An Investigation into the Effect of Regression to the Mean in Road Transport Safety

Lucy Robinson, 110191784
Supervisor: Dr Lee Fawcett
MAS8391 MMathStat Project

In this report we consider two case studies. For our first case study we have data on casualty counts and several explanatory variables for many sites across the Northumbria Police Force area, supplied by the Northumbria Safety Camera Partnership. Treated sites have a before and after period, where high casualty values in the before period led to implementation of mobile speed cameras. We will try to identify whether regression to the mean has affected casualty counts at these sites, by using an accident prediction model constructed using data from control sites. We will use both an Empirical Bayes and Full Bayes analysis, and assess the differences between these, also considering different prior specifications, determining which of these is the most appropriate. For the second case study, we have accident count data from the city Halle, in Germany, with data spanning nine years, supplied by PTV Group, a traffic accident mapping software company. We use all the data to construct an accident prediction model, then use a Fully Bayesian procedure to estimate average accident counts. We can then use the Bayesian posterior predictive distribution to forecast which of these sites will be dangerous in the future, so we can take a proactive approach towards the implementation of road safety schemes.

Contents

Chapter 1

Introduction

1.1 Aims of this project

In this report we investigate the role of statistical modelling in the field of road safety. We consider two case studies:

1. A before/after study aimed at assessing the effectiveness of mobile road safety cameras (speed cameras) in the Northumbria Police Force area in the UK;

2. An analysis of completely untreated, potentially dangerous road safety 'hotspots' in and around the city of Halle, in the state of Saxony-Anhalt, Germany, aimed at predicting the true future road safety of these locations and thus identifying which sites are genuine 'hotspots'.

As we will describe, the data in both case studies are vulnerable to the effects of selection bias, or regression to the mean (RTM), which we attempt to quantify within the Bayesian framework. Case study 1 is retrospective - here, road safety cameras have been applied as a treatment to a selection of 56 sites considered dangerous, and we have figures relating to the number of casualties before and after treatment, as well as several other measured variables. The data were supplied by the Northumbria Safety Camera Partnership, to whom we are grateful. In case study 2, no treatment has yet to be applied; we have annual road traffic accident counts (and various other measured variables) over a period of 9 years (2004 to 2012) for 734 sites in and around the city of Halle, in Germany - the aim here is to identify sites worthy for treatment based on their historic accident record, but also allowing for any potential site-specific RTM effects. Typically (as was the case in case study 1) local authorities would take a *reactive* approach to treatment, implementing a road safety scheme once a high threshold of casualties or accidents has been

observed at a particular location. The aim of case study 2 is to consider a more *proactive* approach to treatment, predicting future levels of safety and acting on these predictions before the threshold has been reached, potentially saving casualties/lives. The data for case study 2 were supplied by PTV, a traffic accident mapping software company based in Karlsruhe, Germany, to whom we are grateful.

In this report we aim to assess the sensitivity of our estimates of regression to the mean to the choice of model structure: we initially use a simple Empirical Bayes technique, later moving on to use MCMC methods to allow for non-conjugate prior distributions. The Empirical Bayes technique is the industry gold standard, however we will show that this method is over-optimistic when identifying regression to the mean effects, as the estimates for casualty/accident rate are too precise due to posterior standard deviations being too small. This method only allows the use of conjugate prior distributions. However, for case study 1, we will show that a non-conjugate specification is most appropriate. For case study 2 we aim to use the posterior predictive distribution within the Bayesian framework to predict future accident counts at the given sites, after accounting for RTM. This will help us to determine which sites will become dangerous 'hotspots' in the next year. The statistical software package R was used for computational work [1], [2].

1.2 Regression to the Mean

1.2.1 Sir Frances Galton and Regression to the Mean

The phenomenon of regression to the mean was first considered by Sir Frances Galton (1822-1911). He published a paper [3] on a study of the heights of parents and their grown children, establishing if and how they were related. We would expect the height of the child to be the average of the two parents; however, this was not the case. He discovered that if the parents were very tall, the child would generally be shorter, and for very short parents, the child would usually be taller. This can be described as an unusually high (or low) value being followed by an average value, which might be closer to the overall mean height of the children. He named this occurrence 'regression to the mean' (RTM), as the heights of the children seemed to 'regress' back to the average value. RTM can be seen in many areas of life - for example Sir Galton also noticed its occurrence in the size of sweet pea seeds. The area I will consider, however is road safety, and we will investigate how the casualty or accident values at particular sites might generally regress back to some underlying mean value.

1.2.2 More examples of RTM

We have seen that RTM is prevalent in Biology - however it can also be seen in the economy, medicine [4], sport [5] and the media [6]. Olympic gold medalists, for example, seem to perform more poorly in the months after the games. This could be because at their peak in the Olympics they have done unusually well, so the attention on their performance is high. If they don't win the next competition, they appear to have gotten worse, rather than just returned to their usual level. It can also be seen in alternative medicinal treatments, for example drinking lemon juice to reduce a headache. The headache is going to disappear over time naturally; however, the lessening of the pain might be attributed to the lemon juice. This is described as a coincidental recovery, whereby the treatment had no real effect. Disappointing film sequels could also be due to regression to the mean. If the original film was a triumph, the expectation would be high for the sequel. If the sequel didn't live up to this success it would be deemed a failure. However if it was an original, it might be considered average quality.

1.2.3 RTM in Road Safety

In 2000, a department for Transport report revealed that each year on Britain's roads there are over 300,000 casualties, around 3,500 of these being fatal. This information was part of a report *Tomorrow's Roads - Safer for Everyone* [7] aimed at reducing road casualties and making the roads safer for all to use. Another incentive to reduce casualties was the financial burden of these accidents, estimated to be around £3bn a year. For these reasons speed cameras were deployed - initially under a two year pilot program involving eight road safety camera partnerships, in April 2000 [8]. Due to the high installation and running costs, they were financed using the money people were fined for exceeding the speed limit. By the end of the year the results looked promising and in 2001 there were safety camera partnerships across the length and breadth of the UK.

Due to the rise in the numbers of speed cameras, and therefore a rise in the number of people being fined, the general public became angry with the scheme. Some people argued that they were "...just another government tax" [9], and speed cameras in some areas were vandalized, seen in Figure 1.1. Opponents of speed cameras were attempting to have the scheme stopped, and their argument about the effectiveness included selection bias and the resulting regression to the mean effects. The argument against the use of speed cameras is as follows.

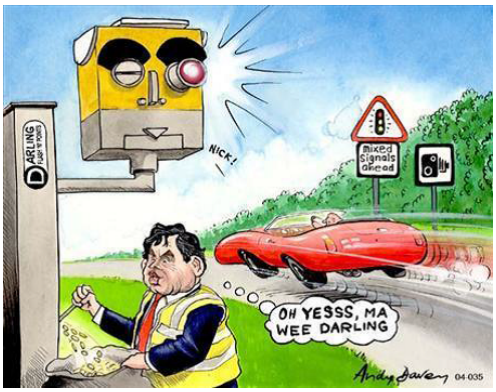

Figure 1.1: Left: Cartoon published in the Independent, April 2001, Right: Vandalism to speed cameras [9]

The number of casualties or accidents in a specific time period is observed, and if that value exceeds some pre-determined 'safety threshold', a safety scheme may be put in place, for example a speed camera. The 'after' period takes place once this scheme is set up, and the same observation on casualty/accident counts made. We find that, more often than not, these counts decrease post-treatment, which can be taken as saying the speed camera has reduced accidents/casualties (perhaps saving lives), so the safety scheme is a worthwhile investment. The main problem with this type of study is the lack of a control group, so we have nothing to compare the casualty rate at the treated sites to. This amounts to using the 'before' period as the control group, so we assume that figures here give the average value of what we would expect to see. However, selection bias has taken place - only the sites with unusually high casualty values have been selected for a safety scheme. This means these sites might not usually be this dangerous, but by chance have had a large number of accidents in the selected time period. We will call these sites 'hotspots'.

Many people argue that these 'hotspots' were at the peak of a 'blip' - they had a high value in the specific 'before' time period, however normally they are not that dangerous. The lower value seen in the 'after' period is, naively, viewed as a result of the treatment; however they believe the casualty value would have reduced to this level naturally, regardless of a speed camera being put in place, due to RTM. We can see the different effects of regression to the mean in before/after studies in Figure 1.2.

We can see that in all cases the casualty value has decreased from the first time it was measured. However here we have historic casualty data, so we can see the underlying mean value, which makes regression to the mean

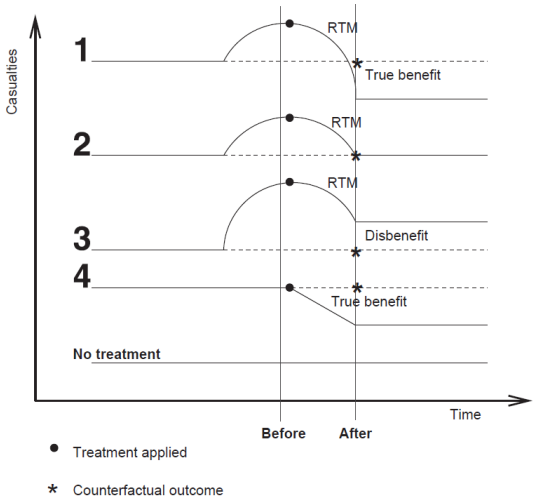

Figure 1.2: Hypothetical outcomes of RTM in before/after studies. [8]

very noticeable. Following line 1, we can see when a safety scheme is put in place the casualty rate decreases past the historic mean value, showing the treatment has had a positive effect in reducing casualties. The second line, however, only reduces from the peak of the 'blip' to the historic mean value, which we would expect over time anyway, so the safety scheme put in place has been a waste of money, as it has not saved any casualties from occuring. Line 3 shows a safety scheme seemingly causing a negative effect on the casualty rate - it has caused the number of casualties to rise higher than the historic rate. Looking at just the 'before' and 'after' values, this is not apparent. The fourth line shows the safety scheme installed when the casualty rate is not at the peak of a 'blip', however the treatment has a positive effect and decreases the casualty rate. Most reports of a before/after study would conclude that this is what occurred, due to the decreased casualty rates, when in reality it could be any of the above options. It has been found in previous studies [8] that the average reduction in road casualties due to speed cameras is around 30%, after accounting for regression to the mean. This low value has led to fewer speed cameras being implemented, and the government looking for different ways to prevent casualties, including road safety advertising and safety talks in schools.

1.3 Modelling Techniques Used

For both case studies, the aim is to produce an accident prediction model (APM), which tells us for each site, what accident/casualty rate we would expect to see at these sites ordinarily, based on observations made at 'similar'

sites. For the Northumbria study, we have a set of control sites which we use to construct the APM; in the Halle study, we use all sites to form the APM and refer to the fitted values at each particular. In both cases, a log-linear model is used to link casualty/accident counts to several explanatory variables.

1.3.1 General Modelling Framework

For the Northumbria analysis we assume a Poisson distribution for casualty frequency Y_j at treated site j with rate m_j. Similarly for the Halle analysis we have a Poisson distribution for accident frequency Y_j at site j, also with rate m_j. To begin with we use a conjugate prior distribution for casualty/accident rates m_j - for a Poisson likelihood we need a Gamma prior. We then move on to different non-conjugate cases, such as the lognormal and Weibull distributions. For all these priors the mean value is $E(m_j) = \mu_j$.

We can then define the log-linear multiple regression model (APM) for μ_j as

$$\mu_j = \exp\{\beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj}\}, \tag{1.1}$$

where $x_{ij}, \ldots, x_{pj}$ represent covariate information, for example average speed. For the Northumbria analysis, this model is estimated using information from a set of control sites, and then applied to the same covariate information at our treated sites. For the Halle analysis we fit the model using information from all sites and then let μ_j be the fitted value at site j.

The sites that are most likely to suffer the effects of regression to the mean will have a large discrepancy between the observed casualty/accident counts and the values given by the accident prediction model. A site could have a much larger casualty count than expected, or in some cases a smaller number of accidents than expected. The difference between the observed values and the posterior mean for m_j is taken to be the RTM effect. The posterior mean value can be shown to be a weighted sum of the observed value and the prior mean - what we *have* seen, and what we *expected* to see.

1.3.2 Empirical Bayes Procedure

This method is the industry standard in the road safety literature, however it only allows us to use conjugate priors, so here we must remain with the Poisson-Gamma formulation. The unconditional distribution for y_j is negative Binomial, so the error structure for μ_j in Equation (1.1) must have the same form. We can use maximum likelihood estimation to estimate the regression coefficients and the negative Binomial over-dispersion parameter

κ in Equation (1.1). In a standard Empirical Bayes analysis, these parameters are then treated as the true values; no acknowledgement is given to the standard errors, and so estimates of μ_j are obtained directly on substitution of covariate information into (1.1). Standard methods (eg. backwards elimination) can be used to discover which of the given explanatory variables are significant.

