



## Predicting cycling accident risk in Brussels: A spatial case–control approach



Grégory Vandenbulcke<sup>a</sup>, Isabelle Thomas<sup>a,b,\*</sup>, Luc Int Panis<sup>c,d</sup>

<sup>a</sup> Center for Operations Research and Econometrics (CORE), Université catholique de Louvain (UCL), 34 Voie du Roman Pays, Louvain-la-Neuve B-1348, Belgium

<sup>b</sup> National Fund for Scientific Research (FRS-FNRS), Belgium

<sup>c</sup> Flemish Institute for Technological Research (VITO), Physical Activity and Activity Based Research on Exposure (PHA<sup>2</sup>RE), Vlasmeer 7, Mol B-2400, Belgium

<sup>d</sup> Transportation Research Institute (IMOB), Universiteit Hasselt (UH), Agoralaan, Diepenbeek B-3590, Belgium

### ARTICLE INFO

#### Article history:

Received 12 November 2012  
Received in revised form 28 June 2013  
Accepted 1 July 2013

#### Keywords:

Bicycle accident  
Accident risk  
Bayesian framework  
Case–control strategy  
Risk factors  
Cycling

### ABSTRACT

This paper aims at predicting cycling accident risk for an entire network and identifying how road infrastructure influences cycling safety in the Brussels-Capital Region (Belgium). A spatial Bayesian modelling approach is proposed using a binary dependent variable (accident, no accident at location  $i$ ) constructed from a case–control strategy. Control sites are sampled along the 'bikeable' road network in function of the potential bicycle traffic transiting in each ward. Risk factors are limited to infrastructure, traffic and environmental characteristics.

Results suggest that a high risk is statistically associated with the presence of on-road tram tracks, bridges without cycling facility, complex intersections, proximity to shopping centres or garages, and busy van and truck traffic. Cycle facilities built at intersections and parked vehicles located next to separated cycle facilities are also associated with an increased risk, whereas contraflow cycling is associated with a reduced risk. The cycling accident risk is far from being negligible in points where there is actually no reported cycling accident but where they are yet expected to occur. Hence, mapping predicted accident risks provides planners and policy makers with a useful tool for accurately locating places with a high potential risk even before accidents actually happen. This also provides comprehensible information for orienting cyclists to the safest routes in Brussels.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

Bicycle use is nowadays promoted by public health and transportation specialists as an effective way to induce a shift to a healthier and environmentally sustainable lifestyle (Buehler et al., 2011; Chapman, 2007; de Nazelle et al., 2011; Elvik, 2009; Jacobsen, 2003; Polis, 2012; Vandenbulcke et al., 2009; WHO, 2002), even if there are adverse health effects due to the exposure to traffic exhaust or accidents (see e.g. Aertsens et al., 2010; Int Panis et al., 2010; Rojas-Rueda et al., 2011; Bos et al., 2013). There is a wide consensus that these risks are dwarfed by health benefits (de Hartog et al., 2010; Rabl and de Nazelle, 2011).

The risk of accident discourages people from cycling (McClintock and Cleary, 1996; Pucher et al., 1999; Parkin et al., 2007; Winters et al., 2011). Except in some countries benefiting

from the 'safety in numbers' effect, the risk for a cyclist to be injured in a road accident is high, compared to motorists (Reynolds et al., 2009). Unfortunately, the risk and consequences of (minor) bicycle accidents are poorly known (Aertsens et al., 2010; Int Panis, 2011; Rabl and de Nazelle, 2011). There is hence the necessity to monitor minor bicycle accidents and to study the risk factors, so that policies can be devised to maximise the health benefits.

In this paper, we aim at: (1) predicting cycling accident risk for a whole network, and (2) identifying how road infrastructure influences cycling safety in the Brussels-Capital Region (Belgium). Within this framework, autocorrelation and multicollinearity are controlled using adequate statistical methods. From a methodological point of view, accident data are coupled with control points that are sampled along the road network, proportionally to an estimation of the bicycle traffic.

This paper is organised as follows. Section 2 presents an overview of the literature into traffic accident research. Section 3 defines preliminary concepts, describes the models and motivates the use of a case–control strategy. Section 4 presents the studied area and provides some figures in terms of bicycle use and accident risks for cyclists. Section 5 describes the data used within the

\* Corresponding author at: Center for Operations Research and Econometrics (CORE), Université catholique de Louvain (UCL), 34 Voie du Roman Pays, Louvain-la-Neuve B-1348, Belgium. Tel.: +32 10 47 21 36.

E-mail address: [isabelle.thomas@uclouvain.be](mailto:isabelle.thomas@uclouvain.be) (I. Thomas).

modelling approach. Finally, Section 6 reports the main results that are further discussed and lead to recommendations in Section 7.

## 2. Overview of the literature into traffic accident research

### 2.1. Traffic accident research in general

Road accidents generally result from the combination and interaction between five categories of factors: human factors (e.g. driver behaviour), vehicle-related factors (e.g. size or state of the vehicle), infrastructure factors (e.g. crossroad design), traffic conditions (e.g. density of traffic) and environmental factors (e.g. weather) (Miaou et al., 2003; Li et al., 2007). Although considerable methodological improvements have been achieved in accident research during the last decades, the lack of information about the human factors, accident mechanisms and driver-related privacy issues have often hampered researchers to get in-depth knowledge about the exact cause and effect relationships with regard to the road accidents as a whole (Lord and Mannering, 2010). The body of the literature hence mainly focuses on examining the factors that affect either the frequency or the severity of accidents. Other studies investigate the association between the type of collision (e.g. rear-end accidents) and factors related to the accident mechanisms (Noland and Quddus, 2004; Lord and Mannering, 2010).

From a methodological point of view, much of the research into traffic accidents may be broadly classified into two groups, depending on the purpose of the study. First, *exploratory methods* may be used as an initial step to 'look at' the data, before performing explanatory methods. They aim at describing the accident data set using basic statistics (i.e. descriptive statistics, test statistics, odds ratios, etc.) and/or various spatial approaches (see e.g. Bailey and Gatrell, 1995; Levine et al., 1995a; Fotheringham et al., 2000; Myint, 2008; Shiode, 2008; Okabe et al., 2009). Second, *explanatory models* are commonly used to estimate the relative importance several factors may have on the occurrence and severity of accidents. Overall, three types of models are generally identified in the literature: the accident-frequency models, the accident-collision models and the accident-severity models. Concretely, the first category of model is applied to compute the probability of observing a definite number of accidents as a function of a set of factors (see e.g. negative-binomial models), while the second and third types of model focus on estimating the probability an accident falls into one definite class of collision or injury severity respectively (see e.g. multinomial or ordered logit models) (Ye and Lord, 2011). See Lord and Mannering (2010) for further information about accident-frequency models.

### 2.2. Focus on cycling accident research and relevant infrastructure factors

Much of the empirical work is recent (90s) and is mainly conducted in social sciences, medical and health care research and transportation (including traffic accident analysis, injury prevention, transport geography and engineering) (see Eluru et al., 2008; Reynolds et al., 2009 for a review of the literature). Examples of accident-frequency models applied to cycling accidents can be found in Wang and Nihan (2004), Hels and Orozova-Bekkevold (2007) and Schepers et al. (2011). On the other hand, empirical works aiming at comparing the impact of factors on different levels of injury severity for cyclists are far more common and can be found notably in Rodgers (1997), Klop and Khattak (1999), Kim et al. (2007) and Eluru et al. (2008). As regards accident-collision models, much of the work is – to our knowledge – quite recent and mainly aims at finding associations between the type of collision/manoeuvre (e.g. door-related and rear-end accidents) and a

set of factors. Relevant examples can be found in Pai (2011) and Yan et al. (2011).

Focusing on the impact of road infrastructures, most of the studies found that the risk of having a cycling accident can be influenced by the road environment as well as by close facilities. In particular, *intersections* are generally known as black spots for cyclists as well as for all road users (Wang and Nihan, 2004; ERSO, 2006; Quddus, 2008; BRSI, 2009a,b; Reynolds et al., 2009; Haque et al., 2010; Pei et al., 2010). They are places where the number of potential conflict points and the risk of having an accident are higher compared to the rest of the network (Wang and Nihan, 2004; Dumbaugh and Rae, 2009). In particular, signalised intersections may lead to an increased risk of slight injury for cyclists, although they are generally associated with reduced risks of being fatally or seriously injured when cycling. At the opposite of the effects observed for other types of road users, roundabouts are also mentioned as having an unfavourable effect on cyclist safety, leading to an increased risk of accident when they replace other types of intersections (Hels and Orozova-Bekkevold, 2007; Daniels et al., 2008, 2009; Møller and Hels, 2008; Reynolds et al., 2009). This effect is even worse when the roundabout replaces a signalised intersection, or when marked bicycle lanes are used instead of other design types (e.g. mixed traffic or grade-separated cycle lanes) (Daniels et al., 2009).

The number and risk of bicycle accidents are generally influenced by the *traffic conditions* (i.e. traffic composition, flows/volumes, etc.) observed at the time of the accident (see e.g. McClintock and Cleary, 1996; Klop and Khattak, 1999; Wang and Nihan, 2004; Hels and Orozova-Bekkevold, 2007; Kim et al., 2007; Eluru et al., 2008; Anderson, 2009). During peak hours, congestion increases not only the number and the risk of non-fatal accidents for cyclists but also the perception of danger (Parkin et al., 2007; Hels and Orozova-Bekkevold, 2007; Møller and Hels, 2008), mainly because of the increased complexity of the traffic situation, the more aggressive driving behaviour and the restricted space left to the cyclists (McClintock and Cleary, 1996; Li et al., 2007; Wang et al., 2009). It however decreases the risk of being seriously or fatally injured in a road accident, owing to a reduced speed differential between slow and fast transport modes (Klop and Khattak, 1999). During off-peak hours, the opposite situation is observed: high vehicle speeds may be achieved, hence increasing the risk of being seriously or fatally injured for cyclists (Klop and Khattak, 1999; Hels and Orozova-Bekkevold, 2007; Kim et al., 2007; Eluru et al., 2008). For example, Kim et al. (2007) found a more than 11-fold increase in the probability of fatal injury as the estimated vehicle speeds pass 65 km/h. The *type of collision partner* (e.g. car user) also plays a key role in the severity of the accident. Depending on their speed, dimension and weight, they may lead to different injury severities. Cars generally account for the largest share of vehicles colliding with cyclists and cause most of injuries for these latter (ERSO, 2006; Chong et al., 2010; Loo and Tsui, 2010), while lorries, buses, vans and sports utility vehicles are more frequently involved in serious and fatal cycling accidents (McCarthy and Gilbert, 1996; ERSO, 2006; Kim et al., 2007; Eluru et al., 2008; BRSI, 2009a; Pei et al., 2010; Yan et al., 2011).

Although there is no consensus about the actual safety effects of the *cycle facilities*, the findings in the literature overall show that it is safer to cycle on-road than on fully segregated cycle facilities (or off-road facilities) (Forester, 1994; Rodgers, 1997; Räsänen and Summala, 1998; Aultman-Hall and Hall, 1998; Aultman-Hall and Kaltenecker, 1999; Pucher et al., 1999; ERSO, 2006). Throughout their review on the safety of urban cycle facilities, Thomas and DeRobertis (2013) also concluded that unidirectional cycle facilities are generally safer than bidirectional cycle facilities at intersections, and that cycle facilities with effective intersection treatments reduce accidents and injuries on busy streets. Regarding *discontinuities in the bicycle network*, Krizek and Roland (2005) found they

introduce high levels of discomfort when they end either on the left side of the street, on parking lots, in large intersections or in a wider width of the curb lane. At the opposite, *contraflow cycling* is expected to be quite safe as motorists and cyclists face each other and keep a continuous eye contact (BRSI, 2013). Kim et al. (2007) indeed show that facing traffic reduces the probability of incapacitating and non-incapacitating injuries for cyclists. Roads with *parking facilities for motorised vehicles* and *garage entrances/exits* are expected to be black zones for cyclists, compared to roads without parking. Parked vehicles indeed restrict sight distances in some specific street and increase the risk of conflict with exiting/parking vehicles or with car doors in the case of parallel or longitudinal parking facilities (Greibe, 2003; Pai, 2011; Rifaat et al., 2011). *Traffic-calming measures* (e.g. 30 km/h areas, pedestrian areas, etc.) are expected to enhance the safety of cyclists in Brussels, as traffic accident research overall shows that high speed is related to a higher accident risk as well as to an increased injury when the cyclists collide with motorised vehicles. Finally, Kim et al. (2007) found that the presence of *institutional areas* (e.g. schools) increased the probability of incapacitating injury for cyclists (whereas it decreased the probability of having other injury severities).

Based on a review of the literature into traffic accident research, it is expected that other infrastructure factors may have an impact on the risk of having a cycling accident. *Bridges and tunnels* are indeed expected to be black spots for cyclists because sudden change may sometimes occur here in terms of infrastructures and road conditions (e.g. reduced space, bridges prone to ice development during winter, lower luminance levels for eyes in tunnels, etc.) (Wang and Nihan, 2004; Khan et al., 2009). Intuitively, the presence of *on-road or crossable tram tracks* is also expected to increase the occurrence of accidents for cyclists: cyclists often declare to get one of their cycle wheels stuck in the tracks, resulting in a loss of control of their bicycle (Cameron et al., 2001; BRSI, 2006). However, no reliable evidence is provided in the literature about such a risk. The presence of *public transport stops* (bus, tram, metro, etc.) is also expected to cause blackspots for cyclists since frequent pedestrian activity generally occurs around these stops (Pei et al., 2010). In particular, previous studies found bus stops and bus transit intensities as being significant factors associated with the presence of bicycle accidents (Quddus, 2008; Cho et al., 2009; Pei et al., 2010).

