Disease mapping models: an empirical evaluation

DISEASE MAPPING COLLABORATIVE GROUP

A. B. Lawson (Aberdeen),* A. B. Biggeri (Florence), D. Boehning (Berlin), E. Lesaffre (KU Leuven),
J-F. Viel (Besancon), A. Clark (Aberdeen), P. Schlattmann (Berlin), F. Divino (Florence)

SUMMARY

The analysis of small area disease incidence has now developed to a degree where many methods have been proposed. However, there are few studies of the relative merits of the methods available. While many Bayesian models have been examined with respect to prior sensitivity, it is clear that wider comparisons of methods are largely missing from the literature. In this paper we present some preliminary results concerning the goodness-of-fit of a variety of disease mapping methods to simulated data for disease incidence derived from a range of models. These simulated models cover simple risk gradients to more complex true risk structures, including spatial correlation. The main general results presented here show that the gamma-Poisson exchangeable model and the Besag, York and Mollie (BYM) model are most robust across a range of diverse models. Mixture models are less robust. Non-parametric smoothing methods perform badly in general. Linear Bayes methods display behaviour similar to that of the gamma-Poisson methods. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

The analysis of geographical variation in rates of disease has many uses, not least in the formulation and assessment of aetiological hypotheses, and also in the area of resource allocation or risk assessment in public health.

The main aims of disease mapping are to:

  (i) describe the spatial variation in disease incidence for the formulation of aetiological hypotheses;
 (ii) identify areas of unusually high risk so that action may be taken;
(iii) provide a 'clean' map of disease risk in a region to allow better resource allocation and risk assessment.

Case studies in disease mapping have covered a wide range of applications, including sudden infant death syndrome [1], lip cancer in Scotland [2], child mortality [3], stomach cancer in

---

* Correspondence to: Dr. A. B. Lawson, Department of Mathematical Sciences, King's College, University of Aberdeen, Old Aberdeen, AB24 3UE, U.K.

Missouri [4], cancer in France [5], lung cancer in Germany [6]. Work in this field has been recently reviewed [7, 8].

The great proliferation of methods which have been proposed for the analysis of disease maps has not been matched by any substantive examination of the relative merits of the methods except in limited situations. For example, Clayton and Kaldor [2] made a comparison of a small set of methods (empirical and full Bayes) and found that while the absolute relative risk estimates varied between methods, the ranking of relative risks across the mapped area (Scottish lip cancer example) remained the same. Some other attempts to assess full Bayesian model sensitivity have been reported [9, 10]. These assessments have been limited to a small range of Bayesian models only.

In this paper we present preliminary results of a large scale study on the behaviour of a range of methods on simulated data, derived from a wide variety of models for the true underlying risk. The models fitted range over simple empirical Bayes and linear Bayes methods, non-parametric smoothing methods, marginal mixture models and full Bayes models. The results presented here represent general goodness-of-fit results using global measures. More detailed analysis of the spatial structure of the model fits is not reported here, although global autocorrelation measures applied to residuals from model fits are examined. Hence, the results reported in this preliminary paper mainly focus on the robustness of the fitted models to the variation of underlying true risk, and the isolation of groups of true models for which the fitted model performs well or badly.

In the next sections we outline the range of models fitted, followed by the simulation procedures and true models examined. Finally we make summary results for the range of fitted models examined.

2. THE BASIC LIKELIHOOD MODEL

In the following, we assume that disease incidence is available in the form of counts of cases within small areas (such as census tracts or postal districts). Hence, our discussion is focused on the analysis of small area count data. We do not consider here the performance of methods for the analysis of case address locations. We also assume that the disease of interest is non-infectious, and so we will examine purely descriptive methods for small area count data.

Within a map of n regions, let O_i denote the observed count of disease in the ith region, E_i denote the expected count in the ith region and θ_i denote the relative risk in that region. We assume that the expected counts are known constants. The observed count in the ith region is assumed to be Poisson distributed with mean $E_i \cdot \theta_i$, and the likelihood of $\{O_i\}$ is given by

$$L(\theta) = \prod^n \frac{\exp(-E_i \cdot \theta_i)}{O_i!} \{E_i \cdot \theta_i\}^{O_i} = \prod_i^n \text{Pois}(O_i; E_i \cdot \theta_i)$$

$$\propto \prod^n \theta_i^{O_i} \cdot \exp \left\{ -\sum^n E_i \cdot \theta_i \right\}$$

and

$$l(\theta) = \ln L(\theta) = \sum O_i \ln \theta_i - \sum E_i \cdot \theta_i, \tag{1}$$

for a constant.

3. MODELS

Often the focus of mapping is to highlight areas of a disease map which deserve further attention, for example, areas with a high relative risk. It is therefore quite sensible to test whether each region's relative risk is greater than one [11], or to estimate the order of the relative risks [12]. However, our examination focuses on a model-based approach where the aim is to provide suitable relative risk estimates in each small area. It is possible to regard the problem of estimation of subregion relative risk as an exercise in parameter estimation under an assumed model. This model can be Bayesian in nature with log-likelihood as given in equation (1) or can be non-parametric in nature.

3.1. Standardized mortality ratio

As a first step, it is often proposed that the standardized mortality ratio (or SMR for short) be computed and mapped. It is defined as

$$\varphi_i = \frac{O_i}{E_i} \tag{2}$$

the ratio of observed to expected counts in each region. This estimator is unbiased, and is frequently used by epidemiologists. However, this estimate is based only on a sample size of one and hence is not really statistically useful; it is a *saturated model* estimate.

While a crude map of $\{\varphi_i\}$ is useful it does have three main disadvantages:

  (i) The variance of φ_i is θ_i/E_i which is large in areas with a small population (and hence a small E_i) and small in areas with a large population; this makes decisions based on the SMR difficult.
 (ii) It does not differentiate between regions when we observe no deaths.
(iii) It makes no attempt to reveal any underlying structure in the data, and is not parsimonious, which is a common flaw of all saturated model estimators.

3.2. Alternative approaches

To address aspects of the above problems a variety of alternative models have been proposed. These can be broadly categorized as follows:

  (i) smoothing models (for examples, non-parametric regression), which try to smooth out noise based on functions of the data in surrounding areas;
 (ii) linear Bayes methods, which are based on a linear function of the SMR,
(iii) Bayesian models, which assume that the relative risks are realizations from some distribution;
(iv) empirical Bayes models, which are similar to Bayesian models but which estimate the prior distribution for the relative risks from the observed data.

3.3. Smoothing methods

An approach to exploratory analysis of data often adopted in univariate problems, is the use of a smoothing method [13] which attempt to reveal the underlying signal or ground truth. A

commonly used method in this class is the *Nadaraya–Watson kernel smoother* (for example, Reference [13], chapter 5), which, when applied to the SMR, is simply a weighted average:

$$\hat{\theta}_i = \sum_{j \neq i} w_j \varphi_j$$

The weights are functions of neighbouring values

$$w_i = \frac{K((\varphi_i - \varphi_j)/h)}{\sum_j K((\varphi_i - \varphi_j)/h)}$$

where the function $K(\cdot)$ is a zero mean radially symmetric probability density function, and h is chosen to minimize some goodness-of-fit measure. Other weighting schemes are available, but do not have the same statistical properties as this estimator. The smoothing constant h can be chosen by using criteria based on cross-validation [13].

Another alternative to smoothing the SMR itself is to separately smooth the numerator and denominator of φ_i. A procedure similar to this has been proposed for estimating the relative risk for case address locations by Kelsall and Diggle [14, 15]. This method was applied to case event and control event location data, where the observation locations were not matched on location, and the authors employed least squares cross-validation to obtain a common smoothing constant for the resulting ratio estimator. It was found that on theoretical grounds the common estimation of the smoothing constant was preferred over the separate estimation of different smoothing constants for numerator and denominator. This method can be applied to count data in small areas and as the counts and expected counts arise within the same areas then it is possible to use deletion-based cross-validation to estimate a common smoothing constant. Note that these methods do not use a likelihood formulation directly.

In our analysis we have applied these methods to simulated data using least squares cross-validation to estimate the smoothing constants.

3.4. Linear Bayes estimators

The application of linear Bayes methodology [16] to disease mapping was introduced by Marshall [3].

It can be shown that the linear Bayes estimator of θ_i based on the SMRs $\{\varphi_i\}$ is

$$\hat{\theta}_i = E_{\Theta}\{\theta_i\} + \frac{\operatorname{cov}(\theta_i, \varphi_i)}{\operatorname{var}_{\Theta}(\varphi_i)} [\varphi_i - E_{\Theta}\{\varphi_i\}] \tag{3}$$

where $E_{\Theta}\{\theta_i\}$ is the expectation over the prior distribution of $\{\theta_i\}$. This estimator is, as the name suggests, the best estimator of the form (in the sense of minimum expected loss):

$$\hat{\theta}_i = \alpha + \beta \varphi_i \tag{4}$$

The above equation leads to an estimator of the form:

$$\hat{\theta}_i = m_i + \frac{v_i}{\dfrac{m_i}{E_i} + v_i} \{\varphi_i - m_i\}$$

where m_i and v_i are, respectively, the prior mean and variance of the relative risk. Hence, if we can estimate the prior mean and variance of θ_i we can easily obtain its posterior estimate. Marshall

[3] proposed two such estimates. Before proceeding we highlight some advantages/disadvantages of this method:

   (i) If E_i is large then the estimate tends to the SMR, and if it is small then the estimate tends towards the prior mean
  (ii) These estimators violate the likelihood principle and are hence questionable on theoretical grounds. The likelihood principle specifies that if two likelihoods are proportional to each other, where the proportionality constant does not depend on the parameter of interest, then the conclusions drawn should be the same.
 (iii) This approach only provides estimated values and methods for the calculation of confidence intervals or hypothesis tests or a measure of the variation, are not provided.

