

2020-12-29

# Bayesian hierarchical models for linear networks

Al-kaabawi, Z

<http://hdl.handle.net/10026.1/16774>

---

10.1080/02664763.2020.1864814

Journal of Applied Statistics

Taylor & Francis (Routledge)

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

# Bayesian Hierarchical Models for Linear Networks

Zainab Al-kaabawi<sup>a</sup>, Yinghui Wei<sup>a</sup>, Rana Moyeed<sup>a</sup>

<sup>a</sup>*Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth,  
PL4 8AA, UK*

---

## Abstract

The purpose of this study is to highlight dangerous motorways via estimating the intensity of accidents and study its pattern across the UK motorway network. Two methods have been developed to achieve this aim. First, the motorway-specific intensity is estimated by using a homogeneous Poisson process. The heterogeneity across motorways is incorporated using two-level hierarchical models. The data structure is multilevel since each motorway consists of junctions that are joined by grouped segments. In the second method, the segment-specific intensity is estimated. The homogeneous Poisson process is used to model accident data within grouped segments but heterogeneity across grouped segments is incorporated using three-level hierarchical models. A Bayesian method via Markov Chain Monte Carlo is used to estimate the unknown parameters in the models and the sensitivity to the choice of priors is assessed. The performance of the proposed models is evaluated by a simulation study and an application to traffic accidents in 2016 on the UK motorway network. The deviance information criterion (DIC) and the widely applicable information criterion (WAIC) are employed to choose between models.

**Keywords:** Hierarchical Models, Bayesian methods, Linear networks, Point processes

---

## 1. Introduction

Traffic crashes have considerable impact on human, economics and the society. To improve road safety, traffic accidents research often seeks to determine prediction methods of traffic accidents. Traditional crash prediction models, such as generalized linear model, are widely used in traffic safety studies. However, multilevel data structure is extensively existed due to technique used to

---

*Correspondence to: Yinghui Wei, Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth, PL4 8AA. Email: [yinghui.wei@plymouth.ac.uk](mailto:yinghui.wei@plymouth.ac.uk).*

*This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Applied Statistics on 29 December 2020, available online: <https://www.tandfonline.com/doi/full/10.1080/02664763.2020.1864814>*

collect or cluster traffic data [24]. Ignoring hierarchical nature of data may produce unreliable estimates of model parameters and statistical inference. Hierarchical modelling is a statistical approach that used to properly take account of multilevel data structure [19, 24]. Currently, hierarchical modelling has been employed in many research fields such as sociology, education, political science and public health. Shankar et al (1998) showed that the explanatory power of crash models had been improved when site-specific random effects and time indicator were incorporated into the negative binomial regression model [44]. Jones and Jørgensen (2003) expounded and discussed possible applications of hierarchical models in road traffic accidents in Norway [28]. The use of hierarchical modelling technique to represent multilevel data structure in crash prediction has been growing since then. In some research, hierarchical models were used to predict crash frequency [39, 15, 37, 29, 34, 42, 26, 21] and in other research, hierarchical models were developed to identify factors affecting crash severity [28, 32, 27]. In previous studies, models were proposed to account for unobserved heterogeneity at the segment level, for example, hierarchical Bayesian binary probit models[43], binary logit models [18], Bayesian multilevel Poisson-lognormal joint models [5], multivariate hierarchical Poisson-lognormal spatial joint models [4], and grouped latent class ordered probit models with class-probability functions[16]. In general, these models accommodated the unobserved heterogeneity arising from the random effects terms, the possibility of systematic variations of unobserved groups across the highway segments or corridors which consist of intersections and roadway segments. In our study, two-level and three-level Bayesian hierarchical models are proposed to capture the unobserved heterogeneity by varying the accident intensity across the motorways and motorway segments.

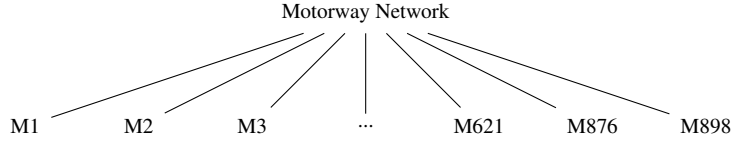
The motorway network is considered as a linear network and road accidents as a spatial point pattern involving the spatial locations of accidents. A linear network  $L$  is defined as the union  $L = \bigcup_{i=1}^n \ell_i$  of a finite collection of line segments  $\ell_1, \dots, \ell_n$  in the plane [7]. The line segment in the plane with endpoints  $u$  and  $v$  is given by  $[u, v] = \{tu + (1 - t)v : 0 \leq t \leq 1\}$ . For a point process on a linear network, an intensity of points along the network is defined as the expected number of points per unit length of network [8].

The novelty of this work is to estimate the intensity function of accidents and study its pattern across the UK motorway network using a Bayesian approach. We proposed hierarchical models to estimate the motorway-specific and segment-specific accidents intensities. In a two-level model, the intensity of accidents was considered homogeneous across segments within each motorway, but was inhomogeneous across motorways. Each motorway consists of junctions that are joined by grouped segments. The intensity of accidents may be inhomogeneous across grouped segments.

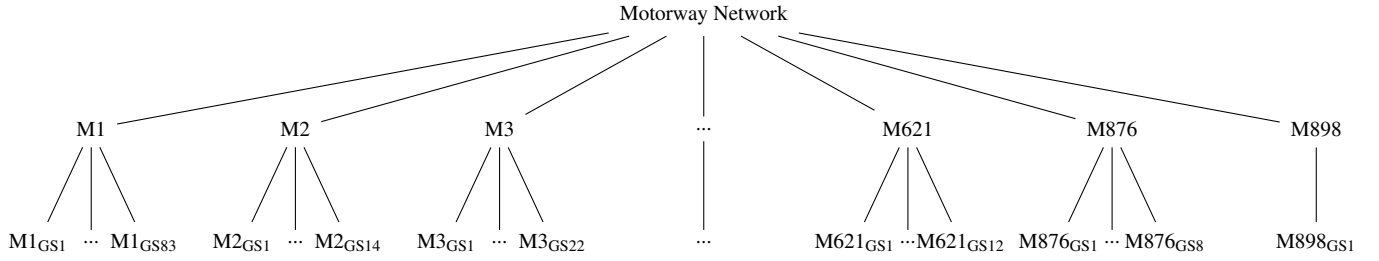
Thus, ignoring the between grouped segments heterogeneity may underestimate the standard error of the accident intensity. In a three-level model, the intensity of accidents was considered homogeneous within grouped segments, but was inhomogeneous among grouped segments and motorways.

## **2. Description of Data**

The accident data are obtained from the Department for Transport in Great Britain [1]. A STATS19 reporting form was used to record accident on the road network by police officers. This form gives details about accident circumstances such as road and weather conditions, time and location of accident, the driver's behaviour and the vehicles involved in a road accident. The traffic accident is registered as the sample point for each link of road in the road network. The road network consists of individual road sections that are segmented in junction to junction link, with different lengths. For this study, the unit of measurement for length is a kilometer. Segments are grouped according count point locations existing on Great Britain's motorway network. In these count points, "traffic estimates are calculated for each link of Great Britain's motorway network, with links' start and end points defined as where the link joins a motorway junction [2, 3]. Each link has a uniquely referenced Count Point (CP), where the traffic is usually counted by enumerators" [2]. There are 51 motorways in the UK and data are recorded on 49 motorways by STATS19. The hierarchical structure of data in this study is shown in Figure 1.



(a) Two-level hierarchical model



(b) Three-level hierarchical model

Figure 1: The hierarchical structure of UK motorway network in 2016.  $M_i$  denotes motorway  $i$ .  $M1_{GS1}$  represents the first grouped segments on M1;  $M1_{GS83}$  represents the eighty third grouped segments on M1 and so on for the rest of the symbols of grouped segments for other motorways. The grouped segments are links joining motorway junctions and they are grouped to incorporate the heterogeneity of the accident intensity across them [2].

Table 1 shows the descriptive statistics for the accident data in 2016 for 49 motorways in the UK.

Motorway	Length of motorway (in Kilometers)	Number of accidents	Number of grouped segments
M1	304.50	593	83
M2	39.39	53	14
M3	95.89	188	22
M4	299.84	475	80
M5	256.22	263	52
M6	366.76	637	90
M8	87.85	163	51
M9	52.60	25	18
M11	80.41	130	17
M18	45.25	48	13
M20	79.54	153	18
M23	26.59	34	7
M25	181.52	657	48
M26	15.46	15	2
M27	47.86	151	24
M32	6.87	17	6
M40	140.21	179	33
M42	64.93	88	21
M45	12.39	3	3
M48	19.30	6	5
M49	8.30	1	3
M50	33.19	5	8
M53	32.21	31	17
M54	35.90	12	13
M55	19.90	19	5
M56	57.89	72	30
M57	16.48	22	11
M58	18.95	7	10
M60	56.75	72	40
M61	36.50	50	12
M62	154.90	253	48
M65	41.88	71	19
M66	13.64	15	8
M67	7.53	4	4
M69	26.18	23	5
M73	11.23	10	8
M74	141.30	55	54
M77	28.93	17	22
M80	41.67	38	33
M90	49.32	23	16
M180	39.68	13	6
M181	3.78	2	1
M271	3.74	8	4
M275	3.98	12	5
M602	6.37	7	2
M606	4.03	14	3
M621	15.05	35	12
M876	11.89	11	8
M898	1.38	2	1

Table 1: Descriptive statistics for accident data on the UK motorways in 2016. The length of motorway is in kilometer.

### 3. Models

#### 3.1. Two-level Bayesian Hierarchical Model (Model 1)

##### 3.1.1. Model Definition

Let  $m$  denote the total number of motorways. The number of accidents  $n_i$  on the motorway  $i$  ( $i = 1, \dots, m$ ) follows a Poisson distribution with mean  $\lambda_i L_i$  where  $L_i$  represents the length of motorway  $i$  and  $\lambda_i$  is the accident intensity on the motorway  $i$  per unit length. Here  $\lambda_i L_i$  is the expected number of accidents on the motorway  $i$  and it can vary from one motorway to another because each motorway could have different conditions and features. Let  $\alpha_i = \log \lambda_i$  denote the log-intensity function and assume it follows a normal distribution  $N(\alpha, \tau^2)$ . The two-level hierarchical model for traffic accidents is given below,

$$\begin{aligned} n_i &\sim \text{Pois}(\lambda_i L_i), \quad i = 1, \dots, m, \\ \alpha_i &\sim N(\alpha, \tau^2). \end{aligned} \quad (1)$$

Here  $\alpha$  is the overall log-intensity and  $\tau^2$  is the between-motorway variance. In this model, we assume that each accident's location follows a uniform distribution on interval  $(0, L_i)$ .

##### 3.1.2. Likelihood Function

Let  $\mathbf{N} = \{n_i, i = 1, \dots, m\}$  represent the accident count and  $\Theta = \{\alpha_1, \alpha_2, \dots, \alpha_m; \alpha; \tau^2\}$  the parameters in the two-level hierarchical model. The likelihood for the two-level hierarchical model is given by,

$$L(\mathbf{N}|\Theta) \propto \prod_{i=1}^m \exp(n_i \alpha_i - L_i \exp(\alpha_i)) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right). \quad (2)$$

##### 3.1.3. Prior Distribution

The specification of the prior distribution depends on available information about unknown parameters. The strategy for specifying prior distributions for the parameters in the hierarchical model includes conjugate, vague and weakly-informative priors. Where possible, conjugate priors are used to ensure that the conditional distributions are of closed forms to ease the simulation; otherwise, we specify non-conjugate priors. In either cases, we specify the values of parameters to have

large variances for vague or non-informative priors, and make use of pre-existing data to construct informative priors. For  $\alpha$ , a conjugate normal prior  $N(\mu_0, \sigma_0^2)$  is assigned. We used a conjugate inverse gamma prior with a shape parameter  $\alpha_0$  and rate parameter  $\beta_0$  for  $\tau^2$ . Alternatively, we specify a uniform prior  $\text{unif}(0, a)$ , with  $a > 0$ , for the between-motorway standard deviation  $\tau$  [31]. Another choice of prior is a half-normal distribution  $\text{HN}(\theta)$  for  $\tau$ , where  $\theta = \frac{\sqrt{\pi}}{\sigma\sqrt{2}}$  and  $\sigma > 0$  as detailed in [30].

