Accident Analysis and Prevention 99 (2017) 262–271

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A novel Bayesian hierarchical model for road safety hotspot prediction

Lee Fawcett[a,*], Neil Thorpe[b], Joseph Matthews[a], Karsten Kremer[c]

[a] School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[b] School of Civil Engineering & Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[c] PTV GROUP, Traffic Software – Safety Development, Haid-und-Neu-Str. 15, 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 29 August 2016
Received in revised form 31 October 2016
Accepted 26 November 2016
Available online 14 December 2016

Keywords:
Accident prediction models
Bayesian statistics
Hotspot prediction
Predictive distribution
Regression-to-mean
Trend

ABSTRACT

In this paper, we propose a Bayesian hierarchical model for predicting accident counts in future years at sites within a pool of potential road safety hotspots. The aim is to inform road safety practitioners of the location of likely future hotspots to enable a proactive, rather than reactive, approach to road safety scheme implementation. A feature of our model is the ability to rank sites according to their potential to exceed, in some future time period, a threshold accident count which may be used as a criterion for scheme implementation. Our model specification enables the classical empirical Bayes formulation – commonly used in before-and-after studies, wherein accident counts from a single before period are used to estimate counterfactual counts in the after period – to be extended to incorporate counts from multiple time periods. This allows site-specific variations in historical accident counts (e.g. locally-observed trends) to offset estimates of safety generated by a global accident prediction model (APM), which itself is used to help account for the effects of global trend and regression-to-mean (RTM). The Bayesian posterior predictive distribution is exploited to formulate predictions and to properly quantify our uncertainty in these predictions. The main contributions of our model include (i) the ability to allow accident counts from multiple time-points to inform predictions, with counts in more recent years lending more weight to predictions than counts from time-points further in the past; (ii) where appropriate, the ability to offset global estimates of trend by variations in accident counts observed locally, at a site-specific level; and (iii) the ability to account for unknown/unobserved site-specific factors which may affect accident counts. We illustrate our model with an application to accident counts at 734 potential hotspots in the German city of Halle; we also propose some simple diagnostics to validate the predictive capability of our model. We conclude that our model accurately predicts future accident counts, with point estimates from the predictive distribution matching observed counts extremely well.

1. Introduction

1.1. Overview

The World Health Organization (World Health Organization, 2015) reports that 1.24 million fatalities and between 20 and 50 million non-fatal injuries occur worldwide every year as a result of road traffic accidents, at huge economic and social cost – typically between 1 and 3% of a nation's GDP. Reducing road traffic accidents is therefore a global challenge with potentially significant benefits to be realized in terms of national economies, society, public services (e.g. healthcare provision) and quality of life. Generally, the current approach for identifying accident hotspots in many countries is *reactive*, taking place after accidents have already occurred and applying remedial treatment at locations observing counts which overtop some pre-determined threshold. It is argued here that action should be *proactive*, to prevent this threshold being overtopped in the first place, through accurate forecasting of accident rates in future years (i.e. a process of *hotspot prediction*). This process relies on the analysis of historical accident data, but these data are prone to confounding effects – principally regression-to-mean (RTM) and trend. This can mislead and cause scarce resources to be targeted inefficiently; for example, wrongly treating sites that are inherently 'safe' simply due to a short term 'blip' (temporary and random increase) in accident rates.

* Corresponding author.
E-mail addresses: lee.fawcett@ncl.ac.uk (L. Fawcett), neil.thorpe@ncl.ac.uk (N. Thorpe), j.matthews2@ncl.ac.uk (J. Matthews), Karsten.Kremer@ptvgroup.com (K. Kremer).

http://dx.doi.org/10.1016/j.aap.2016.11.021

Phenomena such as trend and RTM can vary at a local level, relative to the effects we might observe globally across the network as a whole, and we believe any proposed strategy for predicting accident counts in future years should have the facility to take such differences into account. Important also, we believe, is the ability for more recent accident counts to inform model-based estimates of safety with greater precision than those observed at time-points further in the past. Within the context of Bayesian modeling, the specification of informative prior distributions provides a means of incorporating such features in a way we deem appropriate for the dataset being studied. We believe the inclusion of such features sets our work apart from recent publications in hotspot identification, including (for example) Cheng and Washington (2005); Wang et al. (2011); Heydari et al. (2013); and Sacchi et al. (2015).

1.2. General methodology

The research reported here aims to develop and apply appropriate statistical methodologies for accounting for the confounding effects of RTM and trend in forecasts of future accident rates. The same basic principles can also be used to evaluate the impact of treatment at sites (i.e. in a before-and-after study), again taking these confounding effects into account, to provide a more realistic 'value-for-money' estimate of scheme benefits. Indeed, the general methodology used in the current paper for handling RTM has its roots in Fawcett and Thorpe (2013), in which the sensitivity of RTM estimates to various approaches for accounting for this phenomenon, in the context of road safety scheme evaluation studies, is explored. In a classical before-and-after analysis, data from a single time period (the 'before' period) are used alongside fitted values generated by an accident prediction model (APM) to estimate counterfactual accident counts in the 'after' period. Our model extends this classical approach to allow the inclusion of data from multiple time periods, enabling variations in historical accident counts to inform predictions of future counts. For example, global trend is incorporated into our analysis by including a corresponding time indicator as a covariate (alongside other variables) in the APM. Informative prior distributions proposed for some parameters within our model then allow local, site-specific adjustments to this trend, where appropriate; they also allow the inclusion of increased uncertainty for model-based estimates of safety further in the past, lending predictions greater influence from data in more recent time periods. Further, we attempt to account for discrepancies between the APM and our observed accident counts due to factors for which we have no data (sometimes accounted for using crash modification factors; for example, see Park et al., 2014).

1.3. RTM and trend

Studies assessing the effectiveness of road safety schemes are notoriously bedeviled by the problem of RTM. In most before-and-after studies reported in the road safety literature (e.g. Mountain et al., 1992; Hirst et al., 2004; Li et al., 2008), remedial measures have been deployed after a period of 'unacceptably high' accident counts; in any subsequent period, we might expect these counts to *regress* towards their underlying mean level anyway simply because they were abnormally high during the 'before' period, random fluctuation being a significant factor contributing to these abnormalities. Given that historical counts at treated sites are not always readily available, practitioners are often faced with the task of estimating the underlying mean accident rate at each site – usually within an "empirical Bayes" or "fully Bayes" framework, in which APMs are utilized (see, for example, Miaou and Lord, 2003; Li et al., 2008; Maher and Mountain, 2009; Fawcett and Thorpe, 2013). We argue that the estimation of RTM also has an important role to play in road safety hotspot prediction – estimates of future accident counts at sites across a network need to be adjusted to account for RTM in past time periods as investment at locations based on unadjusted predictions could be unjustified. Previously observed accident counts might also need to be adjusted for RTM to help get a clearer picture of the underlying average level of safety at a location. In the current work, we make use of a global APM within our Bayesian hierarchical model to help smooth for RTM, although modeling accident counts across several time-points also helps to adjust for RTM based on what we observe locally.