1.3.3 Fully Bayesian Techniques

A fully Bayesian analysis is potentially more realistic in estimating RTM effects, as we can allow for uncertainty in estimation of all parameters by adopting prior distributions for the regression coefficients and negative Binomial over-dispersion parameter. These prior distributions could be chosen by an expert in road safety, who could give their opinion on how these coefficients should be distributed; however, here we will use vague priors in the absence of such knowledge.

The Markov Chain Monte Carlo (MCMC) techniques [10] used in this section allow us to make inferences on the parameter vector $\boldsymbol{\theta}$ by simulating realisations from the posterior distribution. This is adapted when using non-conjugate prior specifications, due to the fact the posterior cannot be calculated analytically, and the respective conjugate distribution might be too restrictive for the situation.

The sampling method used in this report is the Metropolis-Hastings technique [10]. This simulates realisations from the posterior distribution via a proposal distribution with density $q(\boldsymbol{\theta^*}|\boldsymbol{\theta})$, which is easy to simulate from. This gives a method of proposing new values $\boldsymbol{\theta^*}$ from the current value $\boldsymbol{\theta}$. The algorithm to produce this is given below.

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\boldsymbol{\theta}^{(0)}$;
2. Generate a proposed value $\boldsymbol{\theta}^*$ using the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$;
3. Evaluate the acceptance probability $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}^*)$ of the proposed move, where
$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{x})q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}|\boldsymbol{x})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right\}; \tag{1.2}$$
4. Set $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$ with probability $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}^*)$, and set $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$ otherwise;

5. Change the counter from j to $j+1$ and return to step 2.

At each stage we generate a new value from the proposal distribution, which is either accepted or rejected. If the value is accepted then the chain moves to a new position, and if rejected stays where it is. This depends on the acceptance probability, of which the optimal value is around 23.4% [10]. If too many values are accepted then the chain will move slowly around the parameter space, and if too few values are accepted then they will correspond to large jumps in values. Both of these would mean the chain takes longer to converge, which means more computational work.

The proposal distribution used in this project is a random walk with Normal innovations. This means the proposed value $\boldsymbol{\theta}$ at stage j is

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(j-i)} + \boldsymbol{w}_j, \tag{1.3}$$

where the $\boldsymbol{w}_j$ are independent and identically distributed random $p \times 1$ vectors, with Normal distributions, where p is the number of covariates used, i.e. $\boldsymbol{w}_j \sim N_p(\mathbf{0}, \Sigma)$. We can then easily simulate an innovation $\boldsymbol{w}_j$ and compute the proposal value given in Equation 1.3. Via this method we can compute realisations from the posterior distribution.

Due to the use of MCMC we can move on to non-conjugate prior distributions, which could be more appropriate for the situation. This leads to the issue of deciding which prior distribution is the 'best' for this model. One way to evaluate this is to consider the Deviance Information Criterion (DIC) [11]. This value takes into account the goodness-of-fit of the model, and has a term that depends on the complexity of the model. In general a more complex model will give the best fit, but may be very complicated and have lots of variables, so the idea is to compromise between simplicity and fit. The lower the value of the DIC the better the model is.

The Full Bayes method hasn't become industry standard yet due to the added computational work, and isn't mentioned much in the applications literature in road safety. A few examples of a Full Bayes approach are [8] and [12], and these papers have more realistic and accurate assessments of the regression to the mean effect. Researchers in the School of Mathematics & Statistics and Transport Operations Research Group (TORG) at Newcastle University have recently been awarded a University Strategic Research Grant to address this issue, the aim of which being to provide local authorities with user-friendly software which performs fully Bayesian analyses of accident/casualty data, to assess the impact of new road safety features.

Chapter 2

Mobile Safety Cameras: A Before/After Study

2.1 Background and Data

The data in this case study were provided by the Northumbria Safety Camera Partnership (NSCP). This body comprises local authorities, the Northumbria Police Force, academics from Newcastle and Northumbria Universities and the local NHS secondary healthcare providers. This partnership joined the national program as mentioned in Scetion 1.2.3 in April 2003, and in February 2004 investigated the role of mobile speed cameras in reducing casualties due to road traffic accidents [8]. Mobile speed cameras are portable, and can be operated from a Partnership vehicle in various locations. This lowers the cost of installing and maintaining fixed speed cameras.

We have two time periods in this study: the 'before' period is from April 2001 to March 2003, and the 'after' period is from April 2004 to March 2006. We have data over this period for 56 sites, for which mobile speed cameras had been implemented in the 'after' period. We will refer to these as the treated sites. We also have data in the 'before' period for 67 sites in the same area, which didn't have speed cameras deployed. We will refer to these as control sites. Our aim is to use these control sites to build an accident prediction model (APM) to apply to the treated sites to use as our prior mean μ_j; see Equation (1.1) This means we can try to estimate what would happen at the treated sites in an average time period, perhaps when casualties were not at the 'peak of a blip' - will regression to the mean occur, and were the speed cameras necessary?

The data for both treated and control sites included seven explanatory variables, and a casualty value. These explanatory variables were: speed limit,

Table 2.1: Summary statistics for casualty counts

	Mean	Median	St. Dev.	LQ	UQ	Min	Max
Contol	4.284	3	4.770	0	18.45	0	24
Treated (Before)	7.786	7	5.379	1	20.63	1	28
Treated (After)	5.321	4	4.023	0	15.25	0	16

average speed, 85th percentile speed, percentage of drivers over the limit, percentage of drivers over 15 mph over the limit, traffic flow, road classification (0, 1, 2 or 3), road type (1 = single carriageway, 2 = dual carriageway or 3 = mixed). We can look at some summary statistics and plots to investigate the data.

We can see from Table 2.1 that the mean casualty value for the treated sites in the 'before' period is much larger than for the control sites. This could suggest the sites here are at the peak of a 'blip', so will be vulnerable to regression to the mean; or they are actually dangerous 'hotspots' that need to be treated. We can see the casualty value decreases for the treated sites from 'before' to 'after' periods, which could show the positive result of speed camera implementation, or the regression to the mean effect. However the standard deviation values are large, leading to very wide 95% confidence intervals. The maximum value has dropped from the 'before' to the 'after' period, once again suggesting the speed cameras have been effective.

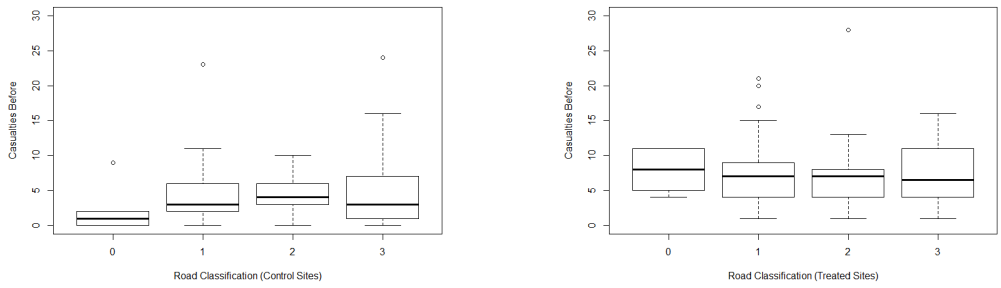

Figure 2.1: Boxplots of Road Classification against 'before' casualty value for Left: control, Right: treated sites

In Figure 2.1 we can see the casualty values against different road classifications in the 'before' period for treated and control sites. For the control sites we can see the median values are all low, however they tend to have longer tails leading to the higher casualty values, showing they are positively skewed. There are a few outliers, but they are not all in one classification,

showing that for the control sites there doesn't seem to be a particularly dangerous road classification. For the treated sites however there are three outliers in one classification, suggesting the treated sites in this class could be more dangerous. There is one very high outlier, but the tails of the plots seem to be more even than for the control sites, suggesting less skew. The median values are all higher, showing the larger number of casualties at these potentially dangerous 'hotspots'.

We can use these data to try estimate our APM, as discussed in Section 1.3.1, but we first need to consider if it is sensible to try and predict casualty rates at the treated sites using the control sites - are these sites exchangeable?

2.2 Assessment of Exchangeability between Treated and Reference Sets

As discussed in Section 1.3.1 our aim is to elicit a value for the mean of the prior distribution for casualties at the treated sites by applying a regression model, constructed using data at the reference sites, to covariate information at the treated sites. To be able to use the control sites to predict what we could expect to see at the treated sites, when casualties are not at the peak of a 'blip', we must check they are exchangeable. We are looking to test the null hypothesis H_0: Sites are exchangeable against H_1: Sites are not exchangeable. This can be done via a permutation test [8].

To test this we need to find the distribution of the test statistic under the null hypothesis. We can do this approximately, by generating multiple samples via random permutations of the data. We estimate the p-value, $\hat{P}$, from comparing the mean value of each explanatory variable used in our model. We can calculate the value δ_p for each of these variables:

$$\delta_p = |\bar{x}_p^{TRT} - \bar{x}_p^{CTR}|, \quad p = 1, \ldots, 7, \tag{2.1}$$

where TRT and CTR denote the treated and control sites respectively.

In this permutation test we randomly re-allocate sites to the treatment and control groups at each permutation. We then find δ_p at each permutation Π_k with $k = 1, \ldots, 10,000$. Next we compare the observed distributions of δ_p from these random permutations with the values of δ_p in the original allocation. We can define the indicator variable I_k as

$$I_k = \begin{cases} 1 & \text{if } \delta_p^{\Pi_k} \geq \delta_p, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

We can now estimate the p-value $\hat{P}$, which tests the null hypothesis. This number is the proportion of values with a value of $\delta_p^{\Pi_k}$ that is at least as large as the original permutation value δ_p:

$$\hat{P} = \sum_{k=1}^{N} \frac{I_k}{N}. \tag{2.3}$$

We perform this calculation for all explanatory variables. All are greater than 0.05; for example, for average observed speed and percentage of drivers over the speed limit, we have $\hat{P}_1 = 0.0767$ and $\hat{P}_2 = 0.5954$ (respectively). Thus there is insufficient evidence to suggest these sites are not exchangeable.

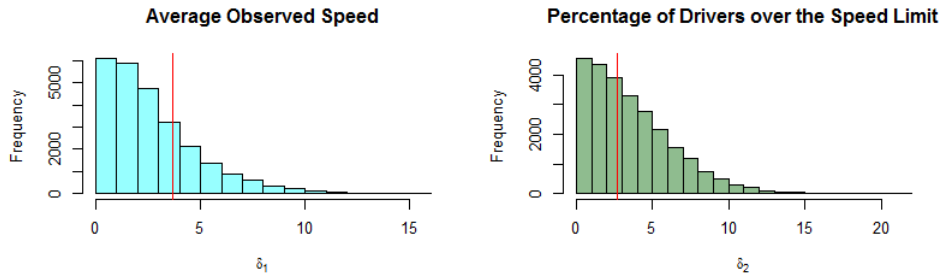

Figure 2.2: Distributions of δ_p for $p = 1, 2$, with the mean value of δ_p for the original permutation superimposed.

We can also look at the distribution of δ_p values, and compare with the original permutation value. If this value is well within the centre of the distribution then we can say that control and treated sites and exchangeable for that variable. This can be seen in Figure 2.2, and we can conclude from this that the sites are exchangeable for both explanatory variables, as the δ_p values are close to the centre of each distribution.

Another method we can use to check exchangeability is a Principal Components Analysis [13]. This method takes a large number of possibly correlated variables and orthogonally transforms them to obtain a small set of linearly uncorrelated principal components. The first principal component carries the highest variance, and the second carries the second highest variance, and so on. This method keeps most of the information and variation in the data, however massively reduces the number of variables, leading to easier visualisation of the data.

Using this method on our data, we obtain the information in Table 2.2. We can see that the first principal component explains around 87% of the data, and a combination of the first two components explains 96% of the data. This

Table 2.2: Principal Component Analysis

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard Deviation	26.158	8.514	5.203	0.994	0.527
Proportion of variance	0.872	0.092	0.034	0.001	0.000
Cumulative proportion	0.872	0.964	0.998	0.999	1.000

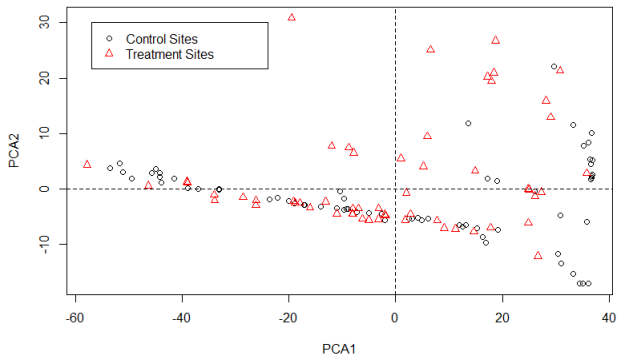

Figure 2.3: Plots of score on first two Principal Components for control and treated sites.

is a huge amount of the variability, so we can ignore the last three principal components, and produce a plot using only the first two, seen in Figure 2.3.