### 3. Methodological framework

#### 3.1. Preliminary notions

The notion of *risk* is the probability that the outcome of interest will occur, following a particular exposure of the population or study group. By analogy, the risk of having an accident for a cyclist is defined as the probability that this accident will occur, following the exposure of the cyclists in the traffic during a specified period of time. *Risk factors* refer to independent variables associated with an increased probability of bicycle accident. They are not necessarily causal factors and can be modified by intervention(s) aiming at reducing the accident probability (e.g. infrastructure modifications).

#### 3.2. Development of a case–control strategy

Accident-frequency and -severity models generally lead to well-known methodological issues (e.g. over- or under-dispersion of accident-frequency data) and require performing proper statistical approaches in order to avoid invalid inferences (Lord and Mannering, 2010). The implementation of these models within a spatial framework also implies that accident data are commonly aggregated over space and/or time (Liu and Jarrett, 2008), which

carries the danger to make wrong inferences about individual-level relationships on the basis of results obtained at an aggregated level of analysis (ecological fallacy). Instead of routinely modelling either accident severity or accident frequency, a case–control methodology is here implemented to model the accident risk for a cyclist along the road network. Except when both accident data and trip patterns are available through e.g. a detailed survey (see e.g. Harris et al., 2011; de Geus et al., 2012, 2014), such an estimation of accident risk presupposes the generation of controls, i.e. the creation of data reflecting the exposure of the population under study (i.e. the cyclists) to the outcome of interest (i.e. the accident). Once generated, such *controls* can be coupled with an accident database in order to produce a binary dependent variable as well as to enable the estimation of accident risks for cyclists through logistic modelling.

However, great care must be taken when generating controls as they are likely to bias the results if they are not rigorously selected. In the literature, *surveys* aiming at estimating such a risk generally raise several questions about their relevance in providing consistent parameter estimates since they often fail to select valid controls (see e.g. attendant comments for Lusk et al., 2011). In the past, many analyses even did not control for exposure and were therefore biased (Thomas and DeRobertis, 2013).

Main issues are the choice, the number and the representativeness of the controls. In general, little is said about why some control sites (e.g. reference streets) are selected rather than others, and to what extent such sites are spatially representative of traffic flows (or background exposure). Furthermore, surveys can be relatively time-expensive, even for small numbers of sampled controls (e.g. 20 controls). Implementing a methodology integrating these issues then matters within the scope of this paper.

#### 3.2.1. Lessons from ecology and epidemiology

Our methodological framework is inspired by research in *epidemiology* and *ecology*, taking advantage of their respective methodological strengths in modelling and especially in case–control strategies. The purpose of these latter is to provide some information on the absence of the outcome of interest (i.e. the disease or the observation of a species), which in turn enables to pair presences and absences in the same database in order to use common regression models based on binary data (e.g. logistic regression). This technique will be adopted here to obtain binary data (presence–absence) of cycling accidents and compute accident risk using logistic regression modelling.

In *case–control studies*, case events are those for which the outcome of interest has been observed (e.g. an accident) and controls are those in the same group/population without the outcome of interest. Controls provide an estimation of the background frequency of an exposure in a study group or population (i.e. the cyclists, or – ideally – the distance or travel time of trips carried out by these cyclists). The use of an appropriate control group matters as a poor choice of controls can lead to wrong inference. According to Grimes and Schulz (2005), controls should be: (1) free of the outcome of the interest, (2) representative of the population at risk, and (3) selected independently of the exposure of interest. For a small number of cases, it is also suggested to draw up to four times more controls than cases in order to improve the power of the study. Studies with case events often sample the controls from unknown or known population groups. In point process spatial models, researchers commonly use the spatial distribution of another common outcome/disease as control group, which is assumed to reflect the spatial distribution of the outcome/disease of interest. For instance, Diggle (1990), Hossain and Lawson (2009) and Lawson (2009) used cases of lung cancer as controls for modelling the spatial distribution of the cases of larynx cancer.

3.2.2. Applying a case–control strategy to accident analysis

Case events are here defined as locations where a bicycle accident occurred on the road network during a finite period of time, while controls are locations where no accident has been officially registered during the same period (2006–2008).

3.2.2.1. Exposure variable. The major problem for adopting a case–control strategy in road safety analysis is the lack of exposure data, from which controls can be selected as point events. Ideally, bicycle traffic would be the best, but it is seldom available at the segment level and for non-motorised transport modes. The only available data is often the number of inhabitants declaring to cycle in the population census, aggregated by areal spatial units. The solution proposed here to estimate an exposure variable is derived from ‘gravity-based’ concepts as conceptualised in accessibility research (see Geertman and Ritsema van Eck, 1995; Geurs and van Wee, 2004):

$$P_s = \sum_{t=1}^T a_t f(c_{st}) \tag{1}$$

where  $P_s$  is the ‘gravity-based index’ at location  $s$ ,  $T$  is the number of spatial units,  $a_t$  are the ‘opportunities’ (e.g. number of cyclists) in location  $t$ ,  $c_{st}$  is a measure of spatial separation between  $s$  and  $t$  (e.g. distance), and  $f(c_{st})$  is the impedance function, denoting the deterrent effect of spatial separation between  $s$  and  $t$  ( $s, t = 1, \dots, T$ ).  $P_s$  is hence a measure of accessibility in  $s$  to all opportunities  $a$  in  $t$ , weighted by the spatial separation between  $s$  and  $t$ . The impedance function here adopts the form of a negative exponential function, which is the most preferred function in the literature (Haynes et al., 2003).

The potential index specification is here adapted for estimating the potential bicycle traffic per spatial unit  $s$ , i.e. the potential exposure of cyclists to accidents and here named ‘Potential Bicycle Traffic Index’ (PBTI). Evidence in the literature lends strong support to the choice of such a potential index as a proxy for the bicycle traffic, as this index is often correlated with trip generation and closely reflects actual behaviours in terms of the induced demand for travel (Haynes et al., 2003; Thill and Kim, 2005).

3.2.2.2. Selection of controls. The random selection of controls is stratified as a function of the PBTI: the number of controls varies from one spatial unit to another and the number of controls will increase with bicycle traffic. Formally, the number of controls  $m_s$  to be drawn in a spatial unit  $s$  is:

$$m_s = \frac{P_s^*}{\sum_{s=1}^T P_s^*}, \quad M_0 = r_s \cdot M_0 \tag{2}$$

where  $P_s^*$  is the PBTI,  $M_0$  the total number of controls which is here four times greater than  $n_{acc}$  the number of geocoded accidents (Grimes and Schulz, 2005), and  $r_s$  is defined as the ‘relative potential index’ at location  $s$  and denotes the relative potential intensity of the bicycle traffic at  $s$  compared with all other locations. Control points are constrained to be drawn on the bikeable road network and outside linear buffered zones around the accidents in order to avoid sampling from these zones. As a consequence, the dependent variable is derived from the combination of case events (occurrence of a bicycle accident at location  $i$  along the network) and controls (no bicycle accident at  $s$ ). This makes the use of a logistic regression possible if risk factors are identified for both cases (1) and controls (0).

3.3. Modelling strategy

A Bayesian modelling approach is preferred for its ability to incorporate prior expert knowledge and to deal with

nuisance/random parameters in complex models (Miaou et al., 2003; Kéry, 2010). Unlike other inferences that generally give point estimations, the Bayesian approach allows the parameters to be characterised as random variables and provides direct probability statements about these (Kéry, 2010; Pei et al., 2010). Probability is hence expressed as the uncertainty we have about the magnitude of a parameter, which makes the Bayesian inference more intuitive compared with the conventional approaches.

3.3.1. Hierarchical modelling and accident risk model

Hierarchical Bayes allows accommodating the inherent stochasticity of some models – e.g. spatial models – owing to its structure in several hierarchical stages (Zhu et al., 2006; Bivand et al., 2008; Ntzoufras, 2009). The prior parameters are supposed to be random variables and may depend on prior distributions that may in turn depend on other parameters (the ‘hyperparameters’) at a second level of the hierarchy. As the dependent variable is binary, a two-stage conditional Bernoulli model with a logistic link is appropriate for predicting the probability of having a bicycle accident at location  $i$  ( $i = 1, \dots, n; n = n_{acc} + M$ ):

$$Y_i \sim \text{Bernoulli}(p_i) \tag{3}$$

$$\text{logit}(p_i) = \log \left[ \frac{p_i}{1 - p_i} \right] = \alpha + x_j \beta \tag{4}$$

where  $Y_i$  is the dependent variable ( $Y_i = 1$  if an accident occurred at location  $i$ ;  $Y_i = 0$  otherwise),  $p_i$  is the probability of having a bicycle accident at location  $i$ ,  $\alpha$  is the intercept of the model,  $\beta$  is the vector of parameters, and  $x_i$  is the vector of risk factors (explanatory variables). This is the first stage of the so-called ‘accident risk model’.

In the second stage, highly uninformative prior distributions are generally assigned to  $\alpha$  and  $\beta$  when there is no prior information about the parameters. In general, normal distributions with mean  $\mu = 0$  and precision  $\tau = 1.10^{-6}$  ( $=1/\text{variance}$ ) are specified for the parameters  $\alpha$  and  $\beta$ .

3.3.2. Autoregressive and autologistic risk models

Incorporating random effects in the model provides a robust basis for inference when spatial autocorrelation and overdispersion are both present (see e.g. Miaou et al., 2003; Zhu et al., 2006; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Haining et al., 2009; Lawson, 2009; Lord and Mannering, 2010). Given that the bicycle accidents spatially concentrate on the network, a spatial Bayesian specification of the accident risk model (including the random effects) is here proposed. An intrinsic conditional autoregressive (ICAR) model was implemented, but did not improve our results (see Vandenbulcke, 2011 for details).

Another spatial specification for predicting accident risk is the autologistic model. This model intends to capture the effect of spatial autocorrelation by including – at the first stage of the Bayesian hierarchy – an additional variable called the ‘autocovariate’ (Flahaut, 2004; Dormann, 2007; Miller et al., 2007). Eq. (4) is hence re-specified as follows:

$$\text{logit}(p_i) = \alpha' + x_i \beta + \lambda S_i \tag{5}$$

where  $S_i$  is the autocovariate for the bicycle accident  $i$  and  $\lambda$  is the parameter for the autocovariate. Such an autocovariate is generally defined as a weighted sum (or average) of the observations in the neighbourhood (Dormann et al., 2007):

$$S_i = \sum_{j \neq i} w_{ij}^* y_j^* \quad (\text{weighted sum}), \quad \text{or}$$

$$S_i = \frac{\sum_{j \neq i} w_{ij}^* y_j^*}{\sum_{j \neq i} w_{ij}^*} \quad (\text{weighted average}) \tag{6}$$

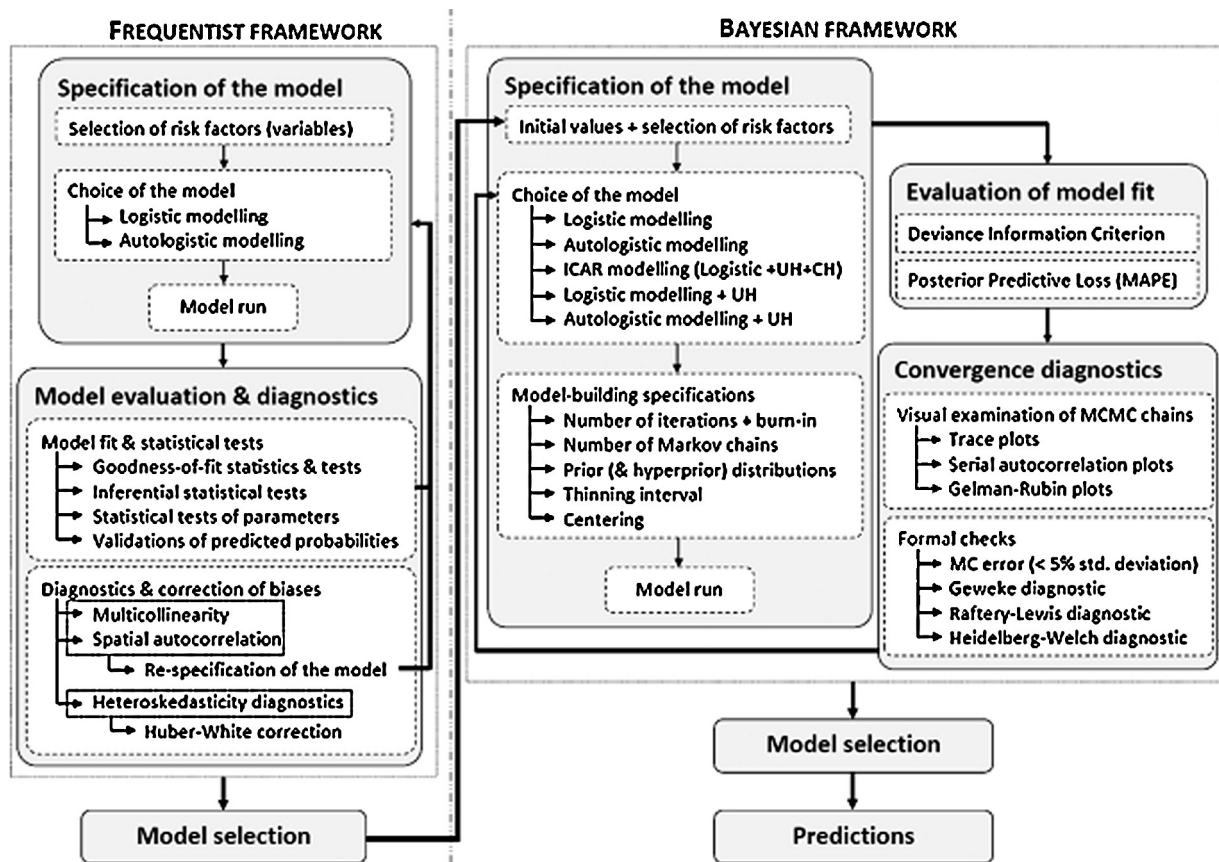


Fig. 1. Modelling strategy (UH: uncorrelated heterogeneity; CH: correlated heterogeneity).

where  $w_{ij}^*$  are the weights assumed to represent the relationship existing between location  $i$  and its neighbours  $j$ , and  $y_j^*$  are the response values observed for these neighbours ( $j$ ). To optimise the model inference, several specifications were tested using different functional forms for the network distance between  $i$  and  $j$ . A spatial weight matrix accounting for a distance-based relationship between the bicycle accident  $i$  and the nearest bicycle accidents  $j$  (1st order neighbourhood) was the best to capture the unexplained variance associated with the presence of spatial autocorrelation:  $w_{ij}^* = e^{-d_{ij}}$  if  $j$  is a 1st order neighbour/accident,  $w_{ij}^* = 0$  otherwise ( $j$  is a 2nd order neighbour/accident or more). Similarly to the logistic specification, normal prior distributions are selected at the 2nd level of the hierarchy.

