Ann. Inst. Statist. Math.
Vol. 43, No. 1, 1–59 (1991)

BAYESIAN IMAGE RESTORATION, WITH TWO APPLICATIONS IN SPATIAL STATISTICS* **

JULIAN BESAG[1]***, JEREMY YORK[1] AND ANNIE MOLLIÉ[2]

[1] *Department of Statistics GN-22, University of Washington, Seattle, WA 98195, U.S.A.*
[2] *Institut Gustave Roussy, INSERM U287, 94805 Villejuif Cedex, France*

(Received November 7, 1989; revised June 18, 1990)

Abstract. There has been much recent interest in Bayesian image analysis, including such topics as removal of blur and noise, detection of object boundaries, classification of textures, and reconstruction of two- or three-dimensional scenes from noisy lower-dimensional views. Perhaps the most straightforward task is that of image restoration, though it is often suggested that this is an area of relatively minor practical importance. The present paper argues the contrary, since many problems in the analysis of spatial data can be interpreted as problems of image restoration. Furthermore, the amounts of data involved allow routine use of computer intensive methods, such as the Gibbs sampler, that are not yet practicable for conventional images. Two examples are given, one in archeology, the other in epidemiology. These are preceded by a partial review of pixel-based Bayesian image analysis.

Key words and phrases: Bayesian restoration, image analysis, spatial statistics, Gibbs sampler, archeology, epidemiology.

1. Introduction

Bayesian image analysis adopts explicit probability models to incorporate general and scene-specific prior knowledge into the processing of degraded images and aims to provide a unified framework within which a wide variety of tasks can be tackled. Its beginnings can be found in a short note of Besag (1983) and, much more significantly, in Grenander (1983) and Geman, S. and Geman, D. (1984).

---

* An earlier version of this article was presented at the symposium on the Analysis of Statistical Information held in the Institute of Statistical Mathematics, Tokyo during December 5–8, 1989.

** This research was carried out partly at the University of Durham, U.K., with the support of an award by the Complex Stochastic Systems Initiative of the Science and Engineering Research Council.

*** Now at Department of Mathematics and Statistics, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, U.K.

The present paper emphasizes *restoration* of images, sometimes considered as a relatively unimportant topic in practice. However, we believe there are many applications, perhaps not conventionally associated with image analysis, where restoration has an important role. In Section 2, we provide an introduction to pixel-based image analysis from a Bayesian perspective, with emphasis on the Gibbs sampler as an inference machine. Sections 3 and 4 describe two applications in spatial statistics, one in archeology, where the aim is to find sites of previous human activity by examining the phosphate level in the soil, and the second in epidemiology, where the risk from a disease over contiguous administrative zones must be estimated from noisy observed incidence or mortality rates.

2. Bayesian formulation

We suppose that a set of records $y = \{y_k : k \in T\}$ is generated by stochastic degradation of a true pixel image $x = \{x_i : i \in S\}$. Here we interpret the term "pixel" liberally, allowing pixel arrays that have no direct connection with "picture elements" and that may be regularly or irregularly distributed. The finite sets S and T may be identical, as in our two applications (except that in Section 3 there are several missing observations), or only loosely related. Thus, in Section 4, x_i represents the underlying log relative risk in zone i and y_i is the corresponding observed incidence or mortality rate of the disease. On the other hand, in single photon emission computed tomography (SPECT), y_k is the Poisson-distributed photon count, registered in the k-th of a bank of detectors surrounding the irradiated tissue, and the task is to *reconstruct* the mean intensity x_i in each pixel i of an arbitrary regular array superimposed on the underlying image; see Geman, S. and McClure (1987), Green (1990) for detailed Bayesian accounts.

