MEASUREMENT ERROR MODELS ON SPATIAL NETWORK LATTICES: CAR CRASHES IN LEEDS

Andrea Gilardi[1], Riccardo Borgoni[1], Luca Presicce[1] and Jorge Mateu[2]

[1] Department of Economics, Management and Statistics, University of Milano - Bicocca, Milan, Italy (e-mail: andrea.gilardi@unimib.it)

[2] Department of Mathematics, Universitat Jaume I, Castellón, Spain

ABSTRACT: Road casualties represent the leading cause of death among young people worldwide, especially in poor and developing countries. This paper introduces a Bayesian hierarchical model to analyse car accidents on a network lattice that takes into account measurement error in spatial covariates. We exemplified the proposed approach analysing all car crashes that occurred in the road network of Leeds (UK) from 2011 to 2019. Our results show that omitting measurement error considerably worsens the fit of the model and attenuates the effects of spatial covariates.

KEYWORDS: CAR, Linear Networks, Network Lattices, Spatial Measurement Error

1 Introduction

As reported by World Health Organisation in 2018, car crashes are responsible for more than 1.35 million casualties each year, representing the leading cause of death among people aged 5-29 years, particularly those living in developing countries. In the last years, several authors developed sophisticated statistical models to analyse the spatial distribution of car crashes at the areal level (e.g. cities or census wards) and help the local authorities define safety measures.

Nevertheless, road casualties represent a classic example of events occurring on a linear network. This paper presents a Bayesian hierarchical model for car crashes developed on a network lattice that takes into account measurement error (ME) in spatial covariates. In particular, a Conditional Auto-Regressive (CAR) prior is introduced to adjust for ME in estimating road traffic volumes within the classical ME model paradigm. The Integrated Nested Laplace Approximation (INLA) framework is adopted for inference. This approach was found particularly convenient for large networks, as the one considered in this paper, while MCMC techniques may be challenging and time-consuming (Muff *et al.*, 2015).

2 Road network and car crashes

The statistical analysis introduced in Section 3 requires a specific data structure that was obtained after several preprocessing steps briefly described hereafter.

The *road network* was built using data extracted from Open Street Map (OSM), an online database that provides open-access geographic rich-attribute data worldwide. We downloaded the street segments that pertain to the most important* roads of Leeds and created a matrix of segments representing the elementary units of the statistical model.

A street network can also be seen as a graph object whose edges represent the road network segments and whose vertices are placed at junctions, intersections, and boundary points (Barthélemy, 2011). We took advantage of the graph representation to contract the street network removing redundant nodes, edges loops, duplicated roads, and several isolated clusters of segments that may create numerical problems (Gilardi *et al.*, 2020). Furthermore, we calculated the weighted edge betweenness centrality, a graph measure correlated with the spatial distribution of commercial activities, which is usually adopted to analyse congestion problems as a proxy for urban traffic (Barthélemy, 2011). Finally, we derived the edges' adjacency matrix, an essential ingredient for the CAR prior used below.

We analysed all car crashes involving personal injuries that occurred in the city of Leeds from 2011 to 2019 and became known to the Police Forces within thirty days from their occurrence. First, we downloaded the data from UK's official road traffic casualty database. Then, we excluded those car crashes that occurred farther than fifty metres from the closest road segment, and, finally, we projected the events to the nearest point of the network and counted the occurrences for each segment. The final sample included 15826 events distributed over 4253 segments covering approximately 1170 km.

3 Statistical methods

Let $y_i,\ i = 1,\dots,n$ represent the number of car crashes that occurred on the ith road segment. Following a classical hypothesis in the road safety literature, we assume that $y_i|\lambda_i \sim \text{Poisson}(e_i\lambda_i)$, where λ_i represents the car crashes rate and e_i is an exposure parameter equal to the geographical length of each segment.

*More precisely, we selected only those segments whose classification range from *Autostrada* (i.e. *Motorway*) to *Strada Comunale* (i.e. *Tertiary Road*).