3.4.1. Global estimates. The so-called *global estimates* are obtained by assuming a constant mean and variance ($m_i = m$ and $v_i = v$) estimated via the method of moments:

$$\hat{m} = \frac{\sum O_i}{\sum E_i}$$

$$\hat{v} = \frac{1}{\sum E_i} \sum E_i \cdot \{\varphi_i - \hat{m}\}^2 - \frac{\hat{m}}{\sum E_i}$$

Obviously the estimate of v can be negative so we adopt the convention that

$$\hat{\theta}_i = \hat{m}$$

if $\hat{v} < 0$.

3.4.2. Local estimates. The so-called *local estimates* are obtained by estimating the prior mean and variance using only those regions which are neighbours of that region:

$$\hat{m} = \frac{\sum_{j \sim i} O_j}{\sum_{j \sim i} E_j}$$

$$\hat{v}_i = \frac{1}{\sum_{j \sim i} E_j} \sum_{j \sim i} E_j \cdot \{\varphi_j - \hat{m}_j\}^2 - \frac{\hat{m}}{\sum_{j \sim i} E_j}.$$

Marshall estimated the standard error of these estimates using the delta method. However, it is better to consider the variance of the posterior distribution. We are not interested in the variance of the estimated values but rather the variance of the posterior distribution of the parameters, that is, we require

$$\text{var}(\theta_i \mid \{O_i\})$$

not

$$\text{var}(\hat{\theta}_i)$$

where $\hat{\theta}_i$ is some summary measure of the posterior distribution, for example, the mean or mode.
   In what follows we have computed both the global and local Marshall estimators using the above estimators for m_i and v_i.

3.5. Empirical Bayes estimators

Empirical Bayes estimators [17] assume that $\{\theta_i\}$ follow some distribution, say $g(\theta)$, which is indexed by a set of parameters, say Ω. These can be estimated from the data by maximizing the marginal likelihood

$$L(\Omega) = \int f(x \mid \theta) g(\theta \mid \Omega) \, \mathrm{d}\theta$$

where $f(x \mid \theta)$ is the joint distribution of the observed cases. The maximization can be done via the EM algorithm, hence not requiring integration, or using an analytical approximation to the integral such as Laplace's method.

3.5.1. Gamma-Poisson model.

The gamma-Poisson model takes $g(\theta)$ to be a gamma distribution with parameters $(\alpha, \beta) \in \mathfrak{R}_+^2 = \Omega$. It is easily shown that the posterior distribution is

$$\theta_i \mid \{O_i\}, \alpha, \beta \sim \mathrm{Gamma}\left(O_i + \alpha, \frac{\beta}{\beta E_i + 1}\right).$$

We can then estimate the $\{\theta_i\}$ by the posterior mean

$$\hat{\theta}_i = (\alpha + O_i) \cdot \frac{\beta}{\beta E_i + 1}.$$

Using this we see that if both the observed and expected counts are large, then this gets pulled towards the SMR, whereas if they are both small, then this tends towards the prior mean. In fact we can write this as a linear combination of the two

$$\hat{\theta}_i = w_i E(\theta_i \mid \alpha, \beta) + (1 - w_i)\varphi_i$$

where $w_i = (\beta E_i + 1)^{-1}$, and $E(\theta_i \mid \alpha, \beta)$ is the posterior expectation of θ_i. Notice that this is simply the global linear Bayes estimate with $m = \alpha\beta$ and $v = \alpha\beta^2$.

As regards the estimation of (α, β), the marginal likelihood is proportional to the negative-binomial distribution

$$L(\alpha, \beta) = \frac{1}{\{\Gamma(\alpha)\}^n} \prod^n \Gamma(O_i + \alpha)\beta^{O_i} \cdot \frac{1}{\{\beta E_i + 1\}^{O_i + \alpha}}$$

and in order to maximize this with respect to (α, β) we need to use a numerical algorithm, see, for example Reference [2].

While the estimated values can be derived from the global linear Bayes estimator, the gamma-Poisson model has some advantages:

  (i) it estimates the full posterior which can be used to give confidence intervals and hypothesis tests whereas the linear Bayes methods do not;
 (ii) even if the actual distribution of the relative risk is a gamma then the gamma-Poisson model estimates it mean and variance via maximum likelihood which is superior to the method of moments used by the linear Bayes methods.

One disadvantage of the gamma-Poisson model is its inability to cope with spatial correlation. However, some recent advances have been made [18].

In what follows we have estimated the gamma parameters from the full posterior marginal distribution via maximization. In a later section, where we examine the change in Bayesian information criterion (ΔBIC) between models [17], we have generated posterior samples of θ_i from the full posterior gamma distribution.

3.5.2. Mixture models (Non-parametric maximum likelihood). It is possible to construct a model where the marginal distribution of counts is governed by a mixture distribution consisting of a set of discrete components. This approach differs considerably from the previous models in that the variation in risk is modelled by a mixture of components and not by a global model for the variation. The mixture model assumes that there are k distinct levels of risk, say $\{\theta_1, \ldots, \theta_k\}$, and our observed count comes from a region with risk θ_j with probability p_j. Thus our likelihood becomes.

$$L = \prod_{i=1}^{n} \sum_{j=1}^{k} p_j \text{Pois}(O_i; E_i, \theta_j)$$

No closed form solution exists for the relative risks in this case, however reliable algorithms are available [19]. These give MLEs for $\{p_j\}$, $\{\theta_i\}$ and also can be used to classify each region with a risk; an alternative is to use the posterior expectation, given by

$$\hat{\theta}_i = \frac{\sum_{j=1}^{k} \hat{\theta}_i \hat{p}_j \text{Pois}(O_i; E_i, \hat{\theta}_i)}{\sum_{j=1}^{k} \hat{p}_j \text{Pois}(O_i; E_i, \hat{\theta}_j)}.$$

We have applied this mixture model in our analysis, the parameter estimation being carried out using the VEM algorithm [20].

3.6. Full Bayes estimators

Here we consider a model proposed by Besag, York and Mollie (BYM) [21]. In our basic model described above the conditional distribution of an observed count O_i given $\{\theta_i\}$ is

$$O_i \mid \{\theta_i\} \sim \text{Pois}(E_i \cdot \theta_i)$$

The model proposed by BYM splits the relative risk parameter into three components (that is $\log(\theta_i) = t_i + u_i + v_i$):

(i) *Trend.* We can allow for spatial trend and/or covariates by having a regression term, say

$$t_i = \sum a_j x_i^j y_i^j + \sum b_k z_k$$

where $\{(x_i, y_i)\}$ are the centroids of the ith region and z is a vector of covariates. It is important that this term does not contain an intercept, as it will not be identifiable. The regression terms could be replace by a non-parametric alternative such as a generalized additive model.

(ii) *Overdispersion (spatial).* It seems plausible that regions close to each other will have similar relative risk, and we can allow for this possibility by having a random variable, say u_i, which is correlated with the other $\{u_j\}$. We usually assume a prior specification such as

$$f(u) \propto \exp\{Q(u_i - u_j)\}$$

where $Q(\cdot)$ is some prespecified function.

(iii) *Overdispersion* (*non-spatial*). By the formulation of the model for the spatially correlated overdispersion, the variance depending on the number of neighbours, independence is not well defined by the model. To combat this we can introduce another term, say v_i, which is a standard uncorrelated overdispersion parameter. Here, the prior distribution is given by

$$v_i \stackrel{\mathrm{ind}}{\sim} \mathrm{N}(0, \sigma_v^2)$$

3.6.1. Choice of $Q(\cdot)$. Although it is popular to choose $Q(\cdot)$ to be quadratic (the *intrinsic Gaussian model*) it is not the only choice possible, and in fact other choices may be preferable for different situations. Here, we have assumed the quadratic form with variance σ_u^2.

3.6.2. MCMC methods. We have implemented the BYM model as defined in the original work [21]. We have used a Gibbs sampler for the $\{\sigma_u^2, \sigma_v^2\}$, while a Metropolis step was used for sampling $\{u_i, v_i\}$. The hyperprior distributions for the variances were inverse exponential with parameter $\varepsilon = 0.05$. For the purposes of this evaluation, we have not included a trend component. This is consistent with many applications of the model and with the original publication. Because it is essential to monitor for convergence of such MCMC algorithms, it was important to assess, prior to implementing the large scale model-fitting exercise, the degree to which the MCMC fitting could be automated. To do this we decided to perform an experiment to assess convergence across a number of data sets. Based on a selection of ten representative data sets we made convergence checks and assessed the variability of the time to convergence. It was found that convergence occurred in all data sets by 18 000 iterations and we then added a 'safety factor' of 2000 iterations. In the model fitting exercise we fixed the iteration length at 20 000 based on this experiment.