#### 3.1.4. Posterior Distribution

##### Posterior if the prior distribution on $\tau^2$ is an inverse gamma distribution

The posterior distribution is the product of the likelihood and the prior distribution. Therefore, the joint posterior density function for the parameters given the observed data is

$$\begin{aligned} \pi(\Theta|\mathbf{N}) &= \prod_{i=1}^m \exp(n_i \alpha_i - L_i \exp(\alpha_i)) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \times \\ &\quad \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} (\tau^2)^{-\alpha_0-1} \exp(-\beta_0/\tau^2). \end{aligned} \quad (3)$$

The conditional posterior distribution of  $\alpha_i$  is given by,

$$\pi(\alpha_i|\alpha, \tau^2, \mathbf{N}) \propto \exp(n_i \alpha_i - L_i \exp(\alpha_i)) \times \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right). \quad (4)$$

The conditional posterior distribution of  $\alpha$  is a normal distribution  $N(\mu_\alpha, \sigma_\alpha^2)$  with mean and variance respectively given below,

$$\mu_\alpha = \frac{\frac{\sum_{i=1}^m \alpha_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{m}{\tau^2} + \frac{1}{\sigma_0^2}} \text{ and } \sigma_\alpha^2 = \frac{1}{\frac{m}{\tau^2} + \frac{1}{\sigma_0^2}}. \quad (5)$$

The conditional posterior distribution of  $\tau^2$  is given by,

$$\pi(\tau^2|\alpha; \alpha_1, \dots, \alpha_m; \mathbf{N}) \propto (\tau^2)^{-(\alpha_0 + \frac{m}{2}) - 1} \exp\left(-\frac{\beta_0 + \frac{\sum_{i=1}^m (\alpha_i - \alpha)^2}{2}}{\tau^2}\right). \quad (6)$$



Hence,  $\tau^2$  has an inverse gamma distribution with shape  $\alpha_0 + \frac{m}{2}$  and rate  $\beta_0 + \frac{\sum_{i=1}^m (\alpha_i - \alpha)^2}{2}$ .

**Posterior if the prior distribution on  $\tau$  is a uniform distribution**

The joint posterior density is

$$\begin{aligned} \pi(\Theta|\mathbf{N}) \propto \prod_{i=1}^m \exp(n_i \alpha_i - L_i \exp(\alpha_i)) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \times \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right). \end{aligned} \quad (7)$$

The conditional posterior distributions of  $\alpha_i$  ( $i = 1, \dots, m$ ) and  $\alpha$  are given in (4) and (5). The conditional posterior density of  $\tau$  is given by,

$$\pi(\tau|\alpha; \alpha_1, \dots, \alpha_m; \mathbf{N}) \propto \left(\frac{1}{\sqrt{2\pi\tau^2}}\right)^m \exp\left(-\sum_{i=1}^m \frac{(\alpha_i - \alpha)^2}{2\tau^2}\right). \quad (8)$$

**Posterior if the prior distribution on  $\tau$  is a half-normal distribution (HN) with parameter  $\theta$**

The joint posterior distribution is

$$\begin{aligned} \pi(\Theta|\mathbf{N}) = \prod_{i=1}^m \exp(n_i \alpha_i - L_i \exp(\alpha_i)) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \times \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right) \times \frac{2\theta}{\pi} \exp\left(-\frac{\tau^2\theta^2}{\pi}\right). \end{aligned} \quad (9)$$

The conditional posterior distributions of  $\alpha_i$  ( $i = 1, \dots, m$ ) and  $\alpha$  are the same as in equations (4) and (5). The conditional posterior distribution of  $\tau$  is

$$\pi(\tau|\alpha; \alpha_1, \dots, \alpha_m; \mathbf{N}) \propto \tau^{-m} \exp\left(-\sum_{i=1}^m \frac{(\alpha_i - \alpha)^2}{2\tau^2} - \frac{\tau^2\theta^2}{\pi}\right). \quad (10)$$

### 3.1.5. Estimation

In equations (5) and (6), the conditional posterior distributions of  $\alpha$  and  $\tau^2$  given other parameters have a known form, but the conditional posterior distributions of  $\alpha_i$  ( $i = 1, \dots, m$ ) in equation (4) do not have known forms. Therefore, the Metropolis-Hastings within Gibbs sampler is used to simulate the Markov chains of  $\alpha_i$  ( $i = 1, \dots, m$ ),  $\alpha$  and  $\tau^2$ . A new value  $\hat{\alpha}_i$  is simulated from a proposal distribution  $q_1(\hat{\alpha}_i, \alpha_i^{(t-1)})$ , which is the normal distribution with mean equalling to current

value  $\alpha_i^{(t-1)}$  and the variance is chosen such that an acceptance rate of  $\hat{\alpha}_i$  is between 0.24 and 0.40 [20]. The proposed value is accepted with the probability

$$r_1(\alpha_i^{(t-1)}, \hat{\alpha}_i) = \min \left\{ \frac{\pi(\hat{\alpha}_i | \alpha^{(t-1)}, \tau^{2(t-1)}) q_1(\alpha_i^{(t-1)}, \hat{\alpha}_i)}{\pi(\alpha_i^{(t-1)} | \alpha^{(t-1)}, \tau^{2(t-1)}) q_1(\hat{\alpha}_i, \alpha_i^{(t-1)})}, 1 \right\}. \quad (11)$$

If the proposed value is rejected, and the current value is taken as the next value in the Markov chain.

The uniform prior distribution on  $\tau$  leads to the conditional posterior distribution of  $\tau$  as given in equation (8). This posterior distribution does not have a closed form, therefore, the Metropolis-Hastings sampler is used. The two steps are defined as follows. Firstly, we draw a proposed value,  $\hat{\tau}$ , from the proposal distribution  $q_2(\hat{\tau}, \tau^{(t-1)})$ , which is a normal distribution with mean equal to the current value  $\tau^{(t-1)}$ . Secondly, the proposed value is accepted with the probability

$$r_2(\tau^{(t-1)}, \hat{\tau}) = \min \left\{ \frac{\pi(\hat{\tau} | \alpha^{(t)}, \alpha_1^{(t)}, \dots, \alpha_m^{(t)}) q_2(\tau^{(t-1)}, \hat{\tau})}{\pi(\tau^{(t-1)} | \alpha^{(t)}, \alpha_1^{(t)}, \dots, \alpha_m^{(t)}) q_2(\hat{\tau}, \tau^{(t-1)})}, 1 \right\}. \quad (12)$$

The conditional posterior distribution for  $\tau$  in equation (10) is produced by using a half-normal prior distribution and does not have a closed form. Therefore, the Metropolis-Hastings sampler is used to simulate  $\tau$ . This sampler includes generating the proposed value  $\hat{\tau}$  from the proposal distribution  $q_2(\hat{\tau}, \tau^{(t-1)})$  and the proposed value is accepted with the probability  $r_2(\tau^{(t-1)}, \hat{\tau})$  described in equation (12). The proposal distribution  $q_2(\cdot, \tau^{(t-1)})$  is a normal distribution with current state  $\tau^{(t-1)}$  as the mean.

### 3.2. Two-level Frequentist Hierarchical Method (Model 2)

In this section, we describe a two-stage approach to modelling the accident data on the UK motorway network using the frequentist method. In stage one, we used the maximum likelihood method to estimate the log-intensity function  $y_i$  and the corresponding standard deviation  $\sigma_i$ . In stage two, the estimated log-intensity functions are combined across motorways to produce an overall estimate. The model can be formulated as follows,

$$\begin{aligned} y_i &\sim N(\alpha_i, \sigma_i^2), \\ \alpha_i &\sim N(\alpha, \tau^2). \end{aligned} \quad (13)$$

Here  $y_i$  represents the estimated intensity at the log scale for motorway  $i$ ,  $\alpha_i$  represents the random effects for the log-intensity and  $\sigma_i^2$  is the within-motorway variance corresponding to  $y_i$ ;  $\alpha$  is the overall intensity at the log scale and  $\tau^2$  represents the between-motorway heterogeneity. The marginal distribution of each estimated log-intensity  $y_i$  is a normal distribution with mean  $\alpha$  and variance  $(\sigma_i^2 + \tau^2)^{-1}$  [22]. Hence the contribution of motorway  $i$  to the likelihood for  $\alpha$  and  $\tau^2$  is given by,

$$L_i(\alpha, \tau^2 | y_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi(\sigma_i^2 + \tau^2)}} \exp\left(-\frac{(y_i - \alpha)^2}{2(\sigma_i^2 + \tau^2)}\right). \quad (14)$$

For  $m$  independent motorways, the likelihood is given by the product of the individual motorway likelihoods as follows,

$$L(\alpha, \tau^2 | \mathbf{y}, \sigma^2) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi(\sigma_i^2 + \tau^2)}} \exp\left(-\frac{(y_i - \alpha)^2}{2(\sigma_i^2 + \tau^2)}\right). \quad (15)$$

### 3.3. Three-level Bayesian Hierarchical Model (Model 3)

#### 3.3.1. Model Definition

In the three-level hierarchical model, the number of accidents within grouped segments follows a homogeneous process but a non-homogeneous process across grouped segments and motorways. Let  $m$  denote the total number of motorways and  $s_i$  ( $i = 1, \dots, m$ ) the number of grouped segments for each motorway  $i$ . Suppose that the intensity of accidents per kilometer is  $\lambda_{ij}$  ( $i = 1, \dots, m$ ;  $j = 1, \dots, s_i$ ), where  $i$  is the index of motorway and  $j$  is the index of grouped segments. The number of accidents  $n_{ij}$  on each grouped segments follows a Poisson distribution with mean  $\lambda_{ij}L_{ij}$ , where  $L_{ij}$  represents the length (in kilometer) of the grouped segments  $j$  on motorway  $i$ . Let  $\alpha_{ij} = \log \lambda_{ij}$  denote the log-intensity function. The three-level hierarchical model is given below,

$$\begin{aligned} n_{ij} &\sim \text{Pois}(\lambda_{ij}L_{ij}), i = 1, \dots, m; j = 1, \dots, s_i, \\ \alpha_{ij} &\sim N(\alpha_i, \tau_i^2), \\ \alpha_i &\sim N(\alpha, \tau^2). \end{aligned} \quad (16)$$

The second level includes the log-intensity of accidents,  $\alpha_{ij}$ , on each grouped segments and the log-intensity of accidents,  $\alpha_i$ , on each motorway as well as the between grouped segments het-

erogeneity,  $\tau_i^2$ . The third level includes the overall log-intensity of accidents  $\alpha$  and the between-motorway heterogeneity,  $\tau^2$ . The intensity of accidents is constant on grouped segments that have the same mark, but it varies across grouped segments and motorways.

### 3.3.2. Likelihood Function

Let  $\Theta_3$  denote model parameters  $\{\alpha_{11}, \dots, \alpha_{ms_m}; \alpha_1, \dots, \alpha_m; \tau_1^2, \dots, \tau_m^2; \alpha; \tau^2\}$  in the three-level hierarchical model. Let  $\gamma$  denote the log-intensity of accidents on grouped segments  $\{\alpha_{11}, \dots, \alpha_{ms_m}\}$ ,  $\alpha = \{\alpha_1, \dots, \alpha_m\}$  and  $\tau^2 = \{\tau_1^2, \dots, \tau_m^2\}$ . It is assumed that the accidents are uniformly distributed within grouped segments.

