

A Novel Bayesian Hierarchical Model for Road Safety Hotspot Prediction

Lee Fawcett^{1a†}, Joe Matthews^{1a}, Neil Thorpe^{1b}, Karsten Kremer², Keith Newman^{1a}

^{1a}School of Mathematics & Statistics and ^{1b}Transport Operations Research Group, Newcastle University, UK; ²PTV Group, Karlsruhe, Germany

†Email: lee.fawcett@ncl.ac.uk



OVERVIEW

- When dealing with traffic accident hotspots authorities are often **reactive**, with countermeasures being implemented once an accident threshold is exceeded.
- We propose a Bayesian hierarchical model for assessing the 'true' level of safety, and predicting future accident counts, at potential accident hotspots.
 - The model is an extension of the classical 'Empirical Bayes' (EB) modeling framework; accident counts from multiple time periods are now used.
 - 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 (the) Mean (RTM)**.
 - Other novel features include a variance inflation device to enable more recent accident counts to inform future predictions with more certainty than counts further in the past, and the ability to estimate the effects of unobserved site-specific factors influencing accident counts.
 - We exploit the **Bayesian posterior predictive distribution** to inform road safety practitioners of *future* levels of safety at potential accident hotspots to allow **proactive** treatment at carefully targeted sites.
- Practitioners can implement our techniques via a web-based application (available now) or a free add-on module to PTV's *Visum Safety* software (coming soon).

THE MODEL: GENERAL STRUCTURE & OUTPUT

- Fig. 1** shows the typical output from our model. It is hoped that practitioners could use this output to make decisions on road safety scheme implementation.
- For accident counts $y_j(t)$ at site j in time period t , we assume

$$y_j(t)|\lambda_j(t) \sim \begin{cases} \text{Poisson}(\text{mean} = \text{var} = \lambda_j(t)), & t \geq 0; \\ \text{NegBin}(\text{mean} = \lambda_j(t), \text{var} > \text{mean}), & t < 0, \end{cases}$$

where $t = 0$ represents the current time period and $t < 0 / t > 0$ represents historic / future time periods respectively.

- We set $\lambda_j(t) = a_j \mu_j(t) \exp\{b_j t\}$, where
 - a global APM is used to infer μ_j from data on infrastructure variables (e.g. major road/minor road, speed limit) and time t ;
 - the parameter a_j is used to account for other unknown site-specific factors;
 - the parameter b_j models any site-specific dependence of counts on time.

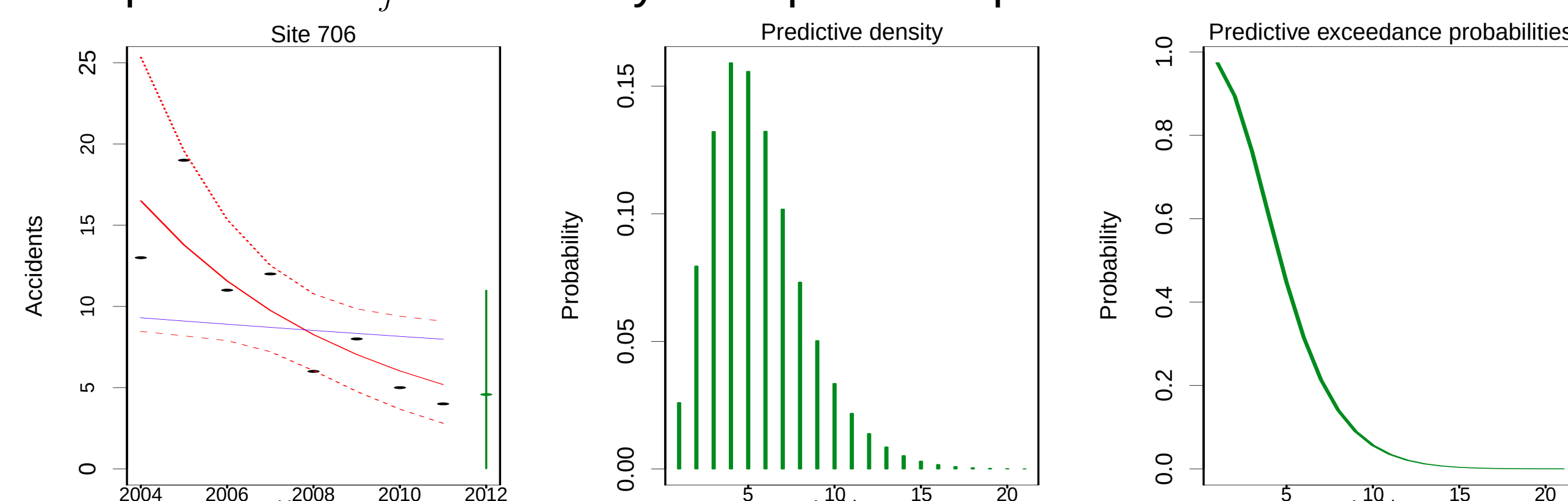


Figure 1: Observed counts; APM (maximum likelihood estimates for $\mu_j(t)$); model-based estimates (posterior means for $\lambda_j(t)$, with 95% confidence intervals); predictive inference for 2012, including 95% prediction interval, full predictive density and posterior predictive probabilities for exceeding y accidents.