In this paper, we confine attention to the estimation of x or of functionals of x, together with an assessment of uncertainty. However, it should be remarked that more sophisticated formulations can be adopted in which the pixel image is augmented by conceptual image attributes that are often of more fundamental importance. For example, in remote sensing of the earth's surface by satellite, each x_i is usually multivariate, with components in several or many spectral bands, and the final aim is not to *restore* the x_i's from the corresponding y_i's but typically to produce a *classification* of pixels into land use or crop type selected from a prescribed list. Similarly, in computer vision, the goal is rarely that of restoring pixel intensities, which may in any case be recorded with near perfection, but to *recognize* objects in the scene or demarcate them by *boundary maps* or perform some other higher-level task. The reader is referred to Geman, S. and Geman, D. (1984), Geman, D. and Geman, S. (1986), Geman, S. and Graffigne (1987), Chow *et al.* (1988) and Geman, D. *et al.* (1990) for the Bayesian approach to such problems, including examples; Besag (1989) provides a partial review.*

As usual, the first stage of a Bayesian analysis is to specify a prior probability density $p(x)$ for each x. We do not necessarily require that typical realizations of $\{p(x)\}$ should resemble the true scene but the distribution should at least support

* A superb review will appear in Geman, D. (1991). See references in authors' reply to Discussion.

the local regularities that are believed to exist. In particular, we usually anticipate that nearby pixel values are likely to be more similar than those further apart. In assessing the local behavior of any prior distribution, the most useful characteristic is the conditional density $p_i(x_i \mid \cdots)$ of x_i occurring at i, given all other pixel values. Usually we want this to depend only on the values at a few pixels in the immediate vicinity of i; these pixels constitute the "neighbors" ∂i of i in the terminology of Markov random fields (Besag (1974)). We postpone further discussion until the discrete and continuous examples in Sections 3 and 4, respectively. However, note that $\{p(x)\}$ may contain unspecified hyperparameters that need to be estimated in addition to x.

The second ingredient of the Bayesian formulation is of course the likelihood $l(y \mid x)$ of an image x for observed records y. This is usually determined by conventional statistical modeling. In many applications, the y_i's can be assumed to be conditionally independent given x and, when $S = T$, y_i may depend only on x_i with a common density f, so that

$$l(y \mid x) = \prod_{i \in S} f(y_i \mid x_i). \tag{2.1}$$

This occurs in both Sections 3 and 4 but we make some further remarks later in this section. Note that $l(y \mid x)$ may introduce some new parameters.

If we assume for the moment that the only unknowns are the x_i's, then inferences about x should be based on the posterior density of x given y; that is,

$$P(x \mid y) \propto l(y \mid x)p(x). \tag{2.2}$$

The most obvious Bayesian point estimate of x is that which maximizes (2.2), namely the maximum *a posteriori* (m.a.p.) estimate x^* of x. This is attractive when (2.2) has a unique maximum but loses its appeal and is extremely difficult to locate if there are many local maxima, as is often the case. Furthermore, the determination of x^* provides no assessment of precision, and the m.a.p. estimate of an arbitrary functional $g(x)$ is not $g(x^*)$ but requires fresh calculation.

As a general rule, we prefer to make inferences empirically by collecting many realizations from the posterior distribution (2.2), using a variant of Metropolis' method called the Gibbs sampler by Geman, S. and Geman, D. (1984). This enables us to tap a major strength of the Bayesian approach, in that it is not concerned merely with point estimates. For example, in restoring a continuous-intensity image, an interval estimate can be assigned to each pixel, indicating the precision of the restoration; in binary classification, a posterior probability can be ascribed to each pixel, rather than mere presence or absence of the attribute; in more general classification, the total area attributed to each class can be supplemented by an interval estimate; and the availability of a catalogue of realizations from the posterior distribution is in itself a valuable aid to visual understanding.

The principle of the Gibbs sampler is very simple. Each image pixel is addressed in turn and, when at pixel i, the current value there is replaced by a new one sampled randomly from the associated univariate conditional density $P(x_i \mid x_{-i}, y)$ given all other current pixel values x_{-i} and the fixed records y.