The above models define a range of possible approaches to the description of the underlying relative risk for any given disease map. Our task was to examine the goodness-of-fit of these models via an empirical evaluation based on simulated data from assumed true relative risk models. In the next section we describe the simulation procedures used for this exercise.

4. SIMULATED MODELS AND METHODS

To assess which of the above models are good at recovering the true spatial variation we decided to simulate data sets from a number of different possible relative risk models and to evaluate the relative merits of each of the fitted models using a variety of criteria. To this end we needed to choose both a suitable map of small areas to simulate relative risks within, and a set of fixed expected cases for these small areas. The intention was to simulate a range of count distributions from a set of true risk models over this set of small areas. To this end we first considered what base map to use for the simulation. Instead of using a regular lattice it was felt important to use a real map, which included a large number of small areas (to reduce edge effects) where the small areas were of similar size and shape, so as to eliminate artefacts of irregular small area form. As a result we chose to use the map of former East Germany, which is relatively large and meets these criteria quite closely. In addition, we also required to use a set of fixed expected counts for the mapped area. These were chosen for a specific disease which had a reasonably regular spatial distribution over the whole area. We decided to use the expected number of deaths from lip cancer for the

period 1980–1989 for this purpose. The total population in 1989 was close to 17 million, so that the total rate is over 170 million person-years. The exact expected rate for the period was 2454.3, and we decided to base our simulations on a rounded total of 2500. Figure 1 displays a choropleth map of these expected rates.

Based on the Eastern Germany map of expected lip cancer counts we proceeded to simulate from a range of models for excess risk (beyond that specified by the expected counts). We decide to simulate 100 data sets from each of the models chosen. Owing to the need to keep the data sets consistent with the expected count it was necessary to constrain the observed counts to sum to the total number of expected cases. Given this conditioning we used the multinomial model:

$$O \sim \text{Multinomial}\left(N_t; \frac{E_i\theta_i}{\sum_j E_i\theta_i}\right)$$

where N_t is the total number of expected cases. Hence, by specifying models for the relative risks $\{\theta_i\}$ $i = 1, \ldots, n$ we can simulate count distributions within the mapped area and subsequently examine the fit of a variety of models.

The models chosen for this examination ranged from simple constant risk models to complex Bayesian random effect models, including trend and differentiated types of heterogeneity (correlated and uncorrelated). In addition, mixture models of different types were examined. Simulation from the multinomial distribution described above was carried out by table look-up methods. As we simulated 100 realizations from each model our goodness-of-fit (GOF) criteria had to be averaged over these realizations. This averaging is discussed in greater detail in a later section. In addition, given that realizations of counts were generated, we required to examine both GOF to the realized counts and to the true relative risk model. For Bayesian ground truth models this is not straightforward, as the parameters of the ground truth models also come from distributions.

In the following section we detail the true relative risk models defined in this study. The models chosen were selected to represent the range of possible underlying risk that might be encountered.
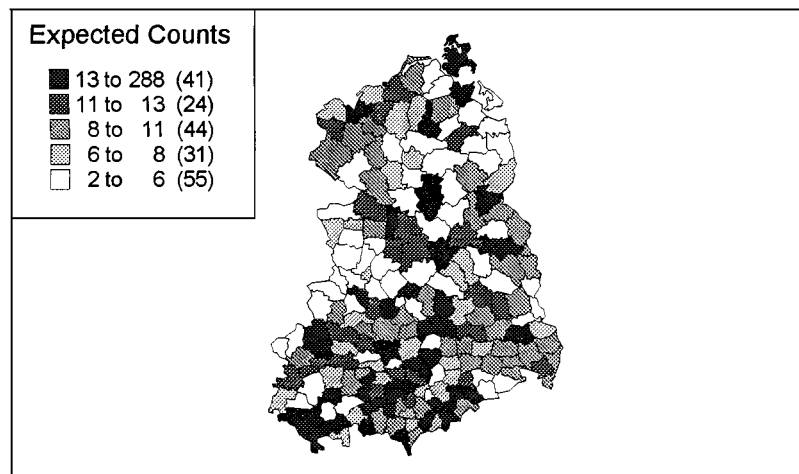

Figure 1. Expected counts for lip cancer in the Eastern Germany region.

Thus, while basic trend models are present, we have also examined random effect and mixture models with trend included. In addition we have introduced specific clusters into some random effect models. The models can be grouped into broad classes: fixed effect models with trend (models 1–3), random effect models with either trend, uncorrelated or correlated heterogeneity (models 4–49); mixture model with a variety of components (such as trend, and random effects) (models 50–109), and models with specific cluster terms and selections of random effects and trend (models 110–151) and finally gamma distribution models with a variety of fixed parameters (models 152–154). These models are detailed in the next sections. The Appendix contains a detailed description of all the simulated models.

The detailed differences between the models within each group were chosen to ensure that an adequate range of true model relative risk were available in the class of models. This served to provide robustness against conclusions based on single model choices, and also coverage of possible ranges of true parameter settings. For example, in the case of trend models (2–3) the first model has a strong y-gradient, while the second has a larger x-gradient. The final choice of coefficient values is of course arbitrary, but the choice is informed by the above balance of robustness (model specificity) and coverage (model generality).

4.1. Fixed effect models

Constant risk-model 1:

$$\theta_i = \theta = 1$$

Linear trend: models 2–3:

$$\theta_i = \exp\{x_i + 2y_i\}$$
$$\theta_i = \exp\{5x_i + y_i\}$$

4.2. Random effect models

Linear trend (models 4, 5)

$$\theta_i = \exp\{ax_i + by_i\}$$

where a and b are random variables with the following distributions:

$$\text{(i)} \quad a \sim N(1, 1); \quad b \sim N(2, 1)$$

$$\text{(ii)} \quad b \sim N(5, 1); \quad b \sim N(1, 1)$$

Quadratic trend (models 6, 7)

$$\theta_i = \exp\{ax_i + by_i + cx_i^2 + dy_i^2 + ex_iy_i\}$$

where a, b, c, d and e are random variables with the following distribution:

(i) $a \sim N(1, 1)$; $b \sim N(2, 1)$; $c \sim N(2, 1)$; $d \sim N(2, 1)$; $e \sim N(1, 1)$
(ii) $a \sim N(1, 1)$; $b \sim N(2, 1)$; $c \sim N(5, 1)$; $d \sim N(0.01, 1)$; $e \sim N(10, 1)$

BYM overdispersion model (*models 8–49*)

$$\theta_i = \exp\{p_1 t_i + p_2 u_i + p_3 v_i\}$$

The $\{p_i\}$ allow us to define the components of the data; t_i is the trend in the ith region, u_i is the correlated heterogeneity and v_i is the uncorrelated heterogeneity. For all of the following, t will be as above and

Correlated

$$u \sim \text{MVN}(0, K)$$

where $k_{i,j} = \sigma^2 \exp\{-d_{i,j}/R\}$, this is the so-called *exponential covariance* model in geostatistics [22]. The values to be used are

$$\ln R \sim \text{N}(\ln(0.5), 0.1)$$

$$\ln R \sim \text{N}(\ln(0.125), 0.1)$$

$$\ln \sigma^2 \sim \text{N}(0, 0.01)$$

$$\ln \sigma^2 \sim \text{N}(\ln(10), 0.1)$$

Uncorrelated

$$v_i \sim \text{N}(0, s)$$

where we use the values

$$\ln s \sim \text{N}(0, 0.1)$$

$$\ln s \sim \text{N}(\log(10), 0.1)$$

Mixture model (*model 50–53*). Define $Q(\lambda)$ to be the one-point distribution at λ, then the mixture model is defined as:

$$\theta_i \sim \sum_{j=1}^{k} w_j Q(\lambda_j)$$

By varying k we change the number of different relative risks in the data set and by varying w we change the proportion of relative risks in the data set.

Mixture with Spatial Structure: discrete overdispersion (*models 54–109*). We can combine the BYM model and the mixture model by defining:

$$\theta_i \sim \exp\{p_1 t_i + p_2 u_i\} \sum_{j=1}^{k} w_j Q(\lambda_j)$$

BYM with non-general clusters (*models 110–151*). We can define a process which allows for both spatially varying correlation and region specific variation but also including specific cluster centres. This can be implemented via the model

$$\theta_i = \exp\{p_1 t_i + p_2 u_i + p_3 v_i\} \left\{ 1 + \sum_{j=1}^{n_c} h(\|x_i + x_{cj}\|) \right\}$$

where x_i is a centroid and x_{cj} are cluster centres. The cluster centres were generated under complete spatial randomness with a Poisson distribution with mean 5.0 for the number of centres.

