LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Knorr-Held, Rasser:

Bayesian Detection of Clusters and Discontinuities in Disease Maps. (REVISED, February 1999)

Sonderforschungsbereich 386, Paper 107 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

gsf

MAX-PLANCK-GESELLSCHAFT

TUM

Bayesian Detection of Clusters and Discontinuities in Disease Maps

Leonhard Knorr–Held and Günter Raßer

Institute of Statistics

University of Munich

Ludwigstr. 33, 80539 Munich

Germany

Email: leo@stat.uni-muenchen.de rasser@stat.uni-muenchen.de

Abstract

An interesting epidemiological problem is the analysis of geographical variation in rates of disease incidence or mortality. One goal of such an analysis is to detect clusters of elevated (or lowered) risk in order to identify unknown risk factors regarding the disease. We propose a nonparametric Bayesian approach for the detection of such clusters based on Green's (1995) reversible jump MCMC methodology. The prior model assumes that geographical regions can be combined in clusters with constant relative risk within a cluster. The number of clusters, the location of the clusters and the risk within each cluster is unknown. This specification can be seen as a change-point problem of variable dimension in irregular, discrete space. We illustrate our method through an analysis of oral cavity cancer mortality rates in Germany and compare the results with those obtained by the commonly used Bayesian disease mapping method of Besag, York and Mollié (1991).

Key words: Cancer atlas; Clustering; Disease mapping; Oral cavity cancer; Relative risk; Reversible jump MCMC.

1 Introduction

Statistical methods for analyzing data on disease incidence or mortality over a set of contiguous geographical regions have gained increasing interest in the last decade. It is still very common in disease mapping to display the standard mortality ratio (SMR), the ratio of observed cases y over expected cases e, for each region either on a relative or an absolute scale. However, these maps can be seriously misleading because the SMR's tend to be far more extreme in less populated regions, especially for rare diseases. Hence, regions with the least reliable data will typically draw the main visual attention. For a thorough discussion of this issue see Clayton and Bernardinelli (1992).

As an example consider Figure 1, which displays the geographical variation of the standard mortality ratios for males and oral cavity cancer, 1986–1990, in all 544 districts of Germany. This dataset will be analyzed later in Section 3. The SMR's vary between 0.16 and 2.63 with a standard deviation of the log SMR's of 0.386. However, the variation of the SMR's is reduced if we only consider highly populated regions. For example, a subsample of regions with more than 50 expected cases has a minimal SMR of 0.58 and a maximal SMR of 1.89. The standard deviation of the log SMR's is now 0.233, indicating that the SMR's tend to be more extreme in less populated regions, but this conclusion is drawn under the assumption of constant risk in the whole of Germany.

Indeed, an unknown part of the variation of the SMR's may be caused by geographically varying unobserved risk factors. For example, in Figure 1 there seem to be areas of higher risk in the north–east and in some parts of the south, especially towards the west, but a naive visual inspection can be seriously misleading and no general conclusion can be drawn from such a map. Therefore, so–called disease mapping methods have been developed to give more reliable estimates of the geographical variation of disease risk. The general goal is to identify the extra–sample variation due to unobserved heterogeneity by filtering the Poisson sample variation.

A well-known method is the empirical Bayes approach of Clayton and Kaldor (1987). Roughly speaking, this method shrinks the SMR's towards a local or a global mean where the amount of shrinkage is determined by the reliability of the data of that particular region. The two smoothing options "local" or "global" seem to be appropriate if unobserved risk factors do or do not have a spatial structure, respectively. However, one of the major goals of disease maps is to identify unobserved risk factors through the geographical variation of the disease so the spatial distribution of those unobserved factors is not known in advance. This led Besag, York and Mollié (1991) to generalize the Clayton and Kaldor method allowing

for both spatially structured and unstructured heterogeneity in one model, which was later called the convolution model by Mollié (1996).

The detection of clusters in diseases is, at first sight, a separate problem. Here the goal is to identify clusters of geographically contiguous regions with elevated (or lowered) risk. Disease clusters may occur not only for infectious diseases, but also for non-infectious diseases, where risk factors do have a spatial structure. In addition one might also be interested in detecting discontinuities in the map, i.e. suspicious differences in relative risk between adjacent regions. However, results from the convolution model are often used to visually identify disease clusters, if the estimated risks exhibit a spatial pattern (e.g. Besag, York and Mollié, 1991, Mollié, 1996). In these cases, the Markov random field (MRF) term, which represents spatially structured heterogeneity, is dominating and the SMR's are essentially spatially smoothed. For that reason, Clayton and Bernardinelli (1992) denote the MRF term the "clustering component".

This paper describes a new approach for the detection of clusters in disease maps. Technically, the method is based on reversible jump MCMC methodology (Green, 1995) and is related to the segmentation of a spatial signal, already tackled in Green. His work has been refined by Arjas and Heikkinen (1997) and Heikkinen and Arjas (1998) who use piecewise constant step functions defined through marked point processes in continuous space. However, in our application space is discrete and irregular, which calls for several changes of the model and the methodology. Basically our prior model assumes that the area considered can be divided into several clusters, i.e. sets of contiguous regions, where each cluster has constant relative risk. The number, the size and the location of the clusters, as well as the risk within each cluster, are unknown. Risks in different clusters are assumed to be independent of each other. The model is therefore able to detect spatial discontinuities. Clusters of size one are not excluded from our model which implies that the model does necessarily

smooth the SMR's. In practice it will always do so, at least to some extend, since there will always be some uncertainty whether a region forms a cluster by itself. However, the sizes of the clusters, which imply the local degree of smoothing, are variable and determined by the data, hence the smoothing is *adaptive*. This is in sharp contrast to MRF priors, where the corresponding smoothing parameter is constant and smoothing is *non-adaptive*.

The method is related to that of Schlattmann and Böhning (1993), who use mixture models within an empirical Bayes framework where each region is assigned to a component of the mixture distribution with constant relative risk. The location of the regions is, however, ignored so that members of a mixture component may be spread over the whole area. In our approach, regions are assigned to clusters with constant risk, too, but all regions in a cluster must be linked. To include location in the model we propose a construction where some regions are marked as so–called cluster centers, each of them defining a cluster. Each of the remaining regions is assigned to the cluster whose cluster center has minimal distance to the region. The distance between two regions is defined as the minimal number of boundaries that have to be crossed to move from one to the other. The construction can be seen as a modification of Voronoi tessellations (see Green, 1995) in discrete, irregular space and ensures that all regions within a cluster are linked.

The output of the algorithm is very rich and can be used for Bayesian inference in several ways. First, the point estimates (mean or median) of the risk of each region incorporate all the posterior uncertainty about the number, the location and the risk level of the clusters. Since all these are variable, the posterior mean estimate will be an average over a large number of piecewise constant step functions and can be seen as essentially nonparametric (Arjas, 1996, Heikkinen and Arjas, 1998). A similar argument holds for all other functionals of the posterior as well, for example for the posterior median. Second, the method provides a large amount of additional probabilistic information. For example, we can calculate the

probability that two or more regions belong to the same cluster. This is especially interesting for two adjacent regions where it gives an intuitive quantification for the location of discontinuities as will be illustrated in our application.

The paper is organized as follows. Section 2 describes our model and gives some features of the implementation by reversible jump MCMC. More details of the sampler are given in the Appendix. Section 3 presents results from an analysis of oral cavity cancer mortality rates of males in Germany, shown in Figure 1. We investigate the location of the clusters and discontinuities which have been identified by our method. We also compare our estimates with those obtained by the method of Besag *et al.* (1991). We close with several comments on alternative model specifications and possible extensions in Section 4.

2 The model

Suppose that data are available in the form of pairs in each of a set of n regions $i = 1, \ldots, n$ giving the number of cases y_i of the disease and the number of expected cases e_i, usually calculated by internal or external standardization with respect to confounding variables.