When each x_i has been updated, a single cycle of the algorithm is complete, as is one step of a Markov chain with stationary transition probabilities. The limit distribution of this chain must be consistent with all the individual conditional distributions and hence with the joint distribution $\{P(x \mid y)\}$ that they determine through the Brook expansion (Besag (1974)). Note that

$$P(x_i \mid x_{-i}, y) \propto l(y \mid x) p_i(x_i \mid \cdots) \tag{2.3}$$

and that only the dependence on x_i in $l(y \mid x)$ is relevant. Thus, if (2.1) holds, $l(y \mid x)$ in (2.3) is substituted simply by $f(y_i \mid x_i)$; whereas in tomography, where each x_i influences very many y_k's, the algorithm becomes ponderous. An intermediate situation occurs when $S = T$ but blur is present (Besag (1986)). However, even if (2.1) holds, it is evident that for "genuine" images, typically containing around 10^5 or 10^6 pixels, it is not yet feasible to obtain large numbers of approximately independent realizations from $\{P(x \mid y)\}$; there may also be storage problems. On the other hand, in unconventional applications, such as those described in Sections 3 and 4, the number of pixels is often no more than 10^3 and may be substantially less. In such situations, the Gibbs sampler is already a workable and powerful tool.

As regards the estimation of additional parameters in $p(x)$ and in $l(y \mid x)$, we illustrate two approaches in Sections 3 and 4. The first is *ad hoc* and avoids awkward computational problems, whereas the second is philosophically more satisfactory but does not always allow easy implementation.

3. Location of archeological sites

Enhanced soil phosphate content, the result of decomposition of organic matter, is often found at sites of known archeological activity. Thus, measurements of phosphate concentration over a study region can provide a useful aid in locating sites that are already known to exist.

Consider a rectangular grid of points ("pixels"), labeled $i = 1, 2, \ldots, n$, at each of which a measurement y_i is available. Suppose $x_i = 1$ or 0, according to whether there is or is not previous activity at i. Buck *et al.* (1988) describe the use of Bayesian change-point analysis to estimate the classification x from the data y, and illustrate their methodology on a 16×16 grid of measurements taken at 10 m intervals in a recent Laconia Survey in Greece. Here, we adopt an image analysis formulation not only to produce a classification, which might in any case be accomplished in many other ways, but also to provide approximate posterior probabilities of previous activity.

As a prior distribution for x, we adopt the somewhat simplistic binary Markov random field, $p(x;\ \beta) \propto e^{\beta v}$, where v denotes the number of like-like adjacencies, horizontally, vertically and diagonally, and β is an unknown parameter. The conditional probability of x_i occurring at i, given all other x_j's, is then

$$p_i(x_i \mid \cdots) \propto \exp\{\beta u_i(x_i)\}, \qquad x_i = 0,\ 1,$$

where $u_i(x_i)$ denotes the number of pixels adjacent to i (neighbors) having value x_i. Thus, $\beta > 0$ encourages any pixel to adopt the value taken by the majority of its neighbors.

As regards the records, we follow Buck *et al.* (1988) in assuming that the logarithms of phosphate concentrations $y_1, \ldots, y_n$ are conditionally independent, given x, and have Gaussian distributions with means $\mu(x_i)$ and common variance κ. Preliminary examination of the data suggested approximate means $\mu(0) = 4.0$ and $\mu(1) = 4.5$ but κ is more difficult to estimate in the absence of the x_i's and is treated as an unknown parameter. Naive classification therefore assigns $x_i = 1$ to pixels i for which $y_i > 4.25$ and $x_i = 0$ otherwise. The result is shown in the first panel of Fig. 1; question marks identify nine missing values at pixels $i \in S \backslash T$. If the naive classification is regarded as correct, then $\hat{\kappa} = 0.310$ is the maximum likelihood estimate of κ and $\hat{\beta} = 0.36$ is the maximum pseudo-likelihood estimate (Besag (1975)) of β; note that, in this particular case, β could have been estimated by (large sample) maximum likelihood but this would not be feasible in more complicated models and may not be desirable for reasons discussed in Besag (1986, 1989).

