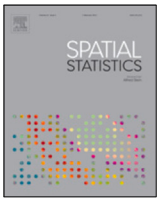
Spatial Statistics 53 (2023) 100722

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



# Spatiotemporal modeling of traffic risk mapping: A study of urban road networks in Barcelona, Spain

Somnath Chaudhuri [a,b], Marc Saez [a,b,*], Diego Varga [a,b,c], Pablo Juan [a,d]

[a] Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Spain
[b] CIBER of Epidemiology and Public Health (CIBERESP), Spain
[c] Department of Geography, University of Girona, Spain
[d] Department of Mathematics, University of Jaume I, Spain

ARTICLE INFO

*Article history:*
Received 21 May 2022
Received in revised form 4 December 2022
Accepted 9 December 2022
Available online 21 December 2022

*Keywords:*
INLA
Network triangulation
SPDE
Traffic accident

ABSTRACT

Accidents on the road have always been a major concern in modern society. According to the World Health Organization, globally road traffic collisions are one of the leading and fastest growing causes of disability and death. The present research work is conducted on ten years of traffic accident data in an urban environment to explore and analyze spatial and temporal variation in the accidents and related injuries. The proposed spatiotemporal model can make predictions regarding the number of injuries incurred on individual road segments. Bayesian methodology using Integrated Nested Laplace Approximation (INLA) with Stochastic Partial Differential Equations (SPDE) has been applied to generate a predicted risk map for the entire road network. The current study introduces INLA- SPDE modeling to perform spatiotemporal predictive analysis on selected areas, precisely on road networks instead of traditional continuous regions. Additionally, the result risk maps act as a baseline to identify the safe routes in a spatiotemporal context. The methodology can be adapted and applied to enhanced INLA-SPDE modeling of spatial point processes precisely on road networks.



* Corresponding author at: Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Spain.
*E-mail address:* marc.saez@udg.edu (M. Saez).

https://doi.org/10.1016/j.spasta.2022.100722

## 1. Introduction

With the growth of population and economic development, the process of urbanization is accelerating, and the number of vehicles in urban cities is increasing. In recent years, road traffic collisions have become a major concern in modern society. According to 2018 global status report on road safety by the World Health Organisation, every year, around 1.2 million people die on the world's roadways, with another 20 to 50 million suffering non-fatal injuries. Traffic accident is declared as one of the leading causes of death for people of all ages (WHO, 2019). Literature suggests road infrastructure and uncontrolled vehicle speed increase accident risk (Briz-Redó et al., 2019). Additionally, temporal factors (time of the day or weekend nights) act as decisive aspects in the number and impact of accidents (Farmer, 2005; Liu and Sharma, 2017).

In the field of road safety, identifying relevant factors and spatio-temporal modeling of traffic accidents have been a major focus of research (Williamson and Feyer, 1995; Prasannakumar et al., 2011; Zhong-xiang et al., 2014; Cantillo et al., 2016; Khulbe and Sourav, 2019). A series of studies (Jegede, 1988; Farmer, 2005; Aghajani et al., 2017; Shafabakhsh et al., 2017) have been conducted to explore historical data in order to identify risk factors and assess the likelihoods of crash-related events in order to categorize spatiotemporal factors influencing traffic accidents. Research in suitable statistical methodologies to analyze traffic accident data is a fundamental line of research in the field of traffic safety analysis. We can focus on detecting areas with a high accident concentration (hot spot detection methods) and focus on modeling traffic accident risk (expressed by raw accident counts or accident rates) depending on a set of predictive covariates. In statistical analysis and prediction modeling, these factors are regarded as significant predictors.

Accessible, and sustainable transport systems in cities are a core target of 2030 sustainable development goals (SDGs) adopted by the United Nations (UNDP, 2021). Thus, there is an opportunity to apply advanced computational techniques to the problem of road safety and traffic management. Various models and techniques have been developed and explored in this domain (Karaganis and Mimis, 2006; Pulugurtha and Sambhara, 2011; Castro et al., 2012; Boulieri et al., 2016; Khulbe and Sourav, 2019). Analyzing traffic safety performance by understanding crash fatality rates and influencing factors is essential for developing well-informed policies and designing effective countermeasures. Understanding the causes of the crashes, identifying appropriate solutions, and, proactively adopting or using them helps improve traffic safety. Studies like, Xu and Huang (2015), Guo et al. (2018) and Wang et al. (2019), highlight the existence of spatial autocorrelation of traffic accident events. In this line to analyze spatial and spatio-temporal distribution of traffic collisions, statistical inference comes along with Bayesian methodology. Boulieri et al. (2016) designed a space–time multivariate Bayesian model to analyze road traffic accidents by severity in different cities of the UK. A Bayesian approach with Markov chain Monte Carlo (MCMC) simulation methods has traditionally been used to fit generalized linear mixed models (GLMM) in a spatial context (Wikle et al., 1998). However, the computation time for MCMC models is considerably high for big datasets (Rue et al., 2009; Smedt et al., 2015). As recommended by Rue et al. (2009), while processing spatial data we can utilize integrated nested Laplace approximations (INLA) in conjunction with SPDE to balance speed and accuracy of the models. But, in the context to traffic accident event modeling, there are limited contributions using INLA-SPDE approach. Recently, Galgamuwa et al. (2019) used Bayesian spatial modeling with INLA in predicting road traffic accidents based on unmeasured information at road segment levels. Similar study by Chaudhuri et al. (2022) explores spatiotemporal modeling of traffic accidents on the road network of London, UK based on an explicit network triangulation using INLA-SPDE.

The aim of this paper is to propose a multi-disciplinary road-safety analysis technique by introducing spatio-temporal modeling of traffic accidents using Bayesian methodology restricted entirely on to the road network. We introduce an advance and realistic computational strategy to construct spatial triangulation restricted only onto the network topology. The proposed model can predict a risk index over individual road segments generating a categorized risk map of the entire road network. The study is conducted on ten years road traffic accidents data from the city of Barcelona, Spain. R (version R 4.1.2) programming language (R. Core Team, 2021) has been used for statistical computing and graphical analysis. All computations were conducted on a quad-core Intel i9-4790 (3.60 GHz) processor with 32 GB (DDR3-1333/1600) RAM.

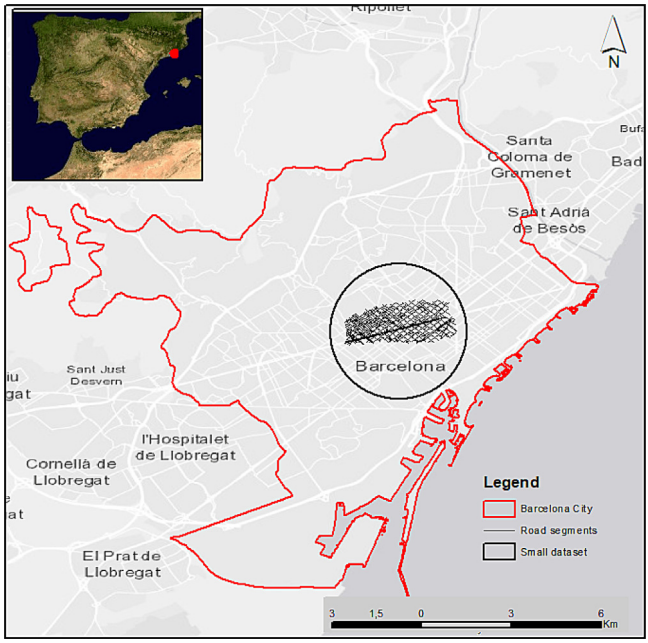



**Fig. 1.** Location of Barcelona city and road network of the study area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The rest of the paper is organized as follows. Section 2 presents the methods adopted in the current study. The beginning of the section reports about the study area and data settings along with official sources of data. A description of the spatio-temporal modeling framework with emphasis on SPDE triangulation designed precisely on road network is discussed in Section 2.2. Section 3 is devoted to present the results of model prediction and risk map analysis. Some discussions are highlighted in Section 4. The paper ends with some concluding remarks in Section 5.