DETAILS AND JUSTIFICATION

- Variance inflation:** We assume $\text{var}(y_j(t < 0)) = \lambda_j(t) \exp\{-t\tau\}$, $\tau > 0$.
 - More recent counts should inform our predictions with more certainty;
 - Unobserved infrastructure changes may have occurred over the observation period – we may be less certain about the contribution of observed counts to our model-based estimates for time periods further in the past.
- Prior specifications:**
 - $a_j \sim \text{Gamma}(\text{mean} = 1)$:
 $\mathbb{E}[\lambda_j(t)] = \mu_j(t)$; for a single count at $t = 0$, classic EB structure recovered.
 - $b_j = b_N b_Z$, where $b_N \sim \text{Normal}(\text{mean} = 0, \text{var} = 0.1)$ and $b_Z \sim \text{Bernoulli}(0.5)$:
Models site-specific trend to potentially offset that detected by the APM - the zero-inflation induced by b_Z ensures this occurs only in obvious cases.
 - $\tau \sim \text{Gamma}(\text{mean} = 0.1)$:
Returns a suitable time-dependent inflation of the variance for $y_j(t)$.
 - The exact informative prior specification for b_j and τ was the result of an extensive and iterative elicitation process.
- Bayesian sampling:**
 - A **Metropolis-within-Gibbs** Markov chain Monte Carlo (MCMC) scheme can be used to update the parameter vector $\theta_j^{(i)} = \{a_j, b_j, \tau, \lambda_j(t)\}^{(i)}$ at each iteration i .
 - MCMC samples can be used to approximate the predictive probability function for future counts $y_j(t = 1) = y$ for $y \in \mathbb{N}^0$, at each site j .

DATA AND APM CONSTRUCTION

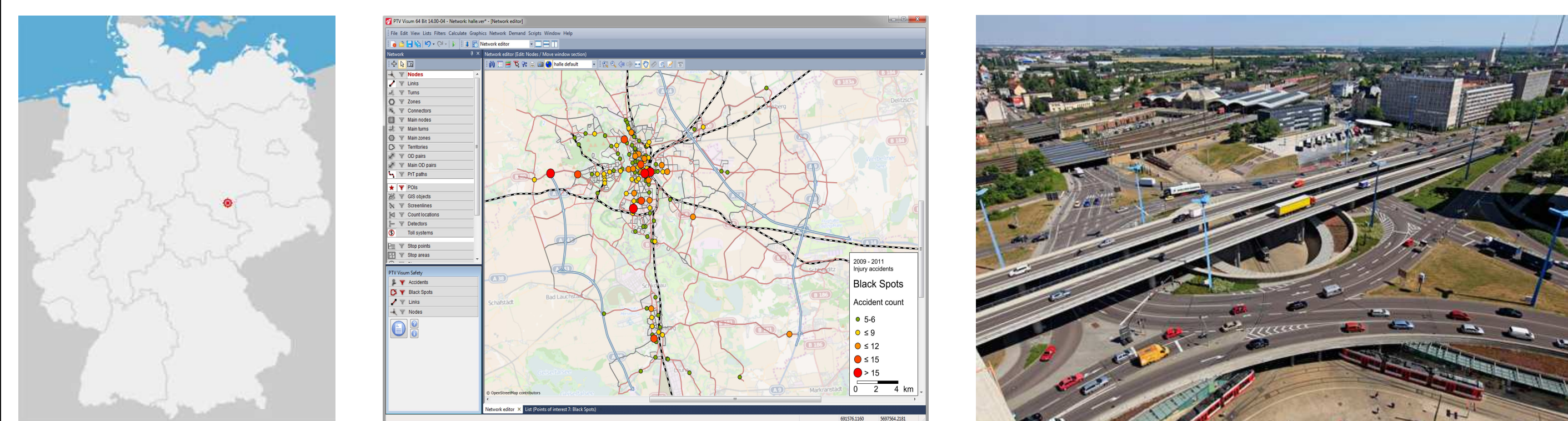


Figure 2: Halle, Germany; data sites, from PTV Visum Safety; 'Riebeckplatz', hotspot site 706.