Ideally in a graph of principal components, for groups that are interchangeable we would see a completely random mix of the two groups. Here we can see that most of the points are gathered along one central line, with a seemingly random mix of control (black circles) and treated sites (red triangles). We have a few treated sites in the middle at the top that are separate from other control sites, so these sites might not be very well represented by the control sites. However for the majority of points there is no clear division between the two groups, so we can conclude again that the sites are interchangeable.

2.3 Empirical Bayes Analysis

We use the posterior disrtibution of the mean casualty rate at each treated site to estimate the effect of RTM, as discussed in Section 1.3.1. For the casualty counts y_j at each treated site j we use a Poisson distribution, with mean m_j. The classic Empirical Bayes analysis adopts a conjugate prior - ie. a Gamma, with mean μ_j and variance μ_j^2/γ. The unconditional distribution of y_j is therefore negative Binomial, with the over-dispersion parameter $\kappa =$

γ^{-1}. We therefore have

$$y_j|m_j \sim \text{Po}(m_j), \tag{2.4}$$

$$m_j \sim \text{Ga}\left(\gamma, \frac{\gamma}{\mu_j}\right). \tag{2.5}$$

2.3.1 Estimating the Accident Prediction Model

We have previously seen the control data has seven explanatory variables, which are: speed limit, average speed, 85th percentile speed, percentage of drivers over the limit, percentage of drivers over 15 mph over the limit, traffic flow, road classification (0, 1, 2 or 3), road type (1 = single carriageway, 2 = dual carriageway or 3 = mixed). To include the factorial variables in our APM, of the form given in Equation (1.1), we must introduce indicator variables. The value for flow is divided by 10,000 to make sure it doesn't dominate the regression equation due to the size of the values.

We now use these control sites to try to estimate the APM. We need to discover which of the explanatory variables significantly influence the number of casualties. This can be done via a backwards elimination procedure, yielding a log-linear model given by

$$\hat{\mu} = \exp\big\{1.93 - 0.04x_1 - 0.01x_2 + 0.44x_3 + 0.67x_{I4} + 0.85x_{I5} + 1.06x_{I6}\big\}, \tag{2.6}$$

where $\hat{\mu}$ denotes the expected casualty value. The significant variables here are x_1 = average speed, x_2 = percentage of drivers over the speed limit, x_3 = flow/10000, x_{I4}, x_{I5} and x_{I6} are indicator variables, signifying road classification A, B, C or U.

$$x_{I4} = \begin{cases} 0 & \text{if road classification = B, C or U} \\ 1 & \text{if road classification = A.} \end{cases}$$

$$x_{I5} = \begin{cases} 0 & \text{if road classification = A, C or U} \\ 1 & \text{if road classification = B.} \end{cases}$$

$$x_{I6} = \begin{cases} 0 & \text{if road classification = A, B or U} \\ 1 & \text{if road classification = C.} \end{cases}$$

The coefficient signs in Equation (2.6) can point towards what might make a road dangerous. For example, the positive coefficient for flow shows that if there is a large number of vehicles on the road each day, the more likely an accident is to occur. We can perform a regression in R using the package 'MASS' [2]. This is given below.

```
Call:
glm.nb(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, init.theta = 2.494232366,
    link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5331  -1.0084  -0.2509   0.5576   1.5056
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.933337   0.533574   3.623 0.000291 ***
x1          -0.041081   0.014733  -2.788 0.005297 **
x2          -0.012686   0.003921  -3.235 0.001215 **
x3           0.444237   0.193423   2.297 0.021636 *
x4           0.674396   0.417056   1.617 0.105870
x5           0.845727   0.422039   2.004 0.045080 *
x6           1.060389   0.380154   2.789 0.005281 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
(Dispersion parameter for Negative Binomial(2.4942) family taken to be 1)
    Null deviance: 117.331  on 66  degrees of freedom
Residual deviance:  75.525  on 60  degrees of freedom
AIC: 327.41
Number of Fisher Scoring iterations: 1
              Theta:  2.494
          Std. Err.:  0.774
 2 x log-likelihood:  -311.406
```

This regression gives us point estimates for the $\hat{\beta}_i,\ i = 1, \ldots, 6$ and also for the value of $\hat{\gamma}$. Here we have

$$\hat{\gamma} = \frac{1}{\hat{\kappa}} = 2.494,$$

where κ is the negative Binomial dispersion parameter. We can see the standard errors for the estimated regression coefficients, however in a typical empirical Bayes analysis these are not used, and the estimated APM is applied directly to the data at the treated sites, the estimated coefficients being used as the 'true' values.

We can now use this accident prediction model, based on data at the 67 control sites, to try and predict casualty values in the 'before' period at the treated sites, if we weren't at the peak of a blip.

2.3.2 Posterior distribution for m_j

We have already formed the Gamma prior distribution and the Poisson likelihood, and due to the fact they are conjugate distributions, the posterior distribution is easy to obtain.

$$\begin{aligned} \pi(m_j|y_j) &\propto \pi(m_j) \times f(y_j|m_j) \\ &\propto m_j^{\gamma-1} e^{-m_j\gamma/\mu_j} \times m_j^{y_j} e^{-m_j} \end{aligned} \tag{2.7}$$

$$\propto m_j^{\gamma+y_j-1} e^{-m_j(1+\gamma/\mu_j)}$$

We can see this is also a Gamma distribution; specifically we have

$$m_j|y_j \sim \text{Ga}\left(\gamma + y_j, \frac{\gamma}{\mu_j} + 1\right), \quad j = 1, \ldots, 56. \tag{2.8}$$

The mean of this distribution is then used as the Empirical Bayes estimate of casualty frequency, which we calculate to be

$$\begin{aligned} \text{E}[m_j|y_j] &= \frac{\gamma + y_j}{\gamma/\mu_j + 1} && (2.9)\\ &= \alpha_j \mu_j + (1 - \alpha_j) y_j, && (2.10)\\ \text{where } \alpha_j &= \frac{\gamma}{\gamma + \mu_j}. \end{aligned}$$

We can see that the Empirical Bayes estimate of casualty frequency is a weighted sum of what we observed at the treated sites, y_j, and what we would expect to see at that site during an average time period μ_j, which is based on information from the control sites.

2.3.3 Results

We can now view results from this analysis, and look in particular at some interesting sites, which we can draw conclusions from. Table 2.3 shows the sites we have chosen to discuss, which all have some interesting features.

Looking initially at site $j = 1$, we see a large observed casualty value in the 'before' period, compared with the very small value for μ_j which we would expect to see at this site, based on the control sites. The posterior mean is a weighted average of these two values. In the 'after' period there were no observed casualties, meaning that without accounting for regression to the mean, speed cameras appear to have saved twenty casualties. However, comparing the 'after' value with the posterior mean, we see the speed cameras have in fact saved around nine casualties, which is one of the most effective safety schemes in this study. This is similar to line 1 in Figure 1.2. This, of course, assumes there are no other casualty reduction factors between the before and after periods, including trend.
Now considering site $j = 5$, simply comparing the 'before' and 'after' values looks as though the safety scheme in place has saved around nine casualties. However, comparing the posterior mean with the 'after' value shows that the speed camera has been ineffective, as after accounting for regression to the mean there has been no decrease in casualty frequency. This is comparable

Table 2.3: Empirical Bayes results for sites $j = 1, 5, 10, 13, 19$. For full table, including standard deviations, see Appendix.

Site	y_j	μ_j	α_j	$E(m_j\|y_j)$	$y_{j,after}$	Observed	After RTM
j=1	20	1.61	0.61	8.81	0	-20	-9
j=5	17	1.67	0.60	7.83	8	-9	0
j=10	6	2.70	0.48	4.41	9	3	4
j=13	12	1.74	0.59	5.95	2	-10	-4
j=19	21	1.40	0.64	8.45	10	-11	1
Total	436			323	298	-138	-25

with line 2 in Figure 1.2, where the entire reduction appears to be attributable to RTM.

Site $j = 10$ actually shows a larger value in the 'after' period than the 'before', which is very unusual, as we would expect the value to decrease. Simply comparing values, we see that the speed camera looks to have caused three casualties, and when accounting for regression to the mean this increases to four. The reason for this is not clear. It could be that trend in casualty counts, which we have not taken into account, in that particular area is increasing, or the 'after' period is actually the peak of the 'blip' and the casualty values will decrease again in the near future.

Site $j = 13$ is quite similar to site $j = 1$, as it appears as though the speed cameras have been effective and prevented casualties; however less effective than they appeared to be by just comparing 'before' and 'after' values.

Finally site $j = 19$ is similar to line 3 in Figure 1.2. The comparison of 'before' and 'after' values looks like the speed camera has been effective, however when taking into account regression to the mean, we can see that the 'after' value is still higher than the posterior mean. This means the speed camera has possibly had a negative impact on the casualty frequency.

We can consider the totalled values to make a rough judgement on the overall effectiveness of the speed cameras. The difference between the 'before' and 'after' values is huge, however the regression to the mean effect can be calculated as around 26%, showing that a large amount of this difference is not due to the speed cameras. However there still appears to be a substantial decrease in casualty values, so the safety schemes put in place appear to have been effective and necessary.

We can now look at some graphs illustrating the differences between the number of casualties, the prior mean μ_j and the posterior mean $\mathrm{E}(m_j|y_j)$.

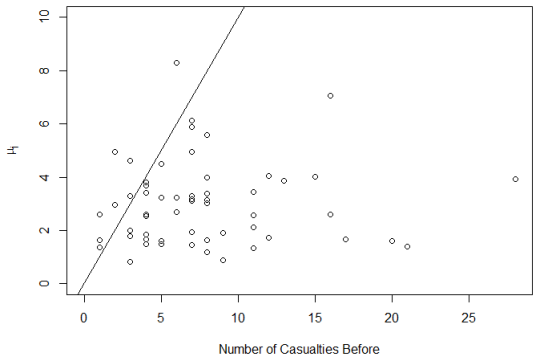

Figure 2.4: Number of casualties in 'before' period against μ_j.

Looking first at Figure 2.4, we see the number of casualties in the before period against the prior mean. We can see most of the values are below the line of equality, showing that most treated sites have a higher number of casualties than we would expect to see. This could be evidence that they are at the peak of a 'blip' and vulnerable to regression to the mean.

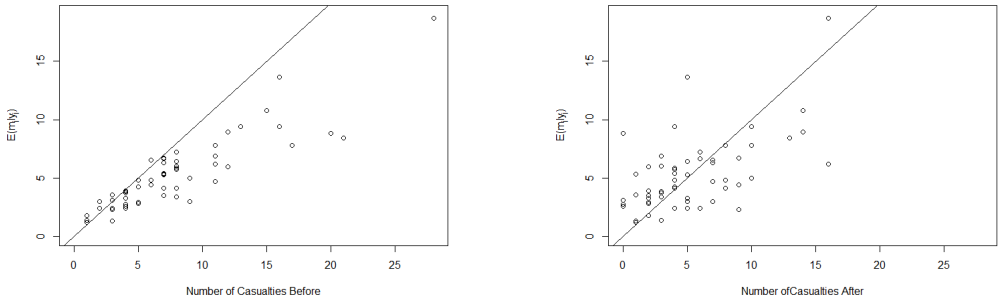

Figure 2.5: Casualty value against E($m_j|y_j$), Left: 'Before', Right: 'After'.

In Figure 2.5 we can see the number of casualties, both 'before' and 'after', against the posterior mean. In the 'before' period we can see that nearly all the values are below the equality line, showing that they have an unusually high casualty value, compared to what we would expect after smoothing by the prior means. However, looking at the 'after' period against the posterior mean, we can see the points are fairly evenly spread around the line of equality. This shows the casualty value has decreased to a level we would expect at that ste, due to a combination of the safety scheme in place and regression to the mean.

There are a few treated sites with very small casualty values, which seems

unusual. The safety scheme could have been put in place for other reasons than very high casualty rates, for example a primary school in that location, or a new road layout.

2.3.4 Problems

The method we have discussed so far is known as the 'gold standard' in the road safety literature, however this method has some limitations. We have used a Gamma prior distribution as it is a conjugate distribution to the Poisson, allowing us to form the Gamma posterior analytically. However, this might not be the most appropriate distribution for the situation, so we could be left with unrealistic conclusions. We have also used point estimates of $\beta_i,\ i = 0, \ldots, 6$ and γ, without any regard to their estimation error. These values in turn lead to the values of μ_j for each treatment site $j = 1, \ldots, 56$. This gives us low and possibly unrealistic values for $\text{SD}[m_j|y_j]$. We will attempt to address these issues in the following section, where we perform a Fully Bayes analysis on the data.