### 3.3.3. Initial values and model selection

Model selection is summarised in Fig. 1. In the first step, logistic and autologistic regressions are performed within a frequentist framework in order to get the initial values. An overall model evaluation is then carried out using: (1) inferential statistical tests (Likelihood ratio and Wald test), (2) statistical tests of the parameters of the risk factors (Wald chi-square statistic), (3) goodness-of-fit statistics (Log Likelihood (LL), Akaike's Information Criterion (AIC)) and tests (Hosmer–Lemeshow (HL) and Le Cessie–Houwelingen tests), and (4) validations of predicted probabilities ( $c$  statistic). Diagnostics for multicollinearity (Variance Inflation Factors and condition indices), spatial dependence of the dependent variable (join-count test statistics under non-free sampling) and spatial autocorrelation of the residuals (Moran's  $I$ ) are also performed and influenced our choice of the risk factors in the models. Heteroskedasticity – when present – is corrected by using the Huber-White method.

In the second step, the initial values of the Bayesian models are determined on the basis of the parameter estimates obtained in the 'best' frequentist models. Starting from initial values that are close to the actual values of the parameters speeds up the convergence of the models.

In the last step, the statistical fit of the Bayesian models is computed to compare their performance and select a better-fitting model. Two goodness-of-fit measures are applied: the Deviance Information Criterion (DIC) and the Mean Absolute Predictive Error (MAPE) (Lawson, 2009). The DIC is a generalisation of the AIC and expresses a trade-off between the model fit and the model complexity (Law et al., 2006; Lawson, 2009; Kéry, 2010). MAPE is a posterior predictive loss measure, aiming at comparing the predictive ability of the models. Models with small DIC and MAPE values are preferred.

### 3.3.4. Convergence diagnostics (Bayesian framework)

The Markov chain starts from an initial value attributed to each parameter, which is either arbitrary or computed from a frequentist method. After a suitable number of iterations, the chain is expected to reach an equilibrium distribution, i.e. the convergence. The first draws obtained before convergence are called the 'burn-in period' and are discarded since they are not representative of the equilibrium distribution (Gelman and Hill, 2007; Bivand et al., 2008; Lawson, 2009; Ntzoufras, 2009; Kéry, 2010). Summary statistics of the posterior distribution are then computed directly from the remaining simulations.

Several diagnostics are run to monitor convergence. They consist of either a simple visual examination of different plots (trace plots, serial autocorrelation plots and Gelman–Rubin statistic plots), or more formal checks that are carried out to monitor

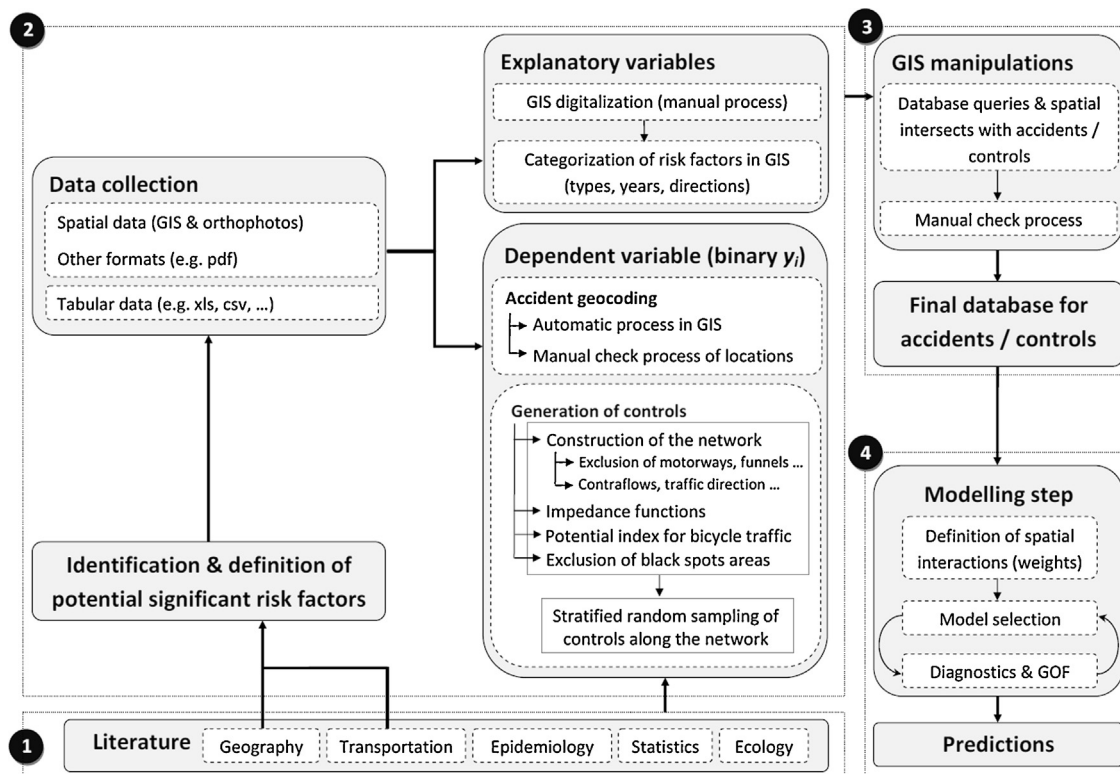


Fig. 2. Data collection – conceptual framework: (1) literature; (2) Data; (3) GIS manipulations; (4) modelling step & outputs.

convergence (Monte Carlo (MC) errors, Geweke statistics, and Raftery–Lewis and Heidelberg–Welch diagnostics).

All frequentist statistics were run in SAS Enterprise Guide 4.2 and R 2.12.1, while Bayesian statistics were computed with R2WinBUGS (Sturtz et al., 2005). Convergence diagnostics and output analyses were performed using CODA (Best et al., 1995).

#### 4. Spatial context: Brussels

The studied area is the Brussels–Capital Region (BCR), which is a highly urbanised area centrally located in Belgium with a population density of about 7,000 inhabitants/km<sup>2</sup> and a wide range of facilities. The city sprawls out of the regional boundaries, into less urbanised municipalities. Such a continuous sub- and peri-urbanisation process is at the root of an increasing motorisation rate that generates car trips converging on the city-centre. Reducing car use is hence one of the major priorities in the Brussels' mobility plan.

However, encouraging a modal shift from car to cycling implies enhancing bicyclists' safety in the city. Since more than 62% of the trips in Brussels are shorter than 5 km, such a shift from the car to the bicycle is far from being unrealistic but also far from being a reality. Indeed, approximately 2% of the inhabitants living in Brussels used the bicycle as an utilitarian mode of transport in 2001 (all purposes, working days only) (Hubert and Toint, 2002). Although bicycle use increased since 2002, only 4% of the total population in Brussels is estimated to use the bicycle in 2012. At the same time, the number of bicycle accidents increased in Brussels, which resulted in a quite stable (but still relatively high) risk of having an accident when biking (Vandenbulcke, 2011).

#### 5. Data collection

Fig. 2 illustrates the data collection step. A review of the literature on risk factors was first conducted (Vandenbulcke, 2011).

Second, data were collected and pre-processed with GIS techniques. While digitising the data, special attention was paid to the direction, year and type of spatial data (e.g. cycle facilities). Bicycle accidents (cases) were geocoded along the road network and were completed using a set of controls. Third, the risk factors result from crossing the digitised data with the binary dependent variable  $Y_i$  into a GIS. Fourth, the final database with accidents, controls and their respective risk factors is used for modelling the risk of bicycle accident along the Brussels' road network. By trial and error, the best models were selected and used to compute predictions for a specific 'bikeable' road trajectory.

Most spatial data related to the road network and to the risk factors are provided by the Brussels Regional Informatics Center (BRIC), using UrbIS<sup>®</sup> databases. The road network is modelled as a connected 'non-planar' graph (the relative heights of links and the direction of travel are considered). In order to account for cyclists living in the Brussels' periphery (and hence avoid edge effects), the road network is 35 km buffered around the BCR boundaries.

##### 5.1. Accident data

###### 5.1.1. Accidents ( $Y_i = 1$ )

In Belgium, road accidents with casualties are registered by the Police and compiled by the Directorate-General for Statistics and Economic Information (DGSEI). 644 bicycle accidents were censused in the BCR over the 2006–2008 period. 93% could successfully be geocoded using a semi-automatic process, after having excluded those for which the street side (of the accident) was totally unknown (see Vandenbulcke, 2011 for details). In Belgium, it however seems that only 15% of the bicycle accidents are officially registered (Doom and Derweduwen, 2005); the registration of cycling accidents is particularly biased by their severity and location (Vandenbulcke, 2011). In particular, single-bicycle accidents are more likely to be underreported and affected by specific types of risk factors (e.g. tram tracks or longitudinal parking facilities). In

such cases, the cyclist often does not feel the need to call the police, and then that there is often no official record of the accident.

### 5.1.2. Controls ( $Y_i = 0$ )

Controls are randomly selected on the 'bikable' road network; their sampling is stratified by spatial unit  $s$  (statistical wards) as a function of the potential bicycle traffic ( $PBTI$ ). Data from the 2001 Socio-Economic Census (DGSEI) are used to estimate the  $PBTI$  for Brussels:

$$PBTI_s^* = a_s + \sum_{t=1}^T a_t \cdot [\zeta_t \cdot \exp(-\delta_t \cdot d_{st}) + \eta_t \cdot d_{st} \cdot \exp(-\varepsilon_t \cdot d_{st})] \quad (7)$$

where  $t$  stands for the statistical wards in the neighbourhood of  $s$  ( $s, t = 1, \dots, T; t \neq s$ ),  $a$  is the number of cyclists commuting to work or school and living in  $s$  (or  $t$ ),  $d_{st}$  is the number of kilometres measured along the bikable road network, between the centroids of  $s$  and  $t$ .  $\zeta_t$ ,  $\delta_t$ ,  $\eta_t$ ,  $\varepsilon_t$  are the parameters of the impedance function attributed to the statistical ward  $t$ . Note that Eq. (7) assumes that all statistical wards  $t$  have the same level of attraction, and thus that there is no preferential direction of travel for cyclists; in other words, routes with segregated cycling facilities are not taken into account in the computation of the  $PBTI$ .

For each municipality  $o$  (each containing several statistical wards  $t$ ), the parameters of the impedance function ( $\zeta_t$ ,  $\delta_t$ ,  $\eta_t$ ,  $\varepsilon_t$ ) are calibrated on the basis of an observed impedance function of cycling trips (i.e. the observed proportion of cycling trips as a function of distance). In other words, the values of the parameters calibrated at the municipal scale are assigned to all statistical wards  $t$  that are contained in  $o$ . Such a method assumes various travel behaviours according to the place of residence of the cyclists. Exploratory analyses (not reported here) suggest that cyclists living in the city centre travel shorter distances than those living in more peripheral locations.

The  $PBTI$  provides an estimation of the potential number of cyclists stopping or transiting in  $s$  during a working day. It refers to the number of cyclists living in  $s$  plus a number of cyclists living in the neighbourhood and being likely to travel the distance between their place of residence  $t$  and  $s$ . A visual check (Fig. 3) suggests that the  $PBTI$  is close to the actual spatial patterns of bicycle traffic reported in surveys. These latter are undertaken yearly by cyclists' advocacy groups and aim at evaluating bicycle traffic at 20 locations selected in Brussels. High  $PBTI$  values are observed in the eastern parts of the 'Pentagon' and 'First Crown', which corresponds to the places where cycling trips are the most common in Brussels (see Vandenbulcke, 2011). At the opposite, low values are obtained for the southern and western parts of the BCR, where few cyclists are observed. Measures of central tendency (mean centre, central feature) and spatial dispersion (standard distance, standard deviational ellipse) confirm the validity of the results.

