>0.0078</td>
</tr>
<tr>
<td>Road type: residential [Ref.: Road type: primary]</td>
<td>-1.0084</td>
<td>-2.5578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigma-phi² [Spatially structured effects]</td>
<td>N/A</td>
<td>700.3172</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigma-theta² [Spatially unstructured effects]</td>
<td>N/A</td>
<td>2.3455</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Performance metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>RMSE</th>
<th>MAE</th>
<th>RMSLE</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.2954</td>
<td>1.3048</td>
<td>0.5569</td>
<td>80.90%</td>
</tr>
<tr>
<td></td>
<td>1.2830</td>
<td>0.4115</td>
<td>0.1727</td>
<td>96.32%</td>
</tr>
<tr>
<td></td>
<td>1.4215</td>
<td>0.4971</td>
<td>0.3140</td>
<td>90.56%</td>
</tr>
<tr>
<td></td>
<td>1.8293</td>
<td>0.4994</td>
<td>0.2390</td>
<td>91.71%</td>
</tr>
</tbody>
</table>

Apostolos Ziakopoulos | Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data
Urban road networks: Harsh braking prediction & transferability

Predictions using Omonoia spatial data

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

<table>
<thead>
<tr>
<th>Performance metrics</th>
<th>GWPR global Poisson</th>
<th>Bayesian Poisson</th>
<th>RCV XGBoost</th>
<th>SPCV XGBoost</th>
<th>Combined Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.9792</td>
<td>1.9804</td>
<td>1.9834</td>
<td>1.8418</td>
<td>1.6114</td>
</tr>
<tr>
<td>MAE</td>
<td>1.0265</td>
<td>1.0290</td>
<td>0.8415</td>
<td>0.7542</td>
<td>0.6645</td>
</tr>
<tr>
<td>RMSLE</td>
<td>0.5508</td>
<td>0.5520</td>
<td>0.5484</td>
<td>0.5189</td>
<td>0.4514</td>
</tr>
<tr>
<td>CA</td>
<td>82.64%</td>
<td>82.74%</td>
<td>83.40%</td>
<td>85.27%</td>
<td>87.55%</td>
</tr>
</tbody>
</table>
Urban road networks: Combined harsh braking predictions

Using combined average, spatial models mitigate their weaknesses and lead to a balanced predictive outcome for harsh brakings.
Urban road networks: Harsh acceleration spatial analyses

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Coefficients</th>
<th>Mean posterior values</th>
<th>CAR Gain values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.4230</td>
<td>-1.2399</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>—</td>
<td>—</td>
<td>0.0588</td>
<td>0.0189</td>
<td>0.0189</td>
</tr>
<tr>
<td>Curvature</td>
<td>9.0471</td>
<td>6.3926</td>
<td>0.0323</td>
<td>0.0309</td>
<td>0.0309</td>
</tr>
<tr>
<td>Neighborhood complexity</td>
<td>—</td>
<td>-0.2308</td>
<td>0.0532</td>
<td>0.0355</td>
<td>0.0355</td>
</tr>
<tr>
<td>Segment length</td>
<td>0.0030</td>
<td>0.0038</td>
<td>0.1058</td>
<td>0.0766</td>
<td>0.0766</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0042</td>
<td>0.0071</td>
<td>0.7184</td>
<td>0.8253</td>
<td>0.8253</td>
</tr>
<tr>
<td>Traffic lights: Yes</td>
<td>—</td>
<td>—</td>
<td>0.3791</td>
<td>0.0026</td>
<td>0.0026</td>
</tr>
<tr>
<td>Pedestrian crossing: Yes</td>
<td>—</td>
<td>—</td>
<td>0.4554</td>
<td>0.0011</td>
<td>0.0011</td>
</tr>
<tr>
<td>Lanes: 2 [Ref.: Lanes: 1]</td>
<td>0.0794</td>
<td>-0.0134</td>
<td>0.0033</td>
<td>0.0027</td>
<td>0.0027</td>
</tr>
<tr>
<td>Lanes: 3 [Ref.: Lanes: 1]</td>
<td>0.4741</td>
<td>-0.1702</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lanes: 4 [Ref.: Lanes: 1]</td>
<td>0.3828</td>
<td>0.4380</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Road type: secondary [Ref.: Road type: primary]</td>
<td>0.7323</td>
<td>0.7202</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Road type: tertiary [Ref.: Road type: primary]</td>
<td>0.3720</td>
<td>0.3610</td>
<td>0.0109</td>
<td>0.0065</td>
<td></td>
</tr>
<tr>
<td>Road type: residential [Ref.: Road type: primary]</td>
<td>-0.6642</td>
<td>-0.6715</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sigma-phi² [Spatially structured effects]</td>
<td>N/A</td>
<td>255.3276</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sigma-theta² [Spatially unstructured effects]</td>
<td>N/A</td>
<td>0.2827</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