The general idea is that the relative risk is *constant* over a set of one or more contiguous regions. This defines a cluster $C_j \subset \{1, \ldots, n\}$, a set of contiguous regions with constant relative risk h_j. The number of clusters k is treated as unknown with $k \in \{1, \ldots, n\}$. Our cluster definition implies that the clusters $C_1, \ldots, C_k$ cover the whole area and that they do not overlap, so $C_1 \cup \ldots \cup C_k = \{1, \ldots, n\}$. Note that in the limiting case $k = 1$ there is constant relative risk over the whole area whereas for $k = n$ not even two (contiguous) regions have the same relative risk.

We postulate the usual Poisson observation model (e.g. Clayton and Bernardinelli, 1992), where y_i has Poisson distribution with mean $e_i h_j$ and h_j is the unknown relative risk in cluster

j with $i \in C_j$. Responses y_i, $i = 1, \ldots, n$, are assumed to be conditionally independent given $H_k = (h_1, \ldots, h_k)$ so the likelihood function of responses $y = (y_1, \ldots, y_n)$ can be written as

$$L(y|H_k) = \prod_{j=1}^{k} \prod_{i \in C_j} \frac{(e_i h_j)^{y_i}}{y_i!} \exp(-e_i h_j).$$

2.1 A prior model for clustering

As a first step in the definition of the clustering model, we mark k regions $g_1, \ldots, g_k$ as *cluster centers*. Each cluster center $g_j \in \{1, \ldots, n\}$ defines a cluster C_j with $g_j \in C_j$. The vector of all cluster centers $G_k = (g_1, \ldots, g_k)$ defines a *cluster configuration*, i.e. an assignment of all regions to one and only one of the clusters. For that purpose, we define a measure of distance $d(i_1, i_2)$ between two regions i_1 and i_2 as the minimal number of boundaries that have to be crossed for moving from i_1 to i_2. This distance measure can be computed from the information if any two regions are adjacent or not, which is usually given in a so–called adjacency matrix. The measure of distance d is used to assign each of the remaining $n - k$ regions to one of the clusters. Region $i \notin G_k$ will be assigned to cluster C_j if it has minimal distance to the corresponding cluster center g_j, i.e. $d(i, g_j) \leq d(i, g_l)$ for all $l \in \{1, \ldots, k\}, \quad l \neq j$. However, this definition is not yet unique, because some regions may have the same distance to two or more cluster centers. To ensure uniqueness we assign those regions to the cluster with the smallest index position of the corresponding cluster center in G_k among all cluster centers with minimal distance to region i. We therefore keep G_k non–ordered, otherwise clusters defined by cluster centers g_i with g_i small would tend to be larger in size than those with g_i large. For example, in our formulation a cluster configuration defined by a cluster center vector $G_2 = (1, 2)$ will in general be different from another one defined by $\tilde{G}_2 = (2, 1)$. Note, that the cluster centers only serve to specify a cluster configuration, they do not have any direct influence on the estimates of the relative risks.

To illustrate the flexibility of the clustering model, Figure 2 gives a cluster configuration of the 544 districts of Germany with $k = 20$. The cluster centers are marked with numbers 1 to 20, the corresponding index positions in G_k. Note that the clusters differ considerably in size and shape. Furthermore it can be seen that, indeed, all regions within each cluster are linked. It is, however, not immediately obvious that this is true in general. Now suppose there is a cluster C_j which breaks down into two or more parts, which are not linked together. Then there must be a region $i_1 \in C_j$ with some distance m to the cluster center g_j and a neighbor i_2 of i_1 with $i_2 \in C_k, k \neq j$, and distance $m-1$ to g_j. Otherwise all regions within C_j must be connected. Because i_1 is a neighbor of i_2 it follows, however, that i_1 must be in C_k and not in C_j which is contradictory to the assumption above and proves our claim.

We now specify a prior distribution for the number of clusters k, the vector of cluster centers G_k and for the vector of relative risks H_k. We assume that the prior for the number of clusters $Pr(k)$, $k = 1, \ldots, n$, is proportional to $(1-c)^k$ with a fixed parameter $c \in [0, 1)$. The limiting case $c = 0$ gives a uniform distribution on $\{1, \ldots, n\}$, whereas $c > 0$ corresponds to a truncated geometric distribution. This choice implies that the prior ratio $Pr(k+1)/Pr(k) = (1-c)$, which penalizes jumps from k to $k+1$, is constant for all k. We typically use small values for c so as to make the prior $Pr(k)$ close to "uninformative". Other choices might be more appropriate but, as Richardson and Green (1997) have noted, results with any prior for k can be converted to those corresponding to other priors without rerunning the algorithm.

For a given number of clusters k we assume that each vector of cluster centers $G_k = (g_1, \ldots, g_k)$ has equal probability

$$Pr(G_k|k) = \frac{(n-k)!}{n!}. \tag{1}$$

One could also introduce weights that take account of specific features so as to support configurations with homogeneous cluster sizes or boundary lengths, for example.

We have made extensive simulations from the prior distribution $Pr(G_k|k) \cdot Pr(k)$ described above. For example, for each region, we have calculated the average size of the cluster the region is assigned to. Figure 3 shows the results for $c = 0.02$, grouping the regions according to the number of adjacent regions. The influence of the number of adjacent regions on the average size of the cluster appears to be minimal. Hence, the degree of smoothing is approximately the same for all regions, a priori. This is in contrast to MRF priors, where there is dependence of the smoothing parameter (the marginal variance) on the number of adjacent areas, see Bernardinelli, Clayton and Montomoli (1995a). We have also calculated the prior probability for each region to form a cluster by itself as well as the probability for being together with a neighbor. These probabilities have some variation, depending mainly on the number of neighbors. In Section 3, we therefore report the corresponding posterior probabilities together with the prior probabilities.

As a prior guess for the relative risks $H_k = (h_1, \ldots, h_k)$ it seems natural to assume that they are symmetrically distributed on the log scale. We therefore adopt a normal distribution for $\log(h_j)$, $j = 1, \ldots, k$, with unknown hyperparameters μ and σ^2. For μ, we assume a diffuse prior (uniform on the whole real line) and for σ^2 a highly dispersed but proper inverse gamma distribution $\mathrm{IG}(a, b)$ with fixed parameters a and b. Independence of components of H_k yields

$$p(H_k|k, \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^k \left(\prod_{j=1}^{k} \frac{1}{h_j}\right) \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^{k} \{\log(h_j) - \mu\}^2\right]. \qquad (2)$$

Conditional independence of H_k and G_k given k defines the prior for the unknown parameters k, G_k, H_k, μ and σ^2 as the product of the prior for k times (1) times (2) times the hyperpriors $p(\mu)$ and $p(\sigma^2)$.

2.2 Implementing reversible jump MCMC

This section gives an informal description of some features of our reversible jump MCMC implementation for sampling from the posterior distribution. In each iteration of the algorithm one of the following six moves is proposed:

Birth: The number of clusters is increased by introducing an additional cluster center.

Death: The number of clusters is decreased by deleting one of the cluster centers.

Shift: One of the cluster centers is moved.

Switch: The positions of two cluster centers in G_k are switched.

Height: The relative risks $h_j, j = 1, \ldots, k$, are changed.

Hyper: The values of the hyperparameters μ and σ^2 are changed.

For a given value of k, each move is proposed with a certain probability. For some values of k certain moves are not possible, for example a death move for $k = 1$. Each move is accepted as the new state of the Markov chain with probability determined by the Metropolis–Hastings–Green ratio (Green, 1995). Below we describe some features of our implementation of these six elementary moves. More details are given in the Appendix. The main reason for choosing those moves was that they appeared to be straightforward to implement, each of them maintaining reversibility. We have included the shift and the switch move in the hope of improved mixing performance, although they seem to be not necessary. In fact, some other MCMC sampler with different proposals or different moves might be more efficient in terms of convergence, mixing or computing time but, in our experience, our algorithm gives reliable results for acceptable run lengths.