## 2. Methods

### 2.1. Data settings

Barcelona is the largest and the capital city of the community of Catalonia, Spain. This second most populous municipality of Spain is located on the coast of northeastern of the country. Fig. 1 shows the location of Barcelona city, and the boundary of the municipality highlighted in red. According to Barcelona's city hall open data service (Open Data BCN) (OpenDataBCN, 2021), the city has a population of 1.6 million and approximately 15,748 inhabitants per square km. Besides being one of the major cultural, economic, and financial center, Barcelona is also a transport hub for entire southwestern Europe. The city municipality maintains an extensive motorway network.

In the current study, we have considered a small area (4.4 square km) from the central part of Barcelona consisting of 2058 road segments as depicted in Fig. 1 inside the black circle. The road network is also accessed from Open Data BCN repository. The police department in the city maintains records of traffic accidents. Detailed records about the circumstances of road accidents on public roads, corresponding casualties and injuries are managed and published annually by Open Data BCN. The data is free and available under the *Creative Commons Attribution 4.0* for public sector information. During the period of January 2010 to December 2019 the study area has records of

11,067 traffic accidents. Fig. A.8 in the Appendix shows the road networks in the study area with traffic accident locations highlighted in red.

Five datasets from Open Data BCN have been accessed for the study, which are referred to the accident itself and interrelated by a record code from 2010 to 2019. The recorded common attributes are unique event id, district and neighborhood, location postal address and geographical coordinates, occurrence day and time, kind of day (working or holiday). Each of dataset contain the following temporal variables: year, month, and time of accident. Related to spatial variables we made few changes. In raw dataset, individual accident locations in most cases are not located exactly on the road segment. We have shifted individual locations to the nearest road segments. In addition, we calculated the on road network distance of nearest bus stop, municipality market, restaurant, school, street market for each accident locations. These distances are used as spatial covariates in the dataset. We report that the traffic intensity records for each road segment are collected from TomTom Traffic Stats (TomTom, 2021).

It is noteworthy to mention that values of traffic intensity for individual road segments is not directly available from TomTom dataset. We used three variables to calculate the traffic intensity such as, *road length* ranging from 3.69 to 186.25 meter, *road type* (values 1 to 7, higher the value less the traffic) and *road speed limit* (18 to 80 km per hour) where roads having 30, 35 and 50 km per hour cover 21%, 28% and 35% of the total. Our proposed formula to calculate the traffic intensity is:

$$Traffic\ Intensity = (Speed\ Limit / Type\ of\ Road) * ln(Road\ Length)$$

where, ln stands for natural logarithm that is log to the base of $e$. Fig. A.9 in the Appendix depicts the calculated traffic intensity for individual road segments of the study area.

We are using the road length in natural logarithm (ln) scale to reduce the range of road length values. Other variables available in the dataset are number of victims, vehicles involved, minor and major injured persons, and number of casualties for each accident records. We have used the number of minor injured person as the response variable in our models. Traffic accidents recorded with only one minor injury comprises the maximum percentage of records (74.8.76%) followed by two minor injuries (15.42%), and 3 or more minor injuries (3.42%). The record shows 6.4% of the accidents are without having any minor injuries. We note that most accidents (99.85%) are having no causality. The number of traffic accidents documented in each of the study years (2010–2019) is similar, with the highest number (1270) recorded in 2016 and the lowest number (847) recorded in 2011. In case of monthly records for the entire study period, January records the minimum accident counts (846) and July has the maximum value (1023). It is worth noting that, almost 50% of all accidents occur during office hours, which are from 8 a.m. to 11 a.m. and from 3 p.m. to 6 p.m.

### 2.2. Statistical analysis

Random spatial events, such as traffic accidents, form irregularly scattered point patterns over regions of interest. Literature shows, spatio-temporal point process models are useful tools for performing focused statistical analysis (Karaganis and Mimis, 2006; Loo et al., 2011; Juan et al., 2012). In this context, Liu et al. (2017) propose that the occurrence of traffic accidents depend on spatio-temporal interacting and triggering factors (Liu et al., 2017). Moreover, we can find recent studies (Galgamuwa et al., 2019; Moradi and Mateu, 2019; Chaudhuri et al., 2021) on spatio-temporal point processes over networks that are able to identify spatial autocorrelations and interactions between points in the pattern. We open the door to consider binomial regression models in combination with a Bayesian framework for the prediction of traffic accidents on individual road segments by aggregating data for the occurrence of accident injuries per road segment given the total traffic intensity. We have used a computationally faster solution for prediction of the marginal distributions for latent Gaussian models and models with a large number of geo-locations by using a Laplace approximation for the integrals with the integrated nested Laplace approximation (INLA) method. (Rue et al., 2009). It focuses on models that can be expressed as latent Gaussian Markov random fields (GMRF) (Rue and Held, 2005). We follow this approach combining a spatio-temporal binomial regression method within a Bayesian framework using INLA and stochastic

partial differential equation (SPDE). In particular, specification of the binomial distribution in INLA for responses $y = 0, 1, 2, \ldots, n$ is represented as

$$Prob\,(y) = \binom{n}{y} p^{y}\,(1-p)^{(n-y)} \tag{1}$$

where, $n$ is the number of trials and $p$ is the probability of success in each trial (Rue et al., 2009).

The mean and variance of $y$ are respectively $\mu = np$ and $\sigma^2 = np\,(1-p)$ and the probability $p$ is linked to the linear predictor by

$$p(\eta) = \frac{exp(\eta)}{1 + exp(\eta)}$$

For the current study, let $Y_{it}$ be the observed number of minor injuries in road traffic accidents on the $i$th road segment and at the $t$th day, $t = 1, \ldots, T$. The average daily traffic intensity for individual road segment is represented by $N_{it}$. We assume that conditional on the relative risk, $\rho_{it}$, the number of observed events follows a binomial distribution

$$Y_{it}|\rho_{it} \sim Binomial\,(N_{it}\,\rho_{it})$$

where the log-risk is modeled as

$$logit\,(\rho_{it}) = \beta_0 + Z_i^T \beta_i + S\,(x_i) + \delta_t + \epsilon_i \tag{2}$$

Here, $S\,(x_i)$ and $\delta_t$ account for the spatially and temporally structured random effects, respectively and $\epsilon_i$ stands for an unstructured zero mean Gaussian random effect and logbeta precision parameters 0.5 and 0.01, defined as penalized complexity priors (Simpson et al., 2017). $Z_i$ represents all covariates included in the model. We assigned a vague prior to the vector of coefficients $\beta = \left(\beta_0, \ldots, \beta_p\right)$ which is a zero mean Gaussian distribution with precision 0.001. All parameters associated to log-precisions are assigned inverse Gamma distributions with parameters equal to 1 and 0.00005.

To compute the joint posterior distribution of the model parameters, we use an INLA-SPDE method, as introduced by Lindgren et al. (2011). SPDE consists in representing a continuous spatial process, such a Gaussian field (GF), using a discretely indexed spatial random process such as a Gaussian Markov random field (GMRF). In particular, the spatial random process represented by S (.) explicitly denote dependence on the spatial field, follows a zero-mean Gaussian process with Matérn covariance function represented as

$$Cov(S(x_i), S(x_j)) = \frac{\sigma^2}{2^{\nu-1}\Gamma\,(\nu)}(\kappa\,\|x_i - x_j\|)^{\nu} K_{\nu}(\kappa\,\|x_i - x_j\|) \tag{3}$$

where $K_{\nu}\,(.)$ is the modified Bessel function of second order, and $\nu > 0$ and $\kappa > 0$ are the smoothness and scaling parameters, respectively. INLA approach constructs Matérn SPDE model, with spatial range r and standard deviation parameter $\sigma$. The parameterized model we follow is of the form

$$(k^2 - \Delta)^{(\alpha/2)}(\tau S(x)) = W(x)$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian operator, $\alpha = (\nu + d/2)$ is the smoothness parameter, $\tau$ is inversely proportional to $\sigma$ and $W\,(x)$ is a spatial white noise and $\kappa > 0$ is the scale parameter, related to range $r$, defined as the distance at which the spatial correlation becomes small. For each $\nu$, empirically derived definition $r = \sqrt{8\nu}/\kappa$ with $r$ corresponding to the distance where the spatial correlation is close to 0.1 (Blangiardo and Cameletti, 2015). Note that we have $d = 2$ for a two-dimensional process, and we fix $\nu = 1$, so that $\alpha = 2$ in our case. We report that, heterogeneity, unobserved factors specific to each accident, although invariant over time, were captured by an identical and independently distributed random effect of zero mean and constant variance.