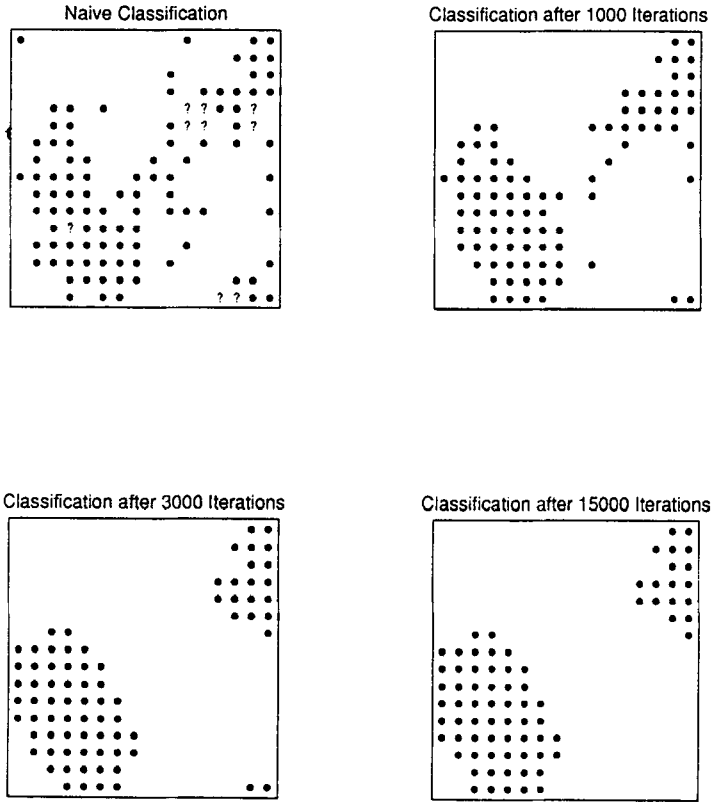

Fig. 1. Classification of sites into high and low phosphate concentrations.

The posterior probability of x given y, with the above values of β and κ, is

$$P(x \mid y) \propto \exp\left\{\hat{\beta} v - \frac{1}{2\hat{\kappa}} \sum_{i \in T} [y_i - \mu(x_i)]^2\right\}. \tag{3.1}$$

Several methods of classification based on (3.1) are available, including exact maximum *a posteriori* estimation (Greig *et al.* (1989)), iterated conditional modes (Besag (1983, 1986)) and iterated conditional expectations (Owen (1989)). However, here we prefer to (approximately) maximize the posterior marginal probabilities $P(x_i \mid y)$ of the individual x_i's using the Gibbs sampler, as first proposed by Grenander (1983). Note that if all facets of the model were correct, this choice would minimize the expected number of misclassifications. Information accumulated from 1000 iterations of the Gibbs sampler produced the classification shown in the second panel of Fig. 1 and represents the majority verdict at each pixel. New estimates of κ and β were then calculated as before and used in a further 1000 cycles of the Gibbs sampler, and so on for a total of fifteen iterations. The third and fourth panels of Fig. 1 show the classification after 3000 and 15000 cycles; the final estimates of κ and β were $\hat{\kappa} = 0.438$ and $\hat{\beta} = 1.04$. More informatively, Fig. 2 provides the estimated posterior probability that $x_i = 1$ for each pixel. Note, for example, that evidence of activity in the bottom right-hand corner is no longer suppressed but that any spatial procedure is necessarily of doubtful value for border pixels, a problem compounded here by the missing data. Overall, we feel that both Figs. 1 and 2 provide useful information for a modest amount of computing and that there are many other situations in which a similar analysis would be helpful. For a different Bayesian perspective, using fuzzy membership models, see Kent and Mardia (1988).