Global Gamma model (*models 152–154*). We can have a random effect equivalent of the equal risk model via the gamma-Poisson model

$$\theta_i \sim \text{Gamma}(\alpha, \beta)$$

with parameter settings

  (i) $\alpha = 1$; $\beta = 2$
 (ii) $\alpha = 2$; $\beta = 1$
(iii) $\alpha = 1$; $\beta = 1$

5. GOODNESS-OF-FIT

Whilst there is a large amount of literature on designing disease mapping models, as yet little attention has been paid to how to assess the goodness-of-fit of these models. We have attempted to employ conventional GOF measures which can be employed across a wide variety of models. For example, we have employed the change of Bayesian information criterion (ΔBIC) ([17], p. 48) to assess the discrepancy between the constant risk model and any specific appropriate fitted model. This gives a relative measure of the GOF of any model and is equivalent asymptotically to the Bayes factor for the two models. This criterion utilizes the likelihood within its formulation and also penalizes parameter rich models. The BIC criterion is available for the BYM, gamma and mixture models fitted. Because not all models fitted to the simulated count data had associated likelihoods, we also adopted more general criteria for GOF. The model fitting was examined for two sets of parameters: the true relative risks underlying the simulated models, and the counts simulated from the models. Assessment of the difference between true risk and fitted risk is important in detecting whether a true model can be recovered. On the other hand the comparison of observed count and fitted count is usually the form of comparison which would be made in any real data analysis situation (without knowing the true relative risk), and hence this comparison may display the expected differences to be found in a model fitting exercise.

To assess the strength of correlation between the fitted model output and these two parameter sets we have examined the Pearson and Spearman correlation coefficient for the fitted count to the simulate count and the relative risks, Pearson's chi-squared measure (RSS) applied to the counts and relative risks, and an autocorrelation coefficient (Moran's I) applied to the (Pearson) residuals from the fit. We also examined a generalized Kullback–Leibler measure of fit, but the results of this measure were highly variable and we did not pursue its use. In the next sections we describe the exact measures used for fitted counts. These measures were also applied to relative risks (except for the ΔBIC which is applied to the simulated count data directly). All results are presented as averages over the 100 simulations.

In the case of the count comparisons, we utilize the observed count (denoted $O_{ij} = \hat{\theta}_{ij} \cdot E_i$). In the relative risk comparison case we define θ_j as the relative risk for the true model for the jth simulation, and $\hat{\theta}_{ij}$ as defined above for the fitted model relative risk. The following measures are described for count data. However, they have also been computed for the relative risk comparisons, using the definitions above (except for the ΔBIC criterion).

5.1. Pearson's Correlation Coefficient

Given the fitted count ($\hat{O}_i = \hat{\theta}_i \cdot E_i$) and the observed count (O_i), for a given simulation we can compute the Pearson correlation coefficient

$$\frac{\sum \hat{O}_i O_i - 1/n \sum \hat{O}_i \sum O_i}{\sqrt{(\{\sum \hat{O}_i^2 - 1/n [\sum \hat{O}_i]^2\} \{\sum O_i^2 - 1/n [\sum O_i]^2\})}}$$

5.2. Spearman's Correlation Coefficient

Given the rank of the fitted count and the rank of the observed count we an compute the Spearman correlation coefficient

$$1 - \frac{6 \sum d_i^2}{n^3 - n}$$

where $d_i = \text{rank}(\hat{O}_i) - \text{rank}(O_i)$.

5.3. Chi-squared (*residual sum of squares RSS*)

Perhaps the oldest goodness-of-fit measure is Pearsons's chi-squared:

$$\sum \frac{(O_i - E(O_i \mid \theta_i))^2}{\text{var}(O_i \mid \theta_i)} = \sum \frac{(O_i - \hat{O}_i)^2}{\hat{O}_i}$$

This a global goodness-of-fit measure and has limitations for comparison between two models. However, we have used this measure to make relative comparisons between models, as it is well known and is related to the standardization of the Pearson residuals when a Poisson model is fitted.

5.4. Moran's I

It is important, in many modelling situations, that any residual correlation be explained by the model. The testing for autocorrelation in residuals is, however, complicated by the correlation structure imposed by the fitted model. We could allow for this correlation by using the parametric bootstrap (or Monte Carlo) on Moran's *I*. This would allow a closer comparison between true models. An analysis of that kind is planned for a future publication. Here, in this preliminary report, we present the relative magnitude of the coefficient only. The coefficient is defined as

$$\frac{n}{2A} \frac{\sum_{(2)} \delta_{ij}(r_i - \bar{r})(r_j - \bar{r})}{\sum (r_i - \bar{r})^2}$$

where we use the Pearson residuals resulting form the chi-squared measure:

$$r_i = \frac{O_i - \hat{O}_i}{\sqrt{\hat{O}_i}}$$

Because we have not made allowance for the correlation induced by the fitted model, all comparisons made with Moran's *I* should be treated with some caution. Comparisons *between*

groups of true models are less reliable than comparisons between fitted models for a given true model.

5.5. Delta-BIC

A common method to choose between Bayesian models is the BIC value [17]. A more natural choice is the difference in BIC values between our fitted model and our null model. The delta-BIC is defined as

$$\Delta\text{BIC} = 2 \ln \left\{ \frac{L(\theta)}{L(\theta_0)} \right\} - (p-1) \ln\{n\}$$

where θ_0 is the vector of estimated relative risks under the null (constant risk) model; p is the number of parameters in the fitted model and n is the number of regions.

6. RESULTS

All of the models were fitted to each simulated data set. We call the model from which the simulated data arose, the true model. The results are given in graphical form and the true models have been ordered 1–154. We have examined the GOF of fitted counts to observed counts, and the GOF of estimated relative risks to true relative risks. The first of these measures is what is usually available in any modelling exercise, and the results of these fits give indications of how well the models fit a realization from the true model. Here we have averaged over the 100 model simulations. Hence, this indicates how well we would expect to do in any given data analysis situation. However, it should be borne in mind that ordinary GOF measures (except the ΔBIC measure) do not penalize parameterization so that ultimately greater parameterizations can lead to closer models. Still, some conclusions concerning relative GOF can be made in this situation. In the case of relative risk comparisons, then we are examining the closeness of the fitted model to the true model. We would expect in general that fitted models do well when recovering their equivalent true models (i.e. BYM models should do well with BYM true models etc.). Plates 1–7 display the results in graphical form for all 154 true models. Plates 1–4 represents the count fits and are respectively, the BIC (Plate 1). Moran's I (Plate 2), Pearson correlation (Plate 3) and residual SS (Plate 4). Plates 5–7 represent the relative risk fits and are respectively: Moran's I (Plate 5), Pearson's correlation (Plate 6), and residual SS (Plate 7). Below we summarize the main features of the results for each GOF measure.

All models performed well on the constant risk model, but some models were unable to cope with autocorrelation in the true relative risks. In some cases, the lack of fit was found to be so great that the results for the method could not be shown (sensibly) in relation to other models. This was true, across all measures for the Kelsall–Diggle method applied to count data. This method resulted in very large residual SS and poor correlations for both count GOF and relative risk GOF. Because of this we have excluded it from the figures, as the relative merits of other methods would be obscured by the scaling. The explanation for this problem, which is mirrored in the (less extreme) behaviour of the kernel smoother, may be due to the considerable smoothing or over-smoothing induced by least squares cross-validation. Inevitably, such smoothing methods will induce autocorrelation in the resulting estimates. We have not presented the Spearman rank correlation here as all the results show similar effects to that of the Pearson correlation.
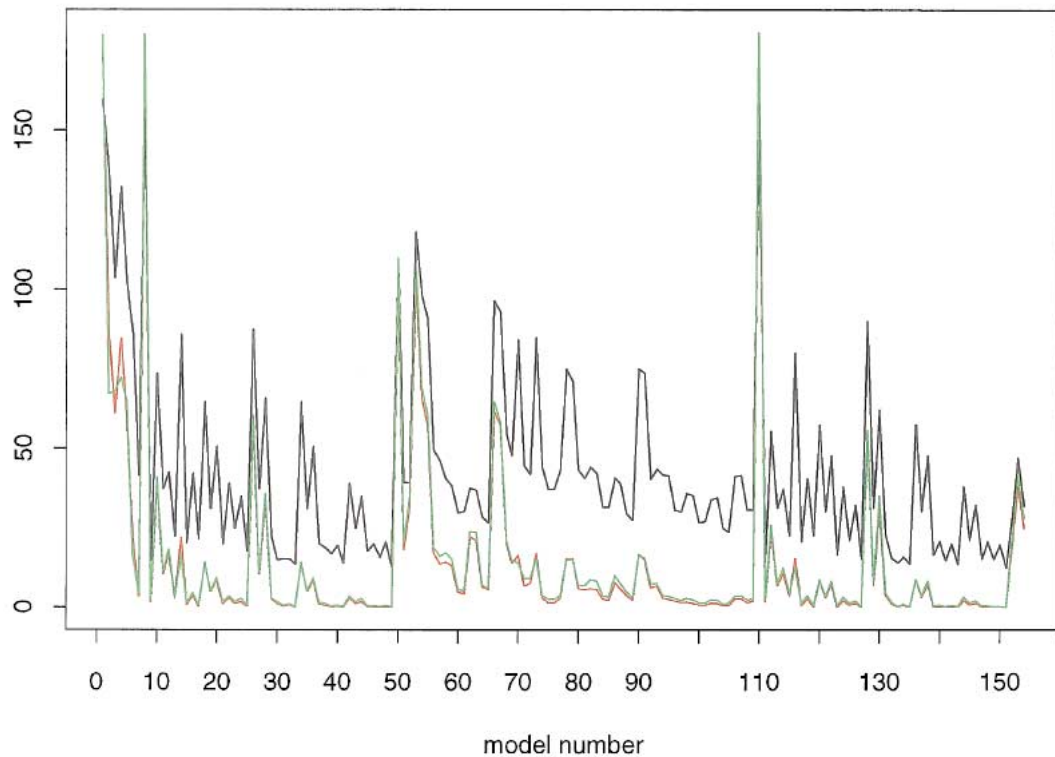
Plate 1. Residual sum of squares (RSS): count data comparison.