Methods for safety scheme evaluation and hotspot prediction should also incorporate information on underlying trends in counts, these trends being a feature of site-specific variation or perhaps globally-acknowledged changes across the network as a whole. Changes in accident rates due to general trend could be a result of many factors such as education programmes (for example driver retraining), more widespread enforcement of laws and regulations and improved safety in highway and vehicle design. Trend might be accounted for by applying a multiplicative factor to the APM (see, for example, Fawcett and Thorpe, 2013), where this factor represents changes in counts we might expect between the 'before' and 'after' period in an evaluation study, after examining local and/or national statistics on changes across the network or region being studied. Or, if counts are available annually across several years (for example), then a time indicator can be used as a covariate in the APM, allowing the explicit modeling of changes through time (these changes could be assumed linear or non-linear); see, for example, Lord and Persaud (2000). In the current work, we investigate the possibility of simultaneously adjusting for globally-identified trend (through an APM which uses a time indicator as a covariate) *and* site-specific trend (through a multiplicative adjustment factor), enabling both sources of trend to inform our predictions at individual sites, hopefully improving the accuracy and reliability of these predictions.

1.4. Aim and structure of this paper

The main aim of this paper is to describe the development and testing of a Bayesian hierarchical modeling procedure for the prediction of road traffic accident hotspots. Section 2 describes our general template for modeling, and the practical role of the various components we include within this template. In Section 3 we describe the application of our model with specific reference to annual accident counts, and associated covariate information, collected at 734 sites in the city of Halle, Germany. In particular, we describe the construction of the APM and some informative prior distributions, our Bayesian sampling scheme and the formulation of the posterior predictive distribution, before presenting some results from our fitted model and outlining some simple validation diagnostics. Finally, Section 4 identifies key conclusions and potential avenues for future research.

2. Hotspot prediction: a Bayesian hierarchical model for accident counts

2.1. General model structure

Generally, our proposed method provides an extension to the classical empirical Bayes framework for analysis, as commonly used in before-and-after studies (e.g. Miaou and Lord, 2003; Li et al., 2008; Maher and Mountain, 2009; Fawcett and Thorpe, 2013), to allow the inclusion of data from multiple time periods in the analysis. We assume that, for each potential hotspot site j, we have data on accident counts $y_j(t)$, t being a discrete time indicator with $t=0$ representing the current time period. For example, if annual counts are available at each site j, then $y_j(t=0)$ represents counts in the

latest year for which complete data are available and $y_j(t<0)$ represents counts in previous years; $y_j(t=1)$ would represent counts for next year, in which we have an interest in making predictions, and $y_j(t>1)$ for subsequent years. As is usually the case, we assume current and future counts $y_j(t\geq 0)$ to be Poisson-distributed with rate, and hence variance, $\lambda_j(t)$. Similarly, for previous years an accident rate of $\lambda_j(t)$ is assumed; however, we allow a larger variance for $y_j(t<0)$. Specifically, the aim is to model historical counts at $t<0$ with variance $\lambda_j(t)c(t)$, $c(t)>0$, for two reasons: (i) to allow more recent counts to inform our predictions with more certainty, and (ii) when adjusting past counts for RTM and trend, it seems sensible for us to be more certain about model-based estimates at more recent historical time-points than for those further in the past. For example, it might be the case that infrastructure changes have occurred at a site level over the observation period, perhaps implying less certainty about the contribution of the observed values to the model-based estimate of accident rates for $t<0$. To allow for this extra-Poisson variability for $y_j(t<0)$ we assume a negative binomial (NegBin) distribution here, giving:

$$y_j(t)|\lambda_j(t)\sim \begin{cases} \text{Poisson}(\lambda_j(t)), & t\geq 0; \\ \text{NegBin}\left(r=\dfrac{\lambda_j(t)}{c(t)-1}, p=\dfrac{1}{c(t)}\right), & t<0; \end{cases}$$

$$c(t)=\exp\{-t\tau\},\ t<0, \tau>0,$$

where the chosen parameterization of the negative binomial distribution has expectation and variance $r(1-p)/p=\lambda_j(t)$ and $r(1-p)/p^2=\lambda_j(t)c(t)$ respectively, as desired, and the choice of the function for c ensures a variance which is increasingly inflated as t decreases; a suitable choice of prior distribution for τ can give an inflation factor which is deemed suitable for the data being analyzed. To allow for trend and RTM in the mean accident rate, we set

$$\lambda_j(t)=a_j\mu_j(t)\exp\{b_j t\}, a_j>0; -\infty<b_j<\infty; t\leq 0, \quad (1)$$

where $\mu_j(t)$ is estimated from a global APM (here, we use a relatively simple log-linear model – see Section 3.2 – although any form of APM could be used), the inclusion of which allows the incorporation of prior beliefs about the underlying mean level of accident counts in time period t at site j. If this APM allows for a change in accident rates through time, then the posterior distribution for $\lambda_j(t)$ has the potential to adjust for both RTM and trend. The role of a_j in Eq. (1) is to account for discrepancies between the APM and our observed accident counts due to factors for which we have not, or cannot, collect data (e.g. poor visibility conditions due to permanent obstacles at the site). In the absence of any site-specific knowledge relating to such factors, a largely uninformative prior could be assigned to this parameter. The inclusion of the $\exp\{b_j t\}$ term in Eq. (1) allows any observed site-specific trend to adjust the global trend modeled in the APM. A carefully selected prior formulation for b_j can give *a priori* weight to the local trend, relative to the global trend, as we see fit. Please see Section 3.2 for our specific choice of priors for τ, a_j and b_j.