The temporal random effect $(\delta_t)$ is assumed to be a smoothed function, in particular a random walk of order one (RW1) (Rue et al., 2009). On the other hand, INLA-SPDE requires a triangulation or mesh structure to interpolate discrete event locations to estimate a continuous process in space (Rue et al., 2017). We use the centroids of each road segment as the target locations over which we

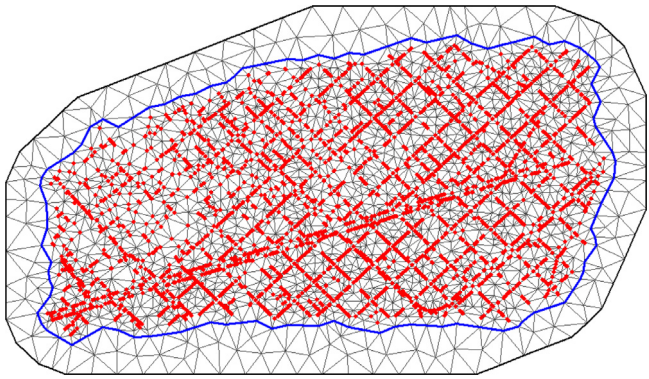


**Fig. 2.** Region mesh with non-convex hull boundary in blue and data locations highlighted as red points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

build the mesh. A detailed description of building a Delaunay's triangulation with emphasis on a network mesh is shown in Section 2.2. Centroids of individual road segments and the triangulations in the mesh are used to generate the projection matrix. We assign penalized complexity priors for the parameters to create INLA-SPDE model object for the Matérn model (Simpson et al., 2017). In the parametrization process we set prior according to hyperparameters for range as *prior.range* (0.01, 0.01) and standard deviation as *prior.sigma* (1, 0.1).

*SPDE network triangulation:*

To begin, we use the traditional SPDE method to triangulate the entire study area by considering the boundary for continuous spatial structure. According to Verdoy (2019), the best fitting mesh should have enough vertices for effective prediction, but the number should be within a limit to have control over the computational time. Following this principle, from a battery of meshes the best fitting mesh is selected having 2352 vertices. Fig. 2 depicts region SPDE mesh with 11,067 traffic accident locations highlighted as red points.

While fitting the mesh (as depicted in Fig. 2) a problem appears. Although the sampled traffic accidents are discrete spatial sites situated exclusively on the road networks, the models fitted with the mesh span the entire study area. So, it is not realistic and ambiguous for the model prediction to provide results in areas without a road network where traffic accidents are unlikely to occur. This leads to the motivation of designing SPDE triangulation precisely on road networks. The technique is carried out in three steps: first, a buffer region is generated for each road segment, then a clipped buffer polygon is created that only includes the area covered by the road network, and finally, SPDE triangulation is applied to the clipped polygon to create *SPDE Network Mesh*. We use R package rgeos (Bivand et al., 2017) to build buffers for individual road segment. While selecting the buffer size we need to balance between number of vertices used to build the triangulated mesh and computational cost (Krainski et al., 2018; Verdoy, 2019). We tested several buffers before deciding on a 15 meters buffer as the optimal one. In the next step, we merge individual buffer segments and convert them into a single polygon clipped within a bounding box covering the study area. The clipped polygon of the buffered segments is depicted in Fig. A.10, with accident locations highlighted as red points.

In the final step of building the proposed network mesh we use the centroids of each segment as initial Delaunay's triangulation nodes on the clipped polygon. Fig. 3 depicts the SPDE mesh precisely created on the road network, with accident locations highlighted in red.

*Risk map design:*

In this section we discuss about the process of designing the traffic accident risk map for the entire road network of the study area. The predicted daily number of minor injuries on individual

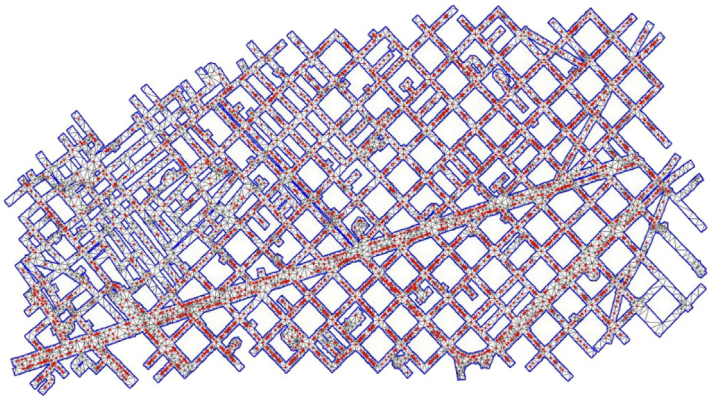


**Fig. 3.** Network mesh with data locations highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

road segment obtained using the binomial model has been considered as the initial Risk Score ($R_{score}$) on that particular segment for that day. According to the literature on road safety before modeling any risk map, a predetermined category range must be determine (Curran-Everett, 2013). Keeping this in mind, we follow the European Road Assessment Programme (EuroRAP) standard to create the risk ratings of the motorways and other national roads in Europe ('Risk mapping for the TEN-T in Croatia, Greece, Italy and Spain: Update', 2016). We implement these 0 to 4 categorical values ranging from low risk to high risk defining what we call a Risk Index. We thus consider a normalization for the raw risk scores following a dynamic normalization technique. Initially, the risk range ($R_{range}$) is calculated as

$$R_{range} = \frac{(max.R_{score} - min.R_{score})}{no.\ of\ risk\ categories}$$

Next, we use $R_{range}$ and $R_{score}$ values in the proposed metric system to calculate the normalized risk index categories. As a relevant example in Table 1, we depict the values of categories in the normalized scale considered as the proposed risk categories to design the traffic risk map. We note that the metric system can be replicated using any other alternative risk index.

We depict an example of how risk index values are calculated using the proposed normalizing metric reported in Table 1. Consider the following scenario: a user wants to calculate the risk index for a specific road segment where four minor injury cases from traffic accidents have been recorded on a given day. So, the $R_{score}$ of that segment is 4. Based on the traffic accident records for all the road segments on that particular day, assume that the maximum $R_{score}$ is 15 and the minimum $R_{score}$ is 0. In the present study, five risk categories have already been considered (i.e., the risk index values 0–4). Using Equation 4, we can calculate $R_{range}$ is 3. Now, we can report that the risk index for the example road segment will match with the second normalize condition, indicating that it is a low–medium risk road segment with risk index value 1 for that particular day considered in the example. In this process, on the same day different segments can have diverse set of risk index values depending on their individual $R_{score}$. Alternatively, the risk index value for the same segment can vary on different days and months.

The risk-index algorithm implemented here has intended to categorize road segments based on the records of number of minor injuries incurred in each segment. As a result, segments having higher minor injury counts are categorized as accident-prone or high-risk roads. Other traffic risk modeling algorithms can use a similar concept.

**Table 1**
Normalization metric for risk index values.

| Normalize condition | Risk index | Safety measure |
|---|---|---|
| $R_{score} = 0$ or $R_{score} < R_{range}$ | 0 | Low risk |
| $R_{range} \leq R_{score} < 2 \times R_{range}$ | 1 | Low-medium risk |
| $2 \times R_{range} \leq R_{score} < 3 \times R_{range}$ | 2 | Medium risk |
| $3 \times R_{range} \leq R_{score} < 4 \times R_{range}$ | 3 | Medium-high risk |
| $4 \times R_{range} \leq R_{score}$ | 4 | High risk |

**Table 2**
Fitted model DIC, WAIC and CPO values.

| Model mesh | DIC | WAIC | CPO |
|---|---|---|---|
| Region mesh | 23687.41 | 23674.25 | 0.3246 |
| ***Network mesh*** | ***23654.73*** | ***23647.06*** | ***0.3243*** |

*Inference:*

Inferences are made following a Bayesian perspective, using the INLA approach (Rue et al., 2009, 2017). We used priors that penalize complexity (called PC priors). These priors are robust in the sense that they do not have an impact on the results and in addition the notion of scale determines the magnitude of effects and simplifies interpretation of results (Simpson et al., 2017).