17	2	1	0	0	0	1	2	2	2	4	0	0	36	96	99
3	0	0	0	0	0	0	0	0	0	0	0	3	55	97	100
1	0	0	0	0	0	0	0	0	1	0	0	17	47	88	97
1	0	0	1	0	0	0	0	0	1	1	17	55	74	83	89
3	1	2	1	0	1	0	0	0	1	1	12	77	69	71	69
10	11	16	13	1	0	0	0	0	2	1	6	19	45	56	57
18	41	63	51	4	1	0	0	0	1	1	9	2	9	11	99
64	76	71	96	93	13	2	0	1	1	13	1	0	0	2	5
82	90	97	100	96	52	9	4	1	1	0	0	0	0	0	13
82	100	100	100	98	75	33	12	1	2	0	0	0	0	0	0
82	98	100	100	100	95	57	21	0	0	1	0	0	0	0	3
73	93	100	100	100	100	87	42	1	0	0	0	0	0	0	1
61	91	99	100	100	100	94	62	3	0	1	0	0	0	0	1
49	73	94	99	99	99	94	55	4	9	0	0	0	0	1	7
41	42	77	96	100	99	88	50	2	0	0	1	1	3	10	16
35	37	56	87	90	93	85	32	1	0	1	4	6	10	44	45

Fig. 2. Estimated posterior probabilities of previous human activity.

4. Mapping the risk from a disease

The mapping of risk for a particular disease, based on observed incidence (or mortality) rates in a moderately large number of contiguous administrative zones ("pixels"), is of importance in the production of cancer atlases and elsewhere. When the disease is sufficiently specific, chance fluctuations in the correspondingly small counts imply that maps based directly on the raw data are at best difficult to interpret and are often misleading when the quantities of real interest are the underlying risks. There are then advantages in applying some form of smoothing,

which may or may not involve a spatial component, and in providing point and interval estimates for the risks.

Let x_i denote the unknown log relative risk in zone i ($i = 1, 2, \ldots, n$) and y_i the corresponding observed number of cases of (or deaths from) the disease during the study period. When the disease is non-contagious and rare, it is usually reasonable to assume that the y_i's, given the x_i's, are independent Poisson variates with means $c_i e^{x_i}$, where c_i is the expected number of cases in zone i assuming constant risk; i.e. based only on the overall incidence rate and the (age-adjusted) population at risk in zone i. We adopt the formulation $x = t + u + v$ for the x_i's. Here t is a standard term, associated with measured covariates that are known or suspected to be relevant to the disease, and is usually in the guise of a linear model $t = A\theta$ with at least A known. The additional terms, u and v, can be interpreted as surrogates for unknown or unobserved covariates; the u_i's represent variables that, if observed, would display substantial spatial structure in that the values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones, whereas the v_i's represent unstructured variables. The inclusion of v is due to Breslow (1984), who noted strong empirical evidence of extra-Poisson variation under the basic model $x = A\theta$. The further inclusion of u is very close in spirit to Clayton and Kaldor (1987) but note that there is a slight logical inconsistency in their detailed formulation; see also, Besag and Mollié (1989) and Mollié (1990). In practice, it will often be the case that either u or v dominates the other but which one will not usually be known in advance. If u, then the estimated risks will display spatial structure; if v, then the effect will be to shrink the estimated risks towards the overall mean. Henceforth, for simplicity and because nothing new is lost, we shall ignore t, though measured covariates have been included in some of our practical investigations.

