Measurement Error Models on Spatial Network Lattices: Car Crashes in Leeds

Presenter: **Andrea Gilardi**[1]

2021-09-11

Conference: **CLADAG 2021**

Coauthor: **Riccardo Borgoni**[1]

Coauthor: **Luca Presicce**[1]

Coauthor: **Jorge Mateu**[2]

1: University of Milano - Bicocca (IT)
2: Universitat Jaume I (ES)

Società Italiana di Statistica

Introduction

- During the last decade, several authors developed car crashes models to support local authorities. They mainly focused on **areal level lattice models**.
- Recently, we observed a surge of interest in statistical analyses developed on a **network lattice**.

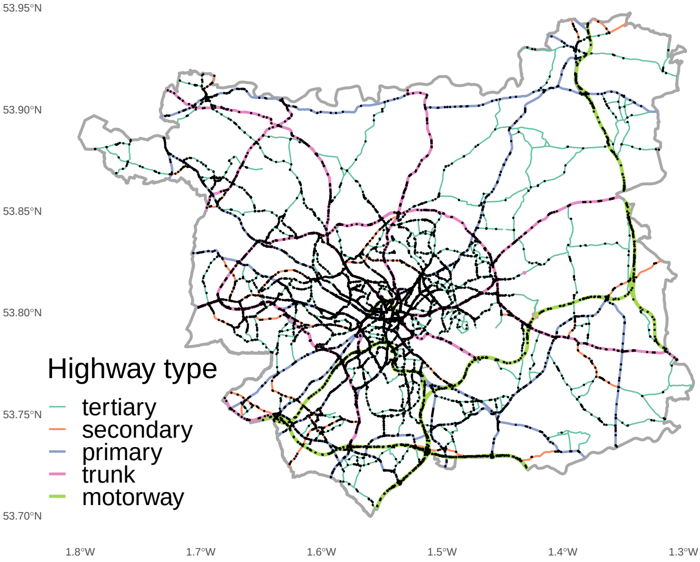
Introduction (cont)

- Several studies suggested that **traffic volumes** represent a key component for estimating road risk indices. Nevertheless, it's difficult to calculate these quantities at the network level.
- Hence, the objective of this paper is to propose a statistical model to estimate **car crashes rates** at the road network level using a **suitable proxy** for the traffic volumes.

Car crashes in Leeds

- We analysed all car crashes that occurred in the major roads of the street network of **Leeds** (UK) from 2011 to 2018. The sample included 13863 events.
- The street network was built using data downloaded from **Open Street Map** (OSM), an open-source online database of vector and raster spatial data.
- First, we filtered only the road segments classified as *motorways*, *trunks*, *primary roads*, *secondary roads*, or *tertiary roads*. Then, we applied several preprocessing steps to remove isolated clusters of road segments. The final sample included 4176 segments.

Car crashes in Leeds (cont)



Weighted edge betweenness centrality measure

The **weighted edge betweenness centrality measure** (WEBCM) is defined as:

$$C_B(e) = \sum_{s,t \in V} \frac{\varsigma(\pi_{s,t}|e)}{\varsigma(\pi_{s,t})},$$

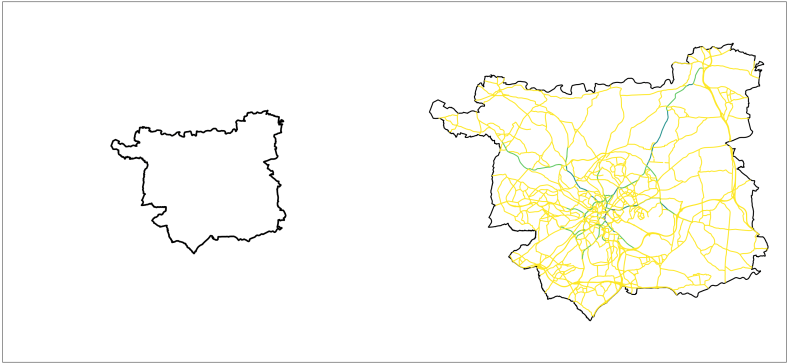
where:

- e is an edge (i.e. a segment of the street network); s, t are two nodes and V is the set of all vertices.
- $\pi_{s,t}$ represents the shortest path between s and t;
- $\varsigma(\cdot)$ computes the sum of the weights for a given path;
- $\pi_{s,t}|e$ means that the eth edge belongs to the shortest path between s and t.

Weighted edge betweenness centrality measure (cont)

- Hence, the WEBCM assigns to each edge a value that is proportional to the sum of the weights of the shortest paths going through that edge.
- The weights are given by the geographical length (in metres) of the segments times a **discount factor** proportional to the road types (to penalise local roads).
- Furthermore, the WEBCM suffers an **edge effect bias** due to the exclusion of some road segments located just outside the city boundary.