The likelihood function for the proposed three-level hierarchical model is given by,

$$\begin{aligned} L(\mathbf{N}|\Theta_3) &= P(\mathbf{N}|\gamma) \times P(\gamma|\alpha, \tau^2) \times P(\alpha|\alpha, \tau^2) \\ &\propto \prod_{i=1}^m \prod_{j=1}^{s_i} \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})) \\ &\quad \times \prod_{i=1}^m \prod_{j=1}^{s_i} \frac{1}{\sqrt{2\pi\tau_i^2}} \exp\left(-\frac{(\alpha_{ij} - \alpha_i)^2}{2\tau_i^2}\right) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right). \end{aligned} \quad (17)$$

### 3.3.3. Posterior Distribution

The same prior distributions for  $\alpha$  and  $\tau^2$  as the ones described in section 3.1.3 are considered.

**Posterior if the prior distribution on  $\tau^2$  is an inverse gamma distribution with shape  $a_0$  and scale  $b_0$**

A joint posterior distribution for parameters  $\Theta_3$  is given by,

$$\begin{aligned} \pi(\Theta_3|\mathbf{N}) &\propto \prod_{i=1}^m \prod_{j=1}^{s_i} \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})) \times \prod_{i=1}^m \prod_{j=1}^{s_i} \frac{1}{\sqrt{2\pi\tau_i^2}} \exp\left(-\frac{(\alpha_{ij} - \alpha_i)^2}{2\tau_i^2}\right) \\ &\quad \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \times \prod_{i=1}^m \frac{b_0^{a_0}}{\Gamma(a_0)} (\tau_i^2)^{-a_0-1} \exp\left(-\frac{b_0}{\tau_i^2}\right) \\ &\quad \times \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} (\tau^2)^{-\alpha_0-1} \exp\left(-\frac{\beta_0}{\tau^2}\right). \end{aligned} \quad (18)$$

The conditional posterior distribution of  $\alpha_{ij}$  is given by,

$$\pi(\alpha_{ij}|\alpha, \tau^2, \mathbf{N}) \propto \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})) \exp\left(-\frac{(\alpha_{ij} - \alpha_i)^2}{2\tau_i^2}\right). \quad (19)$$

The conditional posterior distribution of  $\alpha_i$  is a normal distribution  $N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$  with mean and variance:

$$\mu_{\alpha_i} = \frac{\frac{\sum_{j=1}^{s_i} \alpha_{ij}}{\tau_i^2} + \frac{\alpha}{\tau^2}}{\frac{s_i}{\tau_i^2} + \frac{1}{\tau^2}} \text{ and } \sigma_{\alpha_i}^2 = \frac{1}{\frac{s_i}{\tau_i^2} + \frac{1}{\tau^2}}. \quad (20)$$

The conditional posterior distribution of  $\tau_i^2$  is given by,

$$\tau_i^2 \sim \text{Inv-Gamma}\left(\frac{s_i}{2} + a_0, \sum_{j=1}^{s_i} \frac{(\alpha_{ij} - \alpha_i)^2}{2} + b_0\right). \quad (21)$$

The conditional posterior distribution of  $\alpha$  is a  $N(\mu_\alpha, \sigma_\alpha^2)$  with mean and variance respectively given below,

$$\mu_\alpha = \frac{\frac{\sum_{i=1}^m \alpha_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{m}{\tau^2} + \frac{1}{\sigma_0^2}} \text{ and } \sigma_\alpha^2 = \frac{1}{\frac{m}{\tau^2} + \frac{1}{\sigma_0^2}}. \quad (22)$$

The conditional posterior distribution of  $\tau^2$  is given by,

$$\tau^2 \sim \text{Inv-Gamma}\left(\frac{m}{2} + \alpha_0, \frac{\sum_{i=1}^m (\alpha_i - \alpha)^2}{2} + \beta_0\right). \quad (23)$$

### Posterior if the prior distribution on $\tau$ is a uniform distribution

The probability density function of the uniform prior distribution on  $\tau$  is constant, so it does not appear in the joint posterior distribution. Hence, the joint posterior distribution is given by,

$$\begin{aligned} \pi(\Theta_3|\mathbf{N}) &\propto \prod_{i=1}^m \prod_{j=1}^{s_i} \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})) \times \prod_{i=1}^m \prod_{j=1}^{s_i} \frac{1}{\sqrt{2\pi\tau_i^2}} \exp\left(-\frac{(\alpha_{ij} - \alpha_i)^2}{2\tau_i^2}\right) \\ &\times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \times \prod_{i=1}^m \frac{b_0^{a_0}}{\Gamma(a_0)} (\tau_i^2)^{-a_0-1} \exp\left(\frac{-b_0}{\tau_i^2}\right) \\ &\times \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(\frac{-(\alpha - \mu_0)^2}{2\sigma_0^2}\right). \end{aligned} \quad (24)$$

The conditional posterior distributions of  $\alpha_{ij}$ ,  $\alpha_i$ ,  $\tau_i^2$  ( $i = 1, \dots, m$ ;  $j = 1, \dots, s_i$ ) and  $\alpha$  are the same in equations (19)-(22). The conditional posterior distribution of  $\tau$  is given by,

$$\pi(\tau|\alpha, \alpha, \mathbf{N}) \propto \left(\frac{1}{\sqrt{2\pi\tau^2}}\right)^m \exp\left(-\sum_{i=1}^m \frac{(\alpha_i - \alpha)^2}{2\tau^2}\right). \quad (25)$$

### Posterior if the prior distribution on $\tau$ is a half-normal (HN) distribution with parameter $\theta$

If we use  $\text{HN}(\theta)$  as the prior for  $\tau$ , the joint posterior distribution is given by,

$$\begin{aligned} \pi(\Theta_3|\mathbf{N}) &\propto \prod_{i=1}^m \prod_{j=1}^{s_i} \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})) \\ &\times \prod_{i=1}^m \prod_{j=1}^{s_i} \frac{1}{\sqrt{2\pi\tau_i^2}} \exp\left(-\frac{(\alpha_{ij} - \alpha_i)^2}{2\tau_i^2}\right) \times \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\alpha_i - \alpha)^2}{2\tau^2}\right) \\ &\times \prod_{i=1}^m \frac{b_0^{a_0}}{\Gamma(a_0)} (\tau_i^2)^{-a_0-1} \exp\left(\frac{-b_0}{\tau_i^2}\right) \times \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(\frac{-(\alpha - \mu_0)^2}{2\sigma_0^2}\right) \\ &\times \frac{2\theta}{\pi} \exp\left(-\frac{\tau^2\theta^2}{\pi}\right). \end{aligned} \quad (26)$$

The conditional posterior distributions of  $\alpha_{ij}$ ,  $\alpha_i$ ,  $\tau_i^2$ , ( $i = 1, \dots, m$ ;  $j = 1, \dots, s_i$ ) and  $\alpha$  are the same in equations (19)-(22). The conditional posterior distribution of  $\tau$  is given by,

$$\pi(\tau|\alpha, \alpha, \mathbf{N}) \propto \tau^{-m} \exp\left(-\sum_{i=1}^m \frac{(\alpha_i - \alpha)^2}{2\tau^2} - \frac{\tau^2\theta^2}{\pi}\right). \quad (27)$$

### 3.3.4. Estimation

Bayesian estimation of the three-level hierarchical model is performed using Metropolis-Hastings within Gibbs sampler. We generate random samples from conditional posterior distributions of  $\alpha$ ,  $\alpha_i$ ,  $\tau^2$  and  $\tau_i^2$  ( $i = 1, \dots, m$ ), respectively. Conditional posterior distributions of  $\alpha_{ij}$ , ( $i = 1, \dots, m; j = 1, \dots, s_i$ ) are not of closed forms. In this case, the Metropolis-Hasting algorithm is used. Normal proposal distributions are specified for  $\alpha_{ij}$ , ( $i = 1, \dots, m; j = 1, \dots, s_i$ ) with mean  $\alpha_{ij}^{(t-1)}$  and variance  $\sigma_{ij}^2$ , where  $t$  ( $t = 1, \dots, M$ ) is the iteration index. The variance  $\sigma_{ij}^2$  is chosen such that an acceptance rate is between 0.24 and 0.40) [20].

A value  $\hat{\alpha}_{ij}$  generated from the proposal distribution  $q_1(\hat{\alpha}_{ij}, \alpha_{ij}^{(t-1)})$  is accepted with probability

$$r_1(\alpha_{ij}^{(t-1)}, \hat{\alpha}_{ij}) = \min \left\{ \frac{\pi(\hat{\alpha}_{ij} | \alpha_i^{(t-1)}, \tau_i^{2(t-1)}) q_1(\alpha_{ij}^{(t-1)}, \hat{\alpha}_{ij})}{\pi(\alpha_{ij}^{(t-1)} | \alpha_i^{(t-1)}, \tau_i^{2(t-1)}) q_1(\hat{\alpha}_{ij}, \alpha_{ij}^{(t-1)})}, 1 \right\}. \quad (28)$$

A new value  $\hat{\tau}$  is generated from a proposal distribution  $q_2(\hat{\tau}, \tau^{(t-1)})$ , which is a normal distribution with mean equalling to current value  $\tau^{(t-1)}$ . The new value  $\hat{\tau}$  is accepted with probability

$$r_2(\tau^{(t-1)}, \hat{\tau}) = \min \left\{ \frac{\pi(\hat{\tau} | \alpha^{(t)}, \alpha_1^{(t)}, \dots, \alpha_m^{(t)}) q_2(\tau^{(t-1)}, \hat{\tau})}{\pi(\tau^{(t-1)} | \alpha^{(t)}, \alpha_1^{(t)}, \dots, \alpha_m^{(t)}) q_2(\hat{\tau}, \tau^{(t-1)})}, 1 \right\}. \quad (29)$$

### 3.4. Three-level Frequentist Hierarchical Method (Model 4)

The maximum likelihood estimation is used in three stages to estimate the model parameters (16). In stage one, the log-intensity of accidents  $\alpha_{ij}$  is estimated for each grouped segments where the relevant part of likelihood function in equation (17) is used,

$$L_{ij}(\alpha_{ij}; \mathbf{N}) = \exp(n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij})). \quad (30)$$

The log-likelihood function is

$$\ell_{ij}(\alpha_{ij}; \mathbf{N}) = n_{ij}\alpha_{ij} - L_{ij} \exp(\alpha_{ij}). \quad (31)$$

The maximum likelihood estimate (M.L.E.) of  $\alpha_{ij}$  is given by,

$$\hat{\alpha}_{ij} = \log n_{ij} - \log L_{ij}. \quad (32)$$

To calculate the standard error of  $\hat{\alpha}_{ij}$ , we use the Fisher information matrix  $I(\hat{\alpha}_{ij})$ , which is a scalar containing the entry

$$I(\hat{\alpha}_{ij}) = -E \left[ H(\hat{\alpha}_{ij}) \right] = -E \left[ \frac{\partial^2 \ell_{ij}}{\partial \alpha_{ij}^2} \right] = L_{ij} \exp(\alpha_{ij}), \quad (33)$$

where  $H(\hat{\alpha}_{ij})$  represents the Hessian matrix. The square root of the inverse of the Fisher information scalar is an estimator of the standard error for  $\alpha_{ij}$

$$SE(\hat{\alpha}_{ij}) = \sqrt{I^{-1}(\hat{\alpha}_{ij})} = \left( L_{ij} \exp(\alpha_{ij}) \right)^{-\frac{1}{2}}. \quad (34)$$