Performance metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>GWPR</th>
<th>CAR</th>
<th>RCV XGBoost</th>
<th>SPCV XGBoost</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.0861</td>
<td>0.7961</td>
<td>0.9128</td>
<td>1.1327</td>
</tr>
<tr>
<td>MAE</td>
<td>0.9125</td>
<td>0.4111</td>
<td>0.3728</td>
<td>0.4891</td>
</tr>
<tr>
<td>RMSLE</td>
<td>0.4704</td>
<td>0.2512</td>
<td>0.3000</td>
<td>0.3504</td>
</tr>
<tr>
<td>CA</td>
<td>84.69%</td>
<td>95.74%</td>
<td>93.32%</td>
<td>89.87%</td>
</tr>
</tbody>
</table>

Positive correlation:
- Segment length
- Pass count
- Curvature
- Road type [Secondary]
- Road type [Tertiary]
- Traffic lights

Negative correlation:
- Road type [Residential]

Marginally positive correlation:
- Pedestrian crossing

Marginally negative correlation:
- Neighborhood complexity

Apostolos Ziakopoulos | Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data
Urban road networks: Combined harsh acceleration predictions

<table>
<thead>
<tr>
<th>Performance metrics</th>
<th>GWPR global Poisson</th>
<th>Bayesian Poisson</th>
<th>RCV XGBoost</th>
<th>SPCV XGBoost</th>
<th>Combined Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.6836</td>
<td>1.6841</td>
<td>1.9834</td>
<td>1.6250</td>
<td>1.5010</td>
</tr>
<tr>
<td>MAE</td>
<td>0.8721</td>
<td>0.8700</td>
<td>0.8415</td>
<td>0.7064</td>
<td>0.6903</td>
</tr>
<tr>
<td>RMSLE</td>
<td>0.5082</td>
<td>0.5071</td>
<td>0.5484</td>
<td>0.4791</td>
<td>0.4316</td>
</tr>
<tr>
<td>CA</td>
<td>87.71%</td>
<td>87.62%</td>
<td>83.40%</td>
<td>87.42%</td>
<td>89.09%</td>
</tr>
</tbody>
</table>
Urban road networks: Main findings

- **Spatial analyses** of harsh braking and harsh acceleration frequencies are feasible using GWPR, CAR, RCV XGBoost and SPCV XGBoost methodologies.

- **Very good model fit** on training spatial data (Chalandri) and precise predictions in the test spatial data (Omonoia).

- The investigated **exposure parameters** (segment length and pass count) are consistently **positively correlated** with harsh braking and harsh acceleration frequencies.

- The presence of pedestrian crossings and traffic lights is **mostly positively correlated** with harsh braking and harsh acceleration frequencies.

- Conversely, **gradient** and **neighborhood complexity** are negatively correlated with harsh acceleration frequency.

- **Curvature** is positively correlated with harsh acceleration frequency.

- **Road type** and **number of lanes** are parameters with unclear influence: However, they offer useful information for the calibration of spatial terms/tree separation limits.
152 road segments with 658 nodes
- 15 road segments with traffic lights
- 21 road segments with pedestrian crossings

Naturalistic driving data:
- Trips between 01-09-2019 & 29-11-2019 – 3 months
- A total of 8756 trips from 314 drivers
- 930,346 driving seconds: average trip duration 221 s
- 1543 harsh brakings
- 1033 harsh accelerations

Variable distributions
- Positive skewness (larger right tails)
- High kurtosis (segment length/neighborhood complexity) and low kurtosis (gradient/curvature) – non-normal distributions
Classification of trip-seconds and traffic data per traffic state