2.2. Inference

Inference first proceeds by estimating $\mu_j(t)$ in Eq. (1) via an APM, which itself might be estimated from a pool of reference data; alternatively, 'off-the-shelf' APMs are sometimes used to estimate $\mu_j(t)$, such as COBA models in the UK (Department for Transport (DfT), 2006) or Ripcord-Iserest models in other parts of Europe (Ripcord-Iserest, 2005). If the APM itself is to be estimated, interest at this stage might lie in the identification of significant covariates and the nature of the dependence between accident counts and these covariates. In particular, the significance of a global trend in accident counts can be ascertained at this stage through the inclusion of t as a covariate. Standard regression methods can be employed to find maximum likelihood estimates for parameters in the APM, or a full Markov chain Monte Carlo (MCMC) procedure could be used to make inferences within the Bayesian framework, given a suitable prior specification (details on MCMC are now extensively published; see (for example) Smith and Roberts, 1993). Then, after a suitable choice of prior distributions for τ, a_j and b_j (and hyperparameters therein), standard MCMC methods can be employed to simulate approximate draws from the marginal posteriors for these parameters and hence the mean accident rates $\lambda_j(t)$. The MCMC output for $\lambda_j(t)$ can then be used to approximate the posterior predictive distribution for future accident rates at each site; that is, $f(y_j(t=T)=y \mid \mathbf{y}_j)$ for future time-point T at site j given all past accident counts at this site $\mathbf{y}_j$. Please see Section 3.3 for a more detailed description of the MCMC methods we employ in our data application, and Section 3.4 for a detailed discussion of the posterior predictive distribution.

3. Model application: accident counts in Halle, Germany

3.1. The data

The data used to demonstrate the approach consist of annual accident counts at 734 sites in the city of Halle, Germany, for the years 2004–2012 inclusive. We reserve data from 2012 for prediction validation purposes (see Section 3.6), so that we have $\{y_j(t); t=-7, \ldots, -1, 0; j=1, \ldots, 734\}$ for years 2004, ..., 2010, 2011, using the notation established in Section 2.1. All 734 sites might be considered as candidate road safety hotspots. Accompanying these figures are observations on several covariates. Specifically, for each site $j=1, \ldots, 734$, we have:

- $x_{1,j}$: Traffic volume (average number of vehicles passing through the site per day in the year).
- $x_{2,j}$: Traffic volume from the major road of the intersection ($=x_{1,j}$ if not an intersection).
- $x_{3,j}$: Traffic volume from the minor road of the intersection (=0 if not an intersection).
- $x_{4,j}$: Speed limit at the site.
- $x_{5,j}$: =1 if the site is in an urban area; 0 otherwise.
- $x_{6,j}$: =1 if the site is at an intersection; 0 otherwise.
- $x_{7,j}$: =1 if the site is at a signalised junction; 0 otherwise.
- $x_{8,j}$: =1 if the site is on a major road; 0 otherwise.
- $x_{9,j}$: =1 if the site is at a major intersection; 0 otherwise.
- $x_{10,j}$: =1 if the site is at a four-legged junction; 0 otherwise.

Fig. 1 shows time series plots for four of the sites over the eight year recording period (2004–2011 inclusive). Table 1 gives some numerical summaries for the variables. The plots in Fig. 1 have been selected to demonstrate the features we are attempting to acknowledge within our analysis; i.e. sites 309 and 706 suggest evidence of temporal trend, and sites 163 and 677 indicate potential RTM effects after 'blips' in the year 2008. Mean accident counts across all sites, as shown in Table 1, could be indicative of a global negative trend in counts across the network, particularly from 2007. In practice, care should be taken of any multicollinearity between the covariates we might use to construct an APM; here, we have no significant correlations (at the 5% level of significance) between any of our covariates.

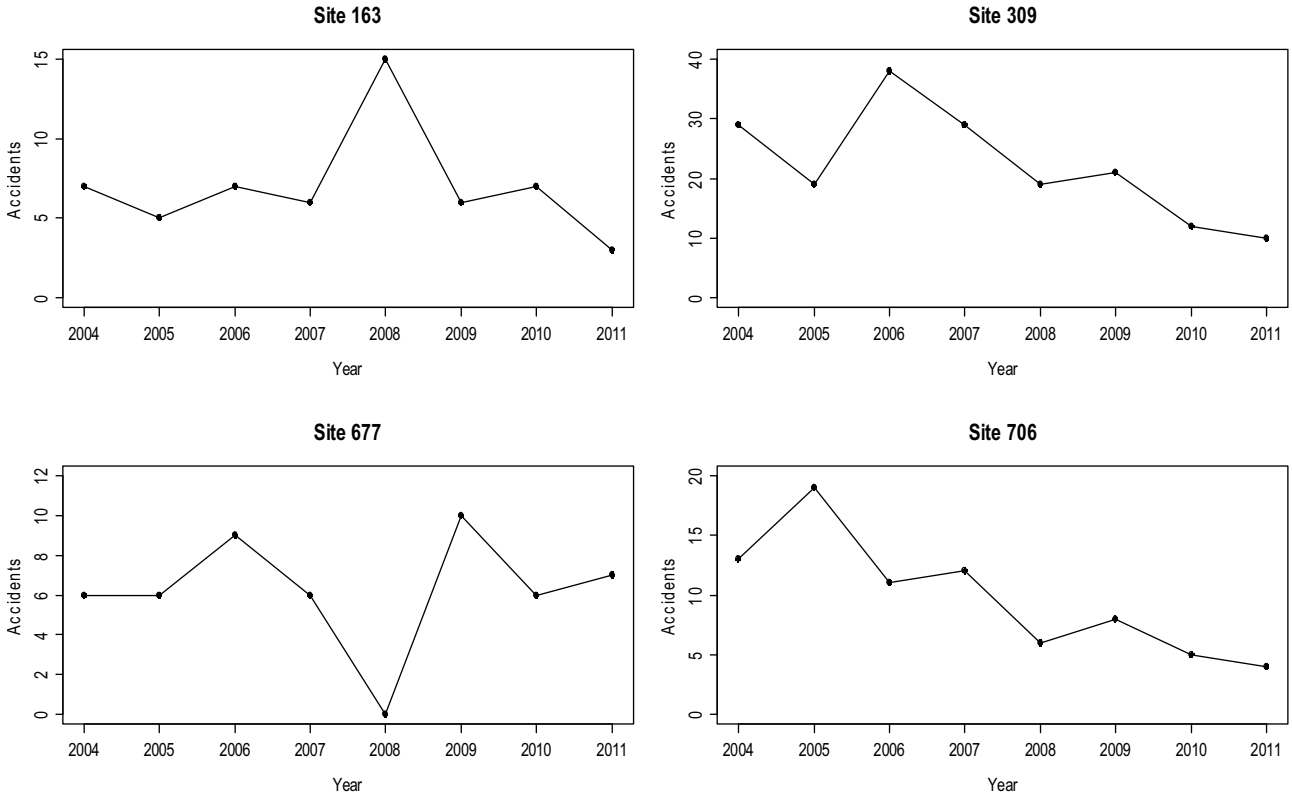

Fig. 1. Time series plots of accidents at sites 163, 309, 677 and 706 (2004–2011).