- We fit our model using annual accident counts at 734 sites in and around the city of Halle, Germany (see **Fig. 2**), for the years 2004–2011. We use counts available for 2012 to validate predictions made for this year from our model.
- For each of these sites, and within each year, we also have information $x_{p,j}$, $p = 1, \dots, n_p$, on $n_p = 10$ infrastructure variables.
- We impose the following log-linear APM for $\mu_j(t)$:

$$\mu_j(t) = \exp \left\{ \beta_0 + \beta_1 t + \sum_{p=1}^{n_p} \beta_p x_{p,j} \right\},$$

with a **negative binomial error structure**.

- We obtain maximum likelihood estimates of the regression coefficients in the APM, resulting in corresponding point estimates of $\mu_j(t)$.
- The estimated APM could be replaced with any other existing APM, such as COBA/RIPCORDER-ITEREST models.

RESULTS AND VALIDATION

- The left-hand-side of **Fig. 1** provides a useful summary of our fitted model for site 706 in Halle. **Table 1** shows numerical results for this site and three others.
- We **validate** our model by comparing predictions with observed counts for 2012. We assess the sensitivity of predictions to the availability of past data (**Fig. 3**).
 - As **Fig. 3** reveals, model predictions compare favourably with observed counts for 2012, regardless of the number of years of historic data (although mean squared errors decrease as the number of years increases).
 - 97.8% of the 95% prediction intervals captured the true counts for 2012; similar levels of agreement are obtained for other levels of confidence.

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

Table 1: Posterior means (95% confidence intervals) for some parameters, at four sites in Halle.

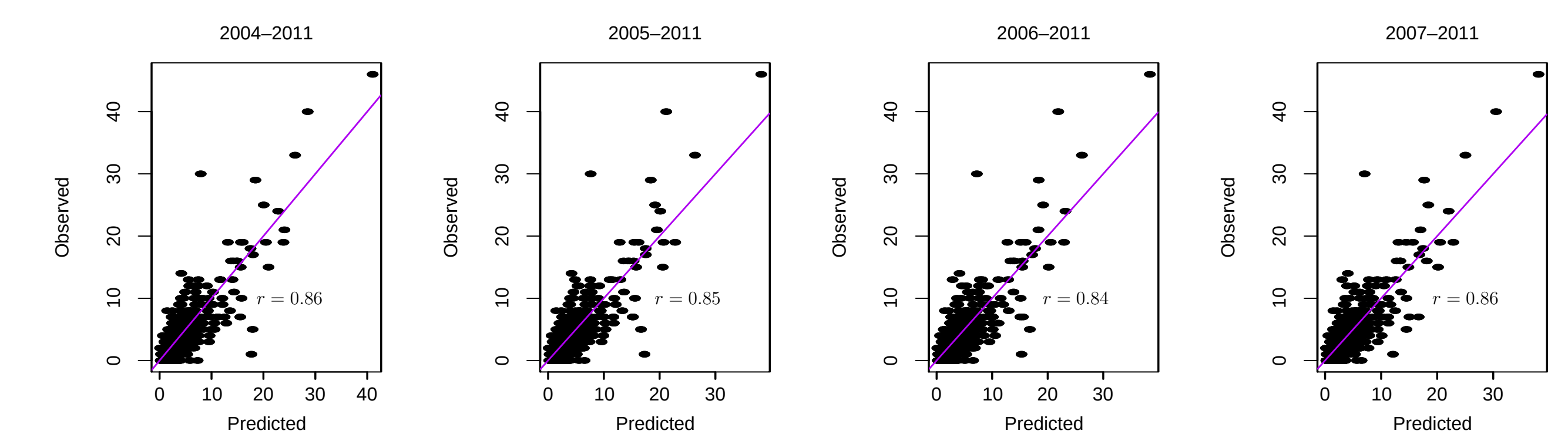


Figure 3: Validation of model predictions for the year 2012, using different length datasets. Predictions shown are the means of the posterior predictive distributions, as shown in Fig. 1 (middle).

SUMMARY, SOFTWARE AND REFERENCES

- We have used a Bayesian hierarchical model to identify annual contributions of RTM and trend to accident counts at 734 sites in and around Halle, Germany; site-specific variations in counts can be used to estimate local offsets, where appropriate.
- A variance inflation component allows us to reduce the weight of observations further in the past on our site-specific estimates of safety.
- Our work is an extension of the classic EB approach to modelling, allowing for multiple time periods and making use of the Bayesian posterior predictive distribution to predict counts in future time periods; our validation checks reveal strong predictive capability.
- We have developed a *Shiny* app which is a prototype to a PTV Visum Safety module; you can try this out: mas-shiny.ncl.ac.uk/hotspot-demo.

Fawcett, L., Thorpe, N., Matthews, J.T. and Kremer, K. (2017). A novel Bayesian hierarchical model for road safety hotspot prediction. *Accid. Anal. Prev.*, **99**, pp. 262–271.

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