In stage two, the log-intensity of accidents  $\alpha_i$  ( $i = 1, \dots, m$ ) is estimated for each motorway. Here, the point estimates  $\hat{\alpha}_{ij}$ , ( $i = 1, \dots, m$ ;  $j = 1, \dots, n_i$ ) are substituted in the likelihood function in equation (17). The relevant part of the likelihood function is given by,

$$L_i(\alpha, \tau^2; \hat{\gamma}) = \prod_{i=1}^m \prod_{j=1}^{s_i} \frac{1}{\sqrt{2\pi\tau_i^2}} \exp \left( -\frac{(\hat{\alpha}_{ij} - \alpha_i)^2}{2\tau_i^2} \right). \quad (35)$$

The log-likelihood function is

$$\ell_i(\alpha, \tau^2; \hat{\gamma}) = \sum_{i=1}^m \left[ -\frac{s_i}{2} \log \tau_i^2 - \frac{\sum_{j=1}^{s_i} (\hat{\alpha}_{ij} - \alpha_i)^2}{2\tau_i^2} \right]. \quad (36)$$

The maximum likelihood estimates  $\hat{\alpha}_i$  and  $\hat{\tau}_i^2$  are respectively given by,

$$\hat{\alpha}_i = \frac{\sum_{j=1}^{s_i} \hat{\alpha}_{ij}}{s_i} \text{ and } \hat{\tau}_i^2 = \frac{\sum_{j=1}^{s_i} (\hat{\alpha}_{ij} - \hat{\alpha}_i)^2}{s_i}. \quad (37)$$



To obtain the standard errors of  $\hat{\alpha}_i$  and  $\hat{\tau}_i^2$ , we use the Fisher information matrix as follows,

$$I(\hat{\alpha}_i, \hat{\tau}_i^2) = -E[H(\hat{\alpha}_i, \hat{\tau}_i^2)] = -E \begin{bmatrix} \frac{\partial^2 \ell_i}{\partial \alpha_i^2} & \frac{\partial^2 \ell_i}{\partial \alpha_i \partial \tau_i^2} \\ \frac{\partial^2 \ell_i}{\partial \tau_i^2 \partial \alpha_i} & \frac{\partial^2 \ell_i}{\partial (\tau_i^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{s_i}{\tau_i^2} & 0 \\ 0 & \frac{s_i}{2\tau_i^4} \end{bmatrix}, \quad (38)$$

where  $H(\hat{\alpha}_i, \hat{\tau}_i^2)$  represents the Hessian matrix. The inverse of the Fisher information matrix is given by,

$$I^{-1}(\hat{\alpha}_i, \hat{\tau}_i^2) = \begin{bmatrix} \frac{\tau_i^2}{s_i} & 0 \\ 0 & \frac{2\tau_i^4}{s_i} \end{bmatrix}. \quad (39)$$

The standard errors of  $\hat{\alpha}_i$  and  $\hat{\tau}_i^2$  are the square root of diagonal elements in (29), and hence are given by,

$$SE(\hat{\alpha}_i) = \sqrt{\frac{\tau_i^2}{s_i}}. \quad (40)$$

$$SE(\hat{\tau}_i^2) = \sqrt{\frac{2\tau_i^4}{s_i}}. \quad (41)$$

In stage three,  $\alpha$  and  $\tau^2$  are estimated by maximising the following likelihood,

$$L(\alpha, \tau^2; \hat{\alpha}_i) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{(\hat{\alpha}_i - \alpha)^2}{2\tau^2} \right\}. \quad (42)$$

The estimated values  $\hat{\alpha}_i$  and  $SE(\hat{\alpha}_i)$  are used as data in (42).

#### 4. Application to UK Motorway Accident Data

##### Non-informative and weakly informative prior distributions

A non-informative prior distribution reflects the lack of prior information about a parameter [33]. A conjugate prior could be non-informative, such as Inv-Gamma(0.001, 0.001) or weakly-informative, such as Inv-Gamma(0.1, 0.1) for  $\tau^2$ . As a sensitivity analysis, we use the uniform prior  $\text{unif}(0, 100)$  and a half-normal prior distribution  $\text{HN}(0.14)$  for  $\tau$ , both are non-informative priors on  $\tau$  [46, 31].

We used Inv-Gamma(0.001, 0.001) as prior for  $\tau_i^2$  ( $i = 1, \dots, m$ ). Finally, a non-informative normal prior distribution was used with mean  $\mu_0 = 0$  and variance  $\sigma_0^2 = 100$  for  $\alpha$ .

### Informative prior distributions

An informative prior describes specific pre-existing information about parameter [33]. The maximum likelihood estimates of  $\tau^2$  and  $\alpha$  and their standard errors of traffic accident data from an earlier year (e.g. 2015) will be used to specify the informative priors in the Bayesian analysis of accident data in the subsequent year. More specifically, the parameters for inverse gamma prior are calculated from solving the following equation:

$$\begin{aligned} E(\tau^2) &= \frac{\alpha_0}{\beta_0} = \hat{\tau}_{ML}^2, \\ \text{var}(\tau^2) &= \frac{\alpha_0}{\beta_0^2} = \text{var}(\hat{\tau}_{ML}^2), \end{aligned} \quad (43)$$

where  $\hat{\tau}_{ML}^2$  is the maximum likelihood estimate, 0.3162 and  $\text{var}(\hat{\tau}_{ML}^2)$  is the variance of  $\hat{\tau}_{ML}^2$ ,  $(0.0738)^2$ , both are obtained from analysing the UK motorways accident data in 2015. Solving the equations in (43), we obtain  $\alpha_0 = 18.36$  and  $\beta_0 = 58.06$ . Thus, the informative prior for  $\tau^2$  is Inv-Gamma(18.36, 58.06). Similarly,  $\mu_0 = -6.65$  and  $\sigma_0^2 = 0.09^2$ .

### Results

Both two-level and three-level hierarchical models are used to analyse the observed accident data in 2016 for the UK motorways. Model parameters include the overall log-intensity of traffic accidents  $\alpha$ , and the between-motorway standard deviation,  $\tau$ . The MCMC simulation requires specifying starting points for the parameters. The initial values 0 and 0.1 are specified for  $\alpha$  and  $\tau$ , respectively. The MCMC algorithm was run on the two-level hierarchical model for 100,000 iterations with a burn-in period of 10,000 and a thinning interval of 10; and on the three-level hierarchical model, 500,000 iterations with a burn-in period of 50,000 and a thinning interval of 100. The number of iterations of the MCMC is different between the two-level and three-level hierarchical models to ensure convergence to the posterior distributions.

Table 2 shows that the two-level Bayesian hierarchical model with different priors gives similar estimates of both the parameters  $\alpha$  and  $\tau$  and their standard deviations as well as their credible intervals. This indicates that the estimates of the model parameters are robust to the choice of prior. The estimated  $\alpha$  and  $\tau$  using the maximum likelihood estimation are similar to those from two-level Bayesian hierarchical model. Under the three-level Bayesian hierarchical model, the posterior mean, standard deviation and 95% credible interval for  $\alpha$  and  $\tau$  are similar when non-informative

and weakly-informative prior distributions are used. The width of the credible interval of  $\alpha$  based on the Bayesian method with an informative prior is narrower than the one estimated with non-informative and weakly-informative priors. In addition, the standard deviation for  $\hat{\alpha}$  based on the Bayesian method with informative priors is smaller than that based on the Bayesian method with other priors. The estimated  $\alpha$  and  $\tau$  are similar between the frequentist and the Bayesian approaches, although the standard deviations are, in general, slightly smaller in the frequentist approach (see table 2), as expected.

Methods	Prior distribution	Parameter	Two-level hierarchical model			Three-level hierarchical model		
			Point estimate	SD	95% CI	Point estimate	SD	95% CI
Bayesian	$\tau^2 \sim \text{Inv-Gamma}(0.001, 0.001)$	$\alpha$	-6.82	0.10	(-7.01, -6.64)	-6.93	0.11	(-7.15, -6.71)
		$\tau$	0.60	0.08	(0.47, 0.77)	0.65	0.09	(0.49, 0.85)
	$\tau^2 \sim \text{Inv-Gamma}(0.1, 0.1)$	$\alpha$	-6.86	0.11	(-7.07, -6.65)	-6.93	0.11	(-7.15, -6.71)
		$\tau$	0.70	0.09	(0.54, 0.88)	0.66	0.09	(0.50, 0.85)
	$\tau \sim \text{unif}(0, 100)$	$\alpha$	-6.86	0.11	(-7.07, -6.64)	-6.93	0.11	(-7.16, -6.71)
		$\tau$	0.70	0.09	(0.55, 0.90)	0.67	0.09	(0.51, 0.87)
	$\tau \sim \text{HN}(0.14)$	$\alpha$	-6.86	0.11	(-7.07, -6.65)	-6.93	0.11	(-7.16, -6.71)
		$\tau$	0.70	0.09	(0.55, 0.89)	0.67	0.09	(0.51, 0.87)
Frequentist	$\alpha \sim \text{N}(-6.65, 0.09^2)$ $\tau^2 \sim \text{Inv-Gamma}(18.36, 58.06)$	$\alpha$	-6.62	0.09	(-6.79, -6.43)	-6.70	0.08	(-6.86, -6.55)
		$\tau$	1.40	0.23	(1.14, 2.08)	1.34	0.11	(1.15, 1.57)
		$\alpha$	-6.81	0.10	(-7.00, -6.62)	-6.64	0.08	(-6.80, -6.49)
		$\tau$	0.64	0.10	(0.54, 0.89)	0.51	0.06	(0.41, 0.66)

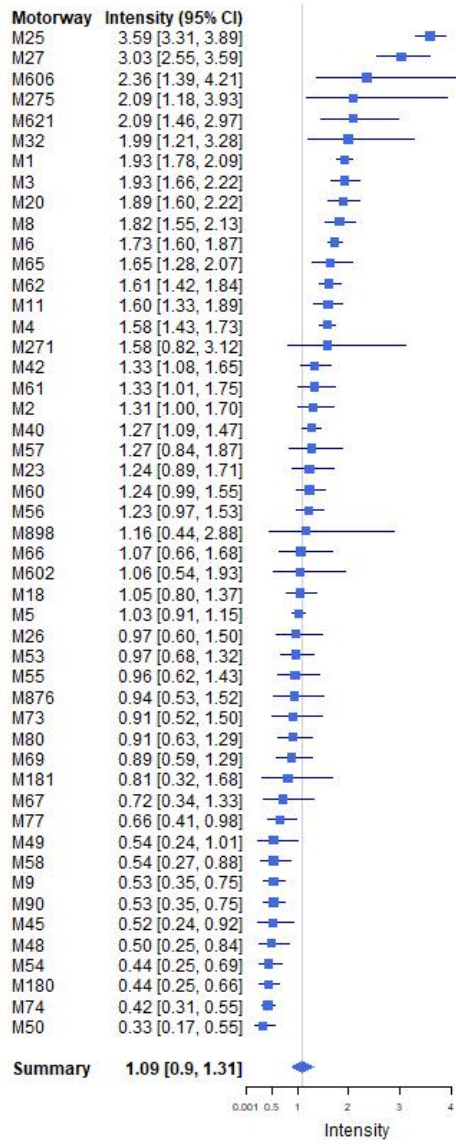
Table 2: Results from the two-level Bayesian hierarchical model, three-level Bayesian hierarchical model, two-level frequentist hierarchical model and three-level frequentist hierarchical model of traffic accidents for 2016. The prior of  $\alpha$  is  $\text{N}(0, 100)$ . HN represents the half-normal distribution. SD: Standard Deviation and CI: Credible Interval or Confidence Interval.