Suppose that in the current configuration k regions are marked as cluster centers. In a birth move one of the remaining $n - k$ regions is chosen randomly as a new cluster center.

The new cluster center g^* is placed randomly among all possible $k+1$ positions in the new vector of cluster centers G^*_{k+1}. A value h^* for the relative risk within the new cluster is inserted at the corresponding position in H^*_{k+1}. In a death move from $k+1$ to k, a randomly selected element of G_{k+1} is deleted. A sequence of a death and a birth move (or vice versa) is therefore able to restore the original configuration. In a shift move, first one of the cluster centers, whose neighbors are not all cluster centers by itself, is picked randomly. This cluster center g_j, say, is then shifted randomly to one of the neighbors that are not already cluster centers. The order in G_k is not changed. Note that the neighbors do not have to be members of the original cluster C_j which would in fact destroy the reversibility of the shift move. A switch move picks out two elements in G_k randomly and switches their position in G_k which will give a slightly different cluster configuration if there are distance ties. A height move proposes new values h^*_j for all elements h_j of H_k, each of them being accepted or rejected separately. Finally, in a hyper move, values of the hyperparameters μ and σ^2 are updated by samples from the corresponding full conditional distributions.

The performance of the algorithm depends on a number of implementation issues. First, the several moves should be designed to have acceptance rates not too low. For moves that involve new values h^*, we therefore use a proposal distribution that approximates the corresponding (fixed-dimension) "full conditional" (the prior for h times the relevant likelihood times a normalizing constant). This device results generally in very good acceptance rates for these moves (height, birth and - indirectly - death). Furthermore, the algorithm is now automatic, as tuning parameters are not involved. Similar proposals might be useful in many other applications of reversible jump MCMC. The shift move will have low acceptance rates, if there is very strong local information in the likelihood. Note, however, that this move is not necessary for convergence of the algorithm and could, in principle, be omitted completely.

A second problem occurs if the posterior is multimodal. This potential problem is inherent in any more complex MCMC application but seems to be of particular concern for reversible jump MCMC if only small dimension changing moves are made. If the simulated chain is trapped in one of the modes, it might be difficult for it to move to some other posterior mode, located somewhere different and clearly separated by an area of low posterior mass. This problem might be even more severe for fixed k, since the birth and death moves are known to improve mixing (Heikkinen and Arjas, 1998). We routinely start several chains with different starting configurations and compare the results. Carefully designed mode jumping moves might also be useful here but require knowledge of the location of the posterior modes.

3 Applications

3.1 Simulations

To see, how well our method works, we have analyzed several artificial datasets. In particular, we have looked how well our method reconstructs a given risk surface, how sensitive our results are to the choices for $Pr(k)$ and $p(\sigma^2)$, and how reliable our algorithm works. The results are generally encouraging and can be found in a supplement paper (Knorr-Held and Rasser, 1999). Based on these results, we recommend to use small, but positive values for c. Sensitivity with respect to $p(\sigma^2)$ was found to be small and we recommend to use $a = 1$ and $b = 0.01$ as default. Of course, sensitivity to the prior should always be studied.

3.2 Results for oral cavity cancer mortality in Germany

We now present results from an analysis of oral cavity cancer of males in Germany. The database records the population size and the number of deaths from oral cavity cancer,

stratified by 16 age bands and 544 districts for the period 1986–1990. The total number of cases is 15,466 ranging between 1 and 501 cases with a median number of 19 cases per district. The overall mortality rate is 40.9 cases per 100,000 males. We have internally standardized the raw data with respect to all 16 age bands by maximum likelihood and have calculated the corresponding standard mortality ratios which are shown in Figure 1.

To examine sensitivity with respect to $Pr(k)$ and $p(\sigma^2)$ we have used $c = 0.0, 0.01$ and 0.02 and $(a, b) = (0.25, 0.00025)$, $(1, 0.01)$ and $(5, 0.125)$ in several combinations. For $(a, b) = (1, 0.01)$, for example, there was some sensitivity for k with respect to $Pr(k)$ with a posterior median of 86 (c=0.0), 75 (c=0.01) and 64 (c=0.02) compared to a prior median of 272, 69 and 35, respectively. However, differences in the log relative risk estimates were found to be small. Results have been even more stable for different choices for $p(\sigma^2)$ with c fixed.

In the analysis presented here, we have set $a = 1$, $b = 0.01$ and $c = 0.02$. A plot of the prior and the posterior for k is given in Figure 5. The results are based on samples of 10,000 realizations, collected by saving the current state after every 10,000th basic update move after a burn–in period of 1,000,000. We have calculated autocorrelations for the corresponding relative risk samples in each region. Mixing was good with a median autocorrelation of only 0.02 for lag 1 and a maximum value of 0.35. For lag 5 and more all values have been around zero. The samples of k are shown in Figure 6. The acceptance rates were around 32% for both the birth and the death move, 26% for a shift, 46% for a switch and 95% for a change of height.

The posterior median estimates of the relative risk vary between 0.71 and 1.59. Figure 4 displays those estimates on the same scale as in Figure 1. Most striking are two large clusters of elevated relative risk, one in the north–east in Mecklenburg–West Pomerania with estimates above 1.2, the other one in the south–west with estimates above 1.5, covering

most parts of Saarland and southern Rhineland–Palatinate along the border to France. In fact, most parts of southern Germany, excluding southern Bavaria, have an elevated relative risk above 1.0.

The most important risk factors for oropharyngeal cancers are tobacco smoking and alcohol abuse (Blot et al., 1994). The Mecklenburg–West Pomerania cluster is consistent with this, because this state has the highest per capita alcohol consumption of whole Germany (Becker and Wahrendorf, 1997). Interestingly, Blot et al. (1994) note that the east–central part of France (Bas-Rhin) along the German border has the highest oral and pharyngeal incidence rate in whole Europe (1983–1987). The south–west cluster is exactly adjoining this area and might therefore continue on the other side of the border.

There are several single regions with conspicuously high risk estimates, compared to their neighbors. West Berlin (estimated relative risk of 1.36), Kiel in the very north (1.30), and Krefeld in the west (1.18) all have elevated relative risks. We have calculated the probability that each of them forms a cluster by itself. The probabilities are 0.16 for West Berlin, 0.70 for Kiel and 0.67 for Krefeld. For comparison, the median probability of all 544 regions is only 0.007. The *prior* probabilities of being alone for these regions are 0.006, 0.03 and 0.02, respectively, compared to a median prior probability of 0.014. This indicates the existence of unobserved risk factors for these regions, possibly related to a higher degree of urbanization.

An interesting feature of Figure 4 is that the map strongly retains the border between former East and West Germany, especially in the south but also for West Berlin. We have therefore calculated the probability that two regions belong to the same cluster for all 1,416 pairs of adjacent regions. Figure 7 compares the distribution of these probabilities for the former east–west border with all remaining ones by boxplots. To avoid a "state border" bias we have stratified the latter group in two subgroups where adjacent regions do or do not belong to the same state, respectively. Figure 7 gives also the corresponding plot for the

prior probabilities. Differences between these subgroups are minimal *a priori*, however, the *posterior* probabilities are lower for the former east-west border. This indicates substantial differences between East and West Germany, either in exposure to relevant risk factors or simply in data quality. There are several hints that the latter is an important factor (Becker and Wahrendorf, 1997). One reason for the apparent differences might be a lack of quality control measures in the former Democratic Republic of Germany in the process of identifying underlying diseases. For example, it might be possible that there is an underreporting of oral cavity cancer due to a relatively high rate of nonidentified cancers "of other and unspecified sites". However, it seems that noncompliance with WHO rules for the identification of underlying disease is not able to explain the differences alone. Other possible reasons are discussed in detail in Becker and Wahrendorf (1997) with relevant references.