All analyses are carried out using the free software R (version 4.1.2) (R. Core Team, 2021), through the INLA package (Rue et al., 2009; Lindgren, 2012; Rue et al., 2017, 'R. INLA Project', 2020). Maps related to study area, Barcelona street network with accident locations and traffic intensity map are designed using ArcGIS Desktop Soft- ware (version 10.8. Redlands) (ESRI, 2021). Other maps to depict SPDE mesh generation technique and plotting risk maps are designed using R package mapview (Appelhans et al., 2016).

## 3. Results

In this section, we present the results of the methodological approach developed in Section 2. We provide results on model fitting, validation, and prediction along with risk maps of accidents. The proposed model (mentioned in Eq. (2)) is fitted using both region mesh as depicted in Fig. 2 and our proposed network mesh as depicted in Fig. 3. Both models are fitted to the daily accident records for the years 2010 to 2018. The remaining accident records of 2019 have been used to test the fitted model. It is worthy to mention that we have executed series of similar models for both categories of mesh, using different dimensions of SPDE meshes with different combinations of covariates. In each case, deviance information criterion (DIC) and the Watanabe–Akaike information criterion (WAIC) are used to assess the performance of the models, and to select the best fitting model by balancing model accuracy against complexity (Spiegelhalter et al., 2002). Models having smaller WAIC value, despite the added complexity, provide a more appropriate fit to sampled data (Blangiardo and Cameletti, 2015). In Table 2 we report the summary results (DIC, WAIC and CPO) related to goodness-of-fit along with computational time (in seconds) for only the best fitted models from individual region mesh and network mesh categories. We note the computational time (in seconds) of the best fit model with region mesh (336) is substantially lower compared to the best fit model with proposed network mesh (3187). This can be explained due to higher number vertices in network mesh (14,368) than the region mesh (2352).

DIC values shown in Table 2 indicate that the WAIC value of model with proposed network mesh (23,647.06) is lower compared to the other. Thus, to model the spatio-temporal structure of traffic accidents on road networks of Barcelona, the binomial model with SPDE network mesh is selected. We additionally note that the model is best by considering random spatial and temporal effects together with set of covariates mentioned in Section 2.1. When the covariates are not considered, the model provided larger DIC and WAIC values.

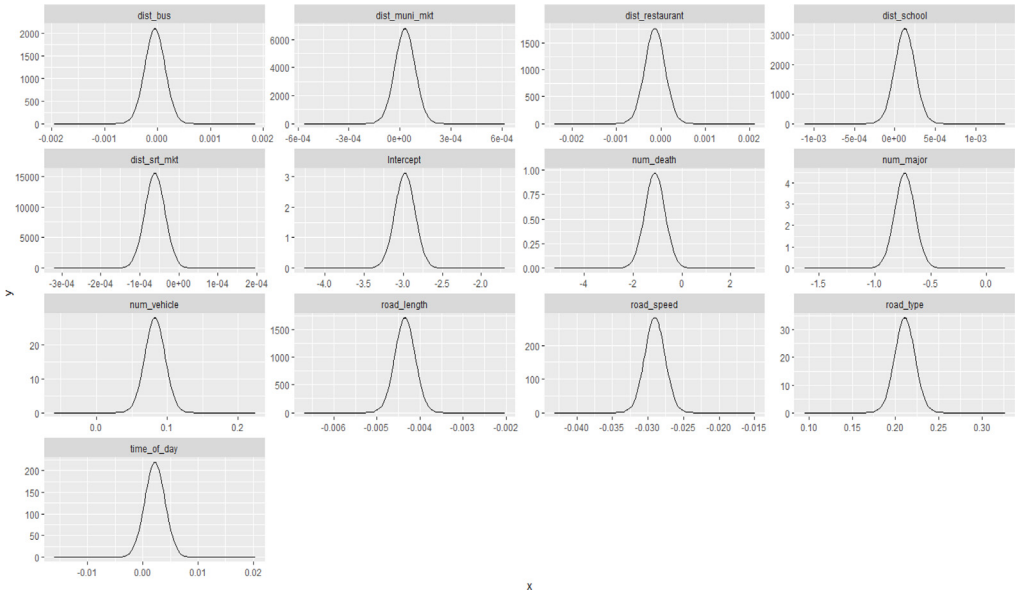



**Fig. 4.** Marginal posterior distributions of covariate coefficients.

The posterior distribution of fixed and random effects included in the model are depicted in Figs. 4 and A.11 (in Appendix). In particular, Fig. 4 shows the marginal posterior distributions of all fixed effects. Additionally, Table 3 depicts the coefficients and 95% of credibility intervals of all fixed effects.

We note that six covariates, namely time of day, distances from nearest bus stop, school, restaurant, municipality, and street markets have no influence in our model. The positive mean values for covariates such as number of vehicles involved in accident and type of road indicate positive influence in the model. It is worthy to note that when the number of causalities and major injuries in individual traffic accident are high then the record of minor injuries is lower compared to other cases, thus there exists a negative association for these two covariates in our model. The covariate associated to number of causalities has the highest negative mean value which indicates strong negative influence on the model. Additionally, road length and speed limit of individual road show negative influence in our model. Indicating when the road length is high and speed limit is also high the number of minor injuries in a collision is comparatively lower than other short roads and low speed tracks.

Moreover, Fig. A.11 in the Appendix depicts the marginal posterior mean of spatial $S(.)$ and temporal $\delta_t$ random effects with 95% credible intervals. The horizontal axis of Fig. A.11 (top) in the Appendix represents the 14,368 triangulation nodes of the SPDE network mesh used in the model. A stronger and significative spatial effect is observed on the vertices of triangles on road segments having higher traffic accident occurrence (highlighted in Fig. 3 as dark red patches). The vertices without accident events show no spatial effect. Similarly, Fig. A.11 (bottom) in the Appendix exhibits the variation of the marginal posterior mean of the daily temporal random effects over the entire study period (2010 to 2019).

We finally report that the spatial effect parameters $\kappa$ and $\tau$ have mean values 92.16 and 0.31 as depicted in Fig. A.12 (in the Appendix) that shows the marginal posterior distributions of the two hyperparameters for the spatial random field. Using $\tau$ and $\kappa$ we can get the value of spatial range $r = 0.055$ km or 55 m. Using the fitted model, we can analyze the goodness-of-fit of the model by

**Table 3**
Marginal posterior mean and credible interval of fixed effects.

| Covariate | Mean | Credible interval |
|---|---|---|
| Nearest bus stop distance | 0.002 | −0.001, 0.003 |
| Nearest municipality market distance | 0.002 | −0.001, 0.004 |
| Nearest restaurant distance | 0.000 | 0.000, 0.000 |
| Nearest school distance | 0.000 | 0.000, 0.000 |
| Nearest street market distance | 0.003 | −0.001, 0.007 |
| ***Number of deaths*** | ***−1.108*** | ***−1.915, −0.301*** |
| ***Number of major injuries*** | ***−0.730*** | ***−0.906, −0.555*** |
| ***Number of vehicles*** | ***0.083*** | ***0.055, 0.110*** |
| ***Road length*** | ***−0.004*** | ***−0.005, −0.004*** |
| ***Road speed limit*** | ***−0.029*** | ***−0.032, −0.026*** |
| ***Road type*** | ***0.211*** | ***0.188, 0.234*** |
| Time of day | 0.002 | −0.001, 0.006 |