Table 1
Basic numerical summaries of accident counts (annually, across all 734 sites) and covariates (across all 734 sites). A full description of the covariates can be found in Section 3.1. 'S.D.' refers to the standard deviation, 'Max.' refers to the maximum and 'Prop.' is the proportion of sites taking certain characteristics.

	Yearly Accident Totals (y)								Flow (Av. Vehicles/day)		
	'04	'05	'06	'07	'08	'09	'10	'11	x_1	x_2	x_3
Mean	3.7	3.7	3.6	3.7	3.6	3.6	3.3	3.1	7432	5759	1673
S.D.	21.4	24.5	25.7	24.9	20.4	23.0	19.3	19.4	11045	8373	4324
Max.	29	35	38	52	29	39	41	48	65484	54919	40141

	Categorical Variables											
	Speed Limit (km/h) (x_4)						x_5	x_6	x_7	x_8	x_9	x_{10}
	30	45	50	60	70	80						
Prop.	0.37	0.12	0.22	0.18	0.04	0.07	0.91	0.86	0.27	0.63	0.20	0.24

3.2. The APM and prior specification

We impose the following log-linear APM to estimate $\mu_j(t)$ in Eq. (1), assuming a negative binomial error structure with over-dispersion γ :

$$\mu_j(t) = \exp\left\{\beta_0 + \beta_t t + \sum_{p=1}^{n_p} \beta_p x_{p,j}\right\}, \quad (2)$$

where $x_{p,j}$ represents covariate information at site j and n_p (=10 in our case) is the number of such covariates based purely on the information available. Although this form of APM is rather simplistic, we use Eq. (2) here to simply demonstrate our overall modeling procedure. A more sophisticated APM could of course be used, according to the user's preference, perhaps incorporating relevant geometric design features and/or link length for the non-intersection sites; a mixture of different APMs could also be used for different site types. Alternatively, techniques based on propensity score matching (PSM; see, for example, Li et al., 2013) could, in principle, be employed to construct site-resembling APMs for subsets of sites.

Analyses such as those in Fawcett and Thorpe (2013), aimed at evaluating road safety countermeasures via before-and-after studies, often use sites from a separate reference pool to estimate (β_t, β_0, β_p) in Eq. (2), before then applying the fitted model to covariate information (t_j, $x_{p,j}$) at site j. However, given a large enough pool of sites of interest, in the context of hotspot prediction we argue that fitting Eq. (2) using data at these sites themselves, and using the fitted values from this model as estimates of $\mu_j(t)$, is a suitable approach. Moreover, in before-and-after studies there is usually reason to suspect that accident counts at the treated sites are from the tail of the distribution of counts across the network generally, and so a separate reference pool is essential for building an APM which will yield realistic fitted values; in the current work,

we assume that none of the 734 sites in Halle have been treated – all sites are *candidate* hotspots.

We obtain estimates of the regression coefficients in Eq. (2), as well as the negative binomial over-dispersion parameter γ, via maximum likelihood, although in the spirit of a truly fully Bayesian analysis prior distributions *could* be specified for these parameters and posterior inferences made. For example, essentially objective priors, with large variances, could be assigned to the regression coefficients (e.g. independent $N(0, 100)$ priors); or with careful thought it might be possible to specify more informative priors here, as in Yu and Abdel-Aty (2013). However, there is a trade-off between a full prior specification, perhaps giving a more realistic assessment of uncertainty in parameter estimation in our model generally, and the computational burden required to incorporate this additional layer of complexity; in our case, with 734 sites and eight years of data, short pilot MCMC runs on the APM suggested this would not be worthwhile.

With reference to the covariates described in Table 1, the most significant in terms of their impact on accident rates were x_3, x_5, x_6, x_7, x_9 and x_{10}, all of these having coefficients significantly greater than 0 at the 0.1% level of significance; the time indicator t was also highly significant, its coefficient indicating a negative (linear) trend in accident counts across the network generally.

After an extensive prior elicitation session, including simulations and a feedback and refinement process, the following prior for τ was settled upon:

$$\tau \sim \text{Gamma}(2, 20),$$

which we believe returns a suitable time-dependent inflation of the negative binomial variance for $y_j(t)$ as t decreases. In the absence of any knowledge about site-specific crash modification factors, we use:

$$a_j \sim \text{Gamma}(\gamma, \gamma).$$

This ensures that the mean of our prior distribution for the accident counts is that suggested by the APM, since $E[a_j] = 1$; also, the variance of the prior distribution for these counts at time $t = 0$ is $\mu_j^2(t)/\gamma$, in keeping with the analyses in Fawcett and Thorpe (2013). Specifically, this particular prior specification ensures that when we have historical data from just one time-point $t = 0$, the same model structure as that used in a typical before-and-after empirical Bayes analysis is recovered, giving us a generalization of such an analysis under our approach with data from multiple time periods. The term b_j is designed to adjust any trend identified by the global APM based on site-specific observations of trend. Here, this is modeled such that:

$$b_j = b_N b_Z,$$

where:

$$b_N \sim N(0, 0.1) \text{ and } b_Z \sim \text{Bernoulli}(0.5). \quad (3)$$

The prior distribution for the zero-inflation component b_Z in (3) in effect penalizes the global trend detected by the APM in Eq. (2) in accordance with our beliefs about site-specific accident trends observed in the city of Halle. Of course, given data from just a single time-point $t = 0$, we would be unable to ascertain any local trend, and so we would set $b_j = 0$ to ensure a strictly zero local trend contribution to the mean accident rate in Eq. (1).

3.3. Summary and Bayesian sampling

To summarize, there are two main stages in our data analysis:

1. The first stage uses an APM (Eq. (2) in our analysis), applied to annual accident counts and associated covariates across all sites in the Halle network, to obtain model-based estimates of accident counts at each site via the fitted values from the APM. Global trend is captured by the APM through the inclusion of time as a covariate. The main result here is the estimation of $\mu_j(t)$ for each site j and each time period $t = \{-7, \ldots, 0\}$, as well as the over-dispersion parameter γ.
2. The second stage assumes a Poisson/negative binomial model for accident counts at each site for time periods $t \geq 0/t < 0$, respectively, where the Poisson/negative binomial mean is *informed* by the APM estimates from the first stage in the analysis, but can be *adjusted* by significant local deviations from the globally-observed trend (through the parameter b_j) and discrepancies between the APM estimates and our observed values due to factors for which we have not collected data (through the parameter a_j, which is constant across each site). We also impose a variance inflation device for observations at increasingly distant historical time periods to capture our uncertainty about model-based estimates of safety as we move further back in time (controlled by our parameter τ). The main aim of this stage of the analysis is to make inferences on (a_j, b_j, τ) via MCMC, leading to posterior inference on the mean accident rates $\lambda_j(t)$ themselves.