2.4 Fully Bayesian Analyses

2.4.1 Full Bayes Analogue of Empirical Bayes Analysis

Initially we will perform the Full Bayes analysis while continuing with the Poisson-Gamma setup, so we can compare the differences between the Empirical and Full Bayes methods like-for-like. However, now we need to define some prior distributions for the β_i, for $i = 0, \ldots, 6$, and for $\kappa = 1/\gamma$. In the absence of any expert prior information regarding these parameters, we use:

$$\beta_i \sim N(0, 100), \tag{2.11}$$

$$\rho = \log(\kappa) \sim N(0, 100), \tag{2.12}$$

where we have introduced the variable ρ to retain the positivity of κ in our MCMC scheme, and log denotes the natural logarithm. These are vague prior distributions, meaning the values for the β_i and ρ could take positive or negative values, and be of small or large magnitude, reflecting our degree of prior uncertainty. The prior distributions for β_i are independent, and if we had more time, an improvement would be to use a multivariate Normal prior to allow dependency between the β_i through a covariance matrix.

We will initialise the Markov Chain at $\beta_i^{(0)}$, the maximum likelihood estimates seen in Equation 2.6, and at $\rho^{(0)} = \log\left({}^{1}/_{2.494}\right) = -0.914$. We do this to

minimise any 'burn-in' period, which would result in us needing to discard some of the iterations.

We will now perform the MCMC for $K = 10,000$ iterations, and at each of these we calculate the total casualty frequency from all sites, given by

$$T^{(k)} = \sum_{j=1}^{56} m_j^{(k)} | y_j, \quad k = 1, \ldots, K. \tag{2.13}$$

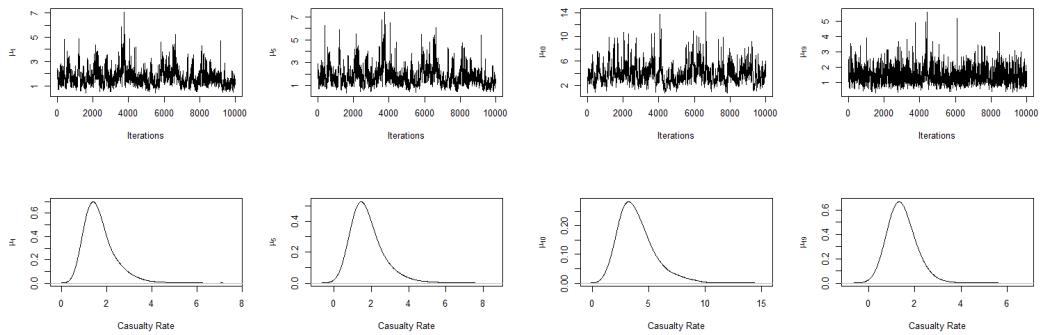

Figure 2.6: Trace plots and densities for μ_j for $j = 1, 5, 10, 19$

We can see some of the trace plots and corresponding densities for μ_j for $j = 1, 5, 10, 19$ in Figure 2.6. We can see the trace plots are remaining at the same level that they were initialised at, and oscillating around this value. The densities look to be positively skewed, suggesting higher casualty values are rare, but expected. The acceptance probabilities for all the parameters were in the range 20% to 40%, which is around the optimal probability of 23%. As a check, we also initialised the chains at several other values to ensure convergence had been attained.

We can now make the main comparison between the two methods of analyses, Empirical and Full Bayes. One way we can compare these is to look at the standard deviation values for m_j for each site $j = 1, \ldots, 56$. This can be seen in Figure 2.7.

We can see that all the points but two lie above the line of equality, which shows that the Full Bayes method is more variable. This is probably due to the parameters β_i and κ being assigned prior distributions. This is one of the reasons the Empirical Bayes method is used more frequently in road safety literature. Comparing standard deviation values for casualty frequency for the Fully Bayesian method, seen in Table 2.4, against Empirical Bayes

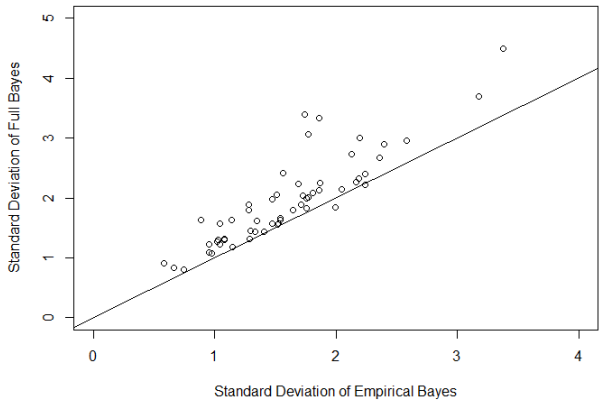

Figure 2.7: Standard deviationsof m_j for Empirical Bayes vs Full Bayes.

method, seen in Appendix, make it appear as though the MCMC method is less accurate. However it is more realistic as it takes more sources of variation into account. The Empirical Bayes method is over-optimistic in it's assessment of variability.

Now we will look again at the sites considered in the Empirical Bayes analysis, and see the shape of the prior and posterior distributions for the sites $j = 1, 5, 10, 19$. These can be seen in Figure 2.8.

Concentrating first on site $j = 1$, we can see the very large 'before' value which is very far from the prior distribution. The posterior is a weighted average of these two, so we see the mean value lies roughly in the middle. The posterior has a larger variance than the prior, due to the vague priors given to the regression coefficients. The 'after' value is much lower than the mean value of the posterior, showing that the speed cameras have had an effect, and the reduction in casualty rate is not solely due to regression to the mean.

For site $j = 5$ we can see the posterior is once again in between the prior and the 'before' value. However at this site the 'after' value is in the area we would expect the posterior mean to be - suggesting that the speed cameras have had little or no effect on decreasing casualty frequency.

Once again we see site $j = 10$ is very unusual. The gap between the 'before' and the prior is only very small, due to a low casualty rate, so the posterior is very similar to the prior, but it has a higher variance. The 'after' value is very high, and is very close to the end of the upper tail of the posterior distribution, which suggests the speed camera has had a negative effect on the casualty frequency.

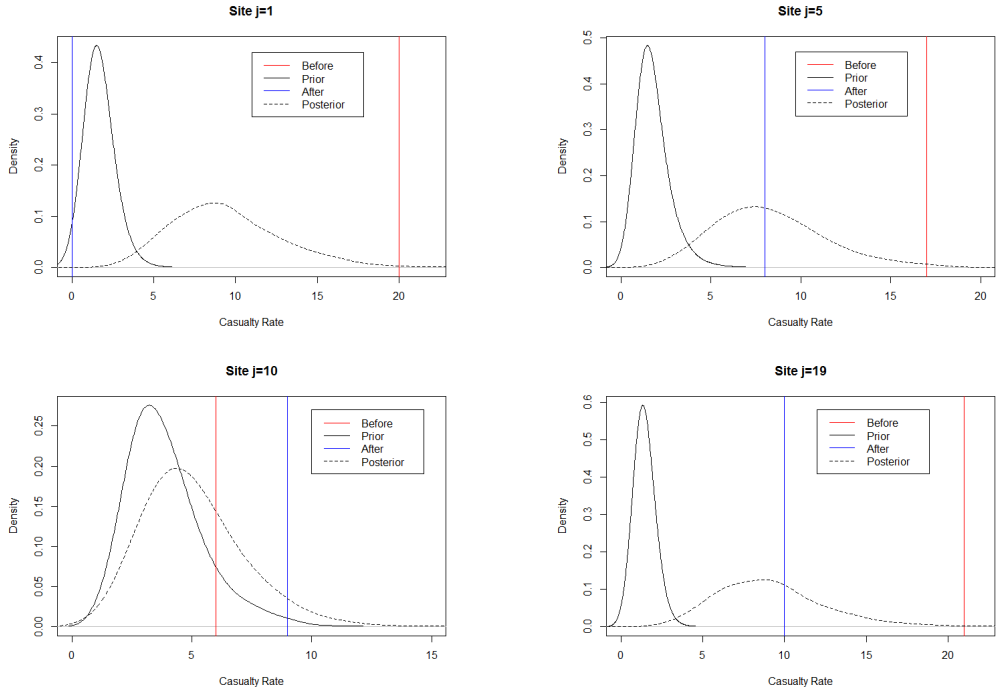

Figure 2.8: Prior and Posterior densities, with 'before' and 'after' values at sites $j = 1, 5, 10, 19$.

The posterior for site $j = 19$ is very low in comparison to the posterior, showing a much larger variance. The 'after' value also lies slightly above the posterior mean, suggesting that, as with site $j = 10$, the speed cameras has had a negative effect on the accident count.
We can compare the values in the 'after' period with the 95% credible intervals for m_j given in Table 2.4. The 'after' value for site $j = 1$ is not included in the confidence interval, as it is too low, suggesting we would not normally see such low casualty frequencies at that site. This could show that the speed cameras are doing a very effective job.

We can see from Table 2.4 that the mean values for the regression coefficients are roughly the same as the maximum likelihood estimates obtained previously; however, they now have larger variability. We can see that none of these intervals contain zero, which suggests again that all the explanatory variables used are important.

One good thing about using MCMC methods is the ability to look at any statistic of interest via inspection of the posterior, for example the median. This could be used as more accurate judgement of the casualty frequency due to the positively skewed shapes of the posterior distributions. However

Table 2.4: Full Bayes results for β_i, γ and sites $j = 1, 5, 10, 19$

		Mean	St.Dev.	Median	95% Credible Interval
	β_0	1.933	0.477	1.935	(1.032, 2.888)
	β_1	-0.043	0.014	-0.043	(-0.071, -0.014)
	β_2	-0.013	0.004	-0.013	(-0.021, -0.004)
	β_3	0.474	0.215	0.475	(0.055, 0.903)
	β_4	0.708	0.435	0.693	(-0.089, 1.597)
	β_5	0.907	0.458	0.887	(0.059, 1.858)
	β_6	1.118	0.404	1.110	(0.374, 1.954)
	$\gamma = e^{-\rho}$	2.197	0.751	2.066	(1.172, 3.990)
μ_j	$j = 1$	1.673	0.692	1.535	(0.727, 3.399)
	$j = 5$	1.778	0.843	1.586	(0.687, 3.934)
	$j = 10$	3.906	1.629	3.620	(1.493, 7.964)
	$j = 19$	1.442	0.505	1.363	(0.712, 2.601)
m_j	$j = 1$	9.466	3.337	9.106	(4.069, 16.816)
	$j = 5$	8.324	3.063	7.963	(3.290, 15.184)
	$j = 10$	5.058	2.048	4.752	(1.936, 9.823)
	$j = 19$	9.220	3.399	8.782	(3.969, 17.313)
	Total	326	33.217	324	(237.284, 362.955)

from Table 2.4 we can see there is very little difference in the total values.

2.4.2 Sensitivity of Results to choice of Prior for m_j

Now we have completed the MCMC for the Poisson-Gamma structure, we can repeat the procedure for a non conjugate prior specification. Two distributions we have chosen to look at are the lognormal and Weibull distributions. To allow some degree of fair comparison of results between the three different priors, we choose the lognormal and Weibull priors such that the prior means and variances are the same as those used in the Gamma prior.

Looking first at the lognormal distribution, we define the location and scale parameters as λ_j and σ^2 respectively. We need to solve two simultaneous equations, one for the mean value and one for the variance. These are given by

$$\frac{\mu_j^2}{\gamma} = \Big(e^{\sigma^2} - 1\Big)e^{2\lambda_j + \sigma^2} \quad \text{and} \quad \mu_j = e^{\lambda_j + {\sigma^2}/{2}}. \tag{2.14}$$

Solving these, we obtain

$$\sigma^2 = \log(1 + \gamma^{-1}), \tag{2.15}$$

$$\lambda_j = \log(\mu_j) - \frac{1}{2}\log(1 + \gamma^{-1}). \tag{2.16}$$

This gives us a lognormal distribution with mean $= \lambda_j$, and variance $= \sigma^2$. We can now perform a MCMC analysis as described in Section 1.3.3 using this as our prior for m_j in Equation (2.4).

For the Weibull prior distribution, we introduce the scale and shape parameters ω and ν_j. Again we solve the simultaneous equations given by

$$\omega \frac{\Gamma(2\omega^{-1})}{\Gamma^2(\omega^{-1})} = \frac{1}{2}(1+\gamma^{-1}) \quad \text{and} \quad \nu_j = \frac{\mu_j}{\Gamma(1+\omega^{-1})}. \tag{2.17}$$

The first of these cannot be calculated analytically, so must be obtained via the 'uniroot' command in R. We can now perform a MCMC analysis for this non-conjugate prior specification. Both the lognormal and Weibull priors for m_j produce MCMC output similar to Figure 2.6, which is not shown here.