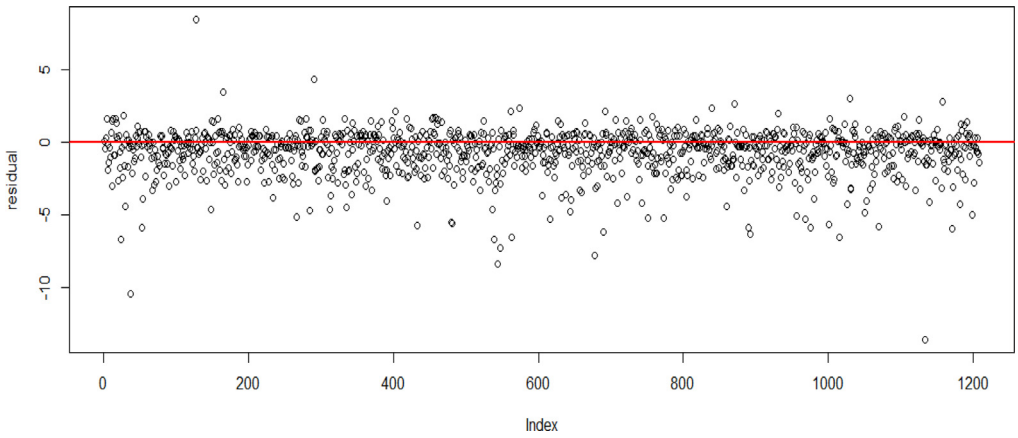



**Fig. 5.** Residual (observed minus predicted) plots.

considering prediction over unsampled locations (Zuur et al., 2017). The fitted model is projected into the mesh at each road segment for this prediction.

The proposed model is tested using accident records for the final study year 2019 combined with the entire model fitting 2010–2018 dataset. We compute predictions as well as their corresponding residuals (observed minus predicted). Fig. 5 depicts such residual plots. We note the residual values are generally close to zero and have no discernible structure.

*Risk map:*

We follow the risk map metrices in Section 2.2 and use safety measure scale shown in Table 1 to calculate the risk index for individual road segments. The predicted values for 2019 are used to construct the normalized risk index values.

Fig. 6 (top) shows the location of 1209 original road accident in 2019. The corresponding projected risk map for 2019 as a whole is shown in Fig. 6 (bottom). The color scales (0–4) used in the risk map follow the same safety measure scale used in Table 1. A visual comparison of the predicted risk map with the original road accident record shows that the road segment containing the observed cluster of accidents is correctly predicted by the risk map as medium to high-risk roads.

Similarly, roads that are predicted to have low or moderate risk are originally roads with no or very few incidents. We report that similar results are observed while comparing the predicted risk map for individual dates in 2019 with the corresponding original accident records.

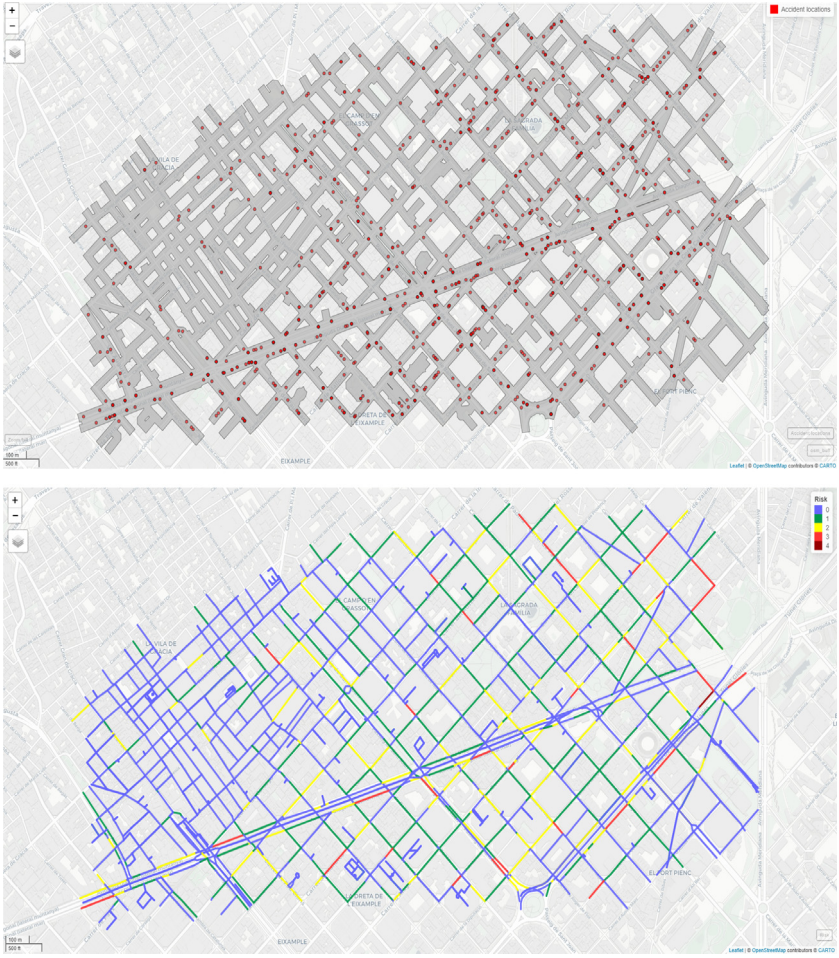



**Fig. 6.** Top: Observed traffic accident events recorded in 2019.
Bottom: Predicted risk map for 2019.

The current results show that the proposed model can produce the road safety index of all road segments, including small details of each junction or sharp turnings. In addition, identifying potentially dangerous roads can serve as baseline for geographic analysis of road safety management. The daily predicted risk maps can have strategic applications in developing GIS analytical tools to identify and depict possible safe routes. For example, in Fig. 7 (top) the start and destination points of a particular user is highlighted by green and red map pins. The user can choose path B which is considered to be shorter in length compared to path A. But in terms of safety measure the user should opt for path A as the cumulative risk index for this path is much less compared to the shorter path B. As the proposed dynamic risk map provides information about the entire road network, it is flexible enough to generate possible alternative safe route(s) between any source and destinations pairs as depicted in another similar example in Fig. 7 (bottom). In this example the length of both the roads between source and destination points are same but in terms of risk index path B is relatively unsafe compared to path A.