We use the function glm.nb within the R package MASS (Ripley, 2016) to estimate $\mu_j(t)$ (for each site $j = 1, 2, \ldots, 734$ and each time period $t = \{-7, \ldots, 0\}$) and the over-dispersion parameter γ in stage 1 of the analysis. For stage 2 we use R-JAGS (Plummer, 2016) to make inferences on (a_j, b_j, τ), and the mean accident rates $\lambda_j(t)$, via MCMC. Specifically, at each iteration i a Metropolis-within-Gibbs MCMC sampler is used to update the parameter vector:

$$\boldsymbol{\theta}_j^{(i)} = \left\{a_j, b_j, \tau_j, \lambda_j(t)\right\}^{(i)},\ j = 1, \ldots, 734;\ t = -7, \ldots, 0,$$

with random walk proposals used for the separate updating of each element within $\boldsymbol{\theta}_j$. Details on MCMC sampling schemes for hierarchical Bayesian models are now widely reported: thus, we do not give full details on the implementation of the sampler for drawing inferences from our model here and the reader is referred to Smith and Roberts (1993) for more information. We run the sampler for 10^5 iterations after discarding the burn-in period. Serial correlation in the samples is removed by thinning to every 10th observation, and the resulting trace plots suggest good mixing and apparent good convergence. Convergence was checked by multiple runs of the chain from various starting-points for each element in $\boldsymbol{\theta}_j$, demonstrating multiple convergence to the same limit for all parameters. The resulting chains for each element in $\boldsymbol{\theta}_j$ are assumed to be samples from the marginal posterior distributions for each of these parameters.

3.4. Prediction

The model structure outlined thus far attempts to adjust current and previous observations on accident counts at each site for RTM and trend. However, the main focus of this research is on prediction: based on predicted counts for future years, the aim is for a proactive approach to road safety scheme implementation. Working within the Bayesian framework provides a suitable vehicle for this via the posterior predictive distribution. Recall that time index t, $t = \{1, 2, \ldots\}$, corresponds to 'future' years $\{2012, 2013, \ldots\}$. Suppose we assume a Poisson distribution with mean $\lambda_j(t)$ for accident counts $y_j(t)$ at site j for $t = \{1, 2, \ldots\}$, just as we do for the current time

Table 2
MCMC and predictive summary for the four sites shown in Fig. 1 for past years 2008 and 2011 and 'future' year 2012. Shown are the posterior/predictive means, with 95% credible intervals in parentheses; values shown for μ_j are the fitted values from the APM in Eq. (2).

	a_j	b_j	2008			2011			2012
site			Observed	$\mu_j(t=-4)$ (APM)	$\lambda_j(t=-4)$	Observed	$\mu_j(t=0)$ (APM)	$\lambda_j(t=0)$	Prediction ($t=1$)
163	2.62 (1.43, 3.64)	−0.01 (−0.13, 0.02)	15	2.37	6.44 (4.56, 8.57)	3	2.17	5.72 (3.46, 7.87)	5.50 (1, 11)
309	1.61 (0.94, 2.61)	−0.10 (−0.20, 0.00)	19	8.23	19.40 (15.49, 23.60)	10	7.54	13.30 (8.55, 19.69)	10.93 (3, 22)
677	2.96 (1.80, 4.16)	−0.01 (−0.10, 0.05)	0	1.90	5.71 (3.95, 7.73)	7	1.75	5.18 (3.32, 7.17)	5.07 (1, 10)
706	0.63 (0.29, 1.20)	−0.12 (−0.25, 0.00)	6	8.61	8.39 (6.05, 11.03)	4	7.90	5.50 (2.89, 9.44)	4.42 (0, 11)

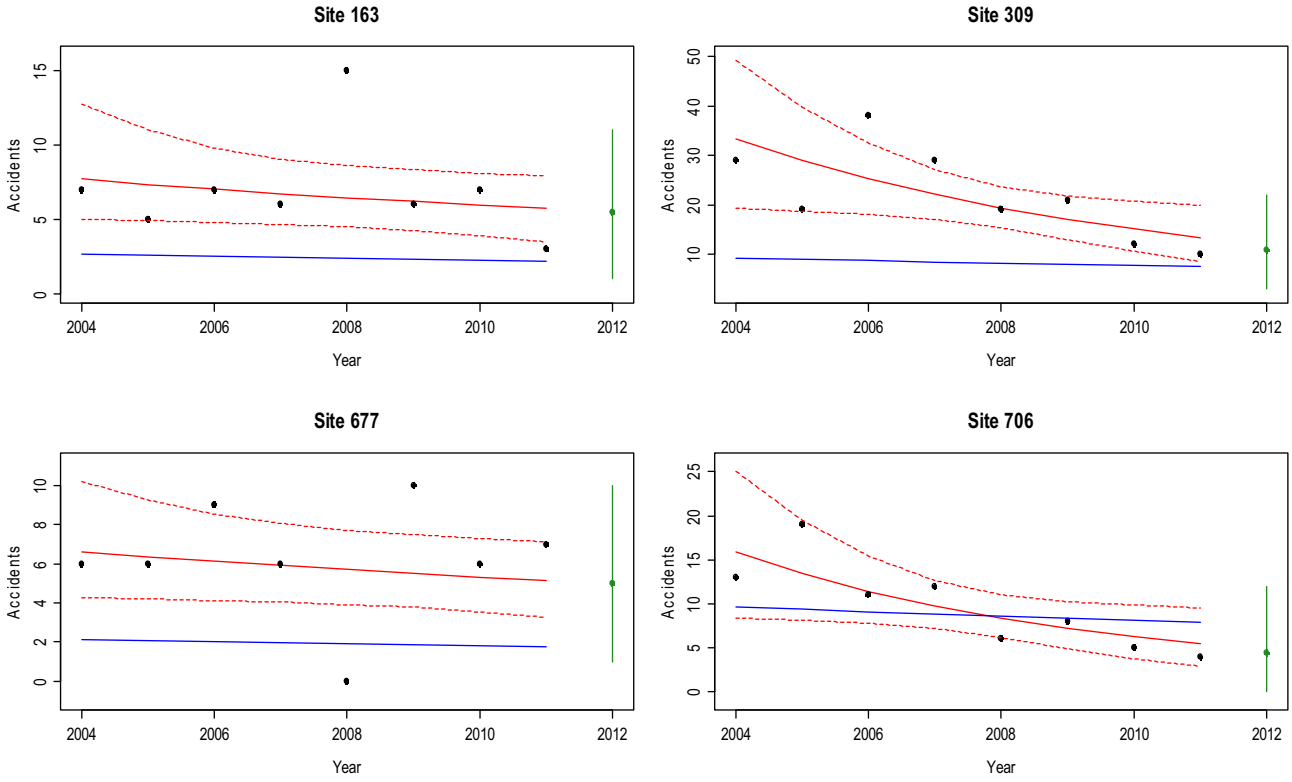

Fig. 2. Observed accidents (black); posterior means for APM estimates (blue); posterior means for model estimate (red solid), with 95% credible intervals (red dotted); predicted future value (predictive mean) with 95% prediction intervals (green; dot for the mean, with the line representing the prediction interval). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