Plate 2. Moran's *I* coefficient (Pearson residuals): count data comparison.
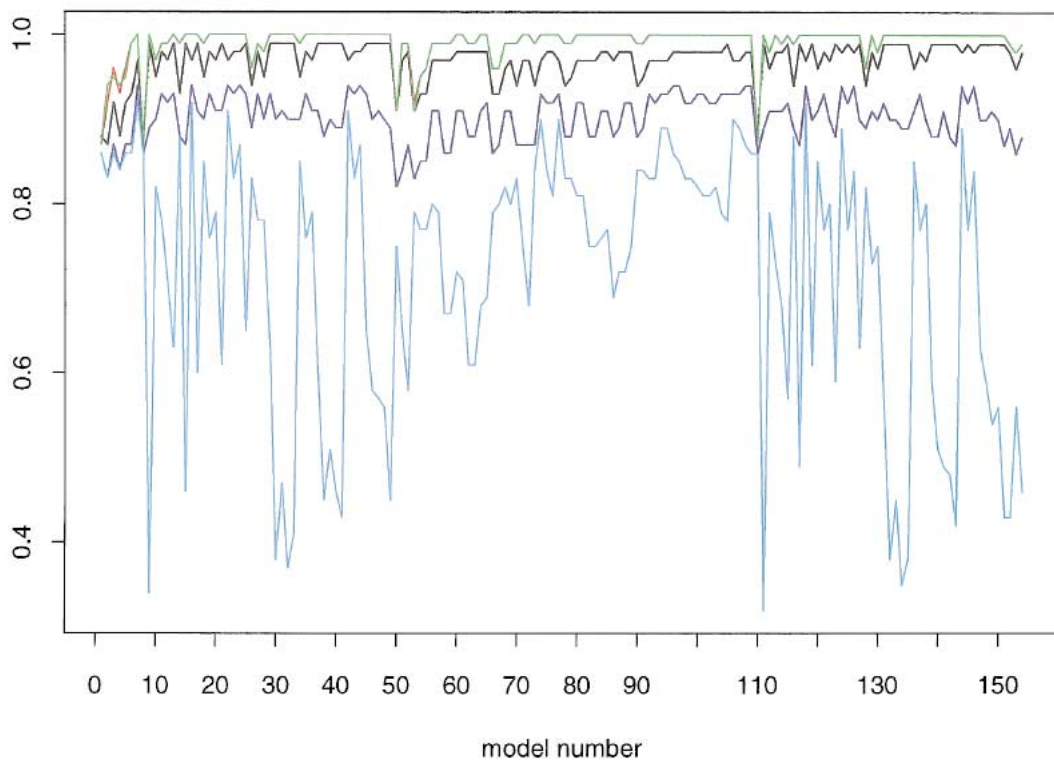
Plate 3. Pearson's count correlation coefficient.