Table 3 shows the estimated overall intensity of accidents per one kilometer. It is clear that results for  $\lambda$  from the three-level Bayesian hierarchical model are similar, except for the informative prior where  $\lambda$  is greater, but close to the estimate from the frequentist method. The two-level Bayesian hierarchical model produces similar results under non-informative and weakly-informative priors of  $\tau$ .

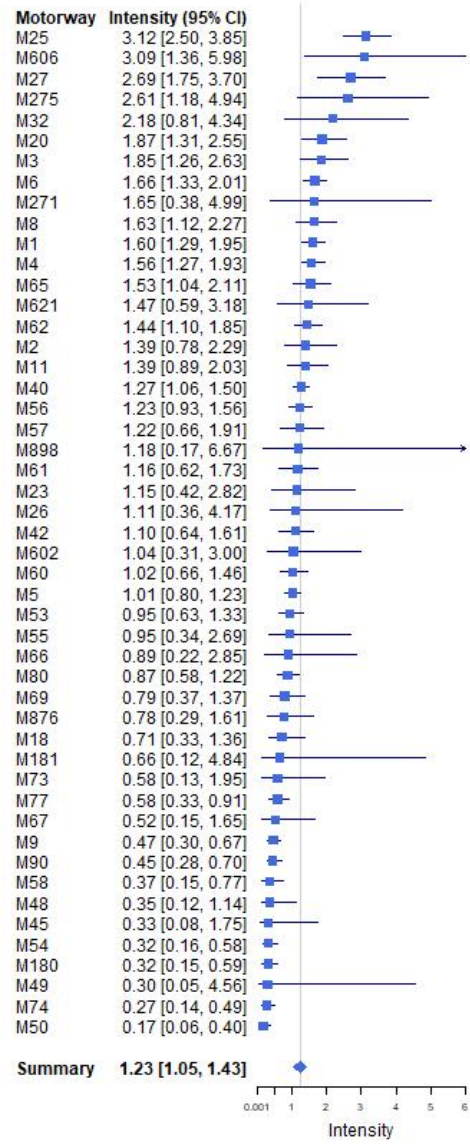
Methods	Prior distribution	Two-level hierarchical model		Three-level hierarchical model	
		Point estimate	95% CI	Point estimate	95% CI
Bayesian	$\tau^2 \sim \text{Inv-Gamma}(0.001, 0.001)$	1.09	(0.90, 1.31)	0.98	(0.79, 1.22)
	$\tau^2 \sim \text{Inv-Gamma}(0.1, 0.1)$	1.05	(0.85, 1.29)	0.98	(0.79, 1.22)
	$\tau \sim \text{unif}(0, 100)$	1.05	(0.85, 1.31)	0.98	(0.78, 1.22)
	$\tau \sim \text{HN}(0.14)$	1.05	(0.85, 1.29)	0.98	(0.78, 1.22)
	$\alpha \sim N(-6.65, 0.09^2)$ $\tau^2 \sim \text{Inv-Gamma}(18.36, 58.06)$	1.33	(1.12, 1.61)	1.23	(1.05, 1.43)
Frequentist		1.10	(0.91, 1.33)	1.31	(1.12, 1.53)

Table 3: Posterior summary and frequentist of overall accident intensity  $\lambda$  per one kilometer for UK motorway network in 2016:  $\lambda = \exp(\alpha)$ . The prior of  $\alpha$  is  $N(0, 100)$ . CI: credible interval or confidence interval. HN represents the half-normal distribution.

Figure 2 shows 49 motorways with intensities of traffic accidents per one kilometer and their corresponding credible intervals for both two-level and three-level Bayesian hierarchical models.



(a) Two-level Bayesian hierarchical model



(b) Three-level Bayesian hierarchical model

Figure 2: Results from the two-level Bayesian hierarchical model and the three-level Bayesian hierarchical model for accident data on the UK motorways in 2016. In the two-level Bayesian hierarchical model, the following prior distributions are used,  $\alpha \sim N(0, 100)$  and  $\tau^2 \sim \text{Inv-Gamma}(0.001, 0.001)$ . In the three-level Bayesian hierarchical model, the prior distributions are  $\alpha \sim N(-6.65, 0.09^2)$  and  $\tau^2 \sim \text{Inv-Gamma}(18.36, 58.06)$ . Results include the posterior mean and the corresponding 95% credible interval for the intensity of accidents  $\lambda_i = 1000 \times \exp(\alpha_i)$  per one kilometer on each motorway and the overall intensity of accidents  $\lambda$  per one kilometer. Square boxes represent posterior means of  $\lambda_i$ , ( $i = 1, \dots, m$ ). The diamond represents the estimated overall intensity of accident  $\lambda$  and its 95% credible interval. Horizontal lines denote 95% credible intervals and the solid vertical line represents the posterior mean of the overall intensity  $\lambda$ .

Based on the estimation results from Figure 2 of the Bayesian hierarchical models for the UK motorway data, the estimated intensities of accidents on the UK motorway network are classified into five categories. Category one ( $\lambda < 0.5$ ) is referred to a very low risk; Category two ( $0.5 \leq \lambda < 1$ ) is referred to a low risk; Category three ( $1 \leq \lambda < 2$ ) is referred to a moderate risk; Category four ( $2 \leq \lambda < 3$ ) is referred to a high risk. Finally, category five ( $\lambda \geq 3$ ) is referred to a very high risk. Based on the results from the Bayesian three-level hierarchical model, the moderate-risk level represents the general intensity of accidents level of the UK motorway network. Based on the results presented in Figures 2 and 3, motorways: M27, M275 and M32 are at high risk, whereas motorways: M25 and M606 form the highest risk motorways, where the expected number of accidents is above 3 per one kilometer for both motorways. On the other hand, motorways: M9, M90, M58, M45, M48, M49, M180, M54, M74 and M50 have the lowest risk such that the expected number of accidents is lower than 0.5 for these motorways.

Based on the results of the two-level Bayesian hierarchical model, Figures 2 and 4 shows that a general level of the intensity of accidents on the UK motorway network is of moderate risk where the moderate risk motorways are M32, M1, M3, M20, M8, M6, M65, M62, M11, M4, M271, M42, M61, M2, M40, M57, M23, M60, M56, M898, M66, M602, M18 and M5. Motorways M25 surrounding almost all of Greater London, England, except North Ockendon, in the United Kingdom and M27 in Hampshire, England, starting west-east from Cadnam to Portsmouth, have a very high-risk level. The expected numbers of accidents are 3.59 per one kilometer of M25 and 3.03 per one kilometer of M27. The motorways M54, M180, M74 and M50 form the lowest risk motorways and their estimated intensities are 4.4, 4.4, 4.2, 3.3 per 10 kilometers. Figure 4, moreover, illustrates that the risk intensity level for motorways M606, M621 and M275 is high and the expected number of accidents is 2.36 per one kilometer of M606 and 2.09 per one kilometer for both M621 and M275.

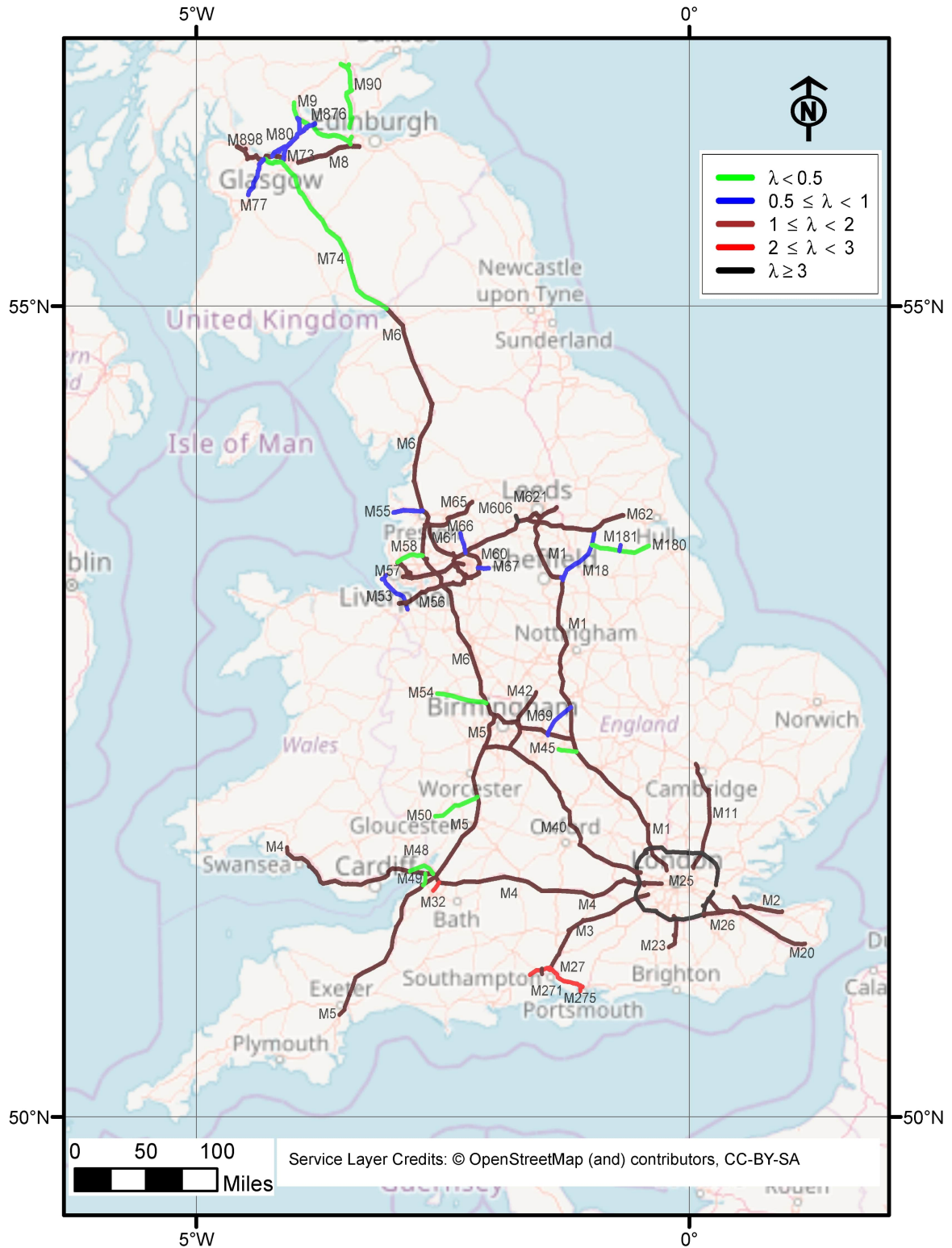


Figure 3: Estimated intensity of traffic accidents per one kilometer on the UK motorway network in 2016. The intensities are estimated using the three-level Bayesian hierarchical model with prior distributions  $\alpha \sim N(-6.65, 0.09^2)$  and  $\tau^2 \sim \text{Inv-Gamma}(18.36, 58.06)$ .