Figure 8 displays the estimated median relative risks of this dataset by the method of Besag *et al.* (1991) with a Gaussian intrinsic prior for the spatial component. The estimates show slightly more variation with values between 0.61 and 1.71. The similarities between Figure 4 and Figure 8 are noticeable and relieving, although there are some apparent differences. In particular, Figure 8 seems to be noisier. This becomes evident from Figure 9, which displays the absolute difference in estimated log relative risk between adjacent regions. Overall, the median absolute difference using the Besag *et al.* model (0.076) is nearly three times as high as with our method (0.027). It seems that the risk variability in some parts of the map induces a considerable overall variability, because smoothing in the convolution model is non-adaptive. Our method, however, is adaptive and therefore the distribution of the absolute differences is much more skewed. An even more pronounced difference can be seen in Figure 9 for absolute differences between regions, where one of the regions has only one or two neighbors. Since the prior marginal variance of the MRF term is considerably larger, smoothing is much less pronounced here and the differences are very large. This can

also be seen from Figure 8, where regions with only one or two neighbors are often in a different risk category than their neighbors. In contrast, our method, where the amount of smoothing is approximately the same for all regions a priori (see Figure 3), does not see much evidence for such large absolute differences, apart from the Kiel cluster. We have also tried a median-based prior instead of the Gaussian in the Besag *et al.* model. The differences in log relative risk have now increased even more to rather unrealistic values for regions with only a few neighbors.

4 Concluding remarks

We have described a novel approach to disease mapping with particular emphasis on the detection of clusters and discontinuities in disease maps. We close now with a few comments on alternative model specifications and possible extensions.

Initially, we have considered a more general cluster model, where every possible partition into k clusters has equal probability a priori, as long as all regions within each cluster are linked. However, if k is treated as unknown, we need to know the number of all possible partitions, say n_k, because this number determines the prior probability $1/n_k$ of a specific partition. These probabilities enter in the prior ratio for any birth or death move. It was and still is far from obvious to us how to calculate the n_k's in irregular space. We have therefore decided to reduce the complexity of the problem by introducing cluster centers. Of course, our model has now the slightly odd feature that, for a given partition, it is difficult to derive its prior probability. But even if this probability is zero, the partition can well be approximated by an average over a set of different configurations, that are supported by our model.

Suppose now that we define a cluster configuration by selecting a few cluster centers and

assigning each of the other regions to one of the clusters based on some distance measure, just as we did. One might argue that other measures of distance as the one we propose might be more appropriate. Indeed, initially, we thought of assigning a specific point to any region, for example the centroid of the region or the location of that region's largest place. The distance between regions could then be defined as the Euclidean distance between the corresponding points. However, such a definition turned out to be not very useful, because clusters will not necessarily be connected. It is in fact rather easy to construct counterexamples, where regions, belonging to a specific cluster, are separated by regions which belong to other clusters. We therefore prefer our distance measure which ensures that clusters are connected and which does fully acknowledge the discrete nature of space.

More generally, our method might be useful for other statistical problems in discrete space. Furthermore, it can be viewed as a module in Bayesian inference for more complex data. For example, in the current context it might be desirable to include covariate information more explicitly in the model. For categorical covariates, one could introduce an additional partition model of unknown dimension (Green, 1995) for the effects of the covariate levels.

Our approach might also be useful in modelling disease risk data in time and space. Such data have been analyzed recently by Bernardinelli et al. (1995b), Waller et al. (1997) and Knorr–Held and Besag (1998). Suppose, that data (y_{it}, e_{it}) are available for n regions $1, \ldots, n$ and T time points $t = 1, \ldots, T$, say years. The obvious extension of our approach would be to define the neighbors of pixel (i, t) as the neighbors in space (all pixels (j, t) where region j is a neighbor of region i) and the neighbors in time (pixels $(i, t-1)$ and $(i, t+1)$ with obvious modifications for the endpoints $t = 1$ and $t = T$). Clusters of constant risk would then be defined over time and space. In particular, such a specification would be able to capture space–time interactions.

Acknowledgements

This research was supported by the German Science Foundation (DFG), SFB 386, and by the European Science Foundation Program on Highly Structured Stochastic Systems (HSSS). The authors express thanks to Nikolaus Becker for access to the dataset and for the computer program for producing the map of Germany, to Iris Pigeot for help on an earlier version of this manuscript, and to the associate editor and three referees for very helpful and encouraging comments.

References

Arjas, E. (1996). Discussion of paper by Hartigan, *in* J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford: University Press, pp. 221–222.

Arjas, E. and Heikkinen, J. (1997). An algorithm for nonparametric Bayesian estimation of a Poisson intensity, *Computational Statistics* **12**, 385–402.

Becker, N. and Wahrendorf, J. (1997). *Atlas of Cancer Mortality in the Federal Republic of Germany 1981–1990*, Berlin: Springer Verlag.

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995a). Bayesian estimates of disease maps: How important are priors?, *Statistics in Medicine* **14**, 2411–2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995b). Bayesian analysis of space–time variation in disease risk, *Statistics in Medicine* **14**, 2433–2443.

Besag, J. E., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion), *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Blot, W. J., Devesa, S. S., McLaughlin, J. K. and Fraumeni, J. F. (1994). Oral and pharyngeal cancers, *in* R. Doll, J. F. Fraumeni and C. S. Muir (eds), *Cancer Surveys: Trends in Cancer Incidence and Mortality, Vol. 19/20*, New York: Cold Spring Harbor Laboratory Press, pp. 23–42.

Clayton, D. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risks, *in* J. Cuzick and P. Elliot (eds), *Small Area Studies in Geographical and Environmental Epidemiology*, Oxford University Press, pp. 205–220.

Clayton, D. G. and Kaldor, J. (1987). Empirical Bayes estimates of age–standardized relative risks for use in disease mapping, *Biometrics* **43**, 671–681.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**, 711–732.

Heikkinen, J. and Arjas, E. (1998). Nonparametric Bayesian estimation of a spatial Poisson intensity, *Scandinavian Journal of Statistics* **25**, 435–450.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space, *Statistics in Medicine* **17**, 2045–2060.

Knorr-Held, L. and Rasser, G. (1999). Bayesian detection of clusters and discontinuities in disease maps: Simulations, *Discussion paper 142*, SFB 386, University Munich. Available at `www.stat.uni-muenchen.de/sfb386/publikationen.html`.

Mollié, A. (1996). Bayesian mapping of disease, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 359–379.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion), *Journal of the Royal Statistical Society B* **59**, 731–792.

Schlattmann, P. & Böhning, D. (1993). Mixture models and disease mapping, *Statistics in Medicine* **12**, 1943–1950.

Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio–temporal mapping of disease rates, *Journal of the American Statistical Association* **92**, 607–617.

Appendix: Details of the sampler

Suppose a cluster configuration with k clusters is defined by a vector of cluster centers $G_k = (g_1, \ldots, g_k)$ and a vector of relative risks $H_k = (h_1, \ldots, h_k)$. In each step of the

algorithm one of the six moves birth, death, shift, switch, height and hyper is proposed with probability $r_B(k)$, $r_D(k)$, $r_{Sh}(k)$, $r_{Sw}(k)$, $r_{He}(k)$ and $r_{Hy}(k)$, respectively. These probabilities have been chosen as $r_B(k) = r_D(k) = 0.4$ and $r_{Sh}(k) = r_{Sw}(k) = r_{He}(k) = r_{Hy}(k) = 0.05$ for $k \in \{2, \ldots, n-1\}$ with appropriate changes for the endpoint cases.