The number of controls to be drawn in each statistical ward  $s$  is weighted as a function of the  $PBTI$ : the higher the  $PBTI$ , the higher the number of controls to be drawn. Given that  $M_0 = 4 \cdot n_{acc} = 2400$  ( $n_{acc}$ : number of bicycle accidents in Brussels), the total number of controls  $m_s = 2416$ . These are then drawn at locations  $i$  along the bikeable road network (stratified per  $s$ ;  $i$  is contained in  $s$ ), from which we removed the black spots of bicycle accidents in order to avoid sampling nearby bicycle accidents. Black spots are obtained by performing a Network Kernel Density Estimation provided by SANET v.4, using a 100 m bandwidth (Okabe et al., 2009). As shown in Fig. 4, cycling accidents are more likely to be observed at intersections (56% of all accidents) and on roads with dense motorised traffic and/or tram tracks. Residential roads (equipped with traffic-calming measures) seem less prone to generate cycling accidents.

As a final step, a year (2006, 2007 or 2008) and a traffic direction are randomly assigned to the controls. This allows associating the

controls with the spatial risk factors, as these latter may be built at a definite moment over the period of interest (e.g. in 2008) or may be reported at one street side only (depending on the location where the control is located). Controls ( $Y_i = 0$ ) are finally appended to the geocoded bicycle accidents ( $Y_i = 1$ ) in the same database. Interestingly, only 18% of controls were sampled at intersections, partly because of the fact that: (1) 56% of the accidents occurred at intersections (so excluding these from sampling, as they are black spots); and (2) random sampling means that controls are more likely to be sampled along road sections.

## 5.2. Risk factors

This paper only focuses on infrastructure and traffic conditions, because the other factors are not available for the controls. Table 1 lists all risk factors used in this study as well as their definition, units and data sources. This table also refers to previous works highlighting the influence of each risk factor on the bicycling safety.

### 5.2.1. Infrastructure factors

Based on the review made in Section 2.2, the following factors were retained: bridges/tunnels, traffic-calming areas, intersections, tram tracks and public transport stops, cycle facilities and discontinuities along the bicycle network, parking facilities and garage entrances/exits (for motorised vehicles), proximity to activities and public services, and streets where contraflow cycling is permitted (see Vandenbulcke et al., 2011 for further details on risk factors). The presence, evolution and street side where the infrastructures are built are controlled over the 2006–2008 period, using orthophotos (BRIC, Google Earth) and accident data (DGSEI). Information on traffic direction in streets where contraflow cycling is allowed is also encoded in order to precise if the cyclist is facing motorised traffic or not.

Important notes must be mentioned here as regards some of these risk factors. First of all, intersections are modelled as 'zones' instead of points, as the probability that a control falls exactly on a point is almost zero. For roundabouts and traffic lights, intersections are manually delineated on the basis of specific features (e.g. stop lines, etc.), while for other types of intersections, zones are defined as 10 m linear buffers around the exact intersection point. Second, a complexity index is computed for each accident/control in line with the Elvik's law on complexity so that it provides a proxy for road legibility (which depends on the number of road signs, markings and transiting road users) (Elvik, 2006). Such an index consists of the sum of all road links radiating outwards (in all possible network directions) from the accident/control location, over a certain distance or 'bandwidth'. Selected bandwidth values are 10, 20, 30, 40, 50, 75 and 100 m.

Third, a time-consuming digitisation process is performed to account for all cycle facilities and discontinuities (e.g. curb extensions) in the parking facilities observed along the road network. Network distances are then computed between each accident/control and the closest cycle and/or parking facility. A crossing risk factor is also constructed through digitising cycle facilities located nearby parking areas (i.e. at a distance < 0.9 m from the closest longitudinal parking area or exit).

Fourth, centroids of garage entrance/exits (modelled as lines into GIS) are computed from BRIC data, which allows computing (1) the number of garage entrances/exits within 100 m (network distance) from the place of the accident/control, and (2) the network distance (m) between each accident and the closest garage entrance/exit. Fifth, three kinds of traffic calming measures are here considered: 30 km/h, residential (20 km/h) and pedestrian areas (prohibited to motorised traffic outside delivery hours, but also to cyclists in some cases). Last but not least, it is not mentioned in the data if the cyclists actually crossed the tram tracks or if they cycled

**Table 1**  
List of risk factors.

Variable	Definition	Categories	Data source	References in the literature (examples)
<i>Infrastructure</i>				
Bridge	1 if the accident/control occurred on a bridge (with safeguards on both sides), 0 otherwise	–	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc) & Google Earth (2004, 2007, 2009)	Khan et al. (2009)
Tunnel	1 if the accident/control occurred in a tunnel or below an elevated infrastructure, 0 otherwise	–	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc) & Google Earth (2004, 2007, 2009)	Wang and Nihan (2004)
Traffic-calming area	1 if the accident/control occurred in a specified type of traffic-calming area, 0 otherwise	1 = 30 km/h area, 2 = pedestrian area, 3 = residential area, 4 = all types of traffic-calming areas (i.e. 1–3)	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, cycling map (BCR 2006 & 2008), Ministry of the Brussels-Capital Region (IRIS 2), City of Brussels (Map of the “comfort area”)	Pucher and Dijkstra (2003), Pucher and Buehler (2008), Rifaat et al. (2011)
Crossroad	1 if the accident/control occurred in a specified type of crossroad, 0 otherwise	0 = no crossroad, 1 = yield/stop signal, 2 = right-of-way, 3 = traffic light, 4 = roundabout, 5 = crossroad with right-turn, 6 = pedestrian light	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc), Google Earth (2004, 2007, 2009)	Wang and Nihan (2004), Hels and Orozova-Bekkevold (2007), Daniels et al. (2008), Eluru et al. (2008), Møller and Høls (2008); Qudus (2008), Daniels et al. (2009), Dumbaugh and Rae (2009), Reynolds et al. (2009), Haque et al. (2010), Pei et al. (2010)
Complexity index	Complexity index at the place of the accident/control, within a specified bandwidth (m)	10, 20, 30, 40, 50, 75, 100 m	Own computation, from BRIC (Brussels UrbIS)	Wang and Nihan (2004), Elvik (2006), Dumbaugh and Rae (2009), Dai et al. (2010)
Tram tracks	1 if the accident/control occurred on or close to a specified type of tram track infrastructure, 0 otherwise	0 = no tram track, 1 = crossing tram tracks, 2 = tram tracks in crossable reserved lanes, 3 = on-road tram tracks	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc), Google Earth (2004, 2007, 2009), STIB-MIVB/BSI	Hedelin et al. (1996); Cameron et al. (2001), Unger et al. (2002)
Cycle facility	1 if the accident/control occurred on a specified type of cycle facility, 0 otherwise	0 = no cycle facility, 1 = unidirectional separated cycle lane, 2 = bidirectional separated cycle lane, 3 = marked cycle lane, 4 = suggested cycle lane, 5 = bus and bicycle lane	Own digitisation and computation, from FPS Economy (2006–2008), BRIC (Brussels UrbIS 2007–2008, GeoLoc), cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)	Forester (1994), McClintock and Cleary (1996), Rodgers (1997), Räsänen and Summala (1998), Aultman-Hall and Kaltenecker (1999), Pucher et al. (1999), de Lapparent (2005), Daniels et al. (2009), Parkin and Meyers (2010)
Parking area	1 if the accident/control occurred close to a specified type of parking area, 0 otherwise	0 = no parking area, 1 = longitudinal parking, 2 = head-in angle parking, 3 = back-in angle parking, 4 = parking perpendicular to the road, 5 = other type of parking area	Own digitisation and computation, from FPS Economy (2006–2008), BRIC (Brussels UrbIS 2007–2008, GeoLoc), Google Earth (2004, 2007, 2009)	Greibe (2003), Krizek and Roland (2005), Parkin et al. (2007), Pai (2011), Rifaat et al. (2011)
Proximity parking-cycle facility	1 if the accident/control occurred on a specified type of cycle facility, very close to a parking area ( $d \leq 0.9$ m, and outside a crossroad), 0 otherwise	1 = unidirectional separated cycle lane, 2 = bidirectional separated cycle lane, 3 = marked cycle lane, 4 = suggested cycle lane, 5 = bus and bicycle lane, 6 = all types of cycle facilities (i.e. 1–5)	Own digitisation and computation, from FPS Economy (2006–2008), BRIC (Brussels UrbIS 2007–2008, GeoLoc), cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)	Pai (2011)
Contraflow cycling	1 if the accident/control occurred in a contraflow cycling and in the opposite direction of motorised vehicles (i.e. in the direction of the contraflow), 0 otherwise	–	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc), cycling map (BCR 2006 & 2008), OneWayMap application), Google Earth (2004, 2007, 2009)	–
Number of garages ( $\leq 100$ m)	Number of garages (in a specified range) over a network distance $\leq 100$ m from the place of the accident/control	0, 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, > 70 garage(s)	Own computation, from BRIC (Brussels UrbIS 2007–2008)	Greibe (2003), Rifaat et al. (2011)



Table 1 (Continued)

Distance garage	Network distance (m) to the closest garage	–	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Greibe (2003)</a> , <a href="#">Rifaat et al. (2011)</a>
Distance discontinuity	Network distance (m) to the closest discontinuity (on cycle facilities)	–	Own digitisation and computation, from BRIC (Brussels UrbIS 2007–2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009))	
Distance parking area	Network distance (m) to the closest type of parking area	1 = park-and-ride, public or private parking area, 2 = delivery parking, 3 = diplomatic corps parking, 4 = disabled parking, 5 = taxi parking, 6 = all types of parkings (i.e. 1–5)	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Greibe (2003)</a> , <a href="#">Krizek and Roland (2005)</a> , <a href="#">Parkin et al. (2007)</a> , <a href="#">Pai (2011)</a> , <a href="#">Rifaat et al. (2011)</a>
Distance public transport stop	Network distance (m) to the closest type of public transport stop	1 = bus stop, 2 = tram stop, 3 = all types of public transport stops (i.e. 1–2)	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Quddus (2008)</a> , <a href="#">Cho et al. (2009)</a> , <a href="#">Pei et al. (2010)</a>
Distance public administration	Network distance (m) to the closest type of administrative building	1 = European administrative building, 2 = regional administrative building, 3 = all types of administrative buildings (i.e. 1–2)	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Levine et al. (1995a,b)</a>
Distance school	Network distance (m) to the closest type of school	1 = primary or secondary school, 2 = international primary or secondary school, 3 = superior school, 4 = all types of schools (i.e. 1–3)	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Kim et al. (2007)</a>
Distance industrial estate	Network distance (m) to the closest industrial estate	–	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Levine et al. (1995a,b)</a>
Distance shopping centre	Network distance (m) to the closest shopping centre/mall	–	Own computation, from BRIC (Brussels UrbIS 2007–2008)	<a href="#">Levine et al. (1995a,b)</a>
Distance playground	Network distance (m) to the closest playground	–	Own computation, from BRIC (Brussels UrbIS 2007–2008)	–
<i>Traffic</i>				
Car traffic (06:00 a.m.–10:59 p.m.)	1 if the accident/control occurred on a road with intensity of car traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	1 = very low car traffic; . . . ; 5 = very high car traffic	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007–2008)	<a href="#">McClintock and Cleary (1996)</a> , <a href="#">Klop and Khattak (1999)</a> , <a href="#">Graw and König (2002)</a> , <a href="#">Wang and Nihan (2004)</a> , <a href="#">Hels and Orozova-Bekkevold (2007)</a> , <a href="#">Kim et al. (2007)</a> , <a href="#">Li et al. (2007)</a> , <a href="#">Parkin et al. (2007)</a> , <a href="#">Eluru et al. (2008)</a> , <a href="#">Møller and Hels (2008)</a> , <a href="#">Quddus (2008)</a> , <a href="#">Anderson (2009)</a> , <a href="#">Daniels et al. (2009)</a> , <a href="#">Dumbaugh and Rae (2009)</a> , <a href="#">Reynolds et al. (2009)</a> , <a href="#">Wang et al. (2009)</a> , <a href="#">Haque et al. (2010)</a> , <a href="#">Pei et al. (2010)</a> , <a href="#">Pai (2011)</a> , <a href="#">Yan et al. (2011)</a>
Van traffic (06:00 a.m.–10:59 p.m.)	1 if the accident/control occurred on a road with intensity of van traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	1 = very low van traffic; . . . ; 5 = very high van traffic	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007–2008)	
Lorry/truck traffic (06:00 a.m.–10:59 p.m.)	1 if the accident/control occurred on a road with intensity of truck traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	1 = very low truck traffic; . . . ; 5 = very high truck traffic	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007–2008)	

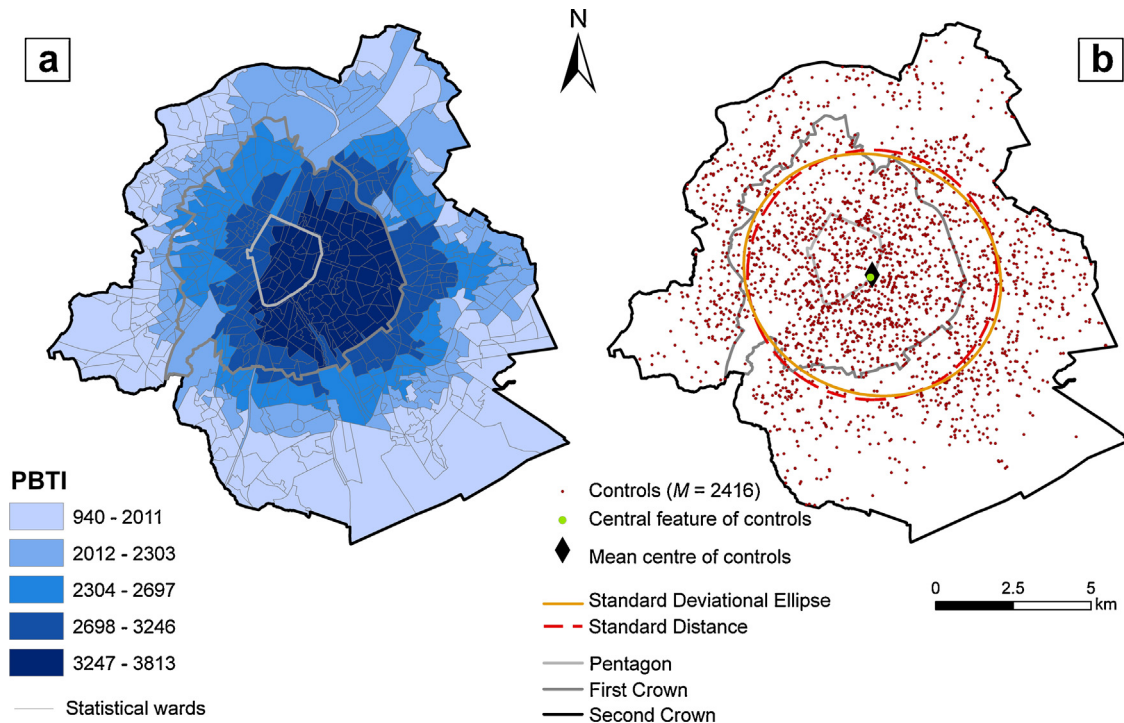


Fig. 3. Spatial distribution of: (a) the exposure variable (PBTI) and (b) the control points.

parallel to these; accident reports often give insufficient information to infer what might have been the cyclist's direction relative to tram tracks.