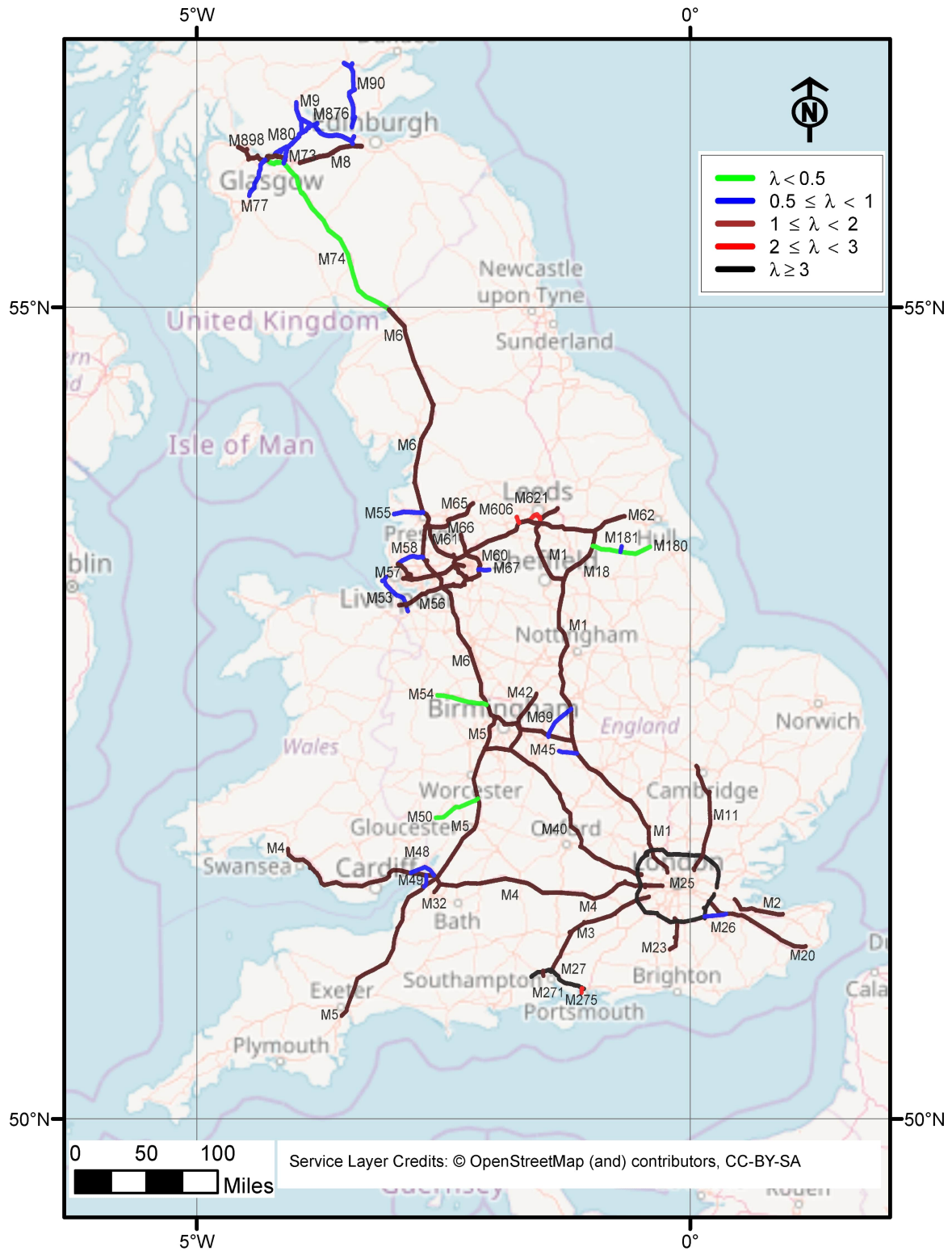


Figure 4: Estimated intensity of traffic accidents per one kilometer on the UK motorway network in 2016. The intensities are estimated using the two-level Bayesian hierarchical model with prior distributions  $\alpha \sim N(0, 100)$  and  $\tau^2 \sim \text{Inv-Gamma}(0.001, 0.001)$ .

## 5. Simulation Study

We conducted a simulation study to assess the performance of the proposed models.

### 5.1. Simulation Design

For the two-level hierarchical model, we considered six scenarios with different true values of parameters  $\alpha$  and  $\tau$ . The true values of  $\alpha$  are set to be -1 and -7. If the overall log-intensity  $\alpha$  is chosen to be lower than -9, the number of accidents on the motorway will be very close to zero. The between-motorway standard deviation  $\tau$  is set to be 0.3, 0.8 and 1.5 to reflect the variation between motorways. A magnitude of 0.3 would indicate that there is not much variation in the motorway specific log-intensity while a magnitude of 1.5 would result in much more variation between motorways. These true values are chosen to be close to the results for the observed data set. The log-intensity  $\alpha_i$  on motorway  $i$  ( $i = 1, \dots, m$ ) is drawn from a normal distribution with mean  $\alpha$  and standard deviation  $\tau$ . For the two-level hierarchical model, the number of accidents  $n_i$  ( $i = 1, \dots, m$ ) on the motorway  $i$  is generated from a Poisson distribution with mean  $L_i \exp(\alpha_i)$ , where  $L_i$  is the length of the motorway  $i$ . We simulated 1,000 data sets for each scenario.

For the three-level hierarchical model, we simulated data according to the following,

$$\begin{aligned}\alpha_i &\sim N(\alpha, \tau^2), \\ \alpha_{ij} &\sim N(\alpha_i, \tau_i^2), \\ n_{ij} &\sim \text{Pois}(L_{ij} \exp(\alpha_{ij})), i = 1, \dots, m; j = 1, \dots, s_i.\end{aligned}\tag{44}$$

Six different scenarios of simulation are considered with  $\alpha = -5, -7$  and  $\tau = 0.3, 0.7, 1.5$  for the three-level models.

The performance of the proposed models is evaluated by comparing the simulated results with the true values using the following metrics, bias, mean square error (MSE) and coverage probability (CP) [40]. The bias in the parameter estimate represents the difference between the average of the estimates over all simulation and the true value. The mean square error is the squared bias plus the variance of the estimated parameter, for example  $\text{MSE}(\hat{\alpha}) = \text{Bias}(\hat{\alpha}, \alpha)^2 + \text{Var}(\hat{\alpha})$ . The coverage probability is the percentage of 95% credible intervals across the 1000 simulated data sets that contain the true value.

## 5.2. Simulation Results

Table 4 shows that the performance of the two-level Bayesian hierarchical model is better than the two-level frequentist hierarchical model in terms of the bias, MSE and coverage probability. In Table 4, the bias and MSE in  $\alpha$  and  $\tau$  for the two-level frequentist hierarchical model is larger than those for the two-level Bayesian hierarchical model for scenarios with true value  $\alpha = -7$ . The bias in  $\tau$  obtained for the two-level Bayesian hierarchical model is insensitive to the choice of priors

Table 5 shows that the three-level Bayesian hierarchical model performs better than the three-level frequentist hierarchical model in terms of the bias, MSE and coverage probability. The three-level frequentist approach produced higher bias for point estimates of  $\alpha$  and  $\tau$  across the six scenarios and a bigger MSE. For scenarios with true value of  $\alpha = -5$ , the bias of the estimated  $\tau$  is sensitive to the specification of the prior distribution. For the true value of  $\alpha = -7$ , the three-level frequentist hierarchical model produced poor coverage probabilities for both the parameters, providing a value of 0 for  $\alpha$ . For the true value of  $\alpha = -5$ , the coverage probabilities are better than those for  $\alpha = -7$ , but the frequentist approach gave lower coverage probabilities for both the parameters. Henderson et al. (2000) showed that the separate analysis using the two-stage method is not performing well compared with the one-stage method [23]. Browne et al. (2006) showed that marginal quasi-likelihood method produced the coverage probability as zero for the random effect variance parameter in random-effects logistic regression model [14]. Our simulation results show that Bayesian hierarchical models appeared better than the frequentist hierarchical models in terms of the bias, MSE and coverage probability.

	True $\tau$	Parameters	$\alpha = -1$					$\alpha = -7$				
			Mean	Bias	MSE	CP	Time	Mean	Bias	MSE	CP	Time
Inv-Gamma(0.001, 0.001)	0.3	$\alpha$	-0.9982	0.0018	0.0018	94.7%	11747	-6.9967	0.0033	0.0028	94.8%	11667
		$\tau$	0.3008	0.0008	0.0009	95.4%		0.2963	-0.0037	0.0021	93.8%	
	0.8	$\alpha$	-0.9929	0.0071	0.0134	95.3%	11679	-6.9806	0.0194	0.0142	94.7%	11664
		$\tau$	0.8041	0.0041	0.0066	95.7%		0.7896	-0.0104	0.0083	95.2%	
	1.5	$\alpha$	-0.9923	0.0077	0.0426	96.4%	18516	-6.9442	0.0558	0.0478	95.0%	11609
		$\tau$	1.5019	0.0019	0.0238	95.5%		1.4470	-0.0530	0.0270	94.7%	
Inv-Gamma(0.1, 0.1)	0.3	$\alpha$	-0.9982	0.0018	0.0018	95.0%	11660	-6.9980	0.0020	0.0028	95.6%	11896
		$\tau$	0.3072	0.0072	0.0009	95.5%		0.3097	0.0097	0.0018	95.5%	
	0.8	$\alpha$	-0.9927	0.0073	0.0134	95.2%	11593	-6.9807	0.0193	0.0141	95.0%	11691
		$\tau$	0.8051	0.0051	0.0066	96.0%		0.7910	-0.0090	0.0082	95.6%	
	1.5	$\alpha$	-0.9920	0.0080	0.0426	96.1%	11903	-6.9441	0.0559	0.0478	94.9%	11683
		$\tau$	1.5005	0.0005	0.0235	95.4%		1.4454	-0.0546	0.0272	94.7%	
unif(0, 100)	0.3	$\alpha$	-0.9983	0.0017	0.0018	94.8%	12492	-6.9975	0.0025	0.0028	95.8%	12486
		$\tau$	0.3050	0.0050	0.0010	95.0%		0.3044	0.0044	0.0021	94.5%	
	0.8	$\alpha$	-0.9927	0.0073	0.0134	95.5%	12707	-6.9816	0.0184	0.0141	95.5%	12432
		$\tau$	0.8131	0.0131	0.0069	95.6%		0.8007	0.0007	0.0083	95.4%	
	1.5	$\alpha$	-0.9923	0.0077	0.0426	96.6%	12580	-6.9452	0.0548	0.0478	94.9%	12465
		$\tau$	1.5181	0.0181	0.0246	95.2%		1.4646	-0.0354	0.0261	95.3%	
HN(0.14)	0.3	$\alpha$	-0.9983	0.0017	0.0018	94.9%	17933	-6.9974	0.0026	0.0028	95.5%	18073
		$\tau$	0.3050	0.0050	0.0010	94.9%		0.3046	0.0046	0.0021	94.7%	
	0.8	$\alpha$	-0.9929	0.0071	0.0134	95.3%	17724	-6.9816	0.0184	0.0141	95.2%	17807
		$\tau$	0.8129	0.0129	0.0069	95.4%		0.8006	0.0006	0.0083	95.4%	
	1.5	$\alpha$	-0.9924	0.0076	0.0426	96.7%	17767	-6.9454	0.0546	0.0477	94.5%	17828
		$\tau$	1.5179	0.0179	0.0245	95.5%		1.4645	-0.0355	0.0261	95.2%	
Frequentist method	0.3	$\alpha$	-0.9982	0.0018	0.0018	94.1%	42	-6.9732	0.0268	0.0034	90.4%	42
		$\tau$	0.2956	-0.0044	0.0009	95.6%		0.2850	-0.0150	0.0021	94.7%	
	0.8	$\alpha$	-0.9928	0.0072	0.0134	94.2%	40	-6.9405	0.0595	0.0166	90.5%	39
		$\tau$	0.7902	-0.0098	0.0065	96.0%		0.7535	-0.0465	0.0096	94.9%	
	1.5	$\alpha$	-0.9925	0.0075	0.0426	95.9%	35	-6.8843	0.1157	0.0551	89.5%	37
		$\tau$	1.4915	-0.0238	0.0234	95.4%		1.3723	-0.1277	0.0386	93.8%	

Table 4: Simulation results of the two-level Bayesian and frequentist hierarchical models (model 1 and 2) under four prior distributions of  $\tau^2$  and the prior distribution  $\alpha \sim N(0, 10^2)$ . Time for running the simulation recorded in seconds. Note: MSE represents mean square error and CP the coverage probability.