To compare the choices of prior distribution we consider graphs similar to Figure 2.8, and then compare the differences in the posterior distributions. These can be seen in Figure 2.9 and 2.10 for lognormal and Weibull respectively.

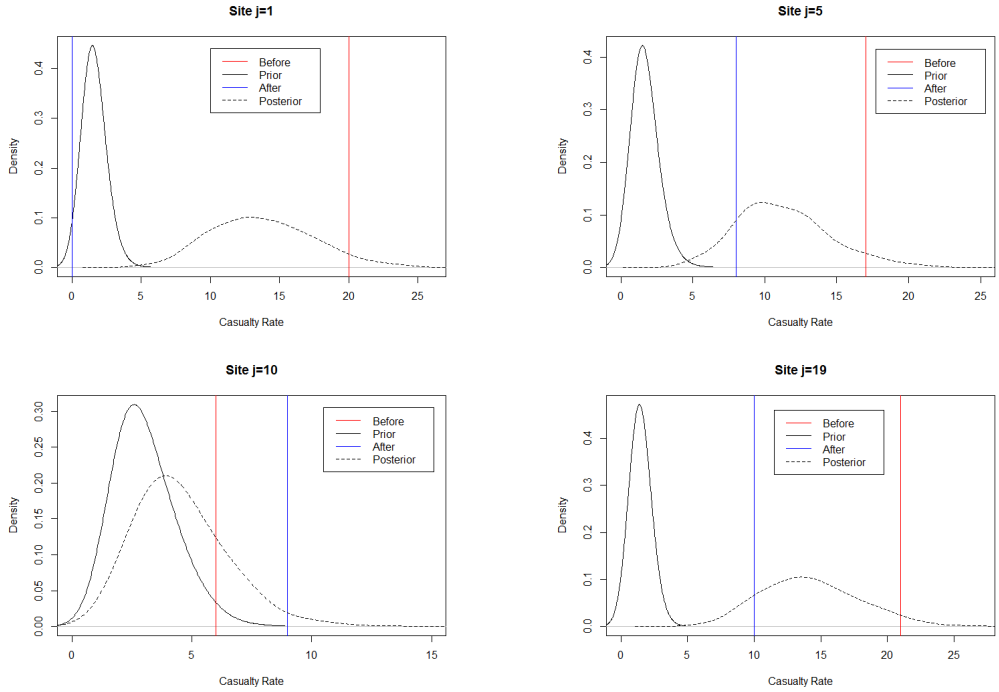

Figure 2.9: lognormal Prior and Posterior distributions, with 'before' and 'after' values at sites $j = 1, 5, 10, 19$.

Comparing Figures 2.8 and 2.9, we can see that for each site the lognormal

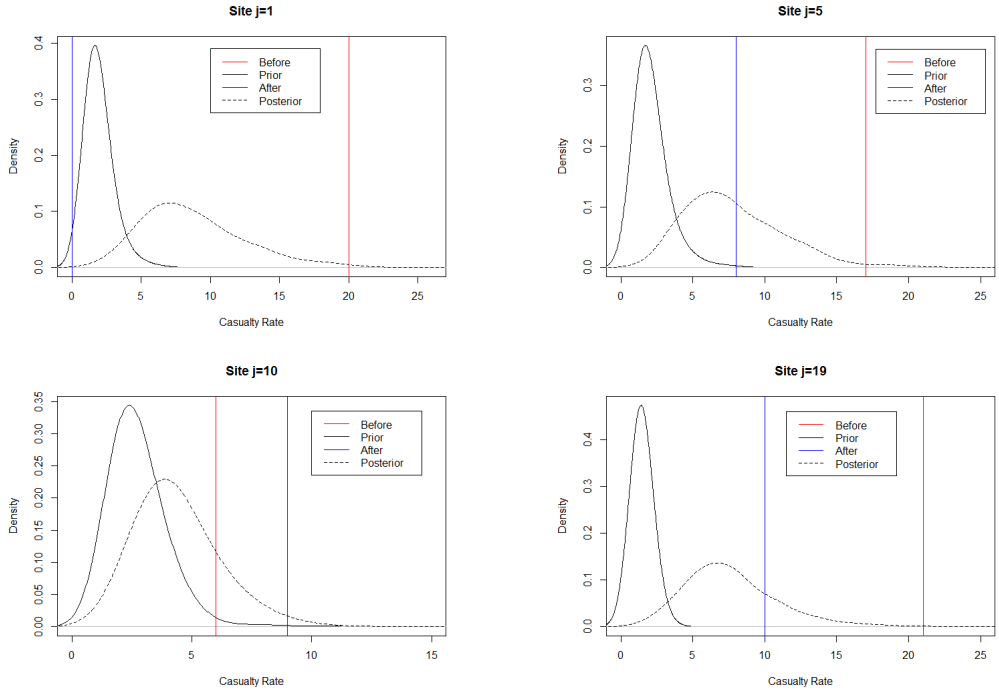

Figure 2.10: Weibull Prior and Posterior distributions, with 'before' and 'after' values at sites $j = 1, 5, 10, 19$.

prior distribution is over a narrower range than that of the Gamma, which results in a higher density value at the peak of the distribution. The priors look to be slightly positively skewed, and they take lower values in general in comparison to the Gamma densities. The resulting posterior distributions are quite different to the Gamma equivalents. They are all shifted to the right, suggesting higher casualty frequencies. The results of this would mean that the speed cameras have had a greater effect, as the 'after' values look lower compared to the posterior means.

Comparing Figures 2.8 and 2.10, we can see the prior densities for sites $j = 1, 5, 19$ are more variable, leading to a lower maximum density value. However for site $j = 10$ the Weibull prior is much narrower than the Gamma prior, indicating it is more certain of the expected value at that site. The resulting posteriors have lower mean values for the casualty rate, however are much more positively skewed than the Gamma posteriors. This suggests looking at a statistic such as the posterior median could be more useful for the Weibull distribution. This shows low casualty values are normal, and high casualty values are rare, but still expected, which seems sensible. The lower values for posterior means however makes it look as though speed cameras

havent had a positive effect at these sites, as all 'after' values are higher than expected values, expect for at site $j = 1$.

We can produce a table similar to Table 2.4 for m_j for each of the new priors. The comparison of sites $j = 1, 5, 10, 19$ can be seen in Table 2.5.

Table 2.5: Full Bayes results for m_j for Gamma, lognormal, Weibull priors

		Mean	St.Dev.	Median	95% Credible Interval
Gamma m_j	$j = 1$	9.466	3.337	9.106	(4.069, 16.816)
	$j = 5$	8.324	3.063	7.963	(3.290, 15.184)
	$j = 10$	5.058	2.048	4.752	(1.936, 9.823)
	$j = 19$	9.220	3.399	8.782	(3.969, 17.313)
	Total	326	33.217	324	(237.284, 362.955)
lognormal m_j	$j = 1$	13.989	3.765	13.647	(7.599, 21.983)
	$j = 5$	11.473	3.466	11.053	(5.751, 18.659)
	$j = 10$	4.730	1.876	4.457	(1.930, 9.211)
	$j = 19$	14.444	3.760	14.064	(7.667, 22.824)
	Total	361	28.096	351	(286.920, 385.645)
Weibull m_j	$j = 1$	8.769	3.725	8.071	(3.230, 18.130)
	$j = 5$	7.846	3.470	7.281	(2.827, 15.769)
	$j = 10$	4.391	1.696	4.126	(1.841, 8.396)
	$j = 19$	7.703	3.349	7.215	(2.607, 15.732)
	Total	334	33.474	326	(249.034, 375.845)

We can see that values for m_j assuming the lognormal priors are larger than those using the Gamma or Weibull priors for each of the mean and the median. This leads us to believe the lognormal distribution is overestimating the number of casualties that could occur at these sites. The posteriors for m_j using Gamma and Weibull priors are very similar. The lognormal 95% credible interval is also much narrower than the other priors.

Table 2.6 compiles some of the important results from each of the methods used. We can see immediately that the biggest advantage of a Full Bayes method is the fact we have access to the posterior via the MCMC, so any summaries are easily obtained directly from the sample. Comparing the three different prior distributions we can see the Gamma and the Weibull produce very similar values, particularly for the median value, which we mentioned might be more appropriate for the analyses using Weibull priors. The lognormal distribution gives much higher values, and the confidence intervals are narrower.

The Deviance Information Criterion (DIC) values can be used as an indicator

Table 2.6: Comparison of Empirical and Full Bayes methods, with different prior distributions.

	Prior	Mean	Median	95% CI	DIC
Empirical Bayes	Gamma	323	-	-	-
Full Bayes	Gamma	326	324	(237.3, 363.0)	692.4
	lognormal	361	351	(286.9, 385.6)	785.6
	Weibull	334	326	(249.0, 375.8)	646.1

of how well the model fits the data, with lower values suggesting a better fit. We can see the lognormal DIC value is much higher than for the other priors, so we can say that the lognormal prior distribution is probably the least appropriate of these three to fit to the data. The Weibull distribution has the lowest DIC value, and so we can say that this is possibly the best fit to the data.

2.5 Discussion

We have discussed two main methods of performing Bayesian inference on the mean casualty rate at sites treated with speed cameras in the Northumbria Police Force area. The results have shown that the best method to do this is a Full Bayes analysis, which uses MCMC methods. This has the flexibility of using non-conjugate prior distributions, and we have looked at two alternatives to the Gamma prior that was initially used: the lognormal and Weibull priors. The use of this method, however, does lead to wide confidence intervals, as we have also given the β_i and κ vague prior distributions. The use of a lognormal prior seemed inappropriate for this analysis due to the large DIC value found, suggesting it was not a good model for the data. In conclusion the most appropriate prior distribution to use is the Weibull or Gamma, as they produce very similar results.

The results we have found seem realistic and accurate, however we have not accounted for trend. This could be a problem, as in recent years there has been a general decline in traffic accidents due to increased numbers of safety schemes and more information about the dangers of speeding. This is something that will be taken into account in the Halle analysis, seen in the next chapter.

Chapter 3

Hotspot identification: A pre-emptive study

3.1 Background and Data

The data in this case study were provided by PTV, a traffic accident mapping software company based in Germany. Researchers in the School of Mathematics and Statistics and the Transport Operations Research Group at Newcastle University are currently working with PTV's Research & Development Team to implement a Fully Bayesian approach to road traffic hotspot identification in their software. Some of the methodology is based on work in Fawcett & Thorpe (2013) [8]. PTV wish to exploit the Bayesian posterior predictive distribution to predict where the 'hotspots' will occur in future years. This is a pro-active response, where speed cameras or safety measures would be put in place before the accidents have risen to a high threshold based on the predictive distribution. This method eliminates the cost of unnecessary safety schemes that might appear necessary, but in fact are not, possibly due to trend and RTM in accident counts. Similarly, sites which, in a particular year, might appear safe could have accident counts migrating upwards towards some underlying mean value - RTM 'in reverse', compared to the effects of RTM in the analyses in Chapter 2.

We have data for 734 sites over the course of nine years, from 2004 to 2012 inclusive, as full records from 2013 are not yet available. These sites are from in and around the city of Halle, which have some of the worst road traffic accident rates in Germany. We will use these data to build an accident prediction model (APM) to use as our prior mean μ_j for accident rates at site j. The data included accident counts for each of the nine years (as opposed to casualty counts studied in Chapter 2), and eight explanatory variables,

which were: traffic volume, urban (yes/no), intersection (yes/no), signalised (yes/no), speed limit, major road (yes/no), major intersection (yes/no) and four legs (yes/no). Volume is the only continuous variable; all the others, excluding speed limit, are indicator variables. Speed limit has six categories (80, 70, 60, 50, 45 or 30 km/h); however all other indicator variables have only two groupings (0=No or 1=Yes). We can look at some summary statistics and plots to investigate the data.

Table 3.1: Mean Accident Values 2004 to 2012 for all 734 sites

Year	Mean	Median	St. Dev.	LQ	UQ	Min	Max
2004	3.65	2	4.63	0	17.35	0	29
2005	3.73	2	4.95	0	19.00	0	35
2006	3.57	2	5.07	0	19.00	0	38
2007	3.71	2	4.98	0	18.00	0	52
2008	3.55	2	4.51	0	16.67	0	29
2009	3.64	2	4.80	0	17.00	0	39
2010	3.29	2	4.40	0	14.67	0	41
2011	3.11	2	4.40	0	15.67	0	48
2012	2.97	1	4.57	0	15.67	0	46

We can see from Table 3.1 that the mean accident rate over these nine years is generally declining. Due to this decreasing trend we will account for the year in our accident prediction model by allowing the time unit $1, 2, \ldots, 9$ to be a explanatory variable. We can see the standard deviation values are quite large, leading to wide 95% confidence intervals.