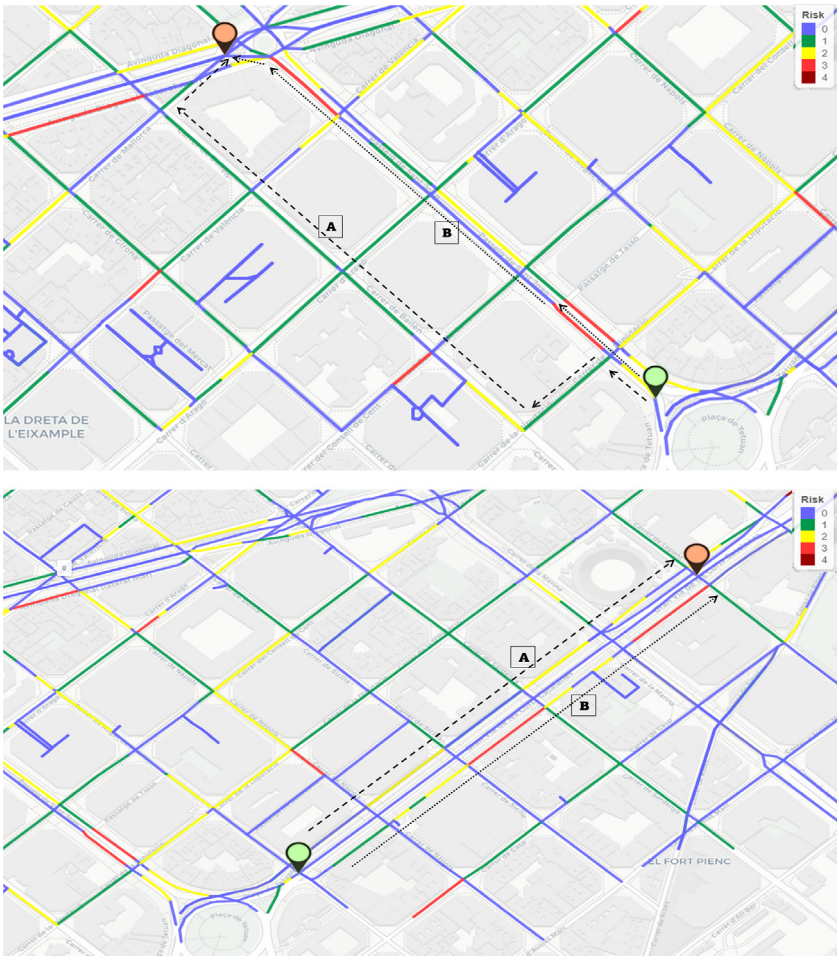



**Fig. 7.** Predicted risk map baseline to identify the safe route.
Top: Example 1, Bottom: Example 2.

## 4. Discussion

In recent years, spatiotemporal modeling of traffic accidents and risk mapping has gained attention, particularly in the domain of multi-dimensional road safety management. INLA-SPDE approach involves projecting the fitted model into the mesh at precise spatial locations, in this case the sampling points (here accidents locations) are located only on the road network. But traditional SPDE mesh is generated for the entire study area which includes road network as well as other regions. In that case, the output of the model may be unavoidably generalized because it will estimate predicted values for regions where there is no chance of an incident occurring. To avoid this ambiguity, we introduce the novel concept of designing the SPDE triangulation precisely on the road network. As a result, instead of performing spatio-temporal predictive analysis on the full continuous region, the proposed INLA-SPDE modeling took a new step by executing it only on selected sections (particularly for road networks).

In this context, Chaudhuri et al. (2022) recently proposed spatiotemporal modeling of road traffic accidents using explicit network triangulation. However, the study only used types and surfaces

of roads as covariates in the model. Significant exogenous variables related to traffic flow, traffic control and temporal variables such as time of accident occurrence and other spatial covariates are ignored in the modeling process. Furthermore, no analysis to compare and identify alternative safer roads is performed in the previous study. Whereas the results of the current study show that, when fitted with selected covariates, the proposed model can generate predicted risk maps of the entire road network for any urban study area. In that sense, the model proposed in this study is dynamic in nature. Additionally, Bayesian methodology is implemented using computationally faster INLA-SPDE approach where the number of covariates can be updated at any stage and the level of significance for individual covariates can be analyzed for further emphasis on the selection of significant factors causing traffic accidents.

Regarding the limitations and future works, we note that the buffer road network used in the current study has complex boundary regions which have some influence on the spatial effect of the model. According to Krainski et al. (2018), the first step in fitting an SPDE model is to create a mesh to represent the spatial process. Building SPDE mesh for a continuous region is relatively easier compared with the proposed SPDE network mesh. In the current study, fine tuning is required to identify the best fit values for minimum allowed distance between vertices and maximum permissible triangle edge length for the inner (and outer) regions. Moreover, additional points around the boundary, or outer extension, must be selected with care. As a rule of thumb, the variance near the boundary is inflated by a factor of 2 along straight boundaries and by a factor of 4 near right-angled corners (Lindgren and Rue, 2015). The complex boundary region of the buffer road network with several right-angled corners (as depicted in Fig. A.10) makes the process critical. The boundaries in the proposed mesh are located inside the mesh and not outside, as in a standard mesh and that creates fictitious spatial structures. Because of this complex boundary nature, it is unavoidable to reduce high boundary effect that might cause a variance twice or four times as great at the border as it is within the domain (Lindgren, 2012; Lindgren and Rue, 2015). Additionally, though the residual diagnostics and predicted risk maps produced by the model matches with the original observed records; but the correlation values of the model indicate room for improvement. Thus, for detailed understanding of the performance of the model, it may be beneficial to analyze further the model fitting phase using INLA-SPDE with a diverse set of spatial and temporal covariates, spatial and temporal structures, and space–time interactions. This paves the way for future research works in this domain and to reduce the boundary effects in the model results.

The final outcomes of the proposed model are predicted risk maps for the entire road network. Thus, using these maps, the road safety index of individual road segments, including small details of each junction point or sharp turn, can be obtained at a glance. This can act as baseline information for geospatial analysis on road safety metrics to design strategic geographical information system (GIS) analytical tools to identify and depict possible safe routes as depicted in Fig. 7. Similar research work by Hannah et al. (2018) considers only spatial traffic variables like speed limits, street junctions, and type of street. But the current study has proposed a more flexible and statistically convincing solution by implementing both spatial and temporal covariates in the predictive model. We note another crucial application of the model is in analyzing change and trend pattern of traffic accidents. Fig. 6 is the combined predicted traffic risk map for all days in 2019. As mentioned in Section 3 similar dynamic risk maps can be developed for individual months, weeks or even days. Using these maps trends in traffic accident risk can be identified for individual roads or, road junctions. A better understanding of these patterns may have implications for road safety measures. Identification of gradual changes in risk patterns and related potential factors, is of interest for future research works on change point detection.

Interestingly, the predicted risk maps can be used as an important guideline for traffic management authorities to identify potentially dangerous roads in any urban region and can take strategic measures and actions to prevent traffic accidents in advance. As a result, risk maps can be used to better understand accident hotspots, improve traffic safety measures, and conceivably can have an impact on public health by reducing traffic accidents. In addition, as the model is flexible and general, it can be applied to a wide range of related problems. The current methodology can be implemented for smart transportation systems by predicting traffic flow and reducing congestion

on roads. This would enable transport authorities to better manage the traffic condition during peak hours and would allow users to choose the best routes to their destinations. In context to smart cities, an intelligent traffic management system based on the proposed model can feasibly control air pollution caused by fine particulate matter emitted by transportation. It can have potential implications to achieve air quality levels for particles in suspension in line with the guideline value of the World Health Organization (WHO, 2019).

Consequently, we report that the proposed modeling approach could be a major step forward in the understanding of road safety measure and can act as a baseline in strategic decision making to control traffic collisions. The novel contribution of this work is that it is able to take advantage of INLA-SPDE approach precisely on road network rather than for continuous region. As a result, the model can predict risk factors for individual road segments and generate dynamic risk maps for the entire network. In conclusion, although it may be complicated to control the boundary effect in the complex network triangulation method, this work is able to present a model that can provide accurate predictions of accident-prone roads and help in identifying alternate safe routes between any source and destination pairs.

## 5. Conclusion

The current study implements Bayesian methodology by including INLA and SPDE to design a dynamic spatio-temporal analysis model predicting the occurrence of traffic accidents in the city of Barcelona, Spain. The use of SPDE network triangulation to estimate the spatial auto-correlation of discrete events is novel in this study. The methodology used in the study is a new step to perform spatio-temporal analysis precisely on road network and contributes to the relatively small amount of literature in this domain. Moreover, the dynamic risk maps of traffic accidents are one of the interesting outcomes. The risk maps can have strategic applications in road safety measures and designing travel risk maps for tourists, corporate travelers, and emergency service providers. The methodology to identify safe routes is dynamic and can be adapted and applied to other locations globally. Furthermore, the current study opens future research scopes to explore the influence of boundary effects on model performance and analyze the variation in spatial effects. We are investigating these anomalies of the spatial impact in a subsequent study project and working on a possible solution to the problem.

## CRediT authorship contribution statement

**Somnath Chaudhuri:** Original idea for the paper, Bibliographic search, Writing of the introduction, Data collection and cleaning, Statistical analysis, Created the tables and figures, Writing – review & editing. **Marc Saez:** Designed the study, Bibliographic search, Writing of the introduction, Statistical analysis, Writing – review & editing. **Diego Varga:** Bibliographic search, Writing of the introduction, Data collection and cleaning, Created the tables and figures, Writing – review & editing. **Pablo Juan:** Designed the study, Bibliographic search, Writing of the introduction, Statistical analysis, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The pre-processed data and R code for implementing the proposed model can be made available upon request.