Weighted edge betweenness centrality measure (cont)



For these reasons, we estimated the edge betweenness centrality considering a buffer of 20 km.

Bayesian Hierarchical Model

- The car crashes counts were modelled using a **full Bayesian hierarchical model** estimated via INLA methodology.

- Let y_i denote the number of car crashes that occurred on the ith road segment. We assumed that

$$y_i|\lambda_i \sim \text{Poisson}(e_i\lambda_i), i = 1, \dots, 4176,$$

where λ_i represents the car crashes rate and e_i is an exposure parameter equal to the geographical length of the segments.

Bayesian Hierarchical Model (cont)

- We specified the following **log-linear model**:

$$\log \lambda_i = \beta_0 + \beta_x x_i + \boldsymbol{\beta}_{\boldsymbol{z}} \boldsymbol{z}_i + \nu_i, i = 1, \ldots, 4176.$$

  β_0 denotes the intercept, x_i is the road traffic volumes (i.e. the unobserved covariate), $\boldsymbol{z}_i$ is a set of socio-demographic regressors, and ν_i are spatially structured error terms modelled via a Conditional AutoRegressive (CAR) model.

- More precisely, we assumed that

$$\nu_i | \left\{ \nu_{i'}, i' \in \partial_i \right\}; \tau_\nu \sim N \left(m_i^{-1} \sum_{i' \in \partial_i} \nu_{i'}, \frac{1}{m_i \tau_\nu} \right)$$

  where τ_ν is the marginal precision of the RE while ∂_i and m_i denote, respectively, the indices and the cardinality of the set of neighbours for road segment i.

Bayesian Hierarchical Model (cont)

- The WEBCM represents a proxy for the road traffic and it has an intrinsic spatial error. Hence, we also defined a **measurement error model** connecting the WEBCM and the traffic values:

  $$w_i = x_i + \xi_i + u_i, i = 1, \ldots, 4176.$$

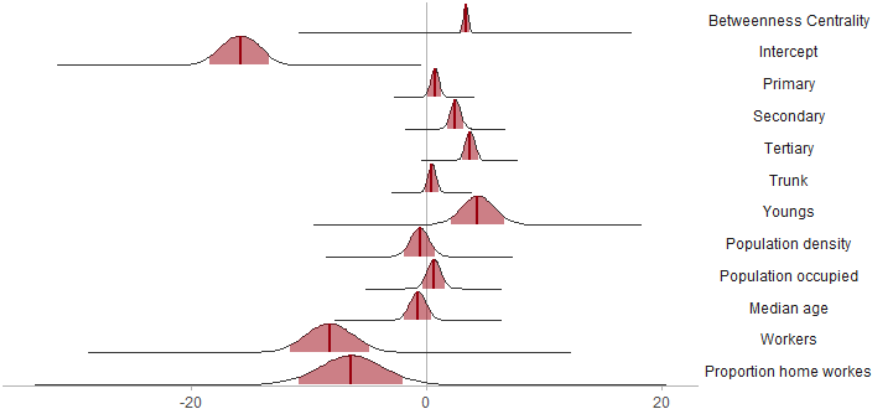
  w_i are the estimates of WEBC, ξ_i are unstructured error terms, i.e. $\xi_i \sim N(0, \tau_\xi^{-1})$, and u_i are spatial random errors, i.e. $u_i | \{u_{i'}, i' \in \partial_i\} \sim$ CAR.

- Finally, we included a third model, usually named **exposure model**, including a relationship between the unobserved covariate and the socio-economic regressors:

  $$x_i = \alpha_0 + \boldsymbol{\alpha_z z_i} + \varepsilon_i, i = 1, \ldots, 4176$$

  The term $\varepsilon_i \sim N(0, \tau_\varepsilon^{-1})$ represents an unstructured error

Results



Betweenness Centrality
Intercept
Primary
Secondary
Tertiary
Trunk
Youngs
Population density
Population occupied
Median age
Workers
Proportion home workes
-20
0
20

Results (cont)

The following table summarises the estimates of β_x, i.e. the coefficient of the error-prone covariate.

	β_x	σ_{β_x}	$\beta_{x(2.5\%)}$	$\beta_{x(97.5\%)}$
No ME	0.066	0.018	0.032	0.101
ME + no buffer	2.564	0.101	2.364	2.761
ME + buffer	3.367	0.201	2.905	3.850

Conclusions

- It seems that the ME model applied to WEBCM can successfully approximate the unknown road traffic values and represent a relevant determinant for the road risk index.
- The suggested methodology could be extended to a spatio-temporal or multivariate framework. Nevertheless, these extensions are not straightforward due to the increasingly burdensome computational costs.
- In the future, we plan to explore different specifications for the measurement error model.