The six moves are now implemented as follows:

1. Birth: A uniformly distributed random variable on all $n - k$ regions, which are not cluster centers, determines the new cluster center g^*. A second uniformly distributed random variable j on $\{1, \ldots, k+1\}$ determines the position of g^* in G^*_{k+1}. A value h^* is generated and inserted into H^*_{k+1} at the corresponding position. The proposal $h^* = h^*_j$ is drawn from a gamma distribution

$$h^*_j \sim \mathrm{G}\left(y_j + \frac{\tilde{\mu}^2}{\tilde{\sigma}^2}, e_j + \frac{\tilde{\mu}}{\tilde{\sigma}^2}\right), \tag{3}$$

where $e_j = \sum_{i \in C^*_j} e_i$, $y_j = \sum_{i \in C^*_j} y_i$, $\tilde{\mu} = \exp(\mu + 0.5\sigma^2)$ and $\tilde{\sigma}^2 = \exp(\sigma^2) \cdot (\exp(\sigma^2) - 1) \cdot \exp(2\mu)$. This proposal distribution is an approximation of the (normalized) "full conditional" $\prod_{i \in C^*_j} L(y_i|h_j) \cdot p(h_j)$, where the lognormal prior $p(h_j)$ is replaced by a gamma distribution $G(\tilde{\mu}^2/\tilde{\sigma}^2, \tilde{\mu}/\tilde{\sigma}^2)$ with the same mean and variance. The birth step is accepted with probability $\alpha = \min\{1, \mathcal{A} \cdot \mathcal{P} \cdot \mathcal{L} \cdot \mathcal{J}\}$, where $\mathcal{A} = Pr(k+1)/Pr(k) \cdot p(h^*)/(n-k)$ is the prior ratio, $\mathcal{P} = r_D(k+1)/r_B(k) \cdot (n-k)/q(h^*)$ is the proposal ratio, $\mathcal{L}$ is the likelihood ratio and $\mathcal{J} = 1$ is the Jacobian. Here $q(h^*)$ denotes the density of the proposal distribution (3), evaluated at h^*.

2. Death: For a death move from $k+1$ to k a uniformly distributed random variable j on $\{1, \ldots, k+1\}$ is generated which determines the cluster center g_j and the corresponding relative risk h_j which are then removed from G_{k+1} and H_{k+1} respectively. The acceptance probability for the death move has the same form as for the corresponding birth move with all ratio terms inverted.

3. Shift: Among the k current cluster centers there are $n(G_k)$ cluster centers which do not only have cluster centers as neighbors. An uniformly distributed random variable j on $\{1, \ldots, n(G_k)\}$ determines a cluster center g_j with $m(g_j)$ "free" neighbors. A second uniformly distributed random variable on $\{1, \ldots, m(g_j)\}$ determines the new cluster center g_j^* which replaces g_j in G_k. The shift step is accepted with probability $\alpha = \min\left\{1, \mathcal{L} \cdot n(G_k)/n(G_k^*) \cdot m(g_j)/m(g_j^*)\right\}$.

4. Switch: For a switch move two random variables i and j, uniformly distributed on $\{1, \ldots, k\}$ with $i \neq j$, are generated. The positions i and j of the corresponding cluster centers g_i and g_j in G_k are now switched. Only the likelihood ratio $\mathcal{L}$ enters in the acceptance probability for the switch move.

5. Height: For each cluster $j = \{1, \ldots, k\}$ a new value h_j^* is proposed from (3) and eventually accepted or rejected separately. The acceptance probability is $\alpha = \min\left\{1, \mathcal{L} \cdot p(h_j^*)/p(h_j) \cdot q(h_j)/q(h_j^*)\right\}$.

6. Hyper: To change the values for μ and σ^2 we use two subsequent Gibbs steps (hence $\alpha = 1$), drawing random variables from the corresponding full conditionals

$$\mu|. \sim \mathrm{N}\left(\frac{1}{k}\sum_{j=1}^{k}\log(h_j),\ \frac{1}{k}\sigma^2\right) \quad \text{and} \quad \sigma^2|. \sim \mathrm{IG}\left(a + \frac{k}{2},\ b + \frac{1}{2}\sum_{j=1}^{k}\{\log(h_j) - \mu\}^2\right).$$

Note that for moves 1-5, the likelihood ratio $\mathcal{L}$ has to be evaluated only for those regions, where the relative risk has changed in the proposal. For example, in a birth move, $\mathcal{L}$ has to be evaluated only for the regions in the new cluster and in a death move only those regions enter in the likelihood ratio that are part of the cluster, which is supposed to be removed.

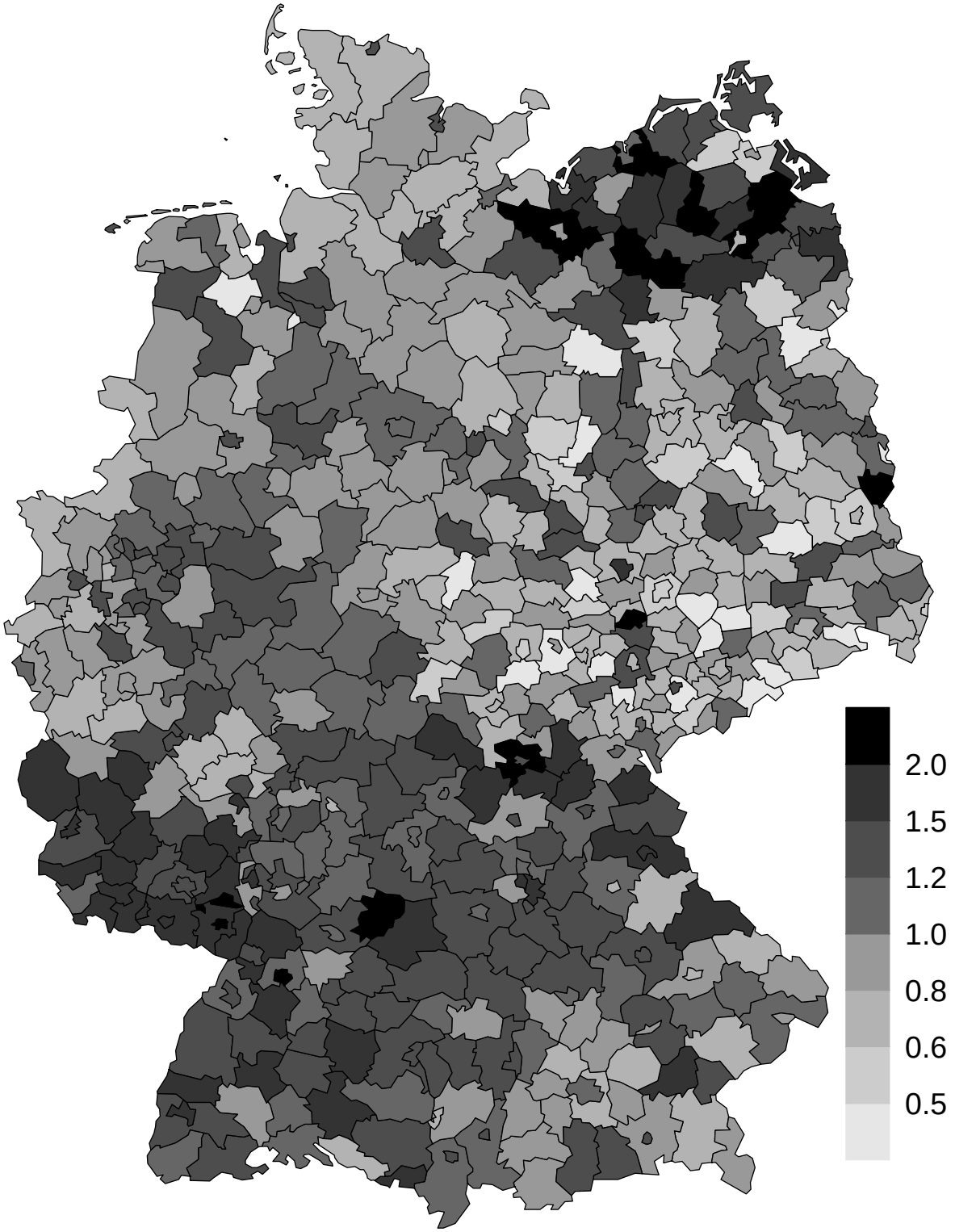

Figure 1: Standard mortality ratios for oral cavity cancer of males in Germany.