## Acknowledgments

We appreciate constructive comments and suggestions from Håvard Rue and Elias T. Krainski. We appreciate the comments of two anonymous reviewers of a previous version of this work who, without doubt, helped us to improve our work. The usual disclaimer applies. All authors reviewed and approved the manuscript.

## Appendix

### A.1. Data settings

See Figs. A.8 and A.9.

### A.2. Methodology

See Fig. A.10.

### A.3. Results

See Figs. A.11 and A.12.

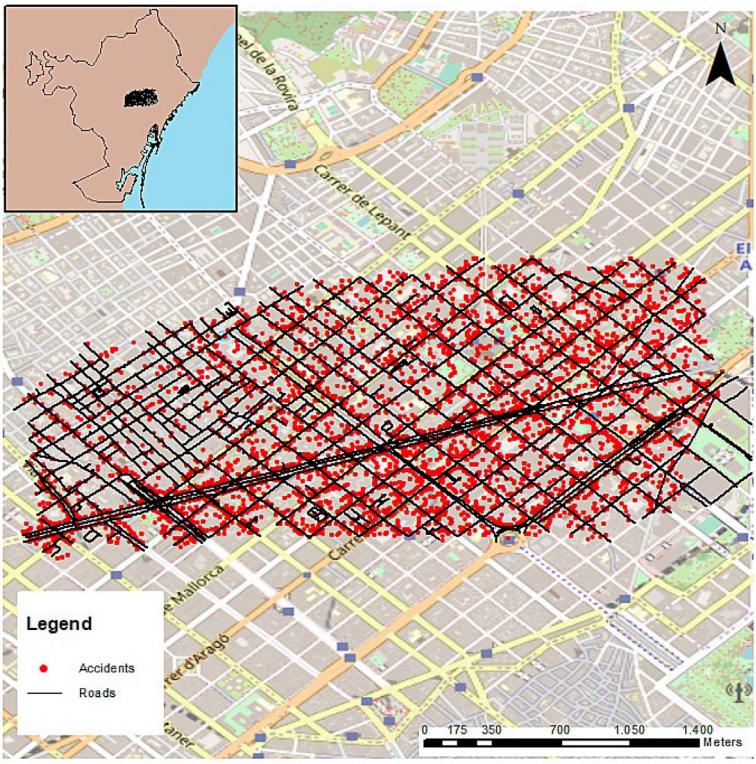



**Fig. A.8.** Location of road network of the study area with traffic accident locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

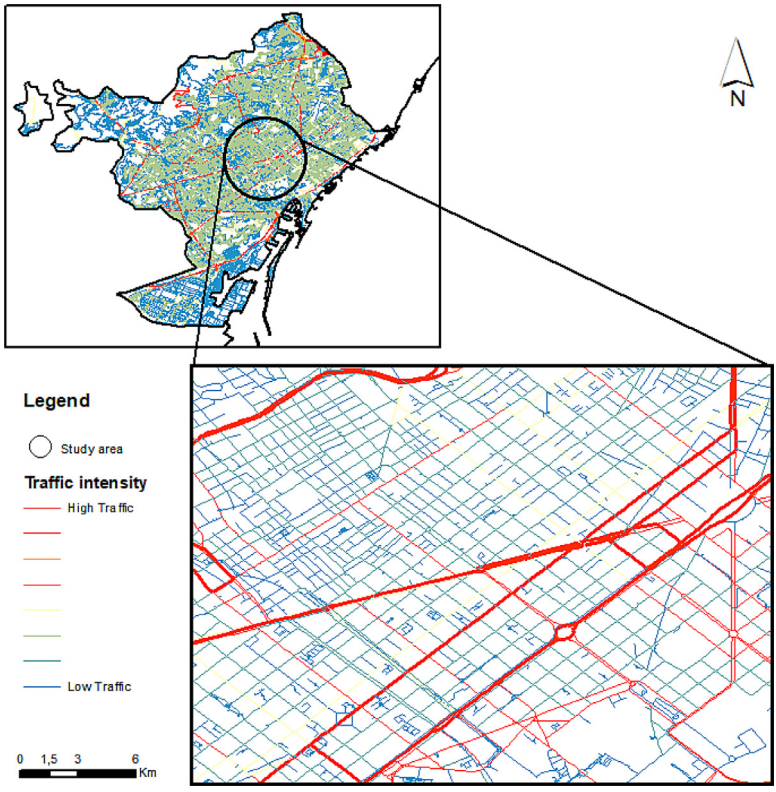



**Fig. A.9.** Traffic intensity on individual road segments of the study area.


**Fig. A.10.** Polygon of buffered road segments, red points indicate traffic accident locations on the road network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

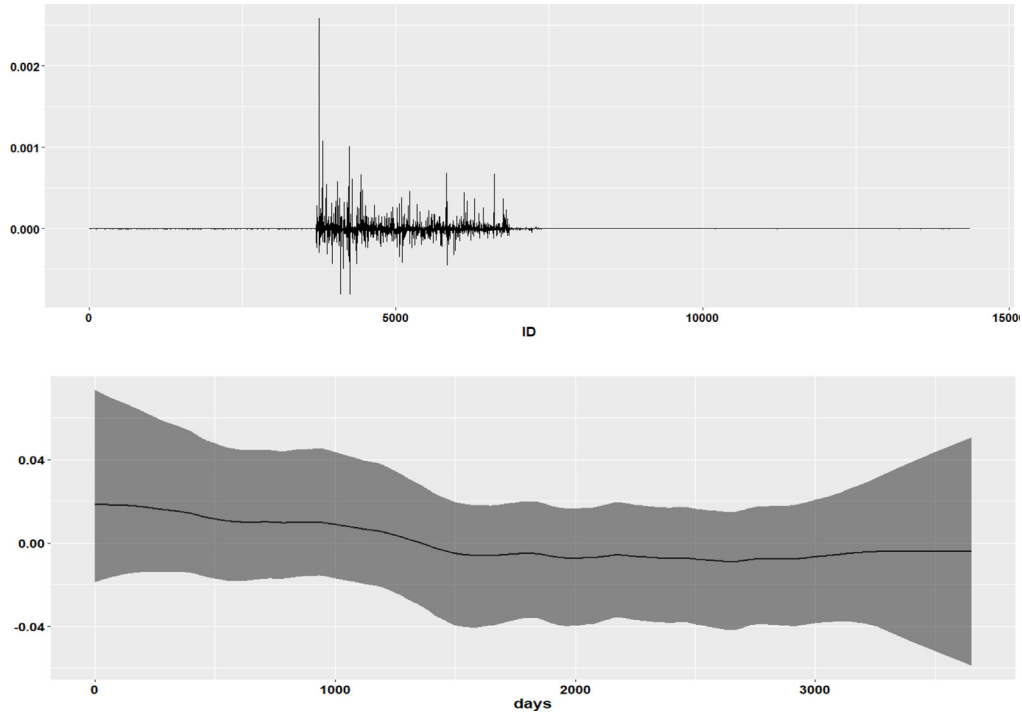



**Fig. A.11.** Top: Marginal posterior mean of the spatial random effect $S(.)$.
Bottom: Marginal posterior mean of the temporal random effect $\delta_t$.

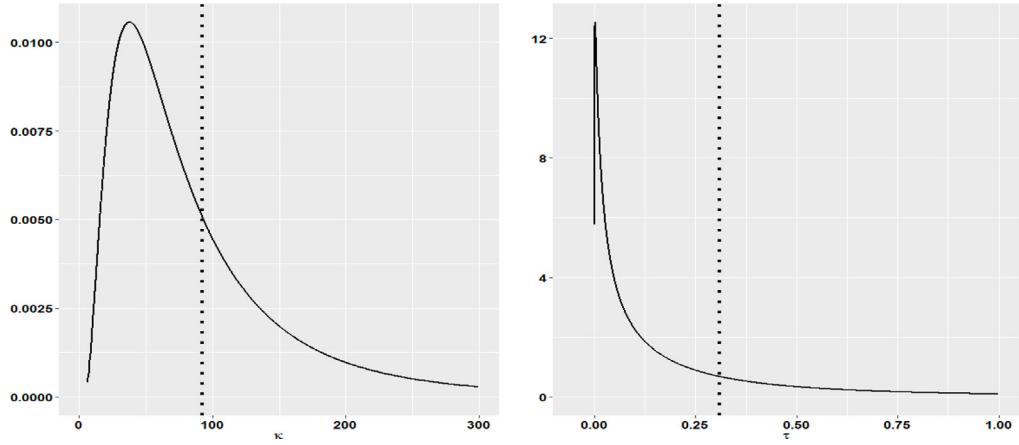



**Fig. A.12.** Marginal posterior distributions of hyperparameters $\kappa$ and $\tau$ for the spatial random field $S(.)$.