5.2.2. Traffic conditions

Data on 2006 motorised traffic volumes are provided by the Brussels' Institute for Environmental Management (IBGE-BIM). They are expressed in terms of private car equivalent units and

are measured for specific vehicle types, road links and time intervals. Given that traffic modelling is computed for major roads only, a categorisation of the data into 5 classes is carried out based on the methodology of Natural Breaks to account for the traffic volume on minor roads (class 1: very low traffic level; ...; 5: very high traffic level; minor roads are systematically classified into class 1). Such a categorisation also allows addressing the potential bias related to traffic modelling. Modelling is indeed not perfect due to the

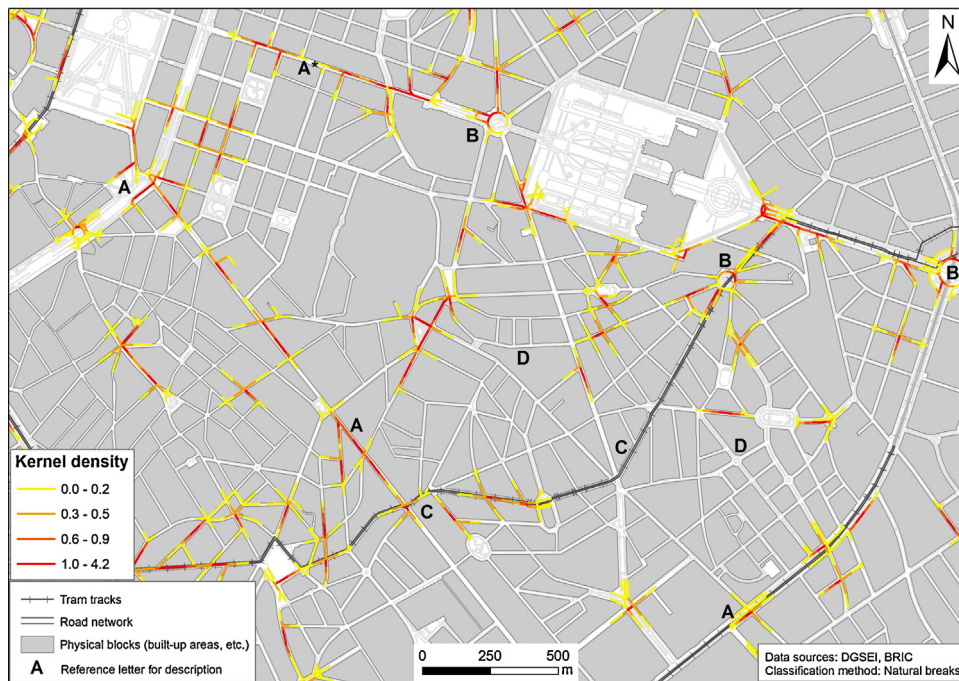


Fig. 4. Black spots of bicycle accidents (2006–2008) in the Brussels' European district. Traffic conditions: A, major roads, with high capacity and dense motorised traffic volume; A\*, *idem*, but with a separated cycle facility; B, large roundabouts, with dense motorised traffic volume; C, road with tram tracks (here: on-road and crossable); D, residential wards, with traffic-calming measures (speed humps, 30 km/h areas, etc.).

non-exhaustiveness of data and the irrational behaviour of road users, and the categorisation into 5 classes allows some variation (and bias) into the traffic values.

Note that three vehicle types are here considered (i.e. car, van and lorry), because they are the most frequent collision partners in accidents with cyclists (see Section 2.2). Also, the street side of the accident/control is taken into account when assigning traffic levels, except at intersections where the maximum traffic level is considered.

### 5.2.3. Interaction variables and ignored risk factors

Most of the previous risk factors are crossed and included in the models through trial and error processes. Crossing two risk factors may be advantageous as it may improve the validity of the models and inference (e.g. by obtaining a significant interaction variable, whereas the two risk factors taken separately may appear as being insignificant). Furthermore, it allows integrating some risk factors that might have been collinear if uncrossed. For instance, crossing road sections/intersections with other risk factors allows integrating combinations of factors within the modelling process, whereas multicollinearity would have been detected if both road sections and intersections were included as such. It then allows providing recommendations focused on specific types (or combinations) of infrastructures (e.g. marked cycle lanes built in roundabouts and road sections with separated cycle lanes built close to parking facilities may be integrated in one analysis).

Interaction variables are here noted using the following notation: '[Risk factor 1] & [Risk factor 2]'. For instance, 'Bridge & no cycle facility' are bridges without any cycle facility. Some infrastructure factors are deliberately ignored due to frequent infrastructure changes (e.g. new advanced stop zones for cyclists were implemented during our study). Also, human, vehicle-related and environmental factors are disregarded since (1) they are not available for controls, (2) they are often erroneously described, and (3) we deliberately focussed on the effect of modifiable risk factors. Finally, the influence of traffic volume and tram tracks is ignored when accidents/controls occur on separated cycle facilities and *outside junctions*, as the cyclists ride on cycle facilities that are physically separated from the road; to our best knowledge, uni- and bidirectional cycle facilities are designed so that they are physically separated from the road and make the crossing or U-turns impossible for motorists outside intersections (through the presence of e.g. trees, barriers, berms, or parking lots). However, in the case where the cyclist is involved in an accident while illegally crossing a street from a cycle facility, the cyclist is considered to cycle on-road; traffic volume is hence logically considered in such a case.

## 6. Results

### 6.1. Model diagnostics and selection

*Logistic modelling* is first performed within a frequentist framework in order to identify which are the most significant risk factors and to get initial values for the parameters of the hierarchical Bayesian models. Goodness-of-fit statistics and inferential statistical tests show that the logistic model fits the data quite well (LL = -1063.1; HL test statistic = 14.1). The measures of association and misclassification indicate that the model correctly predicts higher probabilities for accidents compared to controls ( $c=0.83$ ;  $D_{xy}=0.66$ ) and misclassifies only 14% of the observations (when setting the cut-off point of classification at 0.5). In a second step, the same logistic model is performed within a Bayesian framework. Table 2 (left columns) shows that the (posterior) values of the parameter estimates are very close to these computed within the frequentist framework (see Vandenbulcke, 2011).

*Autologistic and random effect specifications* are then implemented within a hierarchical Bayesian framework in order to deal with the presence of spatial autocorrelation detected in the previous logistic models. Table 3 lists the best models and compares them with the corresponding null model. The autologistic formulation turns out to be the best (model IX); it not only provides evidence for robust convergence, but also results in the smallest DIC value (2118.1). On the contrary, specifications with random effects did not succeed in converging and provided insignificant parameter estimates for both the uncorrelated and correlated random effects. A spatial weight matrix such as that defined for the autologistic model seems to be the most appropriate form to account for the presence of spatial autocorrelation.

Table 2 (right columns) presents the results of the autologistic model. It shows that almost all risk factors are significant and that the MAPE is quite small, indicating a low misclassification under the fitted model.

### 6.2. Predictions of the risk for a specific road trajectory: a tool for planners?

Mapping the predicted risk of accident validates the results of the autologistic model, and provides a useful tool for planners, decision makers and cyclists' advocacy groups. As an illustration, predictions are computed for sampled points located every 10 m along a road trajectory and further interpolated along this latter using the approximate spline curve method from SANET v.4 (bandwidth = 100 m; cell width = 2 m). Fig. 5 identifies the most 'risky' parts of the trajectory, and hence the places where road users should be more careful and/or where the infrastructures might be modified in order to improve bicyclist's safety.

Fig. 6 exhibits the contribution of 4 of the risk factors to the total risk. Streets with contraflow cycling (Fig. 6b) and tram tracks (Fig. 6a) contribute (positively or negatively) to the total risk of accident at a very local scale, whereas van and truck traffic volumes (Fig. 6c) and the autocovariate component (Fig. 6d) have a more spatially loose effect along the trajectory. It is here assumed that such an autocovariate component captures the unobserved/unidentified risk factors that are specific to each location. Interestingly, Figs. 5 and 6 suggest that the risk is higher for 'complex' intersections (i.e. those numbered 1°, 8°, 10–11, 15–16, 18, 20°), roundabouts with marked cycle lanes (1°), roads with on-road tram tracks and tram crossings (8°–9, 12°, 14), as well as roads with dense van and truck traffic volumes (1°, 2°, 4, 6°, 8°, 11, 13–14, 18, 20°). At the opposite, the lowest accident risks are observed for streets located in residential wards (characterised by low traffic volumes (5°, 9, 16)), where contraflow cycling is allowed (5°, 7, 17), or where no garage is observed in the close vicinity (1°, 5°, 12°, 19°).

Predicting cycling accident risk on the entire network provides several important advantages over black spot methods (Fig. 4). Modelling methods indeed exploit *all the available information* from the accident data set (and from all accidents) to compute a predicted risk of accident *for every point of the network*. On the contrary, black spot methods do not take advantage of using such information to infer locations where accidents might have been unreported, as they only identify spatial concentrations of registered accidents. Comparing Figs. 4 and 5 for the same road trajectory shows that the risks of accident are far from being negligible in points where there is actually no reported cycling accident but where they are yet expected to occur (due to e.g. a dense traffic, or the presence of tram tracks). For instance, locations 3 and 10 correspond to major intersections where there is no reported accident but where it is quite doubtful that it is actually the case in view of the local traffic conditions.

Finally, black spot methods do not take into account the building year of the infrastructure, which could be an issue when working