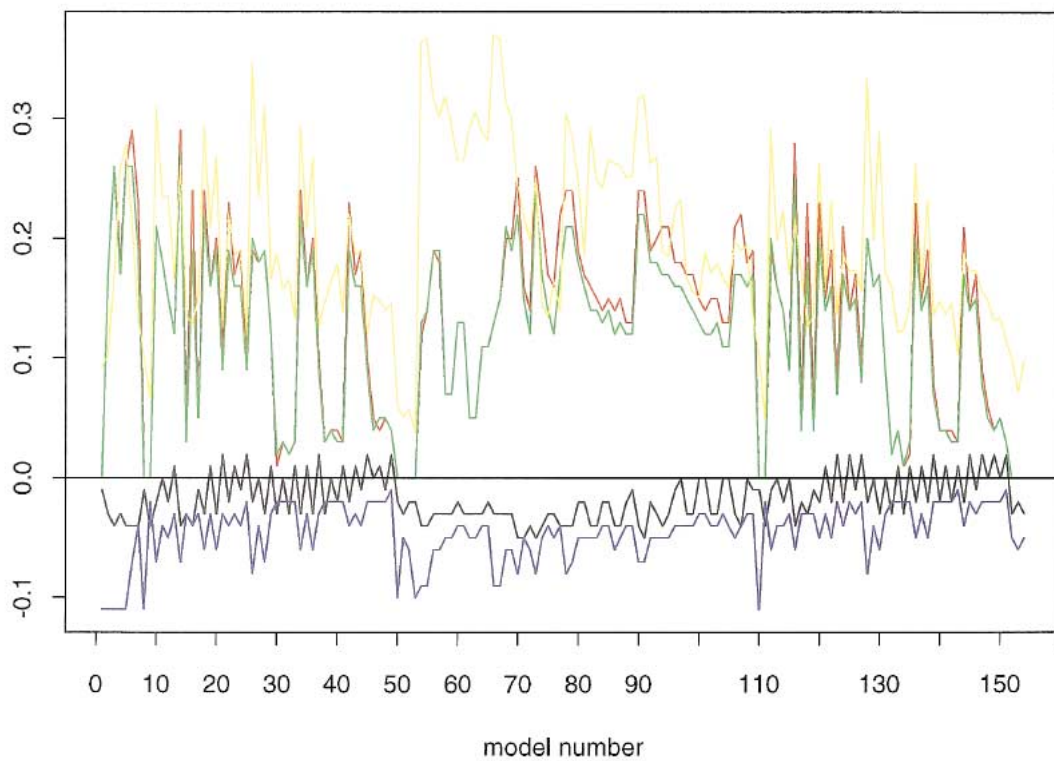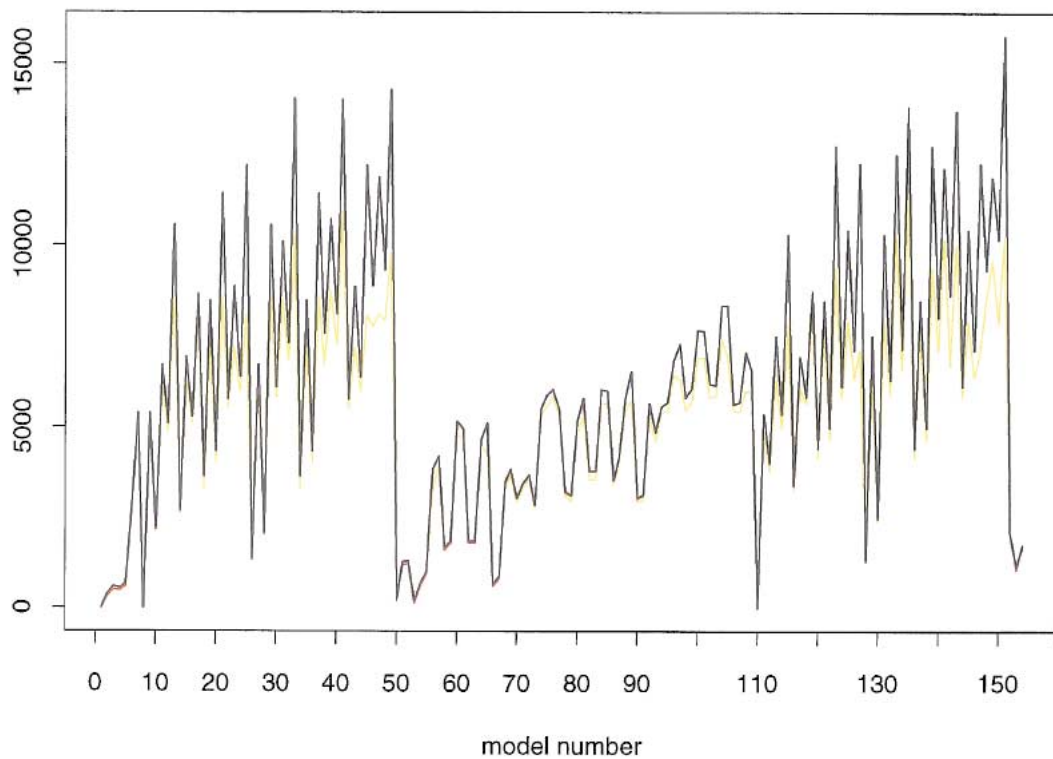Plate 4. ΔBIC criterion: BYM, mixture and gamma-Poisson models.
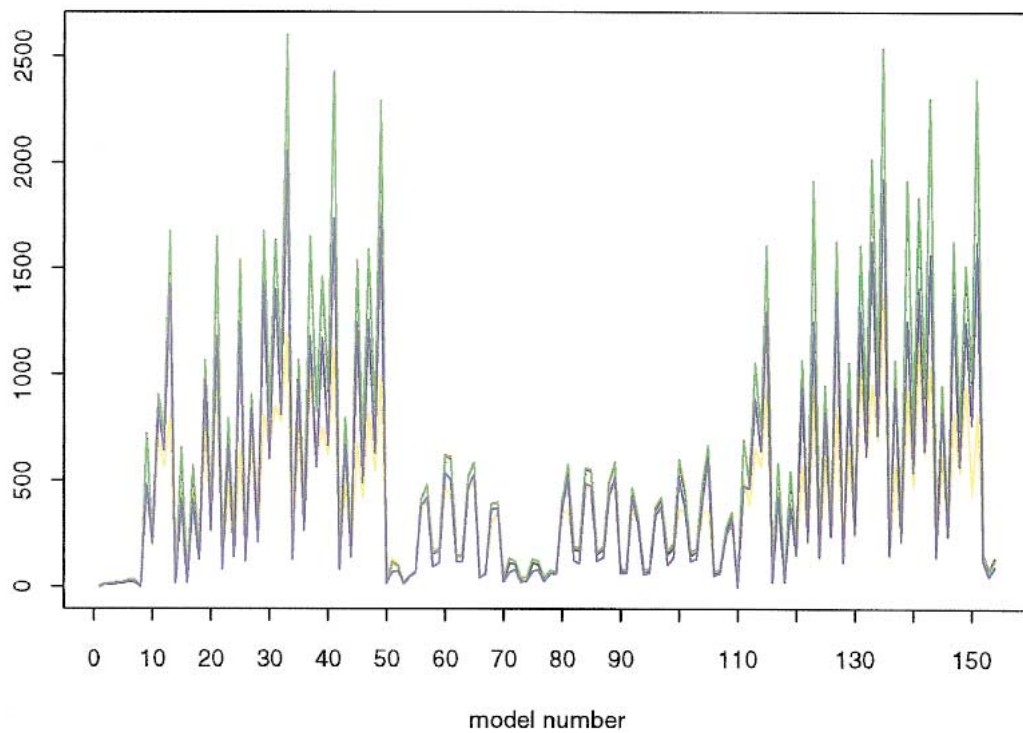
Plate 5. Residual sum of squares (RSS): relative risk comparison.
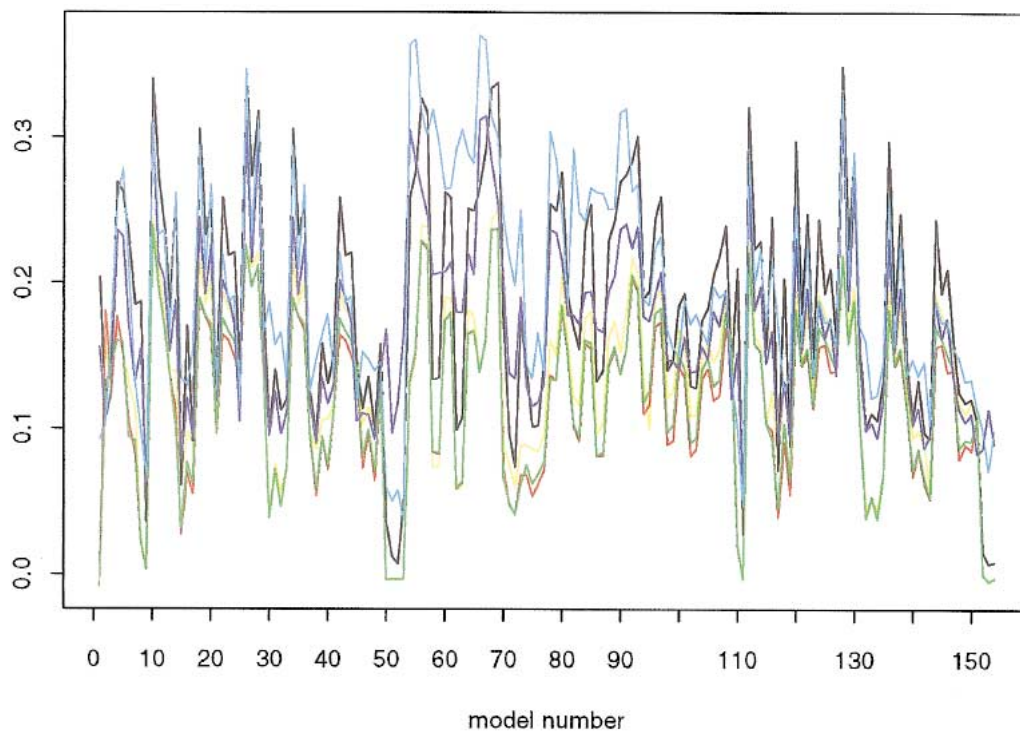
Plate 6. Moran's *I* coefficient (Pearson residuals): relative risk comparison.
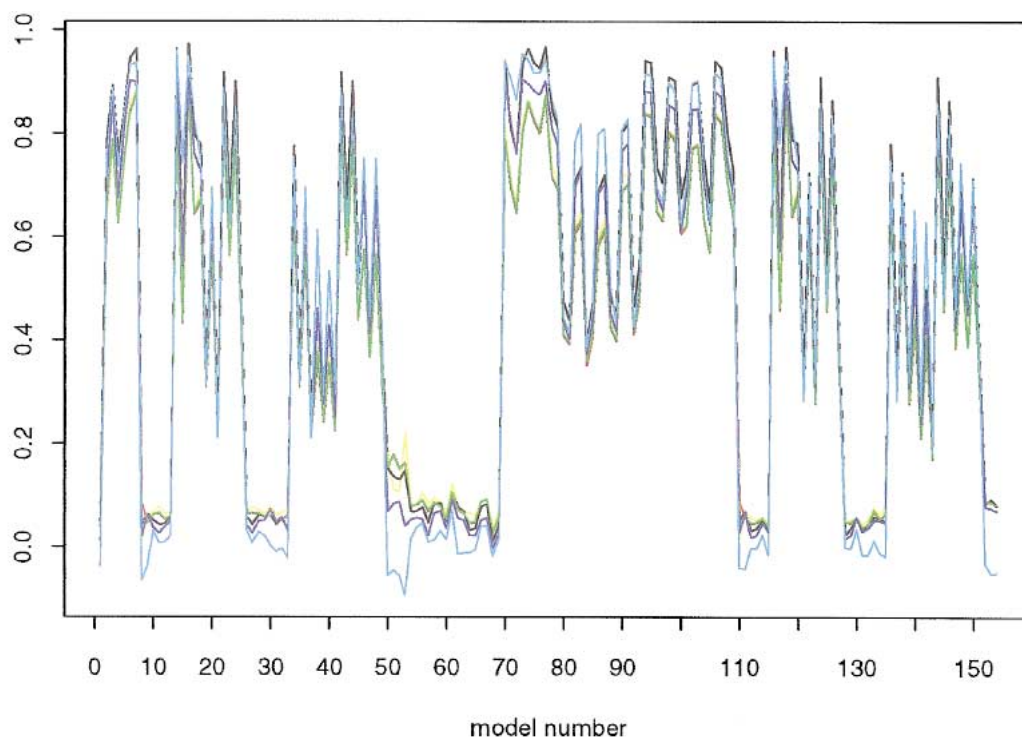
Plate 7. Pearson's relative risk correlation coefficient.

In general, the gamma-Poisson model and BYM appear to behave similarly across a range o f models, producing high correlations across all models for counts but more variable results for relative risks. The BYM model achieves highest ΔBIC values across a range of models. It should also be noted that across all models the true model has a greater effect on the GOF than the type of method used.

7. RESULTS FOR GOF MEASURES BASED ON COUNTS

Plates 1–4 display the results for count data. Overall, the BYM, gamma-Poisson and linear Bayes fitted models paper to perform well, with the mixture fitted models slightly less robust. Smoothing methods are highly variable, however.

7.1. Residual sum of squares

Models with greater degrees of smoothing do not fit the observed data well. This is because the more complicated the model the more smoothing that takes place, hence the estimated relative risks will be further from the SMR (which has zero RSS).

We note that the RSS is highly variable with the gamma-Poisson model consistently giving lower values than the others, this possibly being due to a lack of smoothing in this case. The smoothing model's variability was greater than the rest, performing well in models with a strong spatial correlation but poorly in models where unstructured heterogeneity dominated. Under simple trend models the lowest RSS for counts are found for the gamma model and the global Marshall.

7.2. Moran's I

Plate 3 reveals that the gamma-Poisson, global and local Marshall estimators leave a large amount of autocorrelation in the residuals. This indicates that neither model is capable of distinguishing between long-term trend and autocorrelation. Further, the BYM estimators had autocorrelation which fluctuated around zero, but goes negative for models with spatial autocorrelation. The smoothing model also leaves negligible amounts of autocorrelation. The Local Marshall estimator consistently produces negative autocorrelation. This should however be treated with caution since the expected value of Moran's I is $-1/(n-1) = -0.005$.

It is clear from the figures that the methods which allow for autocorrelation appear to capture this autocorrelation when present in the true model.

7.3. Pearson's correlation coefficient

From this graph we see that both the gamma-Poisson and the global Marshall estimators have a very high correlation with the observed counts. However, this is to be expected since these are the models whose estimated relative risks will be close to the SMR (which has correlation 1). In addition the BYM and mixture models also achieve high correlations. The smoothing model is much more erratic in its behaviour.

However, most of the models do not alter the ranking of the regions to any great extent. This said, when the rankings change it is usually within the regions with a low observed count.

7.4. Analysis of delta-BIC criteria

This criterion is only available for the BYM, gamma and mixture models. All the fitted models appear to display similar behaviour, with only slight differences appearing for particular true model types. Overall, the lowest ΔBIC values appear for the fixed trend, constant risk and simple mixture true models. Models with true extra-variation lead to higher ΔBIC values for all fitted models. Overall, the highest values are achieved for the BYM models, even when applied to a variety of true mixture models. The fitted gamma model appears to behave closely to that of the mixture model across a range of true models.