## References


Aghajani, M.A., Dezfoulian, R.S., Arjroody, A.R., Rezaei, M., 2017. Applying GIS to identify the spatial and temporal patterns of road accidents using spatial statistics (case study: Ilam province, Iran). Transp. Res. Proc. 25, 2126–2138.

Appelhans, T., Detsch, F., Reudenbach, C., Woellauer, S., 2016. Mapview - Interactive viewing of spatial data in R. In: EGU General Assembly Conference Abstracts. EPSC2016-1832.

Bivand, R., Rundel, C., Pebesma, E., 2017. RGEOS: Interface to geometry engine-open source (GEOS). R package version 0.3-26.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons Ltd.

Boulieri, A., Liverani, S., Hoogh, K.de., Blangiardo, M., 2016. A space–time multivariate Bayesian model to analyse road traffic accidents by severity. J.R. Stat. Soc.: Ser. A (Stat. Soc.) 180 (1), 119–139.

Briz-Redó, Á., Martınez-Ruiz, F., Montes, F., 2019. Identification of differential risk hotspots for collision and vehicle type in a directed linear network. Accid. Anal. Prev. 132, 105278.

Cantillo, V., Garcés, P., Márquez, L., 2016. Factors influencing the occurrence of traffic accidents in urban roads: A combined GIS-empirical Bayesian approach. DYNA 83 (195), 21–28.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transp. Res. B 46 (1), 253–272.

Chaudhuri, S., Juan, P., Mateu, J., 2022. Spatio-temporal modeling of traffic accidents incidence on urban road networks based on an explicit network triangulation. J. Appl. Stat. 1–22.

Chaudhuri, S., Moradi, M., Mateu, J., 2021. On the trend detection of time-ordered intensity images of point processes on linear networks. Comm. Statist. Simulation Comput. 1–13.

Curran-Everett, D., 2013. Explorations in statistics: The analysis of ratios and normalized data. Adv. Physiol. Educ. 37 (3), 213–219.

ESRI, 2021. ArcGIS, Environmental System Research Institute (ESRI). https://www.esri.com/.

Farmer, C.M., 2005. Temporal factors in motor vehicle crash deaths. Inj. Prev. 11 (1), 18–23.

Galgamuwa, U., Du, J., Dissanayake, S., 2019. Bayesian spatial modeling to incorporate unmeasured information at road segment levels with the INLA approach: A methodological advancement of estimating crash modification factors. J. Traffic Transp. Eng. (Engl. Ed.).

Guo, Y., Osama, A., Sayed, T., 2018. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accid. Anal. Prev. 113, 38–46.

Hannah, C., Spasić, I., Corcoran, P., 2018. A computational model of pedestrian road safety: The long way round is the safe way home. Accid. Anal. Prev. 121, 347–357.

Jegede, F., 1988. Spatio-temporal analysis of road traffic accidents in Oyo state, Nigeria. Accid. Anal. Prev. 20 (3), 227–243.

Juan, P., Mateu, J., Saez, M., 2012. Pinpointing spatio-temporal interactions in wildfire patterns. Stoch. Environ. Res. Risk Assess. 26, 1131–1150.

Karaganis, A., Mimis, A., 2006. A spatial point process for estimating the probability of occurrence of a traffic accident. In: ERSA Conference Papers. European Regional Science Association.

Khulbe, D., Sourav, S., 2019. Modeling severe traffic accidents with spatial and temporal features. In: ICONIP.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H., 2018. Advanced Spatial Modeling with Stochastic Partial Differential Equations using R and INLA. Chap- man; Hall/CRC.

Lindgren, F., 2012. Continuous domain spatial models in R-INLA. ISBA Bull. 19 (4), 14–20.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63 (19), http://dx.doi.org/10.18637/jss.v063.i19.

Lindgren, F., Rue, H., Lindstroötm, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.

Liu, C., Sharma, A., 2017. Exploring spatio-temporal effects in traffic crash trend analysis. Anal. Methods Accid. Res. 16, 104–116.

Liu, C., Zhang, S., Wu, H., Fu, Q., 2017. A dynamic spatiotemporal analysis model for traffic incident influence prediction on urban road networks. ISPRS Int. J. Geo-Inf. 6 (11), 362.

Loo, B.P.Y., Yao, S., Wu, J., 2011. Spatial point analysis of road crashes in Shanghai: A GIS-based network kernel density method. In: 2011 19th International Conference on Geoinformatics.

Moradi, M.M., Mateu, J., 2019. First-and second-order characteristics of spatio-temporal point processes on linear networks. J. Comput. Graph. Statist. 1–21.

OpenDataBCN, 2021. Open data BCN — Ajuntament de Barcelona's open data service. https://www.opendata-ajuntament.barcelona.cat/en.

Prasannakumar, V., Vijith, H., Charutha, R., Geetha, N., 2011. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. Procedia - Soc. Behav. Sci. 21, 317–325.

Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43 (1), 439–446.

R. Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

R. INLA Project, 2020. https://www.r-inla.org.

Risk mapping for the TEN-T in Croatia, Greece, Italy and Spain: Update, 2016. https://eurorap.org/risk-mapping-for-the-ten-t-in-Croatia-Greece-Italy-and-Spain-update/.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall. In: CRC Monographs on Statistics and Applied Probability), Chapman; Hall/CRC.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.
Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: A review. Annu. Rev. Stat. Appl. 4 (1), 395–421.
Shafabakhsh, G.A., Famili, A., Bahadori, M.S., 2017. GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. J. Traffic Transp. Eng. (Engl. Ed.) 4 (3), 290–299.
Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Stat. Sci. 32 (1), http://dx.doi.org/10.1214/16-sts576.
Smedt, T.D., Simons, K., Nieuwenhuyse, A.V., Molenberghs, G., 2015. Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. Arch. Public Health 73 (S1), http://dx.doi.org/10.1186/2049-3258-73-s1-o2.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (4), 583–639.
TomTom, 2021. Historical traffic data traffic stats. https://www.tomtom.com/products/traffic-stats/.
UNDP, 2021. Sustainable development goals: United nations development programme. https://www.undp.org/sustainable-development-goals.
Verdoy, P.J., 2019. Enhancing the SPDE modeling of spatial point processes with INLA, applied to wildfires. Choosing the best mesh for each database. Commun. Stat.-Simul. Comput. 1–34.
Wang, W., Yuan, Z., Yang, Y., Yang, X., Liu, Y., 2019. Factors influencing traffic accident frequencies on urban roads: A spatial panel time-fixed effects error modell (Y. Guo, Ed.). PLoS One 14 (4), e0214539.
WHO, 2019. Global Status Report on Road Safety 2018. World Health Organization.
Wikle, C.K., Berliner, L.M., Cressie, N., 1998. Hierarchical Bayesian space–time models. Environ. Ecol. Stat. 5, 117–154.
Williamson, A.M., Feyer, A.-M., 1995. Causes of accidents and the time of day. Work Stress 9 (2–3), 158–164.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.
Zhong-xiang, F., Shi-sheng, L., Wei-hua, Z., Nan-nan, Z., 2014. Combined prediction model of death toll for road traffic accidents based on independent and dependent variables. Comput. Intell. Neurosci. 2014, 1–7.
Zuur, A.F., leno, E.N., Saveliev, A.A., 2017. Beginners Guide to Spatial, Temporal and Spatial–Temporal Ecological Data Analysis with R-INLA: Using GLM and GLMM, Vol. 1. Highland Statistics Ltd.