**Table 2**  
logistic (non-spatial) and auto-logistic models (spatial) – results from the Bayesian framework.

Variables	Logistic model						Autologistic model					
	Estimate			Credible interval (95%)		OR	Estimate			Credible interval (95%)		OR (OR <sub>100m</sub> )
	Mean	SD	MC error	2.50%	97.50%		Mean	SD	MC error	2.50%	97.50%	
Intercept <sup>a</sup>	-2.29 <sup>***</sup>	0.09	0.001	-2.47	-2.12	0.10	-2.29 <sup>***</sup>	0.09	0.001	-2.46	-2.12	0.10
Autocovariate variable	-	-	-	-	-	-	2.15 <sup>***</sup>	0.14	0.001	1.89	2.42	8.61
<i>Infrastructure</i>												
Complexity index												
Bandwidth = 10 m	0.15 <sup>***</sup>	0.01	0.000	0.13	0.17	1.16	-	-	-	-	-	-
Bandwidth = 40 m	-	-	-	-	-	-	0.02 <sup>**</sup>	0.00	0.000	0.01	0.02	1.02 (4.79)
Bridge & no cycle facility	0.86	0.58	0.006	-0.29	2.00	2.37	0.88	0.59	0.005	-0.26	2.03	2.42
Contraflow cycling & no crossroad	-0.69 <sup>†</sup>	0.35	0.003	-1.42	-0.05	0.50	-0.89 <sup>†</sup>	0.36	0.003	-1.64	-0.23	0.41
Cycle facility & crossroad												
Fac. 1 (unidir.) & Crossr. 1 (yield/stop)	2.25 <sup>**</sup>	0.92	0.009	0.63	4.27	9.53	2.02 <sup>**</sup>	0.90	0.008	0.44	3.99	7.56
Fac. 2 (bidir.) & Crossr. 1 (yield/stop)	2.88 <sup>**</sup>	1.38	0.013	0.66	6.02	17.78	3.36 <sup>**</sup>	1.38	0.012	1.15	6.56	28.85
Fac. 3 (mark.) & Crossr. 3 (traff. light)	1.96 <sup>**</sup>	0.94	0.009	0.32	4.01	7.10	1.85 <sup>*</sup>	0.91	0.007	0.25	3.79	6.35
Fac. 3 (mark.) & Crossr. 4 (round.)	2.76 <sup>*</sup>	1.52	0.013	0.18	6.13	15.83	2.83 <sup>*</sup>	1.56	0.013	0.13	6.22	16.91
Fac. 4 (sugg.) & Crossr. 2 (right-of-w.)	3.13 <sup>**</sup>	1.42	0.012	0.87	6.46	22.90	3.74 <sup>**</sup>	1.37	0.011	1.60	7.05	42.22
Fac. 0 (no facility) & Crossr. 4 (round.)	1.02 <sup>***</sup>	0.30	0.003	0.43	1.61	2.78	0.67 <sup>†</sup>	0.32	0.002	0.03	1.30	1.96
Fac. 3 (mark.) & Crossr. 0 (no crossr.)	0.73 <sup>*</sup>	0.33	0.003	0.06	1.35	2.07	-	-	-	-	-	-
Tram tracks												
Class 1 (crossing tram tracks)	0.86 <sup>†</sup>	0.44	0.004	0.01	1.75	2.37	1.16 <sup>**</sup>	0.46	0.004	0.29	2.09	3.20
Class 2 (crossable reserved lanes)	0.83 <sup>†</sup>	0.33	0.003	0.17	1.47	2.30	-	-	-	-	-	-
Class 3 (on-road tracks)	1.06 <sup>***</sup>	0.23	0.002	0.60	1.51	2.87	0.82 <sup>**</sup>	0.23	0.002	0.36	1.28	2.27
Number of garages (for d ≤ 100 m)												
Range 0 (no garage)	-0.61 <sup>†</sup>	0.28	0.003	-1.18	-0.07	0.54	-0.60 <sup>†</sup>	0.28	0.002	-1.17	-0.07	0.55
Distance public administration <sup>b</sup>												
Public administration 2 (regional)	1.08 <sup>***</sup>	0.22	0.002	0.65	1.52	2.95	-	-	-	-	-	-
Distance shopping center <sup>b</sup>	-	-	-	-	-	-	0.86 <sup>***</sup>	0.24	0.002	0.38	1.33	2.36 (2.17)
Proximity parking-cycle facility												
Parking & facility 1 (unidirectional)	1.28 <sup>**</sup>	0.45	0.004	0.37	2.14	3.59	1.15 <sup>†</sup>	0.48	0.004	0.18	2.08	3.16
Parking & facility 2 (bidirectional)	2.07 <sup>*</sup>	1.16	0.011	-0.22	4.40	7.95	1.76	1.30	0.011	-0.88	4.27	5.78
<i>Traffic</i>												
Van & truck traffic (6 a.m.–10:59 p.m.)												
Class 2 (low)	1.01 <sup>***</sup>	0.15	0.001	0.71	1.30	2.73	0.92 <sup>***</sup>	0.15	0.001	0.64	1.21	2.52
Class 3 (moderate)	1.32 <sup>***</sup>	0.16	0.001	1.01	1.63	3.75	1.20 <sup>**</sup>	0.16	0.001	0.89	1.51	3.32
Class 4 (high)	1.24 <sup>***</sup>	0.22	0.002	0.80	1.68	3.46	1.26 <sup>***</sup>	0.22	0.002	0.82	1.70	3.53
Class 5 (very high)	2.60 <sup>***</sup>	0.35	0.003	1.93	3.29	13.46	2.13 <sup>***</sup>	0.36	0.003	1.43	2.84	8.38
Deviance	2149 <sup>***</sup>	6.92	0.060	2137	2164	-	2097 <sup>***</sup>	6.70	0.052	2086	2112	-
MAPE	0.21 <sup>***</sup>	0.00	0.000	0.20	0.22	-	0.21 <sup>***</sup>	0.00	0.000	0.20	0.21	-

OR, odds ratio; OR<sub>100m</sub>, odds ratio computed for a 100 m increase (rather than 1 m). *Interaction variables*. Bridge & no cycle facility: bridge = 1 and cycle facility = 0. Contraflow cycling & no crossroad: contraflow cycling = 1 and crossroad = 0. Van & truck traffic (6 a.m.–10:59 p.m.): maximum class value of van and truck traffic. Cycle facility & crossroad: cycle facility = X and Crossroad = Y (X = 1, . . . , 5; Y = 1, . . . , 6). Cycle facility 3 (marked) & no crossroad: cycle facility = 3 and crossroad = 0.

<sup>a</sup> Intercept value resulting from centering.

<sup>b</sup> Exponentially transformed variables ( $e^{-0.001x}$ ).

<sup>†</sup> Significant at 95%.

<sup>\*\*</sup> Significant at 99%.

<sup>\*\*\*</sup> Significant at 99.9%.

**Table 3**  
Model-building specifications and model comparison.

Model ID	Model	Iterations	Burn-in	Thin	$p_D$	DIC	Converged?
I	Null model	15,000	10,000	1	1.00	3011.53	Yes
II	Null model + UH	250,000	50,000	100	25.10	3016.88	No
III	Null model + Autocov	50,000	10,000	10	2.00	2572.26	Yes
IV	Null model + UH + Autocov	36,000	9000	60	3.27	2572.3	No
V	Null model + CH	–	–	–	–	–	No
VI	Null model + UH + CH	–	–	–	–	–	No
VII	Fixed effects only	<b>50,000</b>	<b>10,000</b>	<b>10</b>	<b>22.21</b>	<b>2171.15</b>	Yes
VIII	Fixed effects + UH	35,000	10,000	5	27.02	2171.23	No
IX	Fixed effects + Autocov	<b>20,000</b>	<b>10,000</b>	<b>2</b>	<b>21.08</b>	<b>2118.08</b>	Yes
X	Fixed effects + UH + Autocov	20,000	12,500	10	22.60	2118.64	No
XI	Fixed effects + CH	–	–	–	–	–	No
XII	Fixed effects + UH + CH	–	–	–	–	–	No

Number of Markov Chains = 3. Null model: model with intercept only (no risk factor included). UH, uncorrelated heterogeneity (unstructured errors); CH, correlated heterogeneity (spatially structured errors, ICAR); Autocov, autocovariate;  $p_D$ , effective number of parameters; DIC, deviance information criterion; converged?, ‘Yes’: the model converged to a perfect equilibrium; ‘No’: the model did not converge to a perfect equilibrium, but the trace plots suggest it almost reached it. In **bold**: selected models (VII = logistic model; IX = auto-logistic model).

on several years (e.g. on a three-year period) since a location may be informed as being ‘dangerous’ whereas some important infrastructure changes may already have been implemented during the period of study. More importantly, most of the black spot methods do not consider the traffic direction and may indicate both sides of a street as being dangerous for cyclists whereas most cycling accidents cluster on one side only (example of location 2\* in Fig. 5). Black spot methods hence fail to identify the ‘dangerous’ street side and may lead to erroneous recommendations about infrastructures.

6.3. Discussion of the results

Before discussing the results, readers should be aware that the following conclusions regarding infrastructure factors can be biased because the exposure measure (PBTI) cannot be well empirically validated (see Section 7.1 for further information).

The complexity index has the largest effect on the risk of having a cycling accident (30% of the explanation, whatever the location). This is probably explained by the fact that cyclists and other road

users are faced with a large amount of information at the same time at locations with an increased complexity (e.g. high number of road legs, road users, etc.). Driving errors are hence more likely to occur there.

Although significant at 93% in the autologistic model, the parameter estimate for *bridges without cycle facilities* suggests that the risk of accident increases at such locations. The sudden change in road width and visibility is expected to explain this increased risk, especially if there is no dedicated cycling facility on the bridge. If well kept and designed, such a facility could probably outweigh/reduce the risks caused by the low long-distance visibility and the narrowing of the road space.

Contrary to popular belief, we show that *contraflow cycling* is associated with a reduced accident risk while cycling. Such a result may be due to a ‘risk compensation effect’, i.e. from the fact that drivers tend to behave in a more cautious way due to an increased perceived risk in streets where contraflow cycling is allowed or be the result of better visibility. This last assumption seems to be confirmed by specialists working into the field of traffic safety (BRSI, 2013).

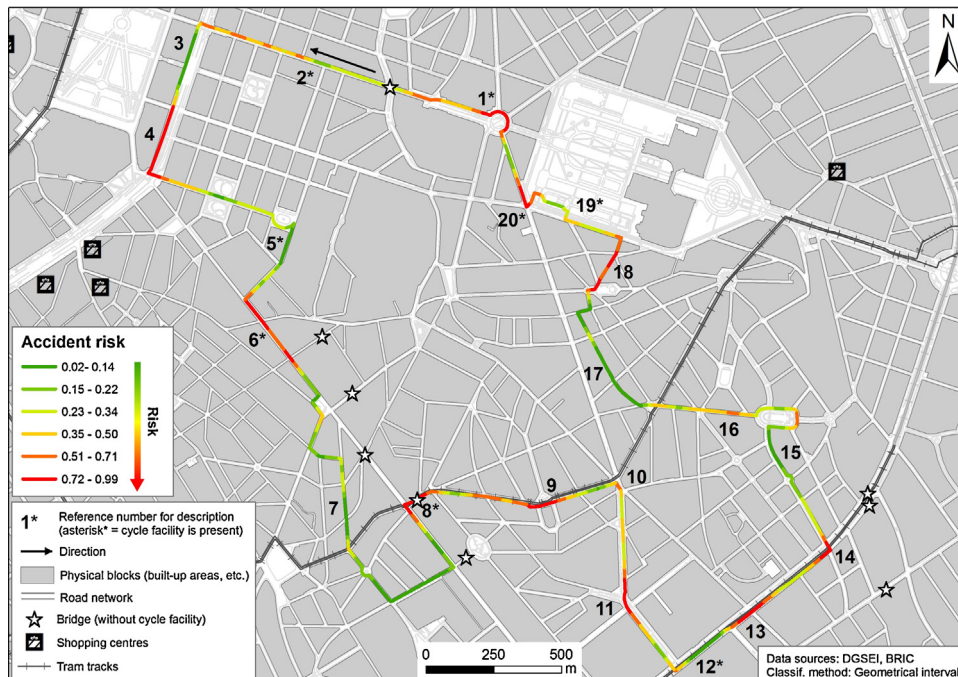
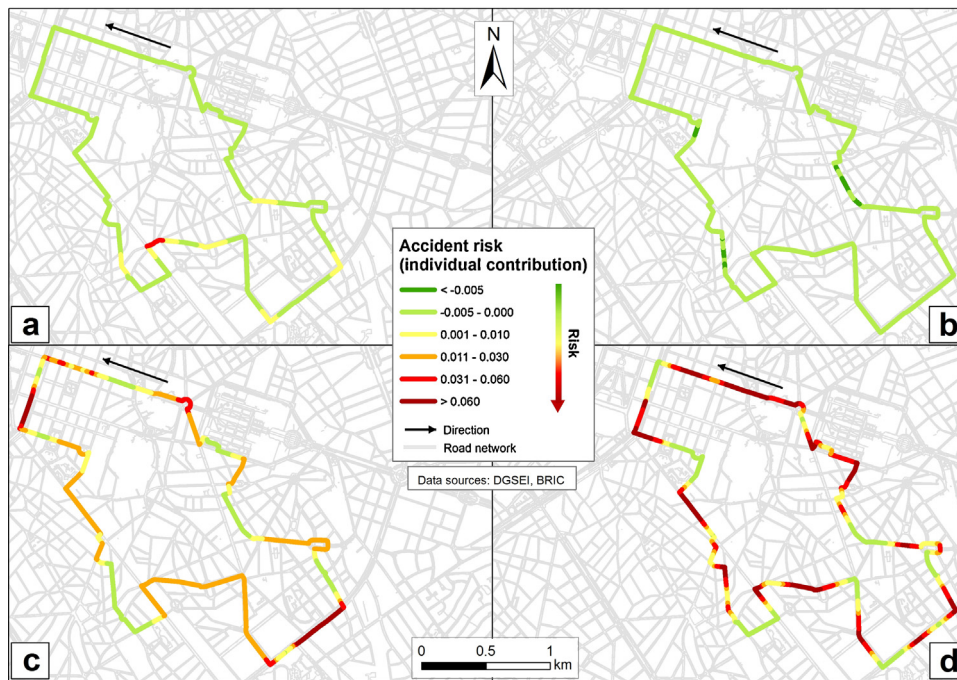


Fig. 5. Predicted cycling accident risk for a specific trajectory in Brussels, computed from the parameter estimates of the autologistic specification.



**Fig. 6.** Predicted risk of having a cycling accident, separately computed for 4 risk factors: (a) tram tracks (on-road and crossings); (b) contraflow cycling (intersections are excluded); (c) van and truck traffic from 6 a.m. to 10:59 p.m. (all classes from 2 to 5); (d) autocovariate component.

Regarding cycle facilities, our results are in line with the literature and indicate that some facilities are associated with an increased risk of accident when associated with a specific type of intersection. In particular, *right-of-way intersections equipped with cycle lanes* lead to the highest accident risk for cyclists, probably because of the non-respect of the right-of-way by motorists (BRSI, 2009a) and the discontinuous character of the facility (e.g. bicycle logos probably make the cyclists less 'visible' for motorists). According to accident data registered in Brussels, collision partners indeed did not give way to the cyclist in about 59% of the accidents. *Yield/stop intersections with separated cycle lanes* also seem to carry a danger, especially when the cyclist rides on a bidirectional facility in the opposite direction of the traffic. Cyclists often have an ill-founded feeling of safety caused by the physical segregation of the facility, while on the other hand motorists often have an inappropriate visual search pattern and do not expect to cross a cyclist coming from the opposite direction (BRSI, 2006).