8. RESULTS OF GOF MEASURES ON RELATIVE RISKS

As with the count GOF measures we have presented the results graphically. Plates 5–7 display the relative risk results. Overall a more variable picture has emerged for the relative risk analysis. All fitted models appear to behave similarly across various models and criteria of GOF with the exception of the residual autocorrelation (Moran's I) where greater differences emerge.

8.1. Residual sum of squares

All models produced a similar pattern, with low values for models 50–120 and 151–154, but high values elsewhere. Most models behave similarly, with low RSS for trend models but higher values for true models with extra-variation. For true mixture models, all the methods appear to produce low RSS, with little differentiation. The worst RSS values are found for true models with extra-variation, and the Marshall estimators appear to have highest RSS for these models.

8.2. Moran's I

Unlike the count analysis, all models produced a positive value of Moran's I, most lying between 0.1 and 0.3. The lowest values were found for fixed trend, simple mixture and gamma true models. Overall there is considerably greater variability between fitted models and between results for types of true models. Generally, the highest residual autocorrelation was found for the fitted smoothing model and BYM model, whereas the lowest correlation was found for the fitted local Marshall estimator and the fitted mixture models. Although the results are locally variable across true models there are no large scale differences apparent for blocks of true models. This was also found for the count residual autocorrelation.

8.3. Pearson's correlation coefficient

For the relative risk analysis, the picture is more variable, although similar variation appears across all the fitted model types. One notable feature is that all models yield low correlations for certain true models. These are the true models with structured overdispersion and mixture structure. Mostly the lowest correlation is achieved by the smoothing model. However, for true models where the fitted models yield high correlations, then the lowest correlation is often found for the fitted mixture model, and in some of these cases the smoothing model achieves high correlation along with the BYM, global Marshall and gamma fitted models. The fact that the

mixture fitted model cannot well recover the relative risks for a range of true mixture models is important.

9. CONCLUSIONS AND FURTHER WORK

While a number of general conclusions can be made for this analysis, it should be borne in mind that a number of artefacts may be apparent due to the fact that we have averaged over 100 simulated data sets. Nevertheless, some general conclusions can be drawn from this simulation study:

 (i) The non-spatial methods leave considerable autocorrelation in the residuals, although they appear to perform reasonably well overall.
 (ii) All models do well in the trivial constant risk case
 (iii) The BYM model achieves the highest ΔBIC and also universally high Pearson correlation for counts and is always high for relative risk fits when the fitted models do well. It does does not do well in estimation of relative risks for certain mixture models and where there is structured overdispersion. Other fitted models do badly in this situation, however.
 (iv) All models appear to perform badly in the relative risk analysis, when there is combined structured and unstructured overdispersion or heterogeneity. However, most models do uniformly well in the count analysis.
 (v) Mixture models do not recover relative risks well even when the true relative risk is from a mixture. Mixture models perform well in the count analysis but are usually bettered by the gamma and BYM models.
 (vi) Of the linear Bayes methods, the global Marshall appears to yield better results than the local version, although it leaves greater autocorrelation in residuals.
 (vii) The use of kernel smoothers cannot be recommended. However, these estimators might perform better applied to the log-relative risk. Kernel smoothers tend to work poorly in the presence of autocorrelation. In addition, the use of the Kelsall–Diggle method (applied to count data) performs very poorly. These can only be recommended for exploratory data analysis.
 (viii) With regards to robustness against misspecification, it would appear that amongst the models fitted, the BYM model with both uncorrelated and correlated overdispersion/heterogeneity included provides the greatest robustness. It is also clear that using only uncorrelated overdispersion (as in the gamma or global linear Bayes models) does not provide such robustness, and will lead to increased residual autocorrelation: a fact that may have been deduced *a priori* from the nature of the model specifications.

Further work is needed to examine small groups of true models in detail and to make more detailed comparisons between fitted models. It would also be of interest to examine the effect of changes in neighbourhood specification, edge effects and spatial residual analysis.

In this study only a small group of fitted models has been examined — many other methods could have been tested. However, we believe that the methods considered yield good indications as to the behaviour of general classes of methods, and it should be possible to relate these results to other methods based on similarities in approach. For example, the poor performance of the smoothing methods examined, may also be predicted for other smoothing methods which do not account for autocorrelation, or the likelihood structure inherent in the data.

APPENDIX

The following list describes the 154 models simulated in the study. Each numbered entry has the following description: number; description; parameter settings.

1 Constant risk

2 linear trend $a = 1$; $b = 2$

3 linear trend $a = 5$; $b = 1$

4 linear trend $a \sim N(1,1)$ $b \sim N(2,1)$

5 linear trend $a \sim N(5,1)$ $b \sim N(1,1)$

6 quadratic trend $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$

7 quadratic trend $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$

8 unstructured overdispersion $v \sim N(0,s)$ ln $s \sim N(0,0.1)$

9 unstructured overdispersion $v \sim N(0,s)$ ln $s \sim N(\ln(10), 0.1)$

10 structured overdispersion $u \sim MVN(0,K)$ ln $R \sim N(\ln(0.5),0.1)$ ln sig $^\wedge$ sq $\sim N(0,0.1)$

11 structured overdispersion $u \sim MVN(0,K)$ ln $R \sim N(\ln(0.5),0.1)$ ln sig $^\wedge$ sq $\sim N((\ln(10),0.1)$

12 structured overdispersion $u \sim MVN(0,K)$ ln $R \sim N(\ln(0.125),0.1)$ ln sig $^\wedge$ sq $\sim N(0,0.1)$

13 structured overdispersion $u \sim MVN(0,K)$ ln $R \sim N(\ln(0.125),0.1)$ ln sig $^\wedge$ sq $\sim N(\ln(10),0.1)$

14 trend + unstructured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N (0,s)$ln $s \sim N(0,0.1)$

15 trend + unstructured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$

16 trend + unstructured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$

17 trend + unstructured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$

18 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

19 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(\ln(10),0.1)$

20 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

21 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.125),0.1)$ln sig $^\wedge$ sq $\sim N(\ln(10),0.1)$

22 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

23 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(\ln(10),0.1)$

24 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.125),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

25 trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(19,1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.125),0.1)$ln sig $^\wedge$ sq $\sim N(\ln(10),0.1)$

26 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

27 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.125),0.1)$ln sig $^\wedge$ sq $\sim N(0,0.1)$

28 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0,K)$ln $R \sim N$ (ln(0.5),0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

29 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0,K)$ln $R \sim N$ (ln(0.125),0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

30 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0,K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N((0,0.1)$

31 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(10),0.1)$ $u \sim MVN(0,K)$ln $R \sim N$ (ln(0.125),0.1)ln sig ^ sq $\sim N(0,0.1)$

32 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(10),0.1)$ $u \sim MVN(0,K)$ln $R \sim N$ (ln(0.5),0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

33 structured + unstructured overdispersion $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0,K)$ln $R \sim N$ (ln(0.125),0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

34 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(0,0.1)$

35 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(0,0.1)$

36 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(\ln(10,0.1)$

37 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

38 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(0,0.1)$

39 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(0,0.1)$

40 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

41 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

42 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(0,0.1)$

43 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(0,0.1)$

44 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

45 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(0,0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

46 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(0,0.1)$

47 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(0,0.1)$

48 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

49 trend + overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ $v \sim N(0,s)$ln $s \sim N(\ln(10),0.1)$ $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125),0.1)$ln sig ^ sq $\sim N(\ln(10),0.1)$

50 mixture weights $= \{0.3,0.7\}$ components $= \{0.7,1.5\}$

51 mixture weights $= \{0.2,0.7,0.1\}$ components $= \{0.3,1.0,2.5\}$

52 mixture weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7,2.6}

53 mixture weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5}

54 mixture + structured overdispersion weights = {0.3,0.7} components = {0.7,1.5} u ˜ MVN (0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(0,0.1)

55 mixture + structured overdispersion weights = {0.3,0.7} components = {0.7,1.5} u ˜ MVN (0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

56 mixture + structured overdispersion weights = {0.3,0.7} components = {0.7,1.5} u ˜ MVN (0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(0,0.1)

57 mixture + structured overdispersion weights = {0.3,0.7} components = {0.7,1.5} u ˜ MVN (0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

58 mixture + structured overdispersion weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(0,0.1)

59 mixture + structured overdispersion weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} u ˜ MVN (0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

60 mixture + structured overdispersion weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(0,0.1)

61 mixture + structured overdispersion weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

62 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7, 2.6} u ˜ MVN(0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(0,0.1)

63 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7, 2.6} u ˜ MVN(0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

64 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7, 2.6} u ˜ MVN(0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(0,0.1)

65 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7, 2.6} u ˜ MVN (0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

66 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2, 1.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(0,0.1)

67 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2, 1.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.5),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

68 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2, 1.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(0,0.1)

69 mixture + structured overdispersion weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2, 1.5} u ˜ MVN(0,K)ln R ˜ N(ln(0.125),0.1)ln sig ^ sq ˜ N(ln(10),0.1)

70 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(2,1) d ˜ N(2,1) e ˜ N(1,1) weights = {0.3,0.7} components = {0.7,1.5}

71 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(2,1) d ˜ N(2,1) e ˜ N(1,1) weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5}