References

Marc Barthélemy. “Spatial networks”. In: *Physics Reports* 499.1-3 (2011), pp. 1–101.

Karim El-Basyouny and Tarek Sayed. “Safety performance functions with measurement errors in traffic volume”. In: *Safety science* 48.10 (2010), pp. 1339–1344.

L Bernadinelli, Cristian Pascutto, NG Best, and WR Gilks. “Disease mapping with errors in covariates”. In: *Statistics in medicine* 16.7 (1997), pp. 741–752.

Julian Besag. “Spatial interaction and the statistical analysis of lattice systems”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (1974), pp. 192–225.

Julian Besag, Jeremy York, and Annie Mollié. “Bayesian image restoration, with two applications in spatial statistics”. In: *Annals of the institute of statistical mathematics* 43.1 (1991), pp. 1–20.

Riccardo Borgoni, Andrea Gilardi, and Diego Zappa. “Assessing the Risk of Car Crashes in Road Networks”. In: *Social Indicators Research* (2020), pp. 1–19.

Wayne A Fuller. *Measurement error models*. Vol. 305. John Wiley & Sons, 2009.

References ii

Andrea Gilardi, Jorge Mateu, Riccardo Borgoni, and Robin Lovelace. "Multivariate hierarchical analysis of car crashes data considering a spatial network lattice". In: *arXiv preprint arXiv:2011.12595* (2020).

Mohamed Bayoumi Kamel and Tarek Sayed. "Cyclist-vehicle crash modeling with measurement error in traffic exposure". In: *Accident Analysis & Prevention* 144 (2020), p. 105612.

Finn Lindgren and Håvard Rue. "Bayesian Spatial Modelling with R-INLA". In: *Journal of Statistical Software* 63.19 (2015), pp. 1–25. URL: http://www.jstatsoft.org/v63/i19/.

Stefanie Muff, Andrea Riebler, Leonhard Held, Håvard Rue, and Philippe Saner. "Bayesian analysis of measurement error models using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series C: Applied Statistics* (2015), pp. 231–252.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

Håvard Rue, Andrea I. Riebler, Sigrunn H. Sørbye, Janine B. Illian, Daniel P. Simpson, and Finn K. Lindgren. "Bayesian computing with INLA: A review". In: *Annual Reviews of Statistics and Its Applications* 4.March (2017), pp. 395–421. URL: http://arxiv.org/abs/1604.00860.

World Health Organization. *Global Status Report On Road Safety 2018*. en. World Health Organization, 2018. ISBN: 978-92-4-156568-4.

References iii

SQ Xie, Ni Dong, SC Wong, Helai Huang, and Pengpeng Xu. "Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure". In: *Accident Analysis & Prevention* 121 (2018), pp. 285–294.

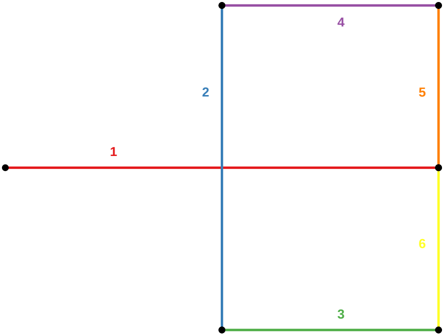
Hong Yang, Kaan Ozbay, Ozgur Ozturk, and Mehmet Yildirimoglu. "Modeling work zone crash frequency by quantifying measurement errors in work zone length". In: *Accident Analysis & Prevention* 55 (2013), pp. 192–201.

Apostolos Ziakopoulos and George Yannis. "A review of spatial approaches in road safety". In: *Accident Analysis & Prevention* 135 (2020), p. 105323.

Supplementary Materials

Spatial network lattice

A street network can be represented at the same time as a spatial object with segments and junctions (left) and a graph object with nodes and vertices (right is the corresponding adjacency matrix).



$$
\begin{array}{c}
\\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix}
\end{array}
\begin{array}{c}
\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\
\begin{pmatrix}
\cdot & \cdot & \cdot & \cdot & 1 & 1 \\
\cdot & \cdot & 1 & 1 & \cdot & \cdot \\
\cdot & 1 & \cdot & \cdot & \cdot & 1 \\
\cdot & 1 & \cdot & \cdot & 1 & \cdot \\
1 & \cdot & \cdot & 1 & \cdot & 1 \\
1 & 1 & 1 & \cdot & 1 & \cdot
\end{pmatrix}
\end{array}
$$