In the first level of the hierarchy, we define a log-linear structure on λ_i, i.e.

$$\log(\lambda_i) = \beta_0 + \beta_z z_i + \beta_x x_i + \theta_i + \phi_i;\ i = 1, \ldots, n, \tag{1}$$

where β_0 denotes the intercept, z_i is an error free covariate representing the road-type of each segment, x_i is an unobservable error prone covariate representing the traffic volumes, while β_x and β_z are the corresponding coefficients. Finally, θ_i and ϕ_i denote spatially structured and unstructured random effects that are modelled using a reparametrisation and a network re-adaptation of Besag-York-Mollié (BYM) prior (Riebler *et al.*, 2016, Gilardi *et al.*, 2020).

The classical spatial ME model assumes that x_i can be observed only via a proxy, say w_i, such that

$$w_i = x_i + u_i + \varphi_i;\ i = 1, \ldots, n.$$

The terms u_i and φ_i represent the ME and denote, respectively, spatially structured and unstructured random effects that are also modelled using the BYM prior. In particular, parameter φ_i adds a spatial smoothing effect to the unobserved covariate x_i. In this paper, we assume that the edge betweenness centrality measure can approximate the unobservable traffic volumes.

At the second stage of the hierarchy, we specified an exposure model that relates x_i with the error-free predictor:

$$x_i = \alpha_0 + \alpha_z z_i + \varepsilon_i;\ i = 1, \ldots, n. \tag{2}$$

The parameter α_0 denotes the intercept, α_z is the coefficient of the error-free covariate, and ε_i is a normally distributed error component. Furthermore, we assigned independent $N(0, 10^3)$ priors to β_0, β_z, α_0, and α_z, i.e. the intercepts and the coefficients assigned to z_i in equations (1) and (2).

The third level completes the specification of the hierarchical model eliciting a $\mathrm{N}(0, 100)$ prior for β_x, i.e. the coefficient of the error-prone covariate, a $\mathrm{Gamma}(1, 5\text{e-}05)$ prior on the precision of u_i and φ_i, and Penalised Complexity priors for the parameters of BYM's re-adaptation (Simpson *et al.*, 2017).

4 Results and conclusions

We estimated the statistical model described in Section 3 using INLA methodology and compared the results with two simpler models: the first one completely ignores ME, while the second one adopts a classical ME without spatial smoothing effects. We found that omitting ME greatly attenuates the importance of traffic volumes, and excluding the spatial smoothing terms worsens

	No ME	ME	Spat. ME
β_x	0.01	1.064	2.95
β_0	-5.307	-9.90	-15.441
β_{primary}	0.61	0.56	0.40
$\beta_{\text{secondary}}$	0.57	0.68	1.05
DIC		33126	30466

Table 1. *Summary of DIC, posterior means of fixed effects, and error-prone covariate.*

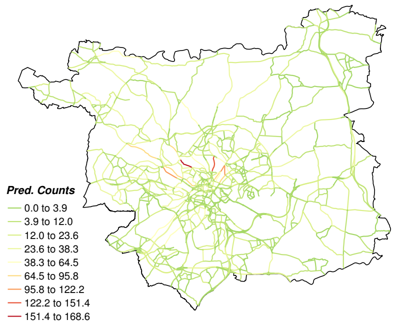

Figure 1. *Map displaying the posterior means of car crashes counts.*

the fit of the model. Motorways were found less prone to car crashes than the other road types, while the posterior distributions of fixed effects and common hyperparameters were found stable among the three models. We report in Table 1 a short summary of fixed effects' posterior means, while Figure 1 displays the posterior means of predicted counts. We can notice that it highlights a few road segments close to the city centre that would require a more detailed statistical analysis.

References

BARTHÉLEMY, MARC. 2011. Spatial networks. *Physics Reports*, **499**(1-3), 1–101.

GILARDI, ANDREA, MATEU, JORGE, BORGONI, RICCARDO, & LOVELACE, ROBIN. 2020. Multivariate hierarchical analysis of car crashes data considering a spatial network lattice. *arXiv preprint arXiv:2011.12595*.

MUFF, STEFANIE, RIEBLER, ANDREA, HELD, LEONHARD, RUE, HÅVARD, & SANER, PHILIPPE. 2015. Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 231–252.

RIEBLER, ANDREA, SØRBYE, SIGRUNN H, SIMPSON, DANIEL, & RUE, HÅVARD. 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research*, **25**(4), 1145–1165.

SIMPSON, DANIEL, RUE, HÅVARD, RIEBLER, ANDREA, MARTINS, THIAGO G, & SØRBYE, SIGRUNN H. 2017. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science*, 1–28.