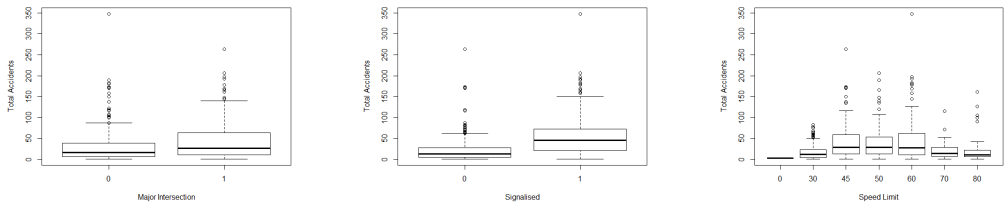

Figure 3.1: Boxplots of total accident value against: Left: Major Intersection, Middle: Signalized, Right: Speed Limit.

In Figure 3.1 we can see some of the indicator variables plotted against total accident count. On the left we can see the variable major intersection. We can see both groups have similar median values, and they are both positively skewed, with many identified outliers. The highest outlier is for a site which

is not a major intersection, suggesting smaller roads may be less likely to have high accident counts. However, the two groups look quite similar so we cannot draw any conclusions regarding the usefulness of this variable as a predictor.

In the middle we can see the signalized groups against total accident value. The sites that are signalised (1) clearly have a much larger variability in accident count, and they have a larger median than the non-signalised sites. This variable also shows positive skew for both groups, with outliers at the top of the accident range. The largest outlier is at a signalised site, suggesting the traffic signal could be causing accidents, or this could be due to the fact traffic signals are usually in place in larger junctions with a higher volume of cars.

On the right we have the six different groups of speed limit values (ignoring the zero value as this site cannot be classified, perhaps due to a variable speed limit) against total accident count. We can see that the median values are highest at speeds 45 to 60 km/h, suggesting most accidents occur on these roads. These categories also have the highest values of outliers and are the most positively skewed, once again suggesting that these could be the most dangerous sites.

We can now use the information at these sites to form our accident prediction model, giving us the prior mean μ_j for accident rates at site j.

3.2 Fully Bayesian Analysis

In Chapter 2 we highlighted the shortcomings of an Empirical Bayes analysis, with posterior estimates of casualty frequency being over-optimistic in terms of their variability. We also illustrated the possibility of using non-conjugate prior specifications within a fully Bayesian analysis, one of which was shown to be superior to the standard conjugate case. With this in mind, we will consider a fully Bayesian analysis of the Halle dataset.

3.2.1 Initial investigations: The Accident Prediction Model

We need to discover which of our explanatory variables are significant predictors of accident count. Most of these variables are indicator variables, one of which has six categories. We will partition the values for speed limit into three new groups: 30, 45-50, 60+. This gives us roughly even groups and means we only need to create two new indicator variables. As suggested by

researchers at PTV, we work with log(volume). After doing this, we perform backwards elimination, which gives a log-linear model given by

$$\hat{\mu} = \exp\{-0.91 - 0.02x_1 + 0.01x_2 + 0.64x_{I3} + 1.10x_{I4} + 0.51x_{I5} + 1.94x_{I6} + 0.29x_{I7} + 0.52x_{I8} - 1.60x_{I9} - 1.67x_{I10}\}, \tag{3.1}$$

where $\hat{\mu}$ denotes the expected accident value, and I denotes an indicator variable. The x_i correspond to the variables given below in order, with x_{I9} and x_{I10} corresponding to the speed indicator variables, given by

$$x_{I9} = \begin{cases} 0 & \text{if speed limit} = 30,\ 60+ \\ 1 & \text{if speed limit} = 45\text{-}50. \end{cases} \tag{3.2}$$

$$x_{I10} = \begin{cases} 0 & \text{if speed limit} = 30,\ 45\text{-}50 \\ 1 & \text{if speed limit} = 60+. \end{cases} \tag{3.3}$$

```
Call:
glm.nb(formula = accidents ~ time + logvolume + urban + intersection +
    signalized + majorroad + majorintersection + fourlegs + speed1 +
    speed2, init.theta = 1.326735172, link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4628  -1.0783  -0.3831   0.3451   3.7792
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.905747   0.078026 -11.608  < 2e-16 ***
time              -0.028152   0.005092  -5.528 3.23e-08 ***
logvolume          0.004962   0.001587   3.126  0.00177 **
urban              0.640265   0.054375  11.775  < 2e-16 ***
intersection       1.100730   0.046387  23.729  < 2e-16 ***
signalized         0.512033   0.032654  15.680  < 2e-16 ***
majorroad          1.942131   0.645278   3.010  0.00261 **
majorintersection  0.294036   0.037121   7.921 2.36e-15 ***
fourlegs           0.519212   0.030793  16.861  < 2e-16 ***
speed1            -1.603662   0.644929  -2.487  0.01290 *
speed2            -1.667385   0.643822  -2.590  0.00960 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
(Dispersion parameter for Negative Binomial(1.3267) family taken to be 1)
    Null deviance: 10011  on 6605  degrees of freedom
Residual deviance:  7286  on 6595  degrees of freedom
AIC: 29094
Number of Fisher Scoring iterations: 1
              Theta:  1.3267
          Std. Err.:  0.0370
 2 x log-likelihood:  -29070.4260
```

We can see that all the variables are significant, so we include them all in our model. The coefficient signs from Equation 3.1 can show us what makes a road dangerous. The coefficients for Intersection and Major road are both large and positive, showing that more accidents would typically happen at a

site that which had either of these criteria. The large negative values for the speed indicator variables suggest that as the speed increases the accident rate decreases. This could be due to the lack of pedestrians on motorways, or due to the fact that these faster roads would generally not have as many traffic signals or intersections. We can see the significant negative trend through time, showing we cannot assume a constant value for μ for each site. There could be some significant interactions, however to keep the model relatively simple, we have not accounted for these.

This regression gives us maximum likelihood estimates for the $\hat{\beta}_i$ for $i = 0, \ldots, 10$, and for $\hat{\gamma}$, the negative Binomial dispersion parameter. In this case study

$$\hat{\gamma} = \frac{1}{\hat{\kappa}} = 1.327.$$

Although we are performing a fully Bayes analysis, we will use these maximum likelihood estimates as starting values in our MCMC, to reduce any burn-in period.

3.2.2 Fully Bayes Setup

As with the work in Chapter 2, we assume a Poisson distribution for accidents at each site j with rate m_j. Although the use of MCMC means we are able to use a non-conjugate prior distribution for m_j, we will initally use a Gamma prior with mean μ_j, μ_j itself being defned as a log-linear function of explanatory variables with regression coefficients and negative Binomial dispersion parameters having the following uninformative, independant priors:

$$\beta_i \sim N(0, 100), \tag{3.4}$$
$$\rho = \log(\kappa) \sim N(0, 100), \tag{3.5}$$

as before. Here $i = 0, \ldots, 10$, and again we use ρ to retain the positivity of κ. Unlike the analysis in Chapter 2, we do not have a set of reference sites to build the APM; here we apply the regression to all sites and estimate μ_j by using the fitted values from the APM. The estimated coefficients in Equation (3.1) will be used as initial values in our MCMC. As seen previously in Section 2.4.1, these are vague prior distributions, as we know very little about what values these variables could take. We will run the MCMC for 10,000 iterations.

3.2.3 Results

We can look at some of the trace plots and densities for μ_j for sites $j = 2, 286, 306, 308, 421, 650$ in Figure 3.2. These sites have been chosen as examples due to interesting characteristics in the accident counts. We can see the trace plots are oscillating around the initial values, showing the lack of a 'burn-in period due to initialising the chain at the maximum likelihood estimates. We can see the densities look roughly Normal, without much skew. The acceptance probabilities for all the variables were in the region 20% to 40%.

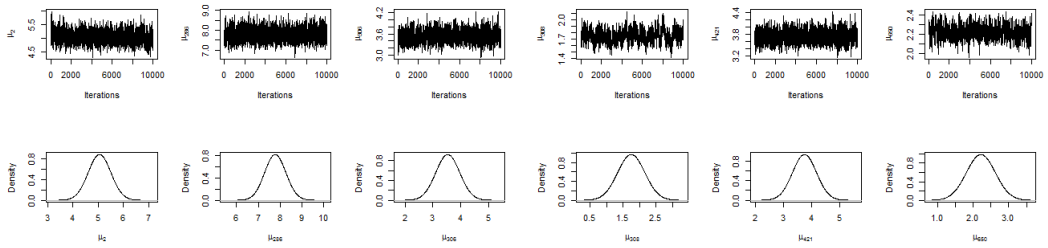

Figure 3.2: Trace plots and densities for μ_j for $j = 2, 286, 306, 308, 421, 650$

In Figure 3.3 we can see observed accident counts, expected accident value μ_j, and the posterior mean, with 95% credible intervals, for six sites over nine years.

We have chosen to look at site $j = 2$ as this was one of the sites with close to the mean total accident value (31.2 accidents) over the nine years. We can see, however, that this site has large accident counts in 2004 and 2005, but in later years appears to be a relatively safe site, with less than four accidents each year. Initially the observed values are very far away from the expected value μ, however from 2010 to 2012 these values are nearly identical, and from 2006 both of these values are encompassed within the 95% confidence interval of the posterior distribution. RTM seems to be affecting this site, with the peak of the 'blip' occurring in 2004 and 2005. We can see that the posterior is a weighted sum of what we expect to see and what we observed, as the posterior mean is between these two values.

Site $j = 286$ is one of the sites with the minimum number of accidents in the nine years (1 accident). We can see the observed values are lower than the values for μ which is very unusual. This suggests that regression to the mean would work in the opposite way to what we have seen so far - this site is seeing an unnaturally low number of accidents, and in the next few years we

would expect it to increase. Both the observed values and μ are contained within the 95% confidence intervals, however the posterior mean is closer to the observed values, suggesting this information is more heavily weighted.

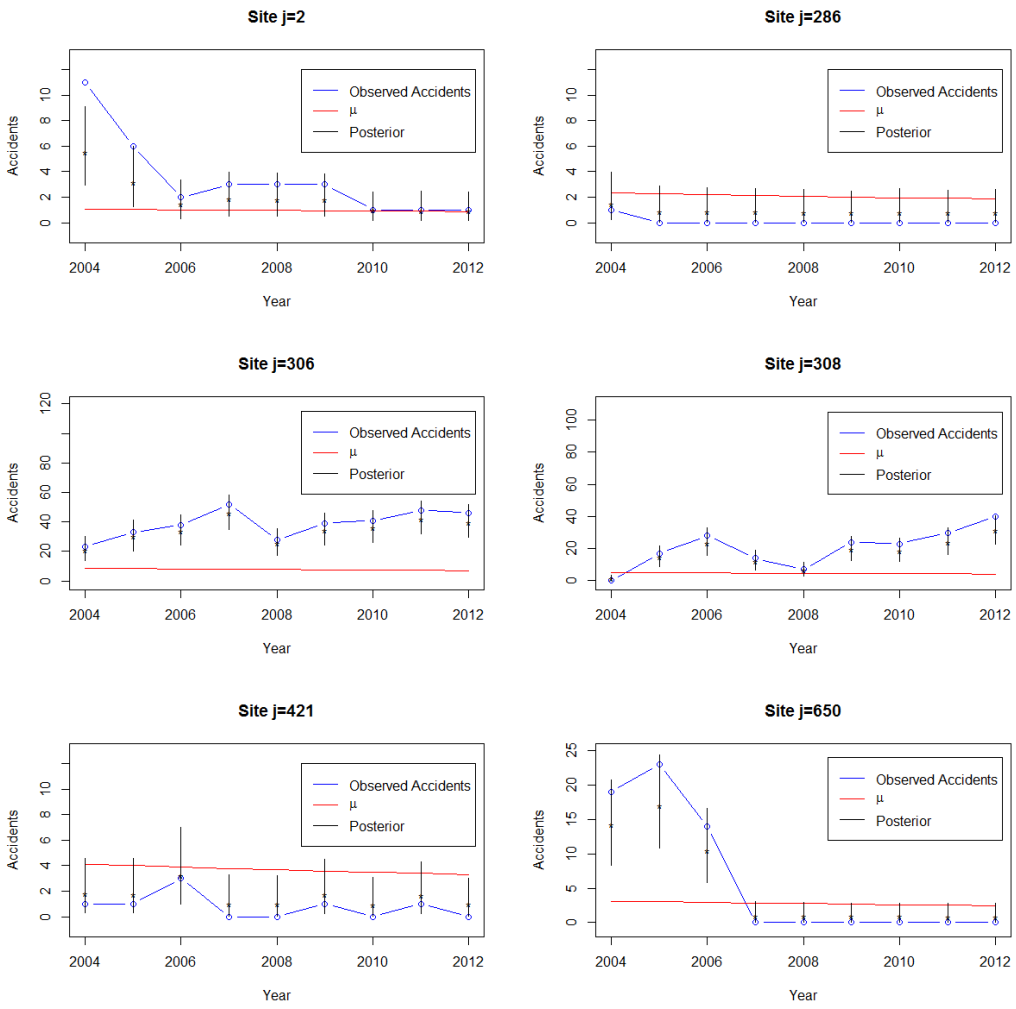

Figure 3.3: Plots showing observed accident values, μ_j, posterior mean and posterior 95% confidence intervals for sites $j = 2, 286, 306, 308, 421, 650$.