	True $\tau$	Parameters	$\alpha = -5$					$\alpha = -7$				
			Mean	Bias	MSE	CP	Time	Mean	Bias	MSE	CP	Time
Inv-Gamma(0.001, 0.001)	0.3	$\alpha$	-5.0032	-0.0032	0.0026	94.7%	559,000	-6.9940	0.0060	0.0037	93.8%	258,760
		$\tau$	0.3001	0.0001	0.0019	95.0%		0.2937	-0.0063	0.0034	93.5%	
	0.7	$\alpha$	-5.0035	-0.0035	0.0116	95.6%	549,926	-7.0003	-0.0003	0.0132	95.0%	436,154
		$\tau$	0.7004	0.0004	0.0066	96.2%		0.6941	-0.0059	0.0085	94.5%	
	1.5	$\alpha$	-5.0052	-0.0052	0.0500	94.8%	542,032	-7.0198	-0.0198	0.0592	93.1%	263,827
		$\tau$	1.4937	-0.0063	0.0262	95.4%		1.5048	0.0048	0.0336	94.9%	
Inv-Gamma(0.1, 0.1)	0.3	$\alpha$	-5.0037	-0.0037	0.0026	95.2%	584,324	-7.0014	-0.0014	0.0037	95.9%	545,479
		$\tau$	0.3131	0.0131	0.0018	95.8%		0.3164	0.0164	0.0025	96.3%	
	0.7	$\alpha$	-5.0035	-0.0035	0.0116	95.6%	577,621	-6.9944	0.0056	0.0136	94.5%	507,931
		$\tau$	0.7025	0.0025	0.0065	96.1%		0.6946	-0.0054	0.0085	94.2%	
	1.5	$\alpha$	-5.0052	-0.0052	0.0501	94.5%	571,022	-7.0187	-0.0187	0.0536	95.2%	569,115
		$\tau$	1.4920	-0.0080	0.0260	95.2%		1.5032	0.0032	0.0340	94.2%	
unif(0, 100)	0.3	$\alpha$	-5.0035	-0.0035	0.0026	95.0%	545,730	-7.0007	-0.0007	0.0037	95.2%	736,853
		$\tau$	0.3082	0.0082	0.0020	94.9%		0.3063	0.0063	0.0033	94.4%	
	0.7	$\alpha$	-5.0041	-0.0041	0.0116	95.6%	533,863	-6.9957	0.0043	0.0137	94.8%	555,450
		$\tau$	0.7111	0.0111	0.0068	96.4%		0.7048	0.0048	0.0089	94.3%	
	1.5	$\alpha$	-5.0062	-0.0062	0.0504	95.5%	580,176	-7.0217	-0.0217	0.0535	95.5%	537,030
		$\tau$	1.5133	0.0133	0.0271	95.8%		1.5287	0.0287	0.0361	93.9%	
HN(0.14)	0.3	$\alpha$	-5.0035	-0.0035	0.0026	95.2%	574,984	-7.0004	-0.0004	0.0037	95.2%	548,584
		$\tau$	0.3082	0.0082	0.0020	95.0%		0.3064	0.0064	0.0033	94.1%	
	0.7	$\alpha$	-5.0041	-0.0041	0.0116	95.6%	584,604	-6.9957	0.0043	0.0136	94.7%	497,680
		$\tau$	0.7111	0.0111	0.0068	96.3%		0.7050	0.0050	0.0089	94.5%	
	1.5	$\alpha$	-5.0061	-0.0061	0.0504	95.5%	546,481	-7.0216	-0.0216	0.0535	95.5%	583,913
		$\tau$	1.5132	0.0132	0.0270	95.7%		1.5286	0.0286	0.0362	93.9%	
Frequentist method	0.3	$\alpha$	-4.9305	0.0695	0.0073	72.4%	537	-6.5961	0.4039	0.1650	0%	479
		$\tau$	0.3082	0.0082	0.0027	81.7%		0.2650	-0.0350	0.0037	89.8%	
	0.7	$\alpha$	-4.9166	0.0834	0.0169	86.0%	528	-6.5297	0.4703	0.2271	0%	547
		$\tau$	0.6602	-0.0398	0.0072	94.1%		0.4942	-0.2058	0.0472	27.1%	
	1.5	$\alpha$	-4.8524	0.1476	0.0593	88.0%	514	-6.3587	0.6413	0.4315	0%	549
		$\tau$	1.3102	-0.1898	0.0532	84.7%		0.9650	-0.5350	0.3012	4.2%	

Table 5: Simulation results of the three-level Bayesian and frequentist hierarchical models (model 3 and 4) under four prior distributions of  $\tau^2$  and the prior coverage  $\alpha \sim N(0, 10^2)$ . Time for running the simulation is recorded in seconds. Note: MSE represents mean square error and CP the coverage probability.

## 6. Model Comparisons

### 6.1. Model Comparisons using Information Criteria

We compare the Bayesian models using the deviance information criterion (DIC) [45] and the Watanabe-Akaike or widely applicable information criterion (WAIC) [47]. Table 6 shows that DIC and WAIC for the three-level Bayesian hierarchical model across different priors are lower than those for two-level Bayesian hierarchical model. This indicates that the three-level Bayesian hierarchical model provides a better fit to the observed data compared with the two-level Bayesian hierarchical model.

Model	Criterion	Prior distribution				
		Gamma(0.001, 0.001)	Gamma(0.1, 0.1)	HN(0.14)	unif(0, 100)	Gamma(18.36, 58.06)
2LBHM	DIC	100408.3	71999.9	71825.6	71950.6	71971.7
	WAIC	89412.6	68222.7	68135.6	68210.9	68223.2
3LBHM	DIC	85204.3	71683.3	71671.0	71671.0	71700.3
	WAIC	1049.9	888.3	895.7	20271.2	891.3

Table 6: *DIC* and *WAIC*. 2LBHM represents the two-level Bayesian hierarchical model and 3LBHM represents the three-level Bayesian hierarchical model.

### 6.2. Model Comparisons using Simulation Study

#### Simulation Design

The term “misspecification” means fitting a wrong model to the data [48]. Model misspecification affects estimation and produces biased estimates. To investigate the effects of model misspecification, we fitted a two-level Bayesian hierarchical model (1) to the same data sets that simulated in section 5.1 via a three-level Bayesian hierarchical model (16). We provide posterior mean, bias, mean square error and coverage probability to investigate whether model (1) is able to analyse data in presence of between grouped segments heterogeneity. The same prior distribution in section 3.1.3 and the same initial values in section 4 are utilized in the Bayesian analysis, and 100,000 iterations were run with a burn-in period of 10,000 and a thinning interval of 10 to obtain posterior samples for  $\alpha$  and  $\tau$ .

### **Simulation Results**

Tables 7 shows that the two-level Bayesian hierarchical model produced biased estimates, with large mean square errors and extremely poor coverage probabilities for both the model parameters. The coverage probability values were 0 or close to 0 for  $\tau$  and exactly equal or close to 100% for  $\alpha$  when the true value for the between-motorway standard deviation was  $\tau = 0.3$  and 0.7. This indicates that the fitted model, two-level hierarchical Bayesian model, is incorrect, when the underlying model has three levels.

	True $\tau$	Parameters	$\alpha = -5$					$\alpha = -7$				
			Mean	Bias	MSE	CP	Time	Mean	Bias	MSE	CP	Time
Inv-Gamma(0.001, 0.001)	0.3	$\alpha$	-4.8199	0.1801	0.0387	100%	20147	-6.8121	0.1879	0.0437	99.9%	19592
		$\tau$	1.6270	1.3270	1.7673	0%		1.6670	1.3670	1.8768	0%	
	0.7	$\alpha$	-4.8165	0.1835	0.0479	99.2%	17955	-6.8004	0.1996	0.0576	98.7%	21103
		$\tau$	1.7362	1.0362	1.0868	0%		1.7856	1.0856	1.1946	0%	
	1.5	$\alpha$	-4.8167	0.1833	0.0841	96.6%	19410	-6.822	0.178	0.0868	97.0%	19959
		$\tau$	2.188	0.688	0.5142	0.37%		2.2409	0.7409	0.5989	0.34%	
Inv-Gamma(0.1, 0.1)	0.3	$\alpha$	-4.8197	0.1803	0.0387	100%	19623	-6.8121	0.1879	0.0438	99.9%	20937
		$\tau$	1.6245	1.3245	1.7607	0%		1.6645	1.3645	1.8699	0%	
	0.7	$\alpha$	-4.8167	0.1833	0.0478	99.6%	18536	-6.8002	0.1998	0.0577	98.4%	20114
		$\tau$	1.734	1.034	1.0821	0%		1.7829	1.0829	1.1887	0%	
	1.5	$\alpha$	-4.8163	0.1837	0.084	96.2%	17885	-6.8219	0.1781	0.0865	97.1%	19965
		$\tau$	2.1836	0.6836	0.5079	03.7%		2.2368	0.7368	0.5926	0.35%	
unif(0, 100)	0.3	$\alpha$	-4.8200	0.1800	0.0388	100%	19398	-6.8129	0.1871	0.0435	100%	21003
		$\tau$	1.6452	1.3452	1.816	0%		1.6869	1.3869	1.9318	0%	
	0.7	$\alpha$	-4.8168	0.1832	0.0478	99.4%	19788	-6.8014	0.1986	0.0573	99.0%	22654
		$\tau$	1.7562	1.0562	1.1289	0%		1.8075	1.1075	1.2430	0%	
	1.5	$\alpha$	-4.8169	0.1831	0.0837	96.7%	22420	-6.8243	0.1757	0.086	97.4%	20769
		$\tau$	2.2136	0.7136	0.5511	0.31%		2.2696	0.7696	0.6437	03.0%	
HN(0.14)	0.3	$\alpha$	-4.8195	0.1805	0.0389	100%	23911	-6.8129	0.1871	0.0435	100%	24285
		$\tau$	1.6455	1.3455	1.8169	0%		1.6869	1.3869	1.9319	0%	
	0.7	$\alpha$	-4.8167	0.1833	0.0478	99.5%	24054	-6.8014	0.1986	0.0573	99.0%	24808
		$\tau$	1.7561	1.0561	1.1287	0%		1.8075	1.1075	1.2429	0%	
	1.5	$\alpha$	-4.8167	0.1833	0.0839	96.5%	24214	-6.8243	0.1757	0.0860	97.4%	24460
		$\tau$	2.2134	0.7134	0.5509	0.32%		2.2697	0.7697	0.6438	3.0%	

Table 7: Simulation results from two-level Bayesian hierarchical model under four prior distributions of  $\tau^2$  and the prior distribution  $\alpha \sim N(0, 10^2)$ . Time for running the simulation is recorded in seconds. MSE represents mean square error and CP the coverage probability.



## 7. Discussion and Conclusions

This study focused on Bayesian hierarchical models for analysing road accidents on the UK motorway network. This work helps to identify the most dangerous motorways in the UK network based on the estimated intensity of traffic accidents. These models have not been used for the UK motorway network before. We modelled the accident data at the motorway level by proposing a two-level hierarchical model to take into account the heterogeneity across motorways. We proposed a three-level hierarchical model to incorporate the heterogeneity not only across motorways but also across grouped segments. We assumed accident intensities are homogeneous within grouped segments but heterogeneous across grouped segments. Using the proposed hierarchical models, we identified motorways with highest and lowest intensities of accidents, classified motorway into different risk categories, and estimated the overall intensity of accidents.

We used both Bayesian and frequentist approaches to estimate the model parameters. In the Bayesian approach, a sensitivity analysis with different prior distributions for  $\tau^2$  has been performed to investigate the effect of the prior choice on the resulting posterior distributions of  $\alpha$  and  $\tau$ . We have used non-informative, weakly-informative and informative priors. In the frequentist approach, the maximum likelihood method has been separately used for each level of the model.