At least 1 trip:
• 100% of road segments under free flow conditions
• 94.74% of road segments under synchronized flow conditions
• 95.39% of road segments under congested flow conditions

Free flow conditions
• 563 harsh brakings & 363 harsh accelerations

Synchronized flow conditions
• 215 harsh brakings & 142 harsh accelerations

Congested flow conditions
• 10 harsh brakings & 4 harsh accelerations
• No sufficient harsh event frequencies to conduct spatial analysis
Urban arterial: Exploratory spatial analyses (1/2)

**Global and local** Moran’s I coefficients (Chalandri area)
1. Distance-based weighting (DB)
2. k nearest-neighbor weighting (kNN)

Interpretation of k nearest-neighbors is more **reasonable:**
- **Rapid reduction** of spatial autocorrelation between segments (fewer neighboring segments)
- **Positive spatial autocorrelation** manifests in harsh event frequencies

**Very few** outlier values appear for local Moran’s I (within 2σ per Anselin, 1995)

**Volatility of the coefficient:** appropriate for preliminary – exploratory analysis

<table>
<thead>
<tr>
<th>Harsh brakings</th>
<th>Correlation threshold</th>
<th>k</th>
<th>Coefficient value</th>
<th>Expectation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
<td>5</td>
<td>0.0913</td>
<td>-0.0066</td>
<td>0.0389</td>
</tr>
<tr>
<td>Harsh accelerations</td>
<td></td>
<td>9</td>
<td>0.1261</td>
<td>-0.0066</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
Merged variograms of semivariance

Exponential theoretical variograms describe harsh event frequencies per road segment with a better fit

Spatial autocorrelation manifests mainly:
• Within 310 m from road segment centroids for harsh brakings
• Within 320 m from road segment centroids for harsh accelerations

In large theoretical road segment samples, harsh events are expected to have:
• Mean values of 9.89 and majority within [0.00, 19.78] for harsh brakings
• Mean values of 7.88 and majority within [0.00, 15.75] for harsh accelerations

Additional observations:
• Greater empirical variogram volatility compared to urban road networks
• Partial geographical cyclicity (wave patterns), denoting patterns of repetition in the data
### Positive correlation:
- Segment length
- Pass count
- Mobile phone use seconds
- Speed difference (driver – traffic)

### Marginally positive correlation:
- Average occupancy

### Marginally negative correlation:
- Bearing [Southbound]
- Standardized traffic flow

### Independent variables

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Mean posterior values</th>
<th>CAR Gain values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.2544</td>
<td>-0.4664</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>-1.1013</td>
<td>-</td>
<td>0.0642</td>
<td>0.0408</td>
</tr>
<tr>
<td>Curvature</td>
<td>-</td>
<td>-</td>
<td>0.0208</td>
<td>0.0183</td>
</tr>
<tr>
<td>Segment length</td>
<td>0.0033</td>
<td>0.0031</td>
<td>0.0454</td>
<td>0.0572</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0023</td>
<td>0.0027</td>
<td>0.0577</td>
<td>0.0364</td>
</tr>
<tr>
<td>Speeding seconds</td>
<td>-</td>
<td>-</td>
<td>0.1374</td>
<td>0.0330</td>
</tr>
<tr>
<td>Mobile use seconds</td>
<td>0.0022</td>
<td>0.0042</td>
<td>0.4583</td>
<td>0.5496</td>
</tr>
<tr>
<td>Speed difference</td>
<td>0.0385</td>
<td>0.0318</td>
<td>0.0755</td>
<td>0.0548</td>
</tr>
<tr>
<td>Average std. current traffic volume</td>
<td>-0.1640</td>
<td>-0.0417</td>
<td>0.0328</td>
<td>0.0469</td>
</tr>
<tr>
<td>Average occupancy</td>
<td>0.0595</td>
<td>-</td>
<td>0.0370</td>
<td>0.0310</td>
</tr>
<tr>
<td>Average driver speed</td>
<td>-</td>
<td>-</td>
<td>0.0387</td>
<td>0.0687</td>
</tr>
<tr>
<td>Average traffic speed</td>
<td>-</td>
<td>-</td>
<td>0.0115</td>
<td>0.0493</td>
</tr>
<tr>
<td>Lane number</td>
<td>-</td>
<td>-</td>
<td>0.0013</td>
<td>0.0027</td>
</tr>
<tr>
<td>Bearing: Southbound</td>
<td>-0.2611</td>
<td>-0.2746</td>
<td>0.0195</td>
<td>0.0067</td>
</tr>
<tr>
<td>Pedestrian crossing: Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0025</td>
</tr>
<tr>
<td>Traffic lights: Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sigma-phi$^2$ [Spatially structured effects]</td>
<td>N/A</td>
<td>0.0662</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigma-theta$^2$ [Spatially unstructured effects]</td>
<td>N/A</td>
<td>0.3796</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Performance metrics