As expected, high accident risk is observed for cyclists riding on *marked cycle lanes built in roundabouts (outer lane)*. In such contexts, collisions often occur when the motorist leaves/enters the roundabout and cuts in on the cyclist riding on the marked facility. Such a design even leads to a higher accident risk for cyclists compared to roundabouts without any cycle facility. Intersections equipped with *traffic lights and marked cycle lanes* are also found to be associated with an increased risk of accident. This is probably due to motorists turning to an adjacent road and cutting in on the cyclist's trajectory on the marked facility; moreover, cycle lanes are often designed in such a way that they position cyclists in the blind spots of the motorised vehicles at signalised intersections. More generally and disregarding the type of facility or intersection, it is not uncommon in Brussels that cycle facilities abruptly stop at intersections, providing no dedicated/safe room for the cyclist within the traffic and hence increasing the accident risk here. Inappropriately designed and/or poorly maintained cycle facilities may also lead to confusing situations.

The close vicinity ( $\leq 0.9\text{m}$ ) between *separated cycle lanes and parking facilities* is also identified here as a significant risk

factor. Cyclists riding on such separated lanes and alongside parked vehicles may indeed run into suddenly opened car doors. Also, the presence of parked vehicles may generate pedestrian activity on the adjacent cycle lane and may potentially lead to an accident. This is all the more true as, in Brussels, the joint presence of parked vehicles and separated cycle lanes is frequently observed alongside major roads, characterised by close attractive activities (e.g. business zones, parks, etc.). Similarly, Table 2 suggests that the presence of *garage/parking driveway* (within a 100 m network distance) is associated with a higher risk of accident. This result may be explained by the fact that motorists leaving/entering a garage driveway may collide with cyclists riding straight ahead on the road (BRSI, 2006).

The presence of *on-road tram tracks* and (*tracks*) *crossings* seems to be related with an increased risk of accident. Cyclists may get stuck in tram tracks, resulting in a loss of control of the bicycle. We also suspect that the presence of on-road tracks forces the cyclist to ride on places that are not especially optimal for his/her own safety. For instance, he/she has to make the difficult choice between riding next to opening doors of parked vehicles and riding between the tracks, i.e. in the middle of the road lane.

The presence of a *shopping centre* in the vicinity of the cyclist's trajectory is also associated with an increased risk. Pedestrian and/or motorised activity is commonly observed in the neighbourhood of a shopping centre, increasing the number of potential conflicting situations.

Last but not least, traffic-related risk factors (levels of van and truck traffic (classes 2–5)) provide the best improvement in model fit: increasing levels of van and truck traffic are associated with higher accident risks. Whatever the type of road user, the complexity of the traffic context increases with traffic density. The road legibility as well as the cognitive capacity of the road user are influenced by the presence of a large amount of information to process in the streetscape (which reduces the ability to detect/carry out appropriate actions to control traffic hazards) (Elvik, 2006). Vans and trucks are also more prone to blind spot when turning and often leave narrow safety margins to cyclists when overtaking; this

increases the risk of accident for cyclists. Furthermore, the large dimensions of vans and trucks may obstruct the view of other road users.

## 7. Conclusions

### 7.1. Contribution of this research

Characteristics of cycling accidents have received little attention in safety research, especially in towns where cyclists represents a small share of all transport modes (Loo and Tsui, 2010). In Brussels, the share of cycling is continuously increasing each year and makes this issue even more important to ensure safe cycling. This paper hence focuses on cycling accidents and – taking advantage of recent research implemented in epidemiology, ecology and transport geography – opens up a new research direction in traffic accident analysis by using a case–control strategy to estimate an ‘accident risk model’, i.e. a model aiming at predicting the risk of having a cycling accident at the level of a whole network. To our best knowledge, this paper is the first attempt to study accident risks for cyclists at the level of a whole network. Our model was applied on the Brussels-Capital Region and used accident data for the years 2006–2008. In order to make possible the use of (auto-)logistic and conditional autoregressive modelling, a binary dataset was constructed by adding controls to the geocoded accident data set. Controls were sampled along the bikeable road network as a function of the bicycle traffic.

Compared to other methodologies in road safety analysis (e.g. models of injury severity or accident frequency), the use of such a case–control strategy has methodological advantages: (1) the estimation of accident risks for cyclists is made possible (e.g. through logistic modelling) despite the absence of data on bicycle traffic exposure; (2) the exhaustive data collection as well as the use of point data avoids the need for arbitrary aggregation of accidents, and hence reduces the risk of ‘ecological fallacy’; (3) compared to black spot methods, predicted values computed from the model provide results for the entire road network and – as a corollary – for locations where cycling accidents are a priori underreported (e.g. on-road tram tracks); (4) if the aim is to obtain a general prediction of accident risk, the use of controls avoids the expensive work of collecting counting data (it is however more questionable if the aim is to provide conclusions on the impact of different road infrastructures); (5) sampling control points only depends on the location of the black spots and on the spatial distribution of cyclists in the area of interest; if this latter remains unchanged throughout the years (and/or follows the same increasing trend over space), the use of out-of-date data does not bias the sampling of controls as the intensity of this sampling is proportional to the PBTI values; (6) last but not least, mapping the predicted values of accident risks along an entire road network would allow cyclists choosing the safest route between an origin (e.g. residence) and a destination (e.g. workplace). Combined with other variables (e.g. topography or exposure to air pollution), optimal paths could be identified for orienting cyclists to the safest, comfortable or healthiest routes.

However, this research is not without weaknesses and limitations: (1) the data collection step is presently time-consuming as it requires collecting additional data for the controls; (2) despite the absence of human- and vehicle-related factors, it is expected that these latter would be correlated with some of the spatial risk factors (e.g. in the case where a distinction is made between child and adult accidents, part of the variance in the model would probably be explained by risk factors referring e.g. to the proximity to schools for children); (3) the quality of the results is strongly constrained by the method of selection of the controls, as well as by the formulation of the potential index (PBTI).

In relation with the latter limitation above mentioned, the fact that the PBTI is based on commuting trips only (data on leisure and shopping trips are not available) and that the availability of cycle facilities is not taken into account in the computation of the PBTI (cyclists may prefer to use some roads equipped with cycle facilities) are likely to lead to systematic errors in the construction of the background variable. The absence of a consistent counting database on bicycle use indeed makes the validation of the PBTI impossible; the visual comparative check of the PBTI against counting data on cycling trips collected each year by the association Pro Velo at 20 locations in Brussels is insufficient to detect systematic deviations/errors due to e.g. extra bicycle traffic near parks or along bicycle infrastructures. Moreover, the estimation of the distribution of traffic using the PBTI is probably biased due to the lack of knowledge about how bicycle facilities and ‘non-commuting’ destinations (e.g. shopping) influence the modal choice among non-commuter cyclists. Safe and comfortable bicycle infrastructures may indeed encourage children and elderly women to cycle (Pucher and Buehler, 2008) and to make detours (i.e. paths with a longer travel time).

The inclusion in the model of an autocovariate and the use of spatial factors taking into account the availability of cycling facilities (among other factors influencing on the route choice) is expected to partly (but not entirely) compensate the presence of systematic errors resulting from the construction of the PBTI. In particular, errors that are spatially aggregated (e.g. due to an underestimation of the bicycle traffic at specific locations) can indeed be corrected by the inclusion of such an autocovariate in the model. Doing this allows picking up the ‘extra’ explanatory power that was improperly attributed to some variables (Anselin, 2005), so limiting the risk of under- or over-estimation when specific infrastructures are present. As illustration, Table 2 shows that most of the mean estimates measured for cycle facilities built at intersections are lower in the autologistic model (suggesting that intersections equipped with cycle facilities are associated with a reduced accident risk). The inclusion of the autocovariate here matters as it is expected that the sampling of controls is partly biased due to the non-inclusion of route choice into the computation of the PBTI. In particular, it is expected that this sampling should be higher on cycle facilities as these latter influence route and modal choice; the accident risk associated with the presence of cycle facilities at intersections is hence probably over-estimated in the logistic model, while it is properly reduced in the autologistic model.

Combining spatial modelling with our case–control strategy is hence strongly supported to reduce systematic errors in our results, and then to obtain better predictions.

### 7.2. Recommendations

Although great caution is required by the fact that our results depend on the validity of the PBTI, we here attempted to derive science-based recommendations for safer urban cycling. Our results first suggest that, in Brussels, special attention should be paid to the cyclist’s safety when designing on-road tram tracks, bridges and/or ‘major’ intersections as these latter all increase the risk of cycling accident (especially when these infrastructures spatially co-occur). In particular, major intersections are generally characterised by a high complexity due to dense crossing traffic as well as many road legs and signs. Whenever possible, they should be made more legible for all road users. Physically segregated tram tracks should also be preferred to on-road tram tracks. It could be profitable not only to cyclists, but also to public transport companies since it may improve the commercial speed of the vehicles. Bridges should also consistently be designed with great care for cyclists in order to offset the increased accident risk caused by the reduced number and/or width of the road lanes. Building adjacent

cycle facilities – separated with physical hurdles – could probably reduce this risk.

Cycle facilities should also be designed with great care, especially at intersections. When investments devoted to the cycle facilities are limited, planners and decision makers should primarily give priority to the provision of high-quality infrastructure (i.e. continuous, visible and well-kept) rather than investing in an extensive network built in haste and carelessly. Separated cycle facilities should, for example, be designed in such a way that motorists and cyclists have a visual contact before arriving at the intersection. Making cycle lanes more visible to motorists (e.g. using coloured pavements) is also expected to reduce the risk of cycle accident, especially when the design is discontinuous.

Outside intersections, building cycle facilities in the ‘door zone’ of parked vehicles (<0.8m) should be avoided as much as possible since cyclists are exposed to a higher accident risk due to the opening of car doors. Last but not least, the reduced risk reported in streets where contraflow cycling is permitted supports for their wider implementation in Brussels; great care should however be taken when designing these as they are associated with an increased risk of accident at intersections. The use of (visible) marked cycle lanes or bicycle logos painted at the entrance of these streets might be useful in warning motorists for the presence of cyclists.

Above all, the implementation of new cycling facilities should not be discouraged by planners and decision makers, but ‘semi-measures’ should be avoided as much as possible. If inappropriately designed, such facilities could lead to adverse effects instead of reducing the risk of accident. At the opposite, high-quality infrastructures might reduce the actual and perceived risk of cycling accident, which might in turn encourage cycling.

## Acknowledgements

This research was partly conducted within the SHAPES research project financed by the Belgian Science Policy (SD/HE/03B). Authors would like to thank the reviewers for their insightful comments and suggestions on earlier drafts of this paper.