We assessed the performance of all proposed models through a simulation study as well as a real application related to the traffic accident data on the UK motorway network in 2016. In the simulation study, different scenarios were explored. We examined three performance criteria, bias, mean square error (MSE) and coverage probability (CP) of parameter estimates. The two parameters  $\alpha$  and  $\tau$  represent the overall accident intensity and between-motorway heterogeneity, respectively. We model all motorways simultaneously using hierarchical models. The performance of different levels of hierarchical models is evaluated through the estimation of these two parameters. The simulation results showed that the performance of the two-level Bayesian hierarchical model is better than the two-level frequentist hierarchical model in terms of the bias and the coverage probability for some simulation scenarios. The performance of both models is similar in terms of mean square errors.

In the real application, the findings from two-level Bayesian hierarchical analysis suggest that three motorways with the highest intensity of traffic accidents are M25, M27 and M606. The M25 has the highest intensity of traffic accidents where the expected number of accidents is 3.59 per one kilometer. The M27 has the second highest intensity of traffic accidents with the expected

number of accidents 3.03 per one kilometer. The M606 is the third most dangerous motorway with the expected number of accidents 2.36 per one kilometer. The three motorways with the lowest intensity of accidents are M50, M74 and M180. The lowest intensity of accidents is the M50 with an expected number of accidents of 3.32 per 10 kilometers. The second lowest intensity of traffic accidents is the M74 with an expected number of accidents of 4.18 per 10 kilometers. The M180 has the third lowest intensity of accidents with an expected number of accidents of 4.35 per 10 kilometers.

The simulation results showed that the three-level Bayesian hierarchical model performed better than the three-level frequentist model in most of the simulation scenarios. The frequentist method failed to attain the required level of actual coverage in some scenarios because of the large bias in the estimates of the overall log-intensity of accidents and the between-motorway standard deviation.

The results of the analysis of the real data using the three-level Bayesian hierarchical model showed that the motorway has the highest intensity of accidents is M25, where the expected number of accidents is 3.12 per one kilometer. The second highest intensity is found on M606 with  $\lambda = 3.09$  per one kilometer. The third highest intensity is found on M27 with an expected number of accidents of 2.69 per one kilometer. In contrast, M50, M74 and M49 have the lowest intensity of accidents with the expected numbers of accidents being 1.7, 2.7 and 3.0 per 10 kilometers, respectively. Some motorways have the similar intensity of accidents, for example, both M2 and M11 have  $\lambda = 1.39$  per one kilometer, both M53 and M55 have  $\lambda = 9.5$  per 10 kilometers, both M73 and M77 have  $\lambda = 5.8$  per 10 kilometers and both M54 and M180 have  $\lambda = 3.2$  per 10 kilometers.

Information criteria (DIC and WAIC) and a simulation study were used to compare between the two-level and three-level Bayesian hierarchical models. The values of DIC and WAIC for the three-level hierarchical model are lower than those for the two-level hierarchical model. This indicates that the three-level Bayesian hierarchical model fits the data better.

Future research can be conducted to investigate the overdispersion of the accident data by employing alternative models, such as negative binomial models [15], zero-inflated Poisson models [36] and Extra-Poisson variation models [13]. The relaxation of the assumption of homogenous accident intensity within grouped segments can be explored to allow for the variation of the accident intensity within grouped segments. The extension could also include incorporating the heterogeneity arising from unobserved factors [41, 18, 6, 12], and spatial correlations at intersections

[25, 35, 9]. In our three-level hierarchical models, we allow parameters of accident intensity to vary across grouped segments. Future research can investigate multivariate random parameters models to account for the spatial correlations [9, 11, 25, 35] and temporal correlations [38], as well as correlations between different types of crash [6]. The incorporation of spatial correlations in crash models has led to improved precision parameter estimates in some data sets [9], and this can be tested for the UK motorway accident data, by taking into account the spatial correlations between grouped segments. The multivariate random parameters models can also be extended to account for the correlations between explanatory variables to further improve the precision of the parameter estimates, due to the possibility of further capturing the underlying unobserved heterogeneity [18]. Variables such as geometric design, traffic conditions, environmental conditions and other variables affecting the accident occurrence in a motorway can be investigated. Models can be developed to include explanatory variables that may attribute to unobserved heterogeneity arising from various sources such as unobserved vehicle characteristics [10], driver characteristics [17, 35], roadway attributes, and environmental factors [12].

## References

- [1] The Department for Transport. <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>.
- [2] The Department for Transport. <http://data.dft.gov.uk.s3.amazonaws.com/road-traffic/all-traffic-data-metadata.pdf>.
- [3] The Department for Transport. <https://roadtraffic.dft.gov.uk/about>.
- [4] Alarifi, S. A., M. Abdel-Aty, and J. Lee (2018). A bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. *Accident Analysis & Prevention* 119, 263–273.
- [5] Alarifi, S. A., M. A. Abdel-Aty, J. Lee, and J. Park (2017). Crash modeling for intersections and segments along corridors: a bayesian multilevel joint model with random parameters. *Analytic methods in accident research* 16, 48–59.
- [6] Anastasopoulos, P. C. (2016). Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic methods in accident research* 11, 17–32.
- [7] Ang, Q. W., A. Baddeley, and G. Nair (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics* 39(4), 591–617.
- [8] Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial point patterns: methodology and applications with R*. CRC Press.
- [9] Barua, S., K. El-Basyouny, and M. T. Islam (2015). Effects of spatial correlation in random parameters collision count-data models. *Analytic Methods in Accident Research* 5, 28–42.
- [10] Behnood, A. and F. Mannering (2017). Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Analytic methods in accident research* 16, 35–47.
- [11] Bhat, C. R., S. Astroza, and P. S. Lavieri (2017). A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. *Analytic methods in accident research* 16, 1–22.
- [12] Bhowmik, T., S. Yasmin, and N. Eluru (2019). A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. *Analytic methods in accident research* 21, 13–31.
- [13] Breslow, N. E. (1984). Extra-poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 33(1), 38–44.
- [14] Browne, W. J., D. Draper, et al. (2006). A comparison of bayesian and likelihood-based methods for fitting multilevel models. *Bayesian analysis* 1(3), 473–514.
- [15] Chin, H. C. and M. A. Quddus (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention* 35(2), 253–259.
- [16] Fountas, G., P. C. Anastasopoulos, and F. L. Mannering (2018). Analysis of vehicle accident-injury severities: a comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. *Analytic methods in accident research* 18, 15–32.
- [17] Fountas, G., S. S. Pantangi, K. F. Hulme, and P. C. Anastasopoulos (2019). The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: a correlated grouped random parameters bivariate probit approach. *Analytic methods in accident research* 22, 100091.

- [18] Fountas, G., M. T. Sarwar, P. C. Anastasopoulos, A. Blatt, and K. Majka (2018). Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. *Accident Analysis & Prevention* 113, 330–340.
- [19] Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- [20] Gelman, A., G. O. Roberts, W. R. Gilks, et al. (1996). Efficient metropolis jumping rules. *Bayesian statistics* 5(599-608), 42.
- [21] Haque, M. M., H. C. Chin, and H. Huang (2010). Applying bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis & Prevention* 42(1), 203–212.
- [22] Hardy, R. J. and S. G. Thompson (1996). A likelihood approach to meta-analysis with random effects. *Statistics in medicine* 15(6), 619–629.
- [23] Henderson, R., P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4), 465–480.
- [24] Huang, H. and M. Abdel-Aty (2010). Multilevel data and bayesian analysis in traffic safety. *Accident Analysis & Prevention* 42(6), 1556–1565.
- [25] Huang, H., F. Chang, H. Zhou, and J. Lee (2019). Modeling unobserved heterogeneity for zonal crash frequencies: A bayesian multivariate random-parameters model with mixture components for spatially correlated data. *Analytic Methods in Accident Research* 24, 100105.
- [26] Huang, H., H. Chin, and M. Haque (2009). Empirical evaluation of alternative approaches in identifying crash hot spots: naive ranking, empirical bayes, and full bayes methods. *Transportation Research Record: Journal of the Transportation Research Board* (2103), 32–41.
- [27] Huang, H., H. C. Chin, and M. M. Haque (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: a bayesian hierarchical analysis. *Accident Analysis & Prevention* 40(1), 45–54.
- [28] Jones, A. P. and S. H. Jørgensen (2003). The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis & Prevention* 35(1), 59–69.
- [29] Kim, D.-G., Y. Lee, S. Washington, and K. Choi (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention* 39(1), 125–134.
- [30] Klaus, B., K. Strimmer, and M. K. Strimmer (2015). Package ‘fdrtool’. *CRAN*. <http://http1.debian.org/pub/CRAN/web/packages/fdrtool/fdrtool.pdf>. Accessed on October 13, 2016.
- [31] Lambert, P. C., A. J. Sutton, P. R. Burton, K. R. Abrams, and D. R. Jones (2005). How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in medicine* 24(15), 2401–2428.
- [32] Lenguerrand, E., J. L. Martin, and B. Laumon (2006). Modelling the hierarchical structure of road crash data—application to severity analysis. *Accident Analysis & Prevention* 38(1), 43–53.
- [33] Lesaffre, E. and A. B. Lawson (2012). *Bayesian biostatistics*. John Wiley & Sons.
- [34] Li, W., A. Carriquiry, M. Pawlovich, and T. Welch (2008). The choice of statistical models in road safety countermeasure effectiveness studies in iowa. *Accident Analysis & Prevention* 40(4), 1531–1542.
- [35] Li, Z., X. Chen, Y. Ci, C. Chen, and G. Zhang (2019). A hierarchical bayesian spatiotemporal random parameters approach for alcohol/drug impaired-driving crash frequency analysis. *Analytic methods in accident research* 21, 44–61.
- [36] Lukusa, M. T. and F. K. H. Phoa (2020). A horvitz-type estimation on incomplete traffic accident data analyzed

- via a zero-inflated poisson model. *Accident Analysis & Prevention* 134, 105235.
- [37] MacNab, Y. C. (2003). A bayesian hierarchical model for accident and injury surveillance. *Accident Analysis & Prevention* 35(1), 91–102.
  - [38] Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic methods in accident research* 17, 1–13.
  - [39] Mitra, S. and S. Washington (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis & Prevention* 39(3), 459–468.
  - [40] Morris, T., I. White, and C. MJ (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine* 38(11), 2074–2102.
  - [41] Pantangi, S. S., G. Fountas, M. T. Sarwar, P. C. Anastasopoulos, A. Blatt, K. Majka, J. Pierowicz, and S. B. Mohan (2019). A preliminary investigation of the effectiveness of high visibility enforcement programs using naturalistic driving study data: a grouped random parameters approach. *Analytic methods in accident research* 21, 1–12.
  - [42] Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of london crash data. *Accident Analysis & Prevention* 40(4), 1486–1497.
  - [43] Rongjie, Y. and M. Abdel-Aty (2014). Using hierarchical bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis & Prevention* 62, 161–167.
  - [44] Shankar, V., R. Albin, J. Milton, and F. Mannering (1998). Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model. *Transportation Research Record: Journal of the Transportation Research Board* (1635), 44–48.
  - [45] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), 583–639.
  - [46] Thompson, S. G., T. C. Smith, and S. J. Sharp (1997). Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in medicine* 16(23), 2741–2758.
  - [47] Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(Dec), 3571–3594.
  - [48] Yoo, W. and E. H. Slate (2005). A simulation study of a bayesian hierarchical changepoint model with covariates. Technical report, Center for Applied Mathematics and Statistics, New Jersey Institute of Technology 2005. Google Scholar.