<table>
<thead>
<tr>
<th>Performance metrics</th>
<th>GWPR</th>
<th>CAR</th>
<th>RCV XGBoost</th>
<th>SPCV XGBoost</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.8905</td>
<td>1.1052</td>
<td>0.4730</td>
<td>0.4730</td>
</tr>
<tr>
<td>MAE</td>
<td>2.0705</td>
<td>0.9002</td>
<td>0.1579</td>
<td>0.1316</td>
</tr>
<tr>
<td>RMSLE</td>
<td>0.6046</td>
<td>0.3565</td>
<td>0.0579</td>
<td>0.2105</td>
</tr>
<tr>
<td>CA</td>
<td>56.58%</td>
<td>84.22%</td>
<td>98.03%</td>
<td>99.34%</td>
</tr>
</tbody>
</table>
Urban arterial: Harsh braking analyses under synchronized flow

Positive correlation:
- Segment length
- Pass count
- Mobile phone use seconds
- Average occupancy

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Coefficients</th>
<th>Mean posterior values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.1012</td>
<td>-2.4520</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>-</td>
<td>1.0782</td>
<td>0.0491</td>
<td>0.0638</td>
</tr>
<tr>
<td>Curvature</td>
<td>71.643</td>
<td>6.6068</td>
<td>0.0084</td>
<td>0.0269</td>
</tr>
<tr>
<td>Segment length</td>
<td>0.0024</td>
<td>0.0019</td>
<td>0.0282</td>
<td>0.0533</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0059</td>
<td>0.0057</td>
<td>0.0833</td>
<td>0.4146</td>
</tr>
<tr>
<td>Speeding seconds</td>
<td>-</td>
<td>0.0020</td>
<td>0.0810</td>
<td>0.0781</td>
</tr>
<tr>
<td>Mobile use seconds</td>
<td>0.0113</td>
<td>0.0134</td>
<td>0.6446</td>
<td>0.2250</td>
</tr>
<tr>
<td>Speed difference</td>
<td>-</td>
<td>-</td>
<td>0.0167</td>
<td>0.0231</td>
</tr>
<tr>
<td>Average std. current traffic volume</td>
<td>-</td>
<td>-0.0195</td>
<td>0.0298</td>
<td>0.0266</td>
</tr>
<tr>
<td>Average hourly traffic volume</td>
<td>-0.0002</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average occupancy</td>
<td>0.0495</td>
<td>0.0371</td>
<td>0.0216</td>
<td>0.0172</td>
</tr>
<tr>
<td>Average driver speed</td>
<td>-</td>
<td>-</td>
<td>0.0219</td>
<td>0.0610</td>
</tr>
<tr>
<td>Average traffic speed</td>
<td>-</td>
<td>-</td>
<td>0.0092</td>
<td>-</td>
</tr>
<tr>
<td>Lanes: 2 [Ref.: Lanes: 1]</td>
<td>-</td>
<td>0.2878</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lanes: 3 [Ref.: Lanes: 1]</td>
<td>-</td>
<td>-0.0207</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lanes: 4 [Ref.: Lanes: 1]</td>
<td>-</td>
<td>-1.9839</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bearing: Southbound [Ref.: Northbound]</td>
<td>-</td>
<td>0.0119</td>
<td>0.0060</td>
<td>0.0095</td>
</tr>
<tr>
<td>Pedestrian crossing: Yes [Ref.: No]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0006</td>
</tr>
<tr>
<td>Traffic lights: Yes [Ref.: No]</td>
<td>-</td>
<td>-</td>
<td>0.0000</td>
<td>0.0002</td>
</tr>
<tr>
<td>Sigma-phi² [Spatially structured effects]</td>
<td>N/A</td>
<td>0.0309</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigma-theta² [Spatially unstructured effects]</td>
<td>N/A</td>
<td>0.3916</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Performance metrics:
- RMSE: 1.6733, 0.7472, 0.2433, 0.3441
- MAE: 0.9404, 0.5206, 0.0461, 0.0472
- RMSLE: 0.4306, 0.2971, 0.0268, 0.0921
- CA: 83.55%, 90.79%, 99.34%, 98.68%
Urban arterial: Harsh acceleration analyses under free flow