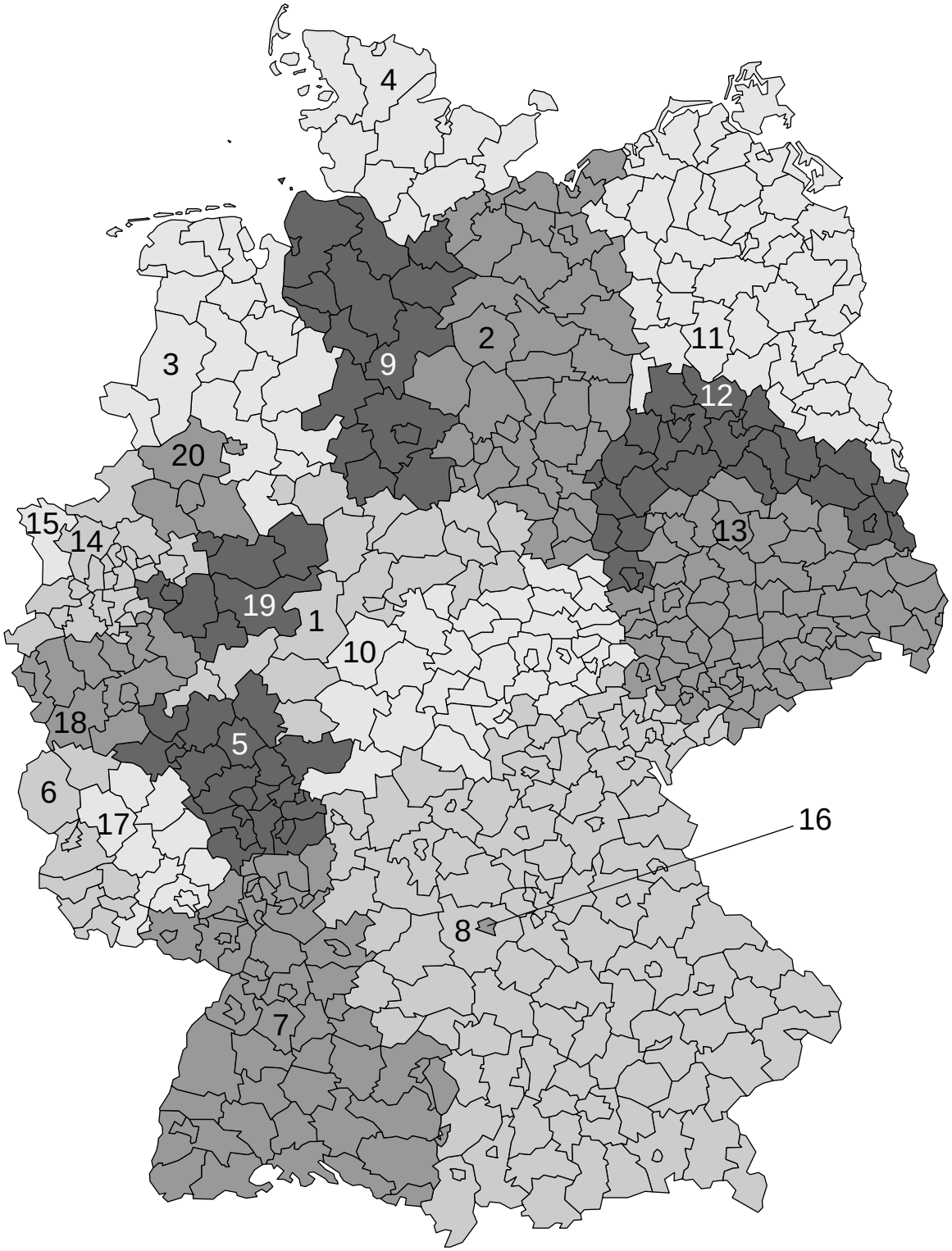

Figure 2: A cluster configuration for Germany with $k = 20$.

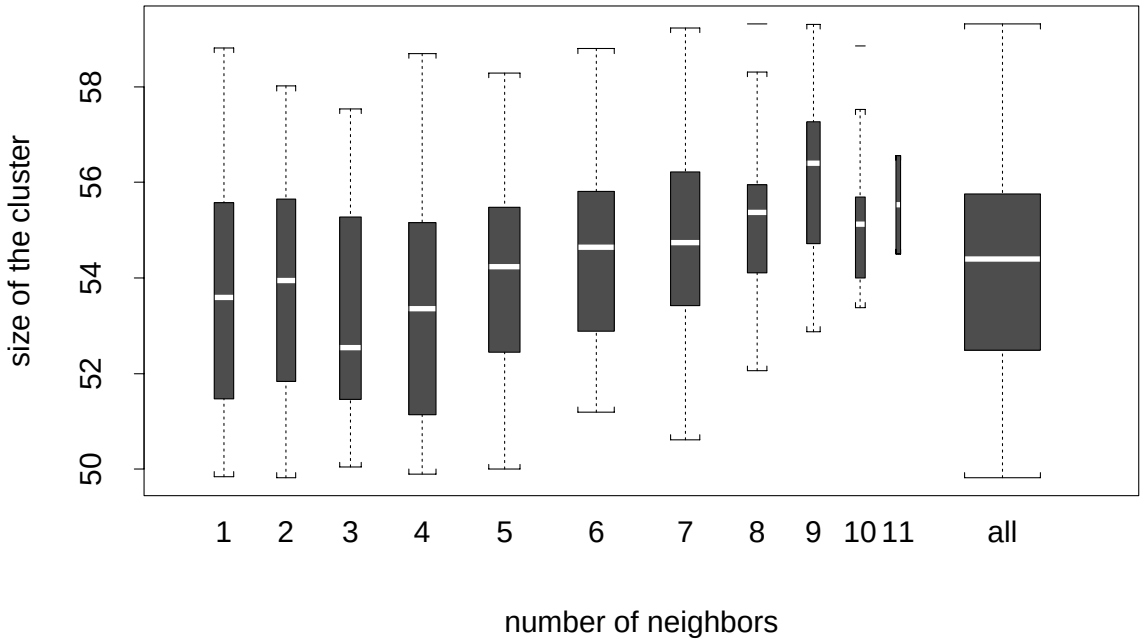

Figure 3: The average size of the cluster the region is assigned to a priori. Grouping is done with respect to the number of neighbors.