We have considered site $j = 306$ as this is the site with the highest accident value (348 accidents). The value for μ suggests we ought to see around 10 accidents there each year, however the observed value is always over twice that. The accident count seems to be increasing with time, with only a slight decrease in 2008. The posterior means and confidence intervals are very close to the observed values, showing the expected value at this site doesn't carry

much weight towards the posterior mean.

The accident value at site $j = 308$ has increased the most over the nine years (increase of 40). In 2004 the accident count was lower than expected, but rises steadily, with a decrease in 2008. These two years are the only ones that the posterior confidence intervals includes both the observed value and μ. The observed values are at the very top end of the confidence intervals, and look to be still increasing.

Site $j = 421$ has close to the average accident change from 2004 to 2012 (0.68 accidents). This particular site has a very low accident count in all years. Similarly to site $j = 286$ we can see the value for μ is higher than the observed value, so we expect the accident value to increase over the next few years, as it regresses to what we expect to see at a site similar to this. The posterior mean values are very close to the observed values, however the confidence intervals seem to be skewed towards larger accident values. The intervals usually include both the observed and expected values.

The last site we consider here is site $j = 650$, which has the largest decrease in accident values over the nine years (decrease of 19). Here we can see the large accident count in the years 2004 to 2006. These values are very far from the values of μ, which means we expect much lower values. Indeed in 2007 onwards the accident value drops below expected, suggesting regression to the mean has already come into effect, with the 'blip occurring until 2006. We can see that had a speed camera been put in at this time that it would have been a waste of money, as the accident count hugely decreased naturally. After 2007 the values for μ and the observed values are both encompassed in the posterior confidence intervals, showing these low values are expected at this site.

We can see that for all sites the value for μ is decreasing, which is due to the declining trend in accidents over time.

In Table 3.2 we can see some summary statistics for the coefficients of the explanatory variables, and the values for μ and m at site $j = 2$ over the nine years. We can see that the credible intervals for all β_i dont include the value zero, showing again that all the explanatory variables are significant. Due to space we have just considered one site, which can be seen in Figure 3.3. The decreasing value of μ can be seen, and we also notice the posterior standard deviation decreases over the nine years. The median values are very similar to the mean, which we observed in the density plots in Figure 3.2 due to the lack of skew. For the value of m we can see a large decrease in the mean value over the first few years, which we also notice in the values for standard deviation. The median values for this site are consistently lower than the

mean, suggesting the posterior distribution has a slightly positively skewed distribution. The total mean accident count is close to 26,000, however due to the large posterior variability, the confidence interval has a range of around 500 accidents.

Table 3.2: Full Bayes results for β_i, γ with μ_j, m_j for site $j = 2$

		Mean	St.Dev.	Median	95% Credible Interval
	β_0	-0.922	0.076	-0.920	(-1.082, -0.786)
	β_1	-0.028	0.005	-0.028	(-0.038, -0.017)
	β_2	0.005	0.002	0.005	(0.002, 0.008)
	β_3	0.646	0.051	0.647	(0.542, 0.741)
	β_4	1.108	0.046	1.105	(1.014, 1.199)
	β_5	0.512	0.033	0.512	(0.447, 0.578)
	β_6	1.366	0.321	1.294	(0.888, 1.981)
	β_7	0.294	0.037	0.294	(0.225, 0.370)
	β_8	0.521	0.030	0.521	(0.462, 0.579)
	β_9	-1.027	0.320	-0.954	(-1.654, -0.553)
	β_{10}	-1.089	0.319	-1.016	(-1.722, -0.609)
	$\gamma = e^{-\rho}$	1.325	0.047	1.324	(1.256, 1.399)
μ_2	2004	1.072	0.057	1.069	(0.965, 1.152)
	2005	1.043	0.053	1.041	(0.943, 1.152)
	2006	1.014	0.050	1.012	(0.920, 1.118)
	2007	0.986	0.048	0.985	(0.896, 1.087)
	2008	0.960	0.046	0.959	(0.872, 1.055)
	2009	0.933	0.045	0.934	(0.848, 1.025)
	2010	0.908	0.045	0.908	(0.824, 0.999)
	2011	0.883	0.045	0.883	(0.799, 0.973)
	2012	0.859	0.045	0.859	(0.774, 0.950)
m_2	2004	5.526	1.635	5.316	(2.920, 9.108)
	2005	3.173	1.225	3.042	(1.273, 5.970)
	2006	1.469	0.931	1.303	(0.319, 3.383)
	2007	1.876	0.924	1.741	(0.531, 3.985)
	2008	1.824	0.896	1.697	(0.530, 3.919)
	2009	1.826	0.908	1.679	(0.504, 3.919)
	2010	0.960	0.711	0.812	(0.157, 2.507)
	2011	0.943	0.679	0.797	(0.153, 2.507)
	2012	0.906	0.635	0.769	(0.143, 2.402)
m_j	Total	26388	1415.48	26354	(26090, 26633)

We will now use try to achieve similar results using a non-conjugate distribution to explain the data. This is easily done due to the use of MCMC in this chapter.

3.2.4 Sensitivity of Results to choice of Prior for m_j

Now we have completed our MCMC procedure for a conjugate prior, we can repeat this method for a non-conjugate prior specification. As in Chapter 2, to allow a like-for-like comparison, we choose the lognormal prior such that the prior mean and variance is the same as that used in the Gamma prior. This gives us the lognormal distribution with mean = λ_j and variance = σ^2, with these given by

$$\sigma^2 = \log(1 + \gamma^{-1}), \tag{3.6}$$

$$\lambda_j = \log(\mu_j) - \frac{1}{2}\log(1 + \gamma^{-1}). \tag{3.7}$$

We can now perform the MCMC analysis described in Section 1.3.3 using this as our prior for m_j. As the results in Table 3.3 show, posterior summaries for the mean accident rate at site $j = 2$ barely change when adopting a lognormal prior for m_j. In fact this was true for all other sites (results not shown here). DIC values for the Gamma and lognormal priors are 945.2 and 1023.4 respectively, showing that the analysis using the Gamma priors, as discussed in Section 3.2.2 to 3.2.3, would be preferred. Thus it is the Poisson-Gamma specification that will be carried forward into the next section.

Table 3.3: Full Bayes results for m_j for site $j = 2$ using lognormal prior

		Mean	St.Dev.	Median	95% Credible Interval
m_2	2004	5.821	1.800	5.563	(3.310, 9.501)
	2005	3.285	1.355	3.002	(1.299, 6.020)
	2006	1.525	1.021	1.442	(0.353, 3.452)
	2007	1.935	1.000	1.852	(0.550, 4.025)
	2008	1.905	0.950	1.805	(0.581, 4.050)
	2009	1.900	0.909	1.752	(0.600, 4.105)
	2010	1.052	0.755	0.959	(0.205, 2.582)
	2011	0.992	0.721	0.851	(0.195, 2.565)
	2012	0.925	0.681	0.890	(0.154, 2.438)
m_j	Total	26402	1583.55	26372	(26110, 26750)

3.3 Predicting accident values for 2013

It is the aim of clients of companies like PTV to predict future accident rates at potential accident hotspots. It is often the case that a local authority waits until an 'accident threshold' has been exceeded at a particular location before a road safety scheme is implemented at that site. However, it would

be preferable for such authorities to adopt a proactive, rather than reactive, approach to safety scheme implementation, acting before an accident threshold has been observed to prevent such high accidents. This is where the Bayesian posterior predictive distribution can be extremely useful. Suppose a future accident rate z_j at site j is Posson distributed with mean m_j, and our posterior, at the current time, is $m_j|y_j \sim \text{Ga}(\gamma + y_j, \gamma/\mu_j + 1)$. Then we can calculate the posterior predictive distribution to be

$$
\begin{aligned}
f(z_j|y_j) &= \int_{M_j} f(z_j|m_j)\pi(m_j|y_j)\ dm_j \qquad (3.8)\\
&= \int_0^\infty \frac{m_j^{z_j}\ e^{-m_j}}{z_j!}\ \frac{(\gamma/\mu_j+1)^{\gamma+y_j}\ m_j^{\gamma+y_j-1}\ e^{-(\gamma/\mu_j+1)m_j}}{\Gamma(\gamma+y_j)}\ dm_j\\
&= \frac{(\gamma/\mu_j+1)^{\gamma+y_j}}{\Gamma(\gamma+y_j)\ z_j!}\int_0^\infty m_j^{z_j+\gamma+y_j-1}\ e^{-m_j(\gamma/\mu_j+2)}\ dm_j\\
&= \frac{(\gamma/\mu_j+1)^{\gamma+y_j}}{\Gamma(\gamma+y_j)\ z_j!}\ \frac{\Gamma(z_j+\gamma+y_j)}{(\gamma/\mu_j+2)^{z_j+\gamma+y_j}}\\
&= \binom{z_j+\gamma+y_j-1}{\gamma+y_j}\left(\frac{\gamma/\mu_j+1}{\gamma/\mu_j+2}\right)^{\gamma+y_j}\left(1-\frac{\gamma/\mu_j+1}{\gamma/\mu_j+2}\right)^{z_j}, \qquad (3.9)
\end{aligned}
$$

where z_j is a predicted future value. In fact, this is negative Binomial distribution with size $= \gamma + y_j$ and probability $= {}^{\gamma/\mu_j+1}/_{\gamma/\mu_j+2}$, we can estimate the probabilities of observing future accident value $z_j = c$, for $c = 0, 1, 2, \ldots,$ in the next year.

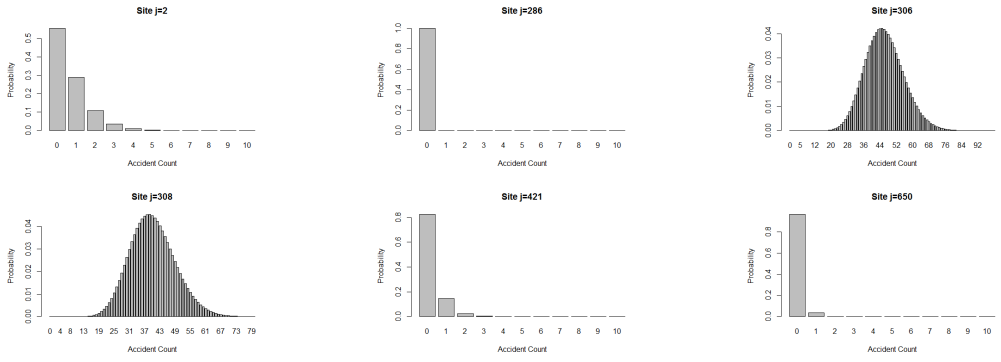

Figure 3.4: Posterior predictive probability distributions for future accident counts (in the year 2013) for sites $j = 2, 286, 306, 308, 421, 650$.

This can be seen in Figure 3.4. We can see for sites $j = 2, 286, 421$ and 650 that the most likely accident value to occur in 2013 is zero. Sites 306 and 308

however look likely to have accident counts of around 50 and 40, respectively. These were the sites with highest overall accidents and biggest increase in accident value, so these sites were quite likely to have high accident values in the future.