Positive correlation:
Pass count
Mobile phone use seconds
Speed difference (driver – traffic)
Average occupancy

Negative correlation:
Speeding seconds

Marginally positive correlation:
Segment length

Marginally negative correlation:
Bearing [Southbound]

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Coefficients</th>
<th>CAR Mean posterior values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.2237</td>
<td>-1.0912</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>-</td>
<td>1.2874</td>
<td>0.0573</td>
<td>0.0756</td>
</tr>
<tr>
<td>Curvature</td>
<td>-</td>
<td>-</td>
<td>0.0337</td>
<td>0.0413</td>
</tr>
<tr>
<td>Segment length</td>
<td>0.0017</td>
<td>0.0011</td>
<td>0.0831</td>
<td>0.0759</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0032</td>
<td>0.0022</td>
<td>0.1097</td>
<td>0.1183</td>
</tr>
<tr>
<td>Speeding seconds</td>
<td>-0.0011</td>
<td>-0.0012</td>
<td>0.0711</td>
<td>0.0631</td>
</tr>
<tr>
<td>Mobile use seconds</td>
<td>0.0027</td>
<td>0.0047</td>
<td>0.2865</td>
<td>0.2212</td>
</tr>
<tr>
<td>Speed difference</td>
<td>0.0528</td>
<td>0.0323</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average std. current traffic volume</td>
<td>-</td>
<td>-</td>
<td>0.0795</td>
<td>0.1281</td>
</tr>
<tr>
<td>Average occupancy</td>
<td>0.0258</td>
<td>0.0328</td>
<td>0.0500</td>
<td>0.0656</td>
</tr>
<tr>
<td>Average driver speed</td>
<td>-</td>
<td>-</td>
<td>0.1457</td>
<td>0.1136</td>
</tr>
<tr>
<td>Average traffic speed</td>
<td>-0.0240</td>
<td>-</td>
<td>0.0585</td>
<td>-</td>
</tr>
<tr>
<td>Lane number</td>
<td>-</td>
<td>-</td>
<td>0.0148</td>
<td>0.0092</td>
</tr>
<tr>
<td>Bearing: Southbound</td>
<td>-0.2434</td>
<td>-0.2327</td>
<td>0.0001</td>
<td>0.0021</td>
</tr>
<tr>
<td>Pedestrian crossing: Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Traffic lights: Yes</td>
<td>-</td>
<td>-</td>
<td>0.0100</td>
<td>0.0058</td>
</tr>
<tr>
<td>Sigma-phi² [Spatially structured effects]</td>
<td>N/A</td>
<td>0.5614</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigma-theta² [Spatially unstructured effects]</td>
<td>N/A</td>
<td>0.3253</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Performance metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>RMSE</th>
<th>MAE</th>
<th>RMSLE</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.2817</td>
<td>1.0912</td>
<td>0.3974</td>
<td>63.16%</td>
</tr>
<tr>
<td></td>
<td>1.5816</td>
<td>0.8612</td>
<td>0.1316</td>
<td>96.18%</td>
</tr>
<tr>
<td></td>
<td>0.6305</td>
<td>0.4286</td>
<td>0.0776</td>
<td>98.68%</td>
</tr>
<tr>
<td></td>
<td>63.16%</td>
<td>86.18%</td>
<td>98.68%</td>
<td>99.34%</td>
</tr>
</tbody>
</table>
### Urban arterial: Harsh acceleration analyses in synchronized flow