period; then the predictive probability function for the future count $y_j(t=1)$ is given by

$$f\left(y|\mathbf{y}_j\right) = \int_{\Lambda} f\left(y|\lambda_j(t=1)\right)\pi(\boldsymbol{\theta}_j|\mathbf{y}_j)d\lambda_j \tag{4}$$

where $\pi(\theta_j|\mathbf{y}_j)$ represents our posterior distribution for all parameters in our model, given all of the data to date ($\mathbf{y}_j$), and Λ is the parameter space for λ_j. Although the right-hand-side of (4) is analytically intractable, it can be approximated using our MCMC chains for $\boldsymbol{\theta}_j$. Specifically, for each site j an estimate of $\mu_j(t=1)$ is obtained via Eq. (2) with $t=1$; then, at each iteration i in the MCMC, the posterior draws for a_j and b_j are used to obtain a draw for $\lambda_j(t=1)$. A draw from the posterior predictive distribution for $y_j(t=1)=y$ is thus given by

$$\frac{\left(\lambda_j^{(i)}(t=1)\right)^y \exp\{-\lambda_j^{(i)}(t=1)\}}{y!},\ y=0,\ 1,\ldots$$

Repeating at all iterations in the MCMC gives a complete (approximate) sample from the predictive distribution for $y_j(t=1)=y$, the mean of which can be used as a point estimate. The full predictive distribution for $y_j(t=1)$ can then be explored for $y \in \mathbb{N}^0$. Sites in our study can then be ranked, and perhaps scored, in terms of their posterior predictive probability of exceeding some pre-determined accident frequency threshold which practitioners might use to help inform safety scheme implementation decisions.

3.5. Results

The model can be estimated, and predictions made, for each of the 734 sites in the Halle dataset. Here, we focus on the four individual sites already introduced in Section 3.1 to demonstrate how our approach manages the confounding issues of RTM and trend.

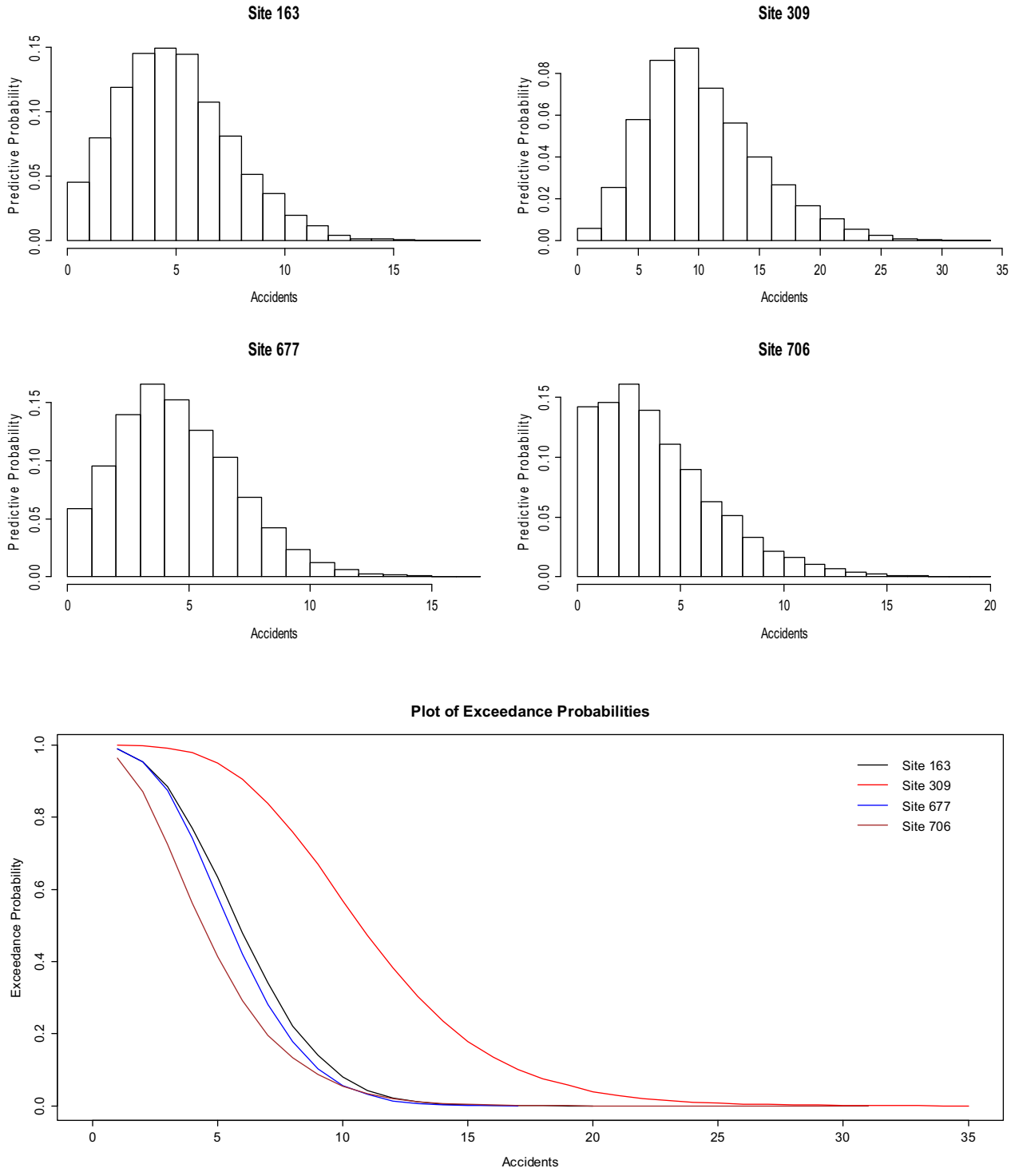

Fig. 3. Full posterior predictive densities for the number of accidents at our four sites of interest in the prediction year 2012 (top) and associated predictive accident exceedance probabilities (bottom).(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2 shows that our zero-inflated trend parameter leads to site-specific trend being identified over and above that observed globally at sites 309 and 706, with wholly negative 95% credible intervals for b_j at these sites. This is evident also in the plots for these sites in Fig. 2, with the model-based estimates of accident rates decreasing at a faster rate than the global APM. Conversely, at the other two sites the model-based estimates of accident rates are almost parallel to those suggested by the APM, and the corresponding posterior means for b_j are very close to zero here. Both Table 2 and Fig. 2 highlight the ability of our model to adjust for RTM. For example, in 2008, the unusually high accident count at site 163 has been down-weighted considerably in our model by the APM, to bring it more in-line with what we might expect to see at this site in this year, with similar observations being made