## References

- Aertsens, J., de Geus, B., Vandenbulcke, G., Degraeuwe, B., Broekx, S., De Nocker, L., Liekens, I., Mayeres, I., Meeusen, R., Thomas, I., Torfs, R., Willems, H., Int Panis, L., 2010. Commuting by bike in Belgium, the costs of minor accidents. *Accident Analysis and Prevention* 42, 2149–2157.
- Agüero-Valverde, J., Jovanis, P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention* 38, 618–625.
- Anderson, T., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention* 41, 359–364.
- Anselin, L., 2005. Exploring Spatial Data with GeoDa: A Workbook. Spatial Analysis Laboratory and Center for Spatially Integrated Social Science, <http://geodacenter.asu.edu/learning/tutorials>
- Aultman-Hall, L., Hall, F., 1998. Ottawa–Carleton commuter cyclist on- and off-road incident rates. *Accident Analysis and Prevention* 30 (1), 29–43.
- Aultman-Hall, L., Kaltenecker, M., 1999. Toronto bicycle commuter safety rates. *Accident Analysis and Prevention* 31 (6), 675–686.
- Bailey, T., Gatrell, A., 1995. *Interactive Spatial Data Analysis*. Longman Scientific & Technical, Burt Hill, UK.
- Best, N., Cowles, M., Vines, S., 1995. *CODA Manual v0.30*. MRC Biostatistics, Cambridge.
- Bivand, R., Pebesma, E., Gomez-Rubio, V. (Eds.), 2008. *Applied Spatial Data Analysis with R*. Springer, New York.
- Bos, I., De Boever, P., Vanparijs, J., Pattyn, N., Int Panis, L., Meeusen, R., 2013. Sub-clinical effects of aerobic training in urban environment. *Medicine & Science in Sports & Exercise* 45 (3), 439–447.
- BRSI (Belgian Road Safety Institute), 2006. *Accidents de cyclistes en contexte urbain – Trois années (1998–2000) d’accidents corporels de cyclistes sur les voiries régionales de la Région de Bruxelles-Capitale*. IBSR/BIVV, Brussels.
- BRSI (Belgian Road Safety Institute), 2009a. *Rapport thématique Cyclistes: accidents de la route impliquant des cyclistes (2000–2007)*. IBSR/BIVV, Brussels.
- BRSI (Belgian Road Safety Institute), 2009b. *Mesure nationale de comportement en matière de vitesse (2003–2007)*. IBSR/BIVV, Brussels.
- BRSI (Belgian Road Safety Institute), 2013. *Le SUL. Cyclistes à contresens dans les sens uniques. Conseils aux usagers*. IBSR/BIVV, Brussels (available online).
- Buehler, R., Pucher, J., Merom, D., Bauman, A., 2011. Active travel in Germany and the U.S. – contributions of daily walking and cycling to physical activity. *American Journal of Preventive Medicine* 41 (3), 241–250.
- Cameron, I., Harris, N., Kehoe, N., 2001. Tram-related injuries in Sheffield. *Injury* 32 (4), 275–277.
- Chapman, L., 2007. Transport and climate change: a review. *Journal of Transport Geography* 15, 354–367.
- Cho, G., Rodríguez, D., Khattak, A., 2009. The role of built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. *Accident Analysis and Prevention* 41, 692–702.
- Chong, S., Poulos, R., Olivier, J., Watson, W.L., Grzebieta, R., 2010. Relative injury severity among vulnerable non-motorised road users: comparative analysis of injury arising from bicycle–motor vehicle and bicycle–pedestrian collisions. *Accident Analysis and Prevention* 42, 290–296.
- Dai, D., Taquechel, E., Steward, J., Strasser, S., 2010. The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban University campus. *Western Journal of Emergency Medicine* XI (3), 294–301.
- Daniels, S., Nuyts, E., Wets, G., 2008. The effects of roundabouts on traffic safety for bicyclists: an observational study. *Accident Analysis and Prevention* 40, 518–526.
- Daniels, S., Brijis, T., Nuyts, E., Wets, G., 2009. Injury crashes with bicyclists at roundabouts: influence of some location characteristics and the design of cycle facilities. *Accident Analysis and Prevention* 40, 141–148.
- de Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2012. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. *Accident Analysis and Prevention* 45, 683–693.
- de Geus, B., Degraeuwe, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Aertsens, J., De Weerd, Y., Torfs, R., Meeusen, R., 2014. Utilitarian cycling in Belgium: a cross-sectional study in a sample of regular cyclists. *Journal of Physical Activity & Health* 11 (7) (online ahead of print 23/05/2013).
- de Hartog, J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives* 118 (8), 109–1116.
- de Lapparent, M., 2005. Individual cyclists’ probability distributions of severe/fatal crashes in large french urban areas. *Accident Analysis and Prevention* 37, 1086–1092.
- de Nazelle, A., Nieuwenhuijsen, M.J., Antó, J.M., Brauer, M., Briggs, D., Braun-Fahrlander, C., Cavill, N., Cooper, A.R., Desqueyroux, H., Fruin, S., Hoek, G., Int Panis, L., Janssen, N., Jerrett, M., Joffe, M., Andersen, Z.J., van Kempen, E., Kingham, S., Kubesch, N., Leyden, K., Marshall, J.D., Matamala, J., Mellios, G., Mendez, M., Nassif, H., Ogilvie, D., Peiró, R., Pérez, K., Rabl, A., Ragetti, M., Rodríguez, D., Rojas, D., Ruiz, P., Sallis, J.F., Terwoert, J., Toussaint, J.F., Tuomisto, J., Zuurbier, M., Lebre, E., 2011. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment International* 37 (4), 766–777.
- Diggle, P., 1990. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society A* 153 (3), 349–362.
- Doom, R., Derveduwen, P., 2005. *Optimalisatie van de verkeersongevallenstatistiekken*. Research contracts CP/02/392, CP/F1/391 (final report) for BRSI, CDO and Belgian Science Policy (available online).
- Dormann, C., 2007. Assessing the validity of autologistic regression. *Ecological Modelling* 207, 234–242.
- Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., Davies, R., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P., Reineking, B., Schröder, B., Schurr, F., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30, 609–628.
- Dumbaugh, E., Rae, R., 2009. Safe urban form: revisiting the relationship between community design and traffic safety. *Journal of the American Planning Association* 75 (3), 309–329.
- Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention* 40, 1033–1054.
- Elvik, R., 2006. Laws of accident causation. *Accident Analysis and Prevention* 38, 742–747.
- Elvik, R., 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention* 41 (4), 849–855.
- ERSO (European Road Safety Observatory), 2006. *Pedestrians & Cyclists*. Retrieved 04.17.11 from [www.erso.eu](http://www.erso.eu)
- Flahaut, B., 2004. Impact of infrastructure and local environment on road unsafety: logistic modelling with spatial autocorrelation. *Accident Analysis and Prevention* 36 (6), 1055–1066.
- Forester, J., 1994. *Bicycle Transportation – A Handbook for Cycling Transportation Engineers*. MIT, Cambridge, USA.
- Fotheringham, A., Brunsdon, C., Charlton, M., 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London, UK.
- Geertman, S., Ritsema van Eck, J., 1995. GIS and models of accessibility potential: an application in planning. *International Journal of Geographical Information Systems* 9 (1), 67–80.
- Gelman, A., Hill, J., 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Geurs, K., van Wee, B., 2004. Accessibility of land-use and transport strategies: review and research directions. *Journal of Transport Geography* 12, 127–140.



- Greibe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention* 35, 273–285.
- Grimes, D., Schulz, K., 2005. Compared to what? Finding controls for case-control studies. *The Lancet* 365, 1429–1433.
- Haining, R., Law, J., Griffith, D., 2009. Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Computational Statistics and Data Analysis* 53, 2923–2937.
- Haque, M., Chin, H., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis and Prevention* 42, 203–212.
- Harris, M., Reynolds, C., Winters, M., Chipman, M., Crompton, P.A., Cusimano, M., Teschke, K., 2011. The bicyclists' injuries and the cycling environment study: a protocol to tackle methodological issues facing studies of bicycling safety. *Injury Prevention* 17 (5), <http://dx.doi.org/10.1136/injuryprev-2011-040071>.
- Haynes, R., Lovett, A., Sünnerberg, G., 2003. Potential accessibility, travel time, and consumer choice: geographical variations in general medical practice registrations in Eastern England. *Environment and Planning A* 35, 1733–1750.
- Hedelin, A., Björnstig, U., Brismar, B., 1996. Trams – A risk factor for pedestrians. *Accident Analysis and Prevention* 28 (6), 733–738.
- Hels, T., Orozova-Bekkevold, I., 2007. The effect of roundabout design features on cyclist accident rate. *Accident Analysis and Prevention* 39, 300–307.
- Hossain, M., Lawson, A., 2009. Approximate methods in Bayesian point process spatial models. *Computational Statistics and Data Analysis* 53, 2831–2842.
- Hubert, J.-P., Toint, P., 2002. *La mobilité quotidienne des belges*. Presses Univ. de Namur.
- Int Panis, L., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., Mishra, V., Thomas, I., Meeusen, R., 2010. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment* 44, 2263–2270.
- Int Panis, L., 2011. Cycling: health benefits and risks. *Environmental Health Perspectives* 119 (3), A114.
- Jacobsen, P., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention* 9, 205–209.
- Khan, G., Santiago-Chaparro, K., Qin, X., Noyce, D., 2009. Application and integration of lattice data analysis network K-functions, and GIS to study ice-related crashes. *Transportation Research Record: Journal of the Transportation Research Board* 2136, 67–76.
- Kéry, M., 2010. *Introduction to WinBUGS for Ecologists. A Bayesian Approach to Regression, ANOVA, Mixed Models and Related Analyses*. Academic Press, Amsterdam.
- Kim, J.-K., Kim, S., Ulfarsson, G., Porrello, L., 2007. Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention* 39, 238–251.
- Klop, J., Khattak, A., 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in North-Carolina. *Transportation Research Record* 1674, 78–85.
- Krizek, K.J., Roland, R., 2005. What is at the end of the road? Understanding discontinuities of on-street bicycle lanes in urban settings. *Transportation Research, Part D* 10 (1), 55–68.
- Law, J., Haining, R., Maheswaran, R., Pearson, T., 2006. Analyzing the relationship between smoking and coronary heart disease at the small area level: a Bayesian approach to spatial modeling. *Geographical Analysis* 38, 140–159.
- Lawson, A., 2009. *Bayesian Disease Mapping Hierarchical Modelling in Spatial Epidemiology*. Chapman and Hall, London.
- Levine, N., Kim, K., Nitz, L., 1995a. Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis and Prevention* 27 (5), 663–674.
- Levine, N., Kim, K., Nitz, L., 1995b. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis and Prevention* 27 (5), 675–685.
- Li, L., Zhu, L., Sui, D., 2007. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography* 15 (4), 274–285.
- Liu, Y., Jarrett, D., 2008. Spatial distribution of road crashes in Great Britain. *Journal of Maps*, 91–107.
- Loo, B., Tsui, K., 2010. Bicycle crash casualties in a highly motorized city. *Accident Analysis and Prevention* 42 (6), 1902–1907.
- Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44, 291–305.
- Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2011. Risk of injury for bicycling on cycle tracks versus in the street. *Injury Prevention* 17 (2), 131–135.
- McCarthy, M., Gilbert, K., 1996. Cyclist road deaths in London 1985–1992: drivers, vehicles, manoeuvres and injuries. *Accident Analysis and Prevention* 28 (2), 275–279.
- McClintock, H., Cleary, J., 1996. Cycle facilities and cyclists' safety – experience from Greater Nottingham and lessons for future cycling provision. *Transport Policy* 3, 67–77.
- Miaou, S.-P., Song, J., Mallick, B., 2003. Roadway traffic crash mapping: a space-time modelling approach. *Journal of Transportation and Statistics* 6 (1), 33–57.
- Miller, J., Franklin, J., Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202, 225–242.
- Møller, M., Hels, T., 2008. Cyclists' perception of risk in roundabouts. *Accident Analysis and Prevention* 40, 1055–1062.
- Myint, S., 2008. An exploration of spatial dispersion, pattern and association of socio-economic functional units in an urban system. *Applied Geography* 28, 168–188.
- Noland, R., Quddus, M., 2004. A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention* 36, 973–984.
- Ntzoufras, I., 2009. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, New Jersey.
- Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science* 23 (1), 7–32.
- Pai, C.-W., 2011. Overtaking, rear-end, and door crashes involving bicycles. An empirical investigation. *Accident Analysis and Prevention* 43, 1228–1235.
- Parkin, J., Meyers, C., 2010. The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis and Prevention* 42, 159–165.
- Parkin, J., Wardman, M., Page, M., 2007. Models of perceived cycling risk and route acceptability. *Accident Analysis and Prevention* 39 (2), 364–371.
- Pei, X., Wong, S.C., Sze, N.N., 2010. A joint-probability approach to crash prediction models. *Accident Analysis and Prevention* 43 (3), 1160–1166.
- Polis, 2012. *Securing the Benefits of Active Travel in Europe*, Position Paper (available online).
- Pucher, J., Komanoff, C., Schimek, P., 1999. Bicycling renaissance in North America? Recent trends and alternatives policies to promote bicycling. *Transportation Research* 33A, 625–654.
- Pucher, J., Dijkstra, L., 2003. Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. *Public Health Matters* 93 (9), 1509–1516.
- Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transport Reviews* 28 (4), 495–528.
- Quddus, M., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity. An analysis of London crash data. *Accident Analysis and Prevention* 40, 1486–1497.
- Rabl, A., de Nazelle, A., 2011. Assessment of health costs and benefits for active transport. *Transport Policy* 19, 121–131.
- Räsänen, M., Summala, H., 1998. Attention and expectation problems in bicycle-car collisions: an in depth study. *Accident Analysis and Prevention* 30 (5), 657–666.
- Reynolds, C., Harris, M., Teschke, K., Crompton, P., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental Health* 8 (47).
- Rifaat, S., Tay, R., de Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis and Prevention* 43, 276–283.
- Rodgers, G., 1997. Factors associated with the crash risk of adult bicyclists. *Journal of Safety Research* 28 (4), 233–241.
- Rojas-Rueda, D., de Nazelle, A., Tainio, M., Nieuwenhuijsen, M.J., 2011. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *British Medical Journal* 343, d4521.
- Schepers, J., Kroeze, P., Sweers, W., Wuüst, J., 2011. Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections. *Accident Analysis and Prevention* 43 (3), 853–861.
- Shiode, S., 2008. Analysis of a distribution of point events using the network based quadrat method. *Geographical Analysis* 40, 380–400.
- Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software* 12 (3).
- Thill, J.-C., Kim, M., 2005. Trip making, induced travel demand, and accessibility. *Journal of Geographical Systems* 7, 229–248.
- Thomas, B., DeRobertis, M., 2013. The safety of urban cycle tracks: a review of the literature. *Accident Analysis and Prevention* 52, 219–227.
- Unger, R., Eder, C., Mayr, J.M., Wernig, J., 2002. Child pedestrian injuries at tram and bus stops. *Injury* 33 (6), 485–488.
- Vandenbulcke, G., 2011. *Spatial Analysis of Bicycle Use and Accident Risks for Cyclists*. Université catholique de Louvain, PhD Thesis, 263p.
- Vandenbulcke, G., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L., 2009. Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy* 16, 77–87.
- Vandenbulcke, G., Dujardin, C., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L., 2011. Cycle commuting in Belgium: spatial determinants and 're-cycling' strategies. *Transportation Research Part A* 45, 118–137.
- Wang, C., Quddus, M., Ison, S., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident Analysis and Prevention* 41, 798–808.
- Wang, Y., Nihan, N., 2004. Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. *Accident Analysis and Prevention* 36, 313–321.
- WHO (World Health Organization), 2002. *A Physically Active Life Through Everyday Transport, with a Special Focus on Children and Older People, and Examples and Approaches from Europe*. Regional Office for Europe, Copenhagen.
- Winters, M., Davidson, G., Kao, D., Teschke, K., 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation* 38, 153–168.
- Yan, X., Ma, M., Huang, H., Abdel-Aty, M., Wu, C., 2011. Motor vehicle bicycle crashes in Beijing: irregular manoeuvres, crash patterns, and injury severity. *Accident Analysis and Prevention* 43 (5), 1751–1758.
- Ye, F., Lord, D., 2011. Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit and mixed logit models. *Transportation Research Record* 2241, 51–58.
- Zhu, L., Gorman, D.M., Horel, S., 2006. Hierarchical Bayesian spatial models for alcohol availability, drug hot spots and violent crime. *International Journal of Health Geographics* 5 (54).