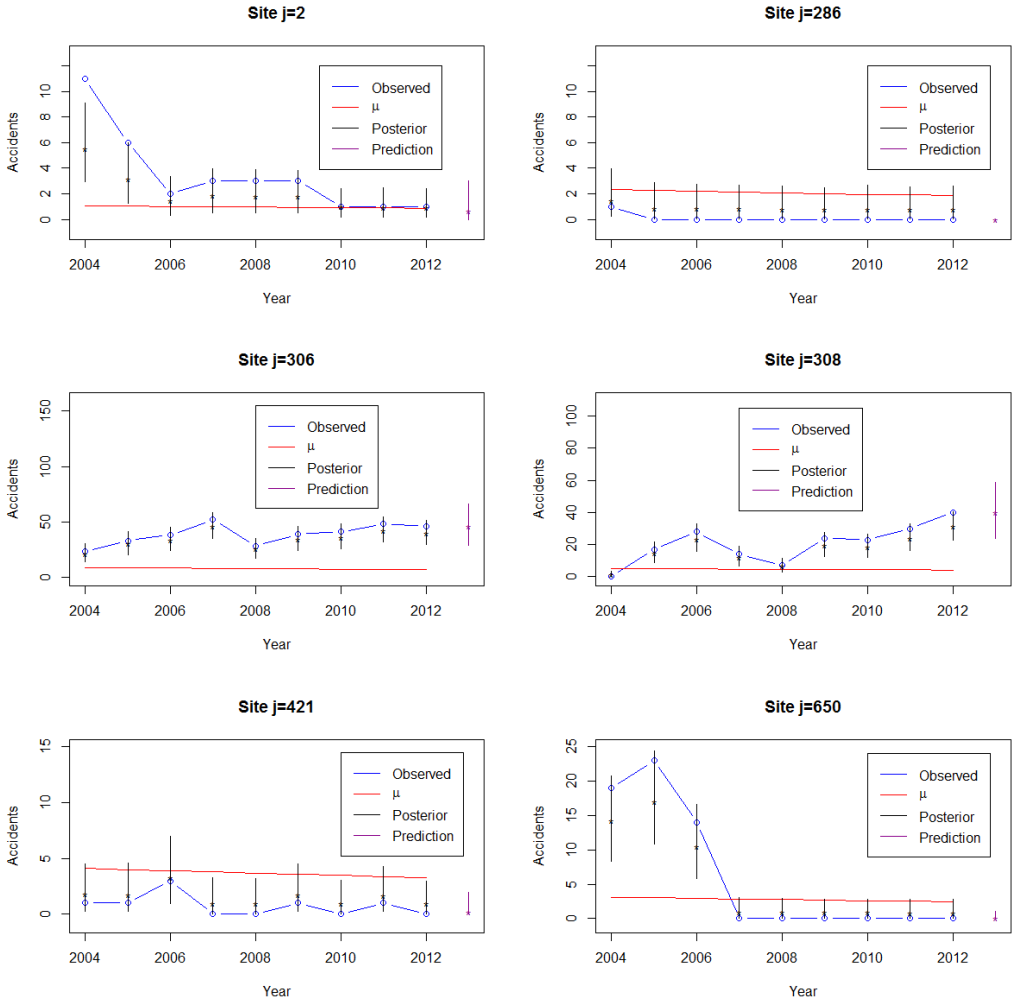

Figure 3.5: Plots showing observed accident values, μ_j, posterior mean, posterior 95% confidence intervals, posterior predictive mean, and posterior predictive 95% confidence intervals for sites $j = 2, 286, 306, 308, 421, 650$.

In Figure 3.5 we can see the posterior predictive mean, with the 95% credible intervals, added to the plots previously seen in Figure 3.3. We can see very low predictive accident counts for sites $j = 2, 286, 421$ and 650, with very narrow confidence intervals. Sites 306 and 308 have predicted mean accident

values that are very similar to the values in the previous year, suggesting a lack of RTM. The predicted accident counts are much higher than all previous values for μ, suggesting these sites have an unusually high accident value compared to other similar sites. These sites also have much wider posterior predictive confidence intervals, suggesting the accident count in the next year could be very variable. Local authorities could use such information to direct future funding of road safety schemes. For example, the plots in Figure 3.4 could be used to determine the predictive probabilites of exceeding a certain accident threshold; if such a probability is deemed high enough, a road safety scheme could be implemented before a high accident count is observed - potentially saving accidents/lives.

3.4 Discussion

In this section we have used MCMC methods to perform Bayesian inference on the accident count at 734 sites in and around Halle in Germany. This method has the advantage of using non-conjugate prior specifications, however, we have found that, via the DIC, the most appropriate is the conjugate Gamma prior. This analysis accounts for trend, and we have found that the average accident rate decreases over time. We have also used the Bayesian posterior predictive interval to predict the accident counts in future years, to allow organistions using PTV software to make judgement on which sites should have a road safety scheme implemented.

Chapter 4

Conclusion

In this report we have considered two case studies; one investigating the effects of speed cameras on road traffic casualties in the Northumbria Police Force area in the UK, and the other attempting to identify road traffic accident hotspots in and around the German City of Halle using the Bayesian posterior predictive distribution. Both studies attempt to filter out the effects of regression to the mean by combining raw casualty/accident counts with carefully elicited prior distributions for the mean casualty/accident rates; the second study also attempts to incorporate trend. Both of these studies used an accident prediction model, which estimates what we expect to observe at these sites when not under the influence of RTM.

In the Northumbria study we have two sets of data; a control set, which had not had speed cameras deployed to, which we used to estimate an accident prediction model; and one treated set, for which we have data from two time periods, both spanning two years: the before and after period, with speed cameras implemented for the 'after' period. We use the APM estimated from the control sites to predict the average casualty value in the 'before' period at the treated sites. We then performed an Empirical Bayes analysis, which is the 'gold-standard' used in the road traffic safety literature. We found that this method was over-optimistic in its assessment of variability of estimates of the mean casualty rate at the treated sites, and a Fully Bayesian analysis, using MCMC methods, was more appropriate. Using this procedure also leads to the possibility of using non-conjugate prior specifications, and we discovered that the most appropriate prior distribution for these data was the Weibull prior. The lognormal distribution was found to be the least appropriate for our analysis.

In the Halle case study we had one set of sites, data from which were used to create the APM. We had data over nine years, from 2004 to 2012 inclusive.

Due to the length of time over which the data were collected, we were able to account for trend. For this study we started with the Fully Bayesian procedure, with the conjugate Gamma prior distribution. We then considered an alternative prior: the lognormal distribution. This was found again to be inferior due to the large DIC value. The conjugacy of the Poisson-Gamma setup meant we could calculate the Bayesian posterior predictive distribution to be a negative Binomial. We can use this distribution to estimate future accident count values, from which local authorities can take a proactive response on the implementation of road safety schemes. We estimated the posterior predictive mean accident value for 2013, and used this to assess which sites could be potentially dangerous. This predictive distribution also gives us the probability of future accident counts at each site, so local authorities can implement a safety scheme if these values are over a certain threshold. This method has the potential to be used in many real life situations, and by using this procedure, accidents can be prevented, rather than fitting a safety scheme as a reactive response.

Chapter 5

Bibliography and Appendix

5.1 Bibliography

[**1**] R Core Team (2015). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[**2**] Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

[**3**] F.Galton, *Regression towards Mediocrity in Hereditary Stature*, Vol. 15 (1886), pp. 246-263

[**4**] Examples of regression to the mean:
http://rationalwiki.org/wiki/Regression_to_the_mean

[**5**] Examples of regression to the mean:
https://explorable.com/regression-to-the-mean

[**6**] Examples of regression to the mean:
https://www-users.york.ac.uk/ mb55/talks/regmean.htm

[**7**] Department for Transport, *Tomorrow's Roads - Safer for Everyone*, The Stationery Office, London, 2000.

[**8**] L.Fawcett and N.Thorpe, *Mobile safety cameras: estimating casualty reductions and the demand for secondary healthcare*, (2013)

[**9**] Introduction to Bayesian Statistics: MAS2317. Case Study 1

[**10**] Bayesian Inference: MAS3321

[**11**] Deviance Information Criteria:
http://onlinelibrary.wiley.com/doi/10.1111/rssb.12062/pdf

[**12**] S.P.Miaou and D.Lord, *Modelling traffic crash-flow relationships for intersections: Dispersion parameter, functional form, and Bayes versus empirical Bayes methods*, Transp. Res. Rec. 1840 (2003), pp. 1310-1340.

[**13**] Multivariate Data Analysis: MAS3325

[**14**] M.J. Maher and L.J. Mountain, *The sensitivity of estimates of regression to the mean*, Acci. Anal. Prev. 41 (2009), pp. 861-868.

[**15**] Department for Transport, *Reports Road Casualties in Great Britain 2011*, The Stationery Office, London, 2012.

5.2 Appendix

Table 5.1: Empirical Bayes results for all 56 treated sites used in Chapter 2, in the Northumbria analysis

	y_j	μ_j	α_j	$E(m_j\|y_j)$	$SD(m_j\|y_j)$	$y_{j,after}$	Observed	After RTM
Site j=1	20	1.61	0.61	8.81	1.86	0	-20	-9
Site j=2	4	1.67	0.60	2.61	1.02	0	-4	-3
Site j=3	9	0.89	0.74	3.02	0.89	5	-4	1
Site j=4	3	3.30	0.43	3.13	1.34	0	-3	-4
Site j=5	17	1.67	0.60	7.83	1.77	8	-9	0
Site j=6	1	1.38	0.64	1.24	0.67	1	0	-1
Site j=7	3	0.82	0.75	1.36	0.58	1	-2	-1
Site j=8	4	1.84	0.58	2.76	1.08	0	-4	-3
Site j=9	7	5.88	0.30	6.67	2.16	6	-1	-1
Site j=10	6	2.70	0.48	4.41	1.51	9	3	4
Site j=11	8	3.16	0.44	5.86	1.81	4	-4	-2
Site j=12	5	1.62	0.61	2.95	1.08	2	-3	-1
Site j=13	12	1.74	0.59	5.95	1.56	2	-10	-4
Site j=14	3	4.63	0.35	3.57	1.52	1	-2	-3
Site j=15	4	2.59	0.49	3.31	1.30	5	1	1
Site j=16	8	3.37	0.43	6.03	1.86	3	-5	-4
Site j=17	6	3.24	0.43	4.80	1.65	8	2	3
Site j=18	11	3.45	0.42	7.83	2.13	10	-1	2
Site j=19	21	1.40	0.64	8.45	1.74	10	-11	1
Site j=20	1	2.59	0.49	1.78	0.95	2	1	0
Site j=21	3	2.00	0.56	2.44	1.04	6	3	3
Site j=22	4	3.82	0.40	3.93	1.54	2	-2	-2
Site j=23	11	1.34	0.65	4.71	1.28	7	-4	2
Site j=24	5	1.49	0.63	2.80	1.02	2	-3	-1
Site j=25	8	5.60	0.31	7.26	2.24	6	-2	-2
Site j=26	9	1.92	0.56	5.00	1.48	10	1	4

	y_j	μ_j	α_j	$E(m_j\|y_j)$	$SD(m_j\|y_j)$	$y_{j,after}$	Observed	After RTM
Site j=27	7	1.94	0.56	4.15	1.35	8	1	3
Site j=28	3	1.79	0.58	2.29	0.98	9	6	6
Site j=29	16	2.59	0.49	9.43	2.19	4	-12	-6
Site j=30	8	3.02	0.45	5.75	1.77	4	-4	-2
Site j=31	4	3.40	0.42	3.75	1.47	3	-1	-1
Site j=32	4	2.55	0.49	3.28	1.29	2	-2	-2
Site j=33	28	3.94	0.39	18.66	3.38	16	-12	-3
Site j=34	13	3.87	0.39	9.42	2.39	10	-3	0
Site j=35	2	4.94	0.34	2.99	1.41	7	5	4
Site j=36	15	4.01	0.38	10.79	2.58	14	-1	3
Site j=37	1	1.64	0.60	1.39	0.74	3	2	1
Site j=38	4	3.70	0.40	3.88	1.52	3	-1	-1
Site j=39	7	1.46	0.63	3.51	1.14	2	-5	-2
Site j=40	8	1.63	0.60	4.15	1.28	4	-4	-1
Site j=41	7	3.12	0.44	5.28	1.71	5	-2	-1
Site j=42	5	3.23	0.44	4.23	1.54	4	-1	-1
Site j=43	6	8.30	0.23	6.53	2.24	7	1	0
Site j=44	7	4.94	0.34	6.31	2.05	7	0	0
Site j=45	2	2.95	0.46	2.44	1.15	5	3	2
Site j=46	8	3.99	0.38	6.46	1.99	5	-3	-2
Site j=47	16	7.07	0.26	13.67	3.18	5	-11	-9
Site j=48	7	3.30	0.43	5.41	1.75	4	-3	-2
Site j=49	11	2.58	0.49	6.87	1.87	3	-8	-4
Site j=50	8	1.19	0.68	3.39	1.05	3	-5	-1
Site j=51	11	2.13	0.54	6.21	1.69	16	5	9
Site j=52	4	1.49	0.63	2.43	0.95	4	0	1
Site j=53	5	4.49	0.36	4.82	1.76	4	-1	-1
Site j=54	12	4.06	0.38	8.98	2.36	14	2	5
Site j=55	7	6.11	0.29	6.74	2.19	9	2	2
Site j=56	7	3.18	0.44	5.32	1.73	1	-6	-5
Total	436			323		298	-138	-25