**Positive correlation:**
- Pass count
- Mobile phone use seconds

**Marginally positive correlation:**
- Traffic flow
  (hourly or standardized)

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>GWPR Coefficients</th>
<th>CAR Mean posterior values</th>
<th>RCV XGBoost Gain values</th>
<th>SPCV XGBoost Gain values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.2573</td>
<td>-2.9399</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gradient</td>
<td>—</td>
<td>1.7464</td>
<td>0.0243</td>
<td>0.0258</td>
</tr>
<tr>
<td>Curvature</td>
<td>—</td>
<td>—</td>
<td>0.0183</td>
<td>0.0180</td>
</tr>
<tr>
<td>Segment length</td>
<td>—</td>
<td>-0.0005</td>
<td>0.0240</td>
<td>0.0259</td>
</tr>
<tr>
<td>Pass count</td>
<td>0.0035</td>
<td>0.0065</td>
<td>0.2169</td>
<td>0.5617</td>
</tr>
<tr>
<td>Speeding seconds</td>
<td>—</td>
<td>-0.0039</td>
<td>0.0017</td>
<td>0.0071</td>
</tr>
<tr>
<td>Mobile use seconds</td>
<td>0.0148</td>
<td>0.0159</td>
<td>0.5723</td>
<td>0.2209</td>
</tr>
<tr>
<td>Speed difference</td>
<td>—</td>
<td>—</td>
<td>0.0332</td>
<td>0.0144</td>
</tr>
<tr>
<td>Average std. current traffic volume</td>
<td>—</td>
<td>—</td>
<td>0.0032</td>
<td>0.022</td>
</tr>
<tr>
<td>Average traffic volume</td>
<td>0.0003</td>
<td>0.0008</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Average occupancy</td>
<td>—</td>
<td>0.0237</td>
<td>0.0457</td>
<td>0.0407</td>
</tr>
<tr>
<td>Average driver speed</td>
<td>—</td>
<td>-0.0224</td>
<td>0.0604</td>
<td>0.0568</td>
</tr>
<tr>
<td>Average traffic speed</td>
<td>-0.0240</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lanes: 2 [Ref.: Lanes: 1]</td>
<td>—</td>
<td>-1.0085</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lanes: 3 [Ref.: Lanes: 1]</td>
<td>—</td>
<td>-1.8321</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lanes: 4 [Ref.: Lanes: 1]</td>
<td>—</td>
<td>-2.8837</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bearing: Southbound [Ref.: Northbound]</td>
<td>0.4721</td>
<td>0.2981</td>
<td>—</td>
<td>0.0067</td>
</tr>
<tr>
<td>Sigma-phi^2 [Spatially structured effects]</td>
<td>N/A</td>
<td>0.0699</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigma-theta^2 [Spatially unstructured effects]</td>
<td>N/A</td>
<td>1.0935</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Performance metrics**

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th>RMSLE</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.2978</td>
<td>0.8258</td>
<td>0.4507</td>
<td>86.84%</td>
</tr>
<tr>
<td></td>
<td>0.5404</td>
<td>0.3904</td>
<td>0.2475</td>
<td>97.36%</td>
</tr>
<tr>
<td></td>
<td>0.5000</td>
<td>0.1711</td>
<td>0.1638</td>
<td>97.37%</td>
</tr>
<tr>
<td></td>
<td>0.281</td>
<td>0.0658</td>
<td>0.0842</td>
<td>98.68%</td>
</tr>
</tbody>
</table>
Urban arterial: Main findings

• **Spatial analyses** of harsh braking and harsh acceleration frequencies are feasible using GWPR, CAR, RCV XGBoost and SPCV XGBoost methodologies per traffic state as well

• **Very good model fit** on study area data, with the exception of GWPR under free flow conditions

• **GWPR** performance seems to be affected by the pronounced geographical **anisotropy** under free flow conditions – the avenue is practically 1-dimensional

• The other methodologies (**CAR, RCV XGBoost** and **SPCV XGBoost**) are not likewise affected

• Results indicate that **different traffic parameters** are correlated with harsh event frequencies per traffic state

• The **exposure parameters** (segment length and pass count) continue to display a **positive influence** on harsh braking and harsh acceleration frequencies

• The creation – examination of directional variograms does not have physical meaning: merged variograms are appropriate
Conclusions of the dissertation (1/5)

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

2. The implementation of both common spatial methods (GWPR, CAR, Moran’s I and variograms) and innovative methods (RCV & SPCV XGBoost) is feasible for spatial analyses of harsh event frequencies on a road segment basis.