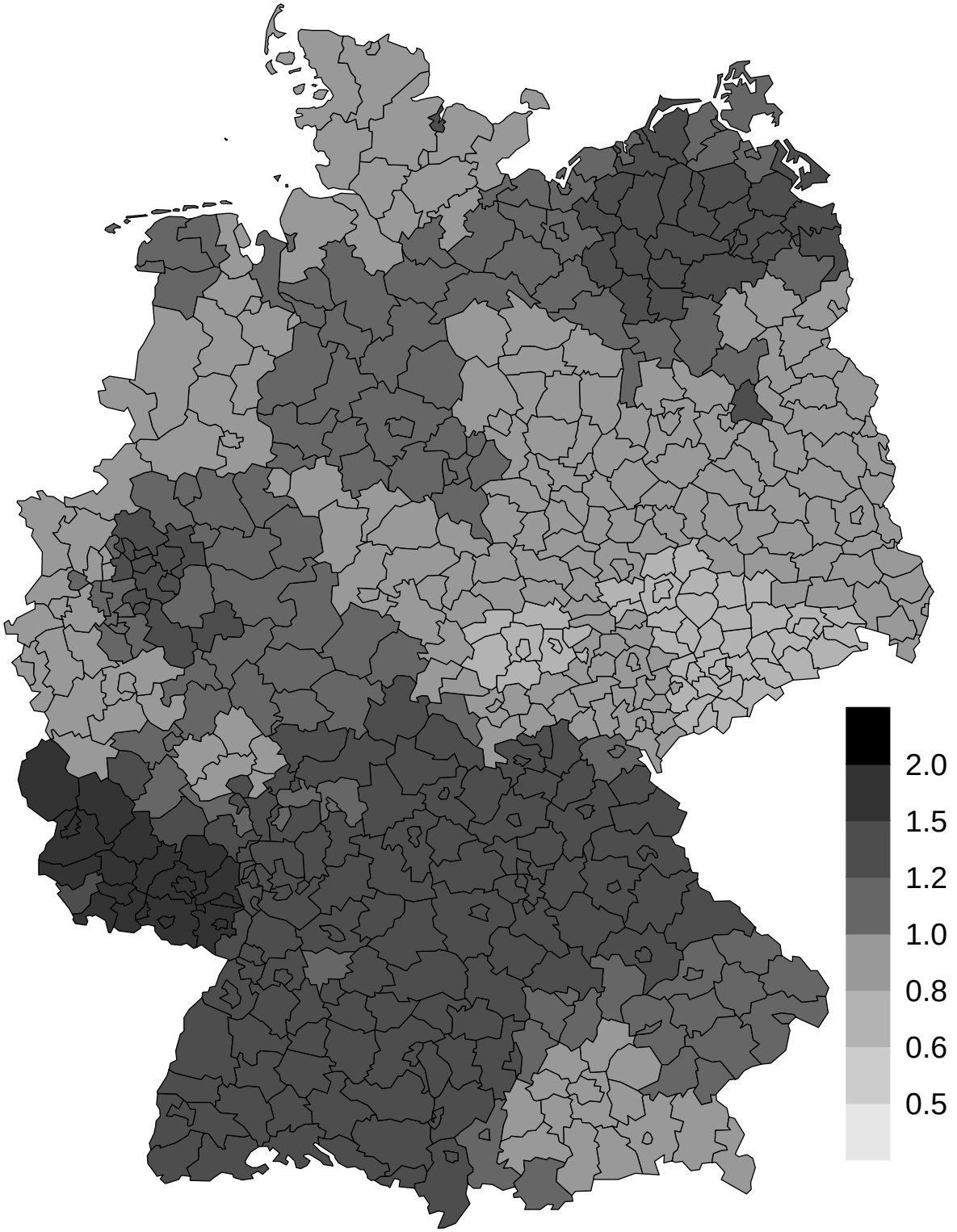

Figure 4: Estimated median relative risks for oral cavity cancer of males in Germany using our reversible jump MCMC algorithm.

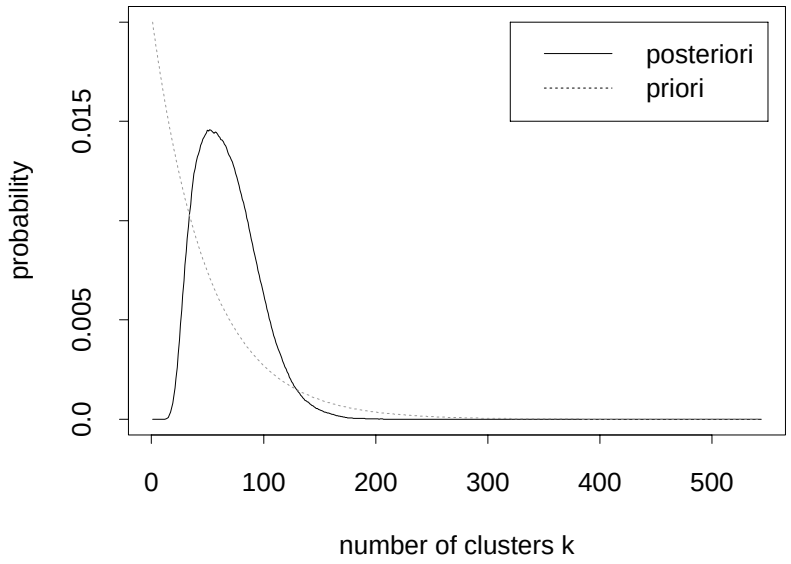

Figure 5: Prior and posterior distribution for the number of clusters k.

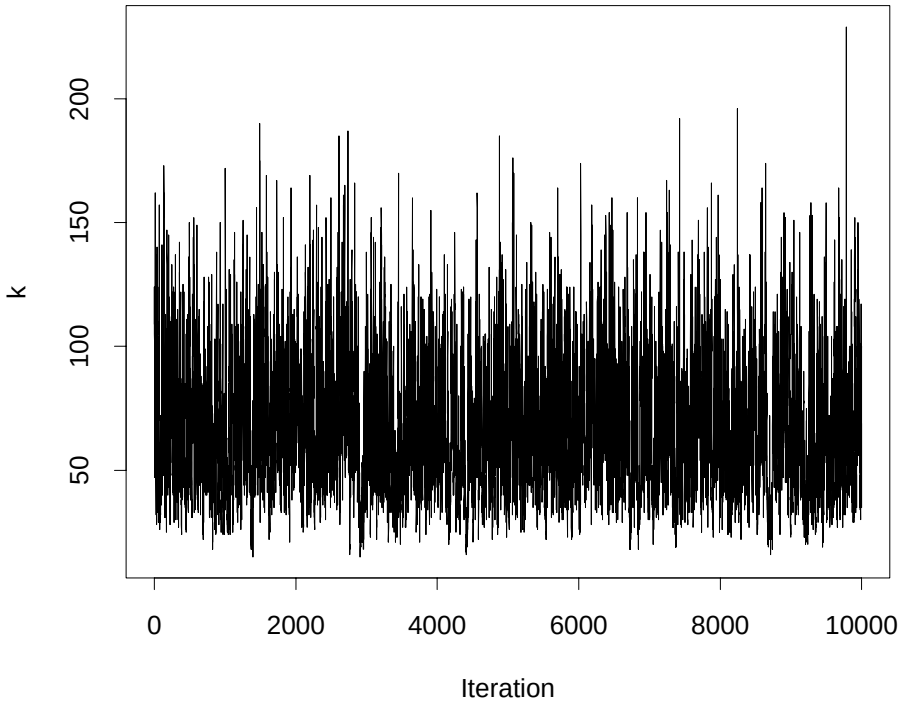

Figure 6: Chain for k versus iteration number.