72 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(2,1) d ˜ N(2,1) e ˜ N(1,1) weights = {0.2,0.3,0.3, 0.2} components = {0.6,0.9,1.7,2.6}

73 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(2,1) d ˜ N(2,1) e ˜ N(1,1) weights = {0.2,0.3,0.3, 0.2} components = {0.7,0.9,1.2,1.5}

74 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(5,1) d ˜ N(0.01,1) e ˜ N(10,1) weights = {0.3,0.7} components = {0.7,1.5}

75 mixture + trend a ˜ N(1,1) b ˜ N(2,1) c ˜ N(5,1) d ˜ N(0.01,1) e ˜ N(10,1) weights = {0.2,0.7, 0.1} components = {0.3,1.0,2.5}

76 mixture + trend $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = {0.2,0.3, 0.3,0.2} components = {0.6,0.9,1.7,2.6}

77 mixture + trend $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5}

78 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.3,0.7} components = {0.7,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

79 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.3,0.7} components = {0.7,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

80 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.3,0.7} components = {0.7,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

81 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.3,0.7} components = {0.7,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

82 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

83 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

84 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0. 125)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

85 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.7,0.1} components = {0.3,1.0,2.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

86 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7,2.6} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

87 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7,2.6} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

88 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7,2.6} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

89 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.6,0.9,1.7,2.6} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.125)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

90 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(0,0.1)$

91 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5} $u \sim MVN(0, K)$ln $R \sim N(\ln(0.5)$, 0.1)ln sig ^ sq $\sim N(\ln(10),0.1)$

92 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.7,0.9,1.2,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln$(0.125), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

93 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(2,1)$ $d \sim N(2,1)$ $e \sim N(1,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.7,0.9,1.2,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N($ln$10),0.1)$

94 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.3,0.7\}$ components = $\{0.7,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

95 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.3,0.7\}$ components = $\{0.7,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N(10),0.1)$

96 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.3,0.7\}$ components = $\{0.7,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

97 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.3,0.7\}$ components = $\{0.7,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N($ln$(10),0.1)$

98 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.7,0.1\}$ components = $\{0.3,1.0,2.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

99 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.7,0.1\}$ components = $\{0.3,1.0,2.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N($ln$(10),0.1)$

100 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.7,0.1\}$ components = $\{0.3,1.0,2.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

101 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.7,0.1\}$ components = $\{0.3,1.0,2.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N($ln$(10),0.1)$

102 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.6,0.9,1.7,2.6\}$ $u \sim MVN(0, K)$ln $R \sim N($ln$(0.5), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

103 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.6,0.9,1.7,2.6\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N($ln$(10),0.1)$

104 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.6,0.9,1.7,2.6\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.125), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

105 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.6,0.9,1.7,2.6\}$ $u \sim MVN(0, K)$ln $R \sim N($ln$(0.125),0.1)$ln sig ^ sq $\sim N($ln$(10),0.1)$

106 mixture + trend + structured overdispersion $a \sim N(1,1)$ $b \sim N(2,1)$ $c \sim N(5,1)$ $d \sim N(0.01,1)$ $e \sim N(10,1)$ weights = $\{0.2,0.3,0.3,0.2\}$ components = $\{0.7,0.9,1.2,1.5\}$ $u \sim MVN(0, K)$ln $R \sim N($ln $(0.5), 0.1)$ln sig ^ sq $\sim N(0,0.1)$

107 mixture + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5} u ~ MVN(0, K)ln R ~ N(ln (0.5), 0.1)ln sig ^ sq ~ N(ln(10),0.1)

108 mixture + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5} u ~ MVN(0, K)ln R ~ N(ln (0.125), 0.1)ln sig ^ sq ~ N(0,0.1)

109 mixture + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) weights = {0.2,0.3,0.3,0.2} components = {0.7,0.9,1.2,1.5} u ~ MVN(0, K)ln R ~ N(ln (0.125), 0.1)ln sig ^ sq ~ N(ln(10),0.1)

110 clustering + unstructured overdispersion v ~ N(0,s)ln s ~ N(0,0.1)

111 clustering + unstructured overdispersion v ~ N(0,s)ln s ~ N(ln(10),0.1)

112 clustering + structured overdispersion u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N (0,0.1)

113 clustering + structured overdispersion u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N (ln(10),0.1)

114 clustering + structured overdispersion u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N (0,0.1)

115 clustering + structured overdispersion u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N (ln(10),0.1)

116 clustering + trend + unstructured a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N (0,s)ln s ~ N(0,0.1)

117 clustering + trend + unstructured a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N (0,s)ln s ~ N(ln(10),0.1)

118 clustering + trend + unstructured a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(0,0.1)

119 clustering + trend + unstructured a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(ln(10),0.1)

120 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) u ~ MVN(0, K)ln R ~ N(0.5),0.1)ln sig ^ sq ~ N(0,0.1)

121 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) u ~ MVN(0, K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

122 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) u ~ MVN(0, K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

123 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) u ~ MVN(0, K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

124 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) u ~ MVN(0, K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(0,0.1)

125 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) u ~ MVN(0, K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

126 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) u ~ MVN(0, K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

127 clustering + trend + structured overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(19,1) u ~ MVN(0, K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

128 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln (0.5),0.1)ln sig ^ sq ~ N(0,0.1)

129 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln (0.125),0.1)ln sig ^ sq ~ N(0,0.1)

130 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln (0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

131 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln (0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

132 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(0,0.1)

133 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

134 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

135 clustering + structured + unstructured v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

136 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(0,0.1)

137 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

138 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

139 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

140 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

141 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

142 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

143 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(2,1) d ~ N(2,1) e ~ N(1,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N (ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

144 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(0, 0.1)

145 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0, 0.1)

146 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

147 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(0,0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

148 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(0,0.1)

149 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(0,0.1)

150 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(ln(10),0.1) u ~ MVN(0,K)ln R ~ N(ln(0.5),0.1)ln sig ^ sq ~ N(ln(10),0.1)

151 clustering + trend + overdispersion a ~ N(1,1) b ~ N(2,1) c ~ N(5,1) d ~ N(0.01,1) e ~ N(10,1) v ~ N(0,s)ln s ~ N(ln(10),0.1 u ~ MVN(0,K)ln R ~ N(ln(0.125),0.1)ln sig ^ sq ~ N(ln(10),0.1)

152 Gamma $a = 1$; $b = 2$
153 Gamma $a = 2$; $b = 1$
154 Gamma $a = 1$; $b = 1$

ACKNOWLEDGEMENTS

REFERENCES

1. Cressie N, Chan NH. Spatial modelling of regional variables. *Journal of the American Statistical Association* 1989; **84**:393–401.
2. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–681.
3. Marshall R. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics* 1991; **40**: 283–294.
4. Tsutakawa RK, Shoop GL, Marienfeld C. Empirical Bayes estimation of cancer mortality ranges. *Statistics in Medicine* 1985; **4**:201–212.
5. Mollie A, Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine* 1991; **10**:95–112.
6. Schlattmann P, Böhning D, Clark A, Lawson A. Lung cancer mortality in women in Germany 1995: a case study in disease mapping. In *Disease Mapping and Risk Assessment for Public Health*, Lawson AB *et al.* (eds). Wiley: New York, 1999; Chapter 31.
7. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel J-F, Bertollini R. *Disease Mapping and Risk Assessment for Public Health*. Wiley: New York, 1999.
8. Lawson AB, Cressie NAC. Spatial statistical methods in environmental epidemiology. In *Handbook of Statistics*: *Bioenvironmental and Public Health Statistics*, vol 18, 357–396, Sen PK, Rao, CR (eds). North Holland, 2000.
9. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: flow important are priors? *Statistics in Medicine* 1995; **14**(21/22):2411–2431.
10. Best N, Waller L, Thomas A, Conlon E, Arnold R. Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6*, Bernardo J *et al.* (eds). Oxford University Press: London, 1998.
11. Schlattmann P, Böhning D. Mixture models and disease mapping. *Statistics in Medicine* 1993; **12**:1943–1950.
12. Conlon E, Louis T. Addressing multiple goals evaluating region-specific risk using Bayesian models. In *Disease Mapping and Risk Assessment for Public Health*, Lawson AB *et al.* (eds). Wiley: New York, 1989.
13. Simonoff JS. *Smoothing Methods in Statistics*. Springer: New York, 1996.
14. Kelsall J, Diggle P. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine* 1995; **14**:2335–2342.
15. Kelsall J, Diggle P. Spatial variation in risk of disease: a non-parametric binary regression approach. *Journal of the Royal Statistical Society*, Series C 1998; **47**(4):559–573.
16. Lee PM. *Bayesian Statistics*: *An Introduction*, 2nd edn. Arnold, 1997.
17. Carlin B, Louis T. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, 1996.
18. Wolpert R, Ickstadt K. Poisson/gamma random field models for spatial statistics. *Biometrika* 1998; **85**:251–267.
19. Böhning D, Schlattmann P, Lindsay. CA.MAN—Computer Assisted Analysis of Mixtures: Statistical Algorithms. *Biometrics* 1992; **48**: 283–303.
20. Bohning D. *Computer Assisted Analysis of Mixtures and Applications*. CRC Press, 1999.
21. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; **43**:1–21.
22. Wackernagel H. *Multivariate Geostatistcs*. Springer Verlag, 1995.