3. Positive autocorrelation is detected in harsh braking and harsh acceleration frequencies in both urban road networks and urban arterials.

4. In urban road networks, spatial autocorrelation manifests mainly within 190 m from road segment centroids for harsh brakings and within 200 m from road segment centroids for harsh accelerations. The respective distances for urban arterials are estimated at 310 m and 320 m.
Urban road networks:

1. The exposure parameters (segment length and pass count) are positively correlated with harsh braking and harsh acceleration frequencies.

2. Conversely, gradient, neighborhood complexity and residential road type are parameters negatively correlated with harsh braking frequencies.

3. Curvature, traffic lights and secondary and tertiary road types are parameters positively correlated with harsh acceleration frequencies.

4. Residential road type is negatively correlated with harsh acceleration frequencies.
Conclusions of the dissertation (3/5)

Urban road networks:

5. **Precise predictions** of harsh event frequencies can be **successfully conducted** via the exploitation of the examined data and using the implemented methods. Custom accuracy achieved: **87.6%** for harsh brakings and **89.1%** for harsh accelerations.

6. Using **combined average**, spatial models **mitigate** their weaknesses and lead to a **balanced** predictive outcome for harsh events. A **more complete** image of hotspots is obtained.
Conclusions of the dissertation (4/5)

Urban arterial:

1. It is meaningful to create and examine spatial data per traffic state in order to analyze harsh event frequencies; as a bridge between road safety and traffic flow disciplines.

2. Results indicate that different variables are correlated with increased harsh event frequencies under free flow conditions compared to synchronized flow conditions.

3. The inclusion of traffic and driver behaviour parameters offers additional capabilities for in-depth examination of causal parameters, without any transferability or prediction capabilities.
Conclusions of the dissertation (5/5)

Urban arterial:

4. The exposure parameters (segment length and pass count) as well as mobile use seconds are positively correlated with harsh braking and harsh acceleration frequencies.

5. The parameter of average occupancy is positively correlated with harsh event frequencies under synchronized flow and marginally positively correlated under free flow.

6. The parameter of speed difference of driver and vehicle is positively correlated with harsh braking frequencies under free flow and with harsh acceleration frequencies (under both conditions).

7. Circumstantial correlations are found with traffic flow parameters (reduction of brakings under free flow, increase of accelerations under synchronized flow).
Innovative contributions

1. Novel methodological research framework
   Conducting road safety spatial analysis for harsh event frequencies using multi-parametric high-resolution data per road segment

2. Inception of a number of purpose-made big-data algorithms
   Implementation of the algorithms for critical functions:
   (i) calculation of additional geometric characteristics
   (ii) data processing and merging
   (iii) map-matching of trip-seconds to road segments

3. Innovative types of spatial analyses
   (i) spatial analyses of urban road networks
   (ii) spatial analyses results were used for successful predictions of harsh event frequencies

4. Spatial analyses with added depth for urban arterials – Separate examinations for free flow and synchronized flow traffic states

5. Original insights and statistical correlations were obtained for the parameters affecting harsh event frequencies
Future research

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

2. **Introduction of temporal dimension**
   Conducting spatio-temporal analyses for the identification of seasonal trends and the detection of any hotspot migration effects

3. **Analyses per driver aggressiveness**
   Driver classification based on their aggressiveness and produced harsh events

4. **Implementation of additional spatial or machine learning models**
   Indicatively: Neural networks, additional CAR priors, spatial lag models

5. **Investigation of additional environments or parameters**
   Rural roads, multiple countries, presence of public transport, low speed zones
THANK YOU
Spatial Analysis of Road Safety and Traffic Behaviour using High Resolution Multi-parametric Data

Apostolos Ziakopoulos
Civil – Transportation Engineer
PhD Candidate – Researcher
National Technical University of Athens

www.nrso.ntua.gr/apziak
apziak@central.ntua.gr

July 2020