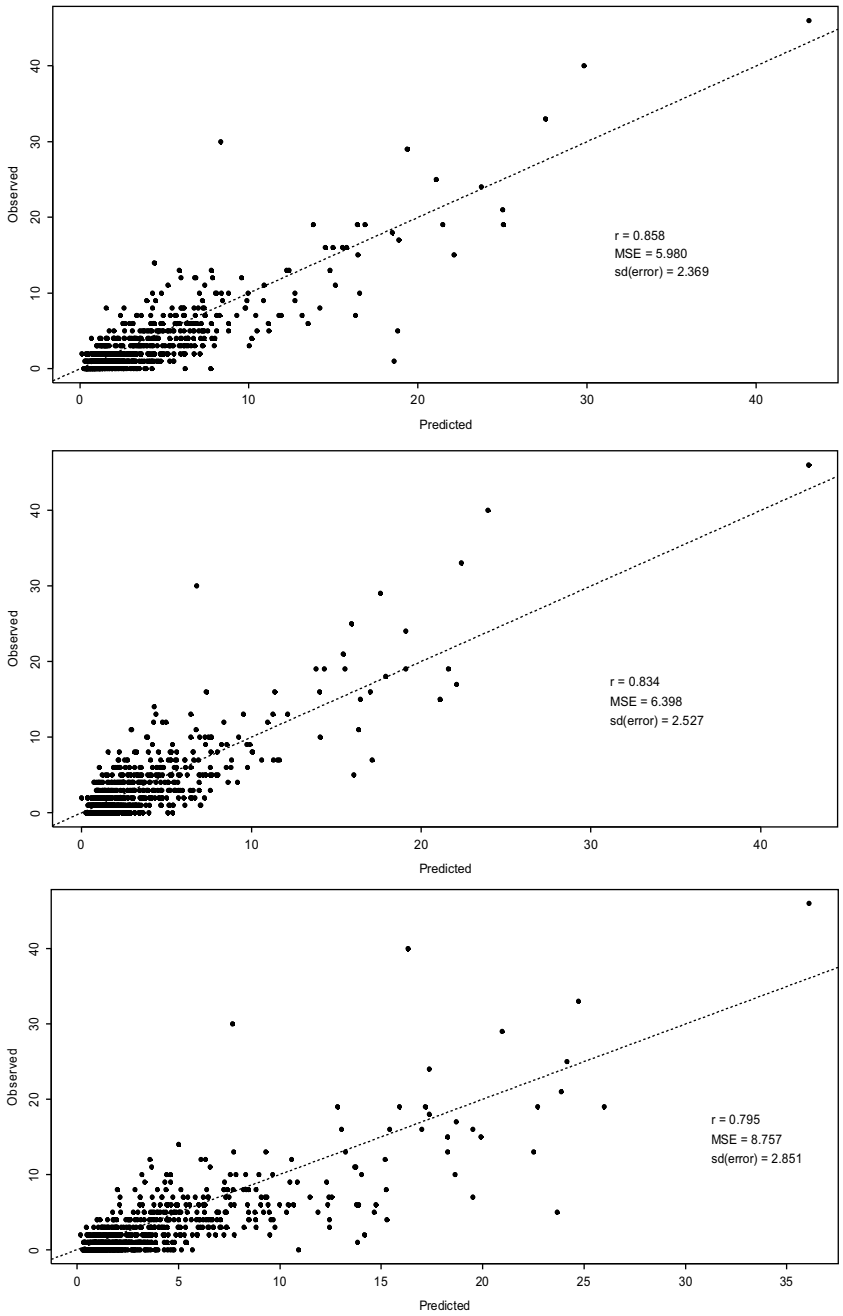

Fig. 4. Posterior predictive distribution means for the final year of data against the corresponding observed accident counts: using data from 2004 to 2011 (inclusive), top; 2007–2011 (inclusive), middle; 2011 only, bottom. Diagonal dotted lines are the lines of equality. In each plot, r is the correlation coefficient between the observed and predicted counts; MSE is the mean squared error; and sd(error) is the standard deviation of the errors.

in 2008 (and perhaps 2009) for site 677. Also shown in Fig. 2 are estimates of accident counts in the prediction year 2012, with 95% prediction intervals. Although these intervals appear rather wide, we would expect these to shrink as more historical data comes on-line year-on-year. Fig. 3 shows the full posterior predictive distributions for the year 2012 at each of our four sites of interest, and a plot of the predictive probabilities of exceeding a range of accident frequencies in this year.

Plots such as these illustrate the potential application of our hotspot predictive model. For example, practitioners could determine the probability of exceeding a given number of accidents the following year and hence use this to aid the decision-making process for implementing preventative measures. To illustrate, if an annual count of ten accidents at a particular location is deemed high enough to warrant action, then according to our analysis – for the four sites we show results for in this paper – we predict that site 309 is by far the most worthy of treatment, followed by sites 163, 706 and 677; see Fig. 3, in which the plot of estimated exceedance probabilities indicates that it is very unlikely that this threshold will be over-topped next year at sites 163, 706 and 677. Specifically, the posterior predictive probabilities of exceeding ten accidents in 2012 for sites 309, 163, 706 and 677 are 0.435, 0.039, 0.032 and 0.022 (respectively).

3.6. Validation of predictions

Recall that our accident counts across the Halle network in 2012 were reserved for prediction validation purposes. In an attempt to validate our model's predictive capability, we obtain full predictive distributions for all 734 sites in the year 2012, as shown for our four illustration sites in Fig. 3. Plots of predicted accident counts (which we take to be the mean of the posterior predictive distributions) against their observed counterparts are given in Fig. 4, along with summary statistics; we show results when all eight years of data have been used (2004–2011 inclusive), but also for when five years of data have been used (2007–2011 inclusive) and when just a single year of data has been used (2011 only, essentially giving results analogous to an empirical Bayes analysis). All plots show a strong association between our predictions and the observations, with high correlations and points scattered around the line of equality. However, we see a general improvement in our predictions as the number of years of data included in the analysis increases: the correlation between predictions and observed values increases, and the mean squared errors and standard deviations of the errors between observed and predicted values decreases. Although not apparent from these plots, the precision of our predictions also increases with the number of years of included data.