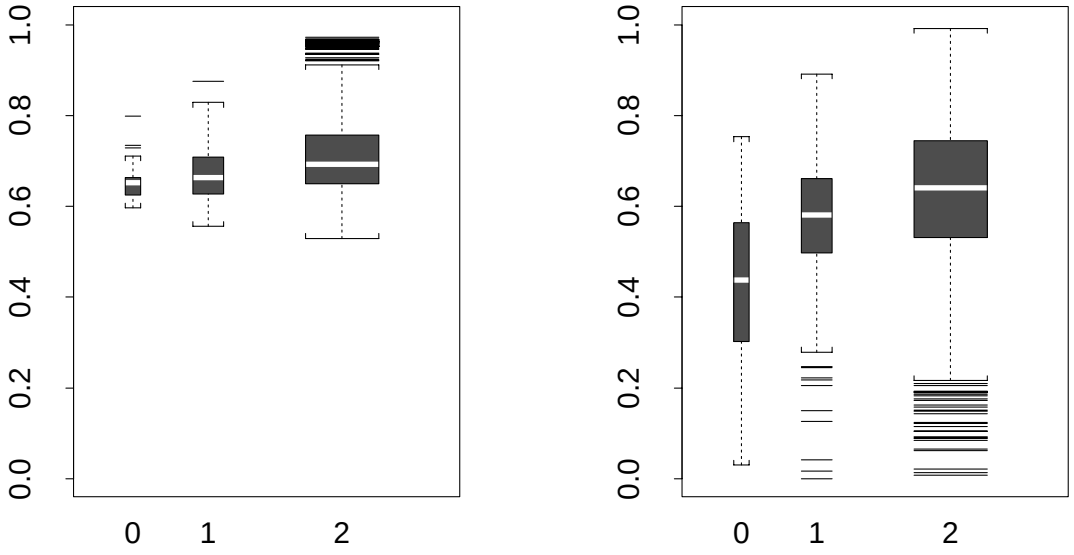

Figure 7: Boxplots of the prior (left) and posterior (right) probabilities that adjacent districts are within the same cluster for 0: former east–west border, 1: all other boundaries between different states, 2: boundaries within states. The width of the boxes is proportional to the number of observations.

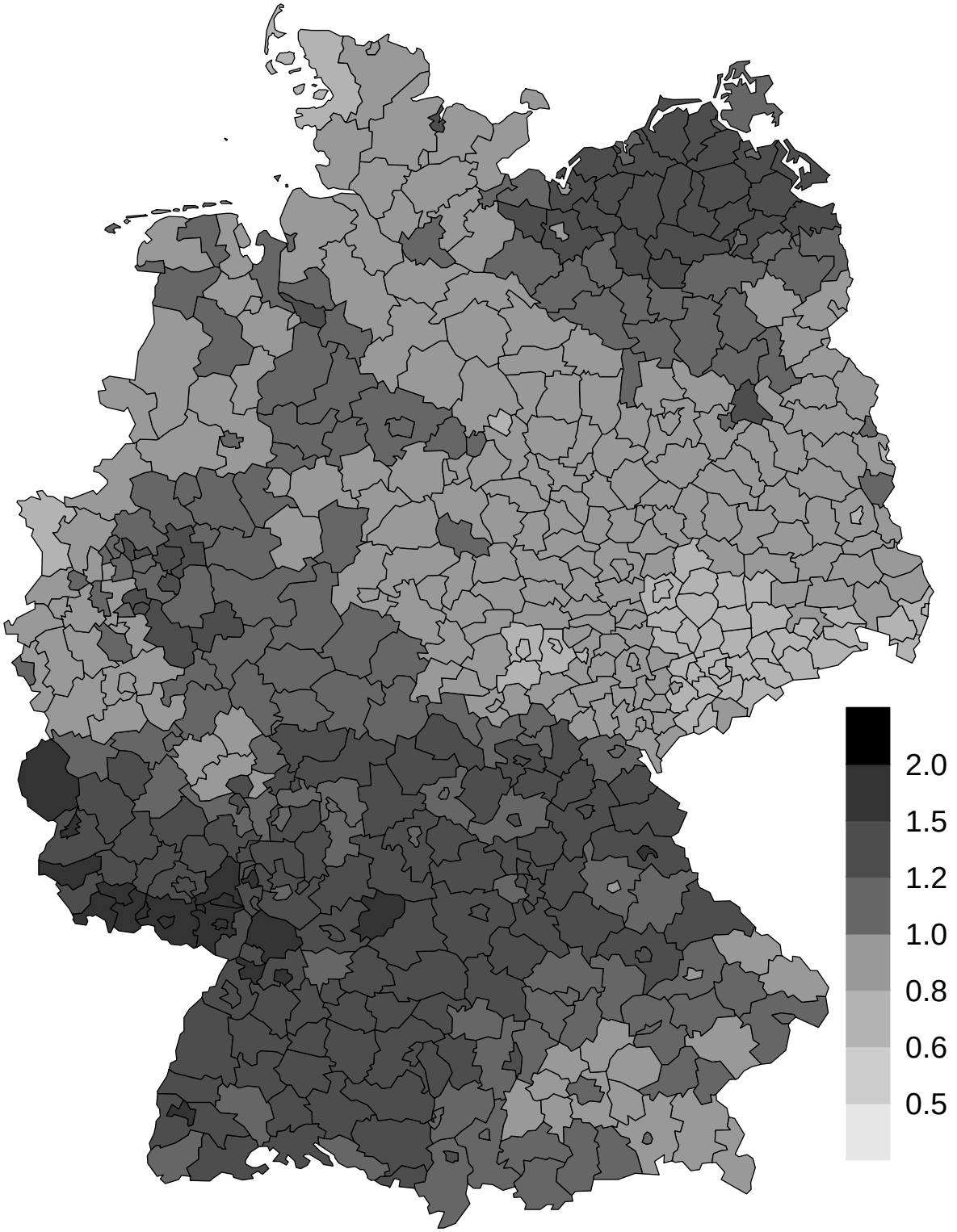

Figure 8: Estimated median relative risks for oral cavity cancer of males in Germany with the method of Besag *et al.*

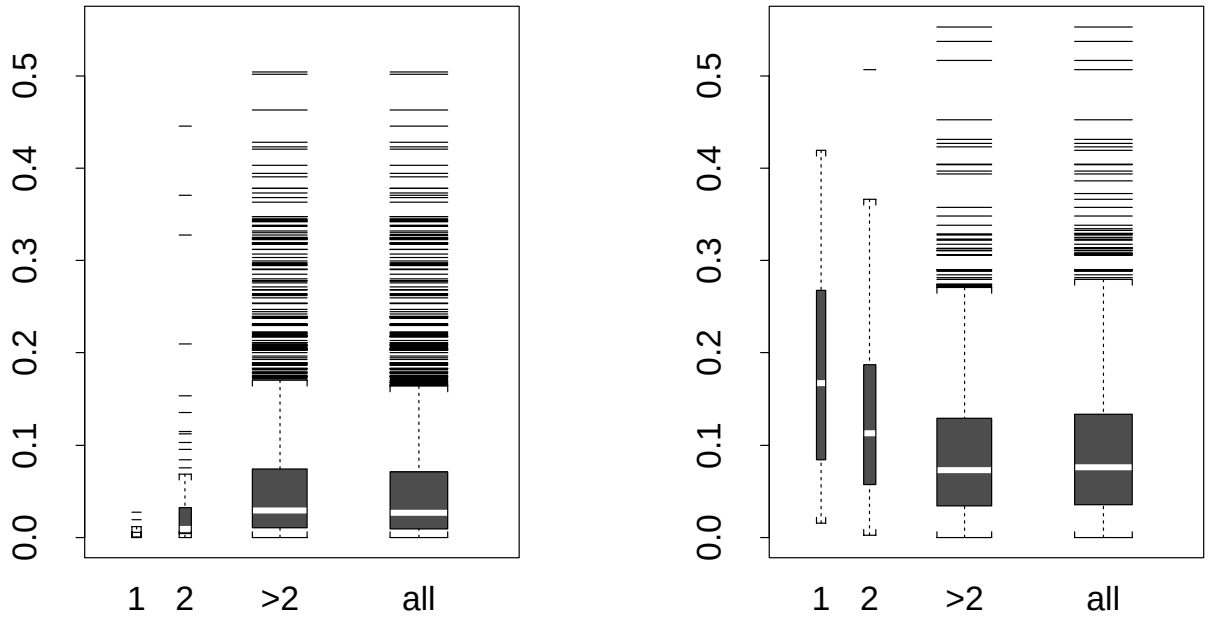

Figure 9: Boxplots of the absolute difference in log relative risk between adjacent districts. Left panel: Our method. Right panel: Method of Besag *et al.* The minimum of the number of neighbors of the two adjacent districts is used for grouping (1, 2, > 2).