As a more detailed validation exercise, for each site we also compare the percentile of the predictive distribution for 2012 corresponding to the observed value in that year. Over the full sample of sites, these percentiles should cover the full range from 0 to 100%. Although not shown here, for each of our three analyses using eight years, five years and one year of data, these percentiles do indeed cover this range, suggesting validity in our predictions. Further to this, our 95% prediction intervals for each prediction made were found to capture the observed value on 97.89% of occasions, with similar levels of agreement observed for other intended prediction interval coverages.

4. Conclusions and further work

In this paper we have outlined a novel Bayesian approach to road traffic hotspot prediction. Specifically, our model allows both locally- and globally-observed trend effects to inform predictions and adjust historical model-based estimates of safety, at each site within a pool of candidate road safety hotspots, whilst also smoothing through observed values to account for the confounding effect of RTM. Further, we allow a variance inflation device to afford greater precision to model-based estimates of accident counts in more recent years. Where practically possible, and appropriate, informative prior distributions are elicited to help maximize the precision of estimates from our model.

Demonstrating our model by fitting to data observed at 734 locations in the city of Halle, we exploit the posterior predictive distribution so that our predictions of accident counts in future years take into account uncertainty in parameter estimation and randomness in future observations. We propose simple methods to validate predictions from our model, and for the Halle dataset we have shown that our model predicts well. Results obtained from fitting the model can be displayed in different ways. For example, the full predictive distribution can be viewed graphically, or summarized using a point estimate (perhaps the mean) with a prediction interval (usually the interval with 95% coverage); focusing more on potential road safety scheme implementation, we can also examine the predictive probability of exceeding a threshold accident count in future years. At this point, we should note that a ranked list of sites according to their predicted hotspot potential could, of course, be sensitive to the choice of point estimate used to summarize the posterior predictive distribution or – if an accident frequency exceedance probability approach is to be used – the choice of threshold. Our results in Figs. 2 and 3 illustrate various approaches, but this is something we feel could warrant further investigation and input from practitioners.

In the future we aim to improve the accuracy of our APM by reducing the size of the reference pool it is built from, perhaps moving from a global APM to one which is much more tailor-made for each site in question, using only a subset of the most similar sites to build the APM for this site. For example, separate APMs could be used for intersections and links; relevant geometric design features could be incorporated as covariates in the APM, and for links, other important features such as link length could be used as covariates. The purpose of the current work is to illustrate our Bayesian hierarchical model and its predictive power, and *not* get distracted with issues of APM selection – hence the use of a rather simple log-linear form for our APM in Eq. (2). However, we acknowledge that a better approach to the APM construction could be used, and the benefits of using more sophisticated/potentially better-fitting APMs within the context of our model is something we are currently investigating. Indeed, a key area of our on-going research is the investigation of an extension to the APM to account for potential spatial dependencies between accident counts. We believe such an extension will be of interest to road safety practitioners as it has the potential to indicate any hotspot *areas* or geographic *regions*, rather than individual site-specific results as are currently returned.

Acknowledgments

We would like to thank our colleagues at the Tyne and Wear Traffic Accident Data Unit (TADU) and the Northumbria Safer Roads Initiative (NSRI) for their continued support in this work. Joe Matthews' work is supported by an EPSRC Doctoral Training Award. Data supporting this publication is openly available under an 'Open Data Commons Open Database License'. Additional metadata are available at: http://dx.doi.org/10.17634/154300-33. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

References

Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37, 870–881.

U.K. Department for Transport (DfT), 2006. The COBA Manual (Available at:. https://www.gov.uk/government/publications/coba-11-user-manual).

Fawcett, L., Thorpe, N., 2013. Mobile safety cameras: estimating casualty reductions and the demand for secondary healthcare. J. Appl. Stat. 40 (11), 2385–2406.

Heydari, M., Miranda-Moreno, L.F., Esteban Amador, L., 2013. Does prior specification matter in hotspots and before-and-after studies in road safety? Transp. Res. Rec.: J. Transp. Res. Board 2392, 31–39.

Hirst, W.M., Mountain, L.J., Maher, M.J., 2004. Sources of error in road safety scheme evaluation: a quantified comparison of current methods. Accid. Anal. Prev. 36, 705–715.

Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accid. Anal. Prev. 40, 1531–1542.

Li, H., Graham, D.J., Majumdar, A., 2013. The impacts of speed cameras on road accidents: an application of propensity score matching methods. Accid. Anal. Prev. 60, 148–157.

Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transp. Res. Rec.: J. Transp. Res. Board 1717, 102–108.

Maher, M.J., Mountain, L.J., 2009. The sensitivity of estimates of regression to the mean. Accid. Anal. Prev. 41, 861–868.

Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form and Bayes versus empirical Bayes methods. Transp. Res. Rec.: J. Transp. Res. Board 8401, 1310–1340.

Mountain, L., Fawaz, B., Sineng, L., 1992. The assessment of changes in accident frequencies on link segments: a comparison of four methods. Traffic Eng. Control 33, 429–431.

Park, J., Abdel-Aty, M., Lee, C., 2014. Exploration and comparison of crash modification factors for multiple treatments on rural multilane roadways. Accid. Anal. Prev. 70, 167–177.
Plummer, M., 2016. JAGS Version 3.4.0 User Manual (Available at: www.stats.ox.ac.uk/~nicholls/MScMCMC15/jags_user_manual.pdf).
Ripcord-Iserest, 2005. Road Infrastructure Safety Management (Available at: https://www.ineco.com/webineco/en/what-we-do/solutions/smart-it/ripcord-iserest).
Ripley, B., 2016. MASS. R Package Version 7., pp. 3–45 (Retrieved Mar. 5, 2016. Available at: https://cran.r-project.org/web/packages/MASS/MASS.pdf).
Sacchi, E., Sayed, T., El-Basyouny, K., 2015. Multivariate full-Bayesian hotspot identification and ranking: a new technique. In: Transportation Research Board: 94th Annual Meeting, Washington. DC.
Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B: Stat. Methodol. 55, 3–24.
Wang, C., Quddus, M., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accid. Anal. Prev. 43, 1979–1990.
World Health Organization, 2015. Global Status Report on Road Safety: Supporting a Decade of Action. World Health Organization, Geneva.
Yu, R., Abdel-Aty, M., 2013. Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. Accid. Anal. Prev. 56, 51–58.