We must now formulate our prior beliefs concerning x as a joint distribution for u and v. In the absence of other information, we assume independence of u and v and that v is a realization of Gaussian white noise with unknown variance λ. For u, we choose a density from among the family

$$p(u) \propto \exp\left\{-\sum_{i<j} w_{ij}\phi(u_i - u_j)\right\}, \quad u \in \mathcal{R}^n, \tag{4.1}$$

based only on pairwise differences among the u_i's; here the w_{ij}'s are prescribed non-negative weights, with $w_{ij} = 0$ unless i and j are contiguous zones, and $\phi(z)$ is a specified even function of z, increasing with $|z|$. The conditional density of u_i is therefore

$$p_i(u_i \mid \cdots) \propto \exp\left\{-\sum_{j\in\partial i} w_{ij}\phi(u_i - u_j)\right\}, \quad u_i \in \mathcal{R},$$

where $w_{ij} = w_{ji}$ defines w_{ij} for $i > j$ and ∂i denotes the zones contiguous to i and hence its "neighbors" in the terminology of Markov random fields (cf. Section 2). The non-zero w_{ij}'s may take account of the features of contiguous zones, such as populations at risk, common boundary length and so on, but here we make

the simplest choice $w_{ij} = 1$. It should be noted that the family (4.1) is strictly improper because it only addresses differences in the u_i's and not their overall level. The impropriety could be easily removed by restricting one or more of the u_i's to any finite interval but is in any case inconsequential to the eventual posterior distribution for u and v.

We have experimented with two choices of ϕ. The simplest is $\phi(z) = z^2/2\kappa$, where κ is an unknown positive constant, in which case (4.1) becomes

$$p(u \mid \kappa) \propto \frac{1}{\kappa^{n/2}} \exp\left\{-\frac{1}{2\kappa}\sum_{i\sim j}(u_i - u_j)^2\right\}, \tag{4.2}$$

where $i \sim j$ denotes i and j are contiguous. This is a Gaussian *intrinsic autoregression*, if we extend Künsch's (1987) terminology to irregular pixel arrays, and has conditional moments

$$E(u_i \mid \cdots) = \bar{u}_i, \qquad \mathrm{Var}(u_i \mid \cdots) = \kappa/n_i, \tag{4.3}$$

where n_i is the cardinality of ∂i and $\bar{u}_i$ is the corresponding mean value. The scaling in (4.3) would be undesirable for marginal variances but has some appeal in the conditional formulation. An interpretation of (4.2) is that it provides a stochastic version of linear interpolation.

Our alternative choice has been $\phi(z) = |z|/\kappa$, where κ is an unknown scale parameter. Then u_i has conditional density

$$p_i(u_i \mid \cdots) \propto \frac{1}{\kappa} \exp\left\{-\frac{1}{\kappa}\sum_{j\in\partial i}|u_i - u_j|\right\}, \tag{4.4}$$

which has its mode at the median rather than at the mean of the contiguous u_i's. This distribution is therefore more appropriate than (4.2) if discontinuities in the risk surface are expected and can be interpreted as a stochastic version of the median filter, so popular in remote sensing applications.

Note that in both the above formulations, $\kappa \downarrow 0$ implies constant u_i's, whereas κ large implies correspondingly large but spatially structured variation. Similarly, $\lambda \downarrow 0$ implies $v = 0$, whereas λ large implies substantial but unstructured extra-Poisson variability. Note also that in the prior distribution of x, induced by those of u and v, the conditional density of x_i depends on all other x_j's, not merely on those in contiguous zones.

For definiteness, we now concentrate on (4.2) as the prior density for u. Then the joint posterior density of u, v, κ and λ on which we base inferences, rather than on that of x, is given by

$$\begin{aligned} P(u, v, \kappa, \lambda \mid y) \propto & \prod_{i=1}^{n}\{\exp(-c_i e^{x_i})(c_i e^{x_i})^{y_i}/y_i!\} \\ & \times \kappa^{-n/2}\exp\left\{-\frac{1}{2\kappa}\sum_{i\sim j}(u_i - u_j)^2\right\} \\ & \times \lambda^{-n/2}\exp\left\{-\frac{1}{2\lambda}\sum_{i=1}^{n} v_i^2\right\} \times \mathrm{prior}(\kappa, \lambda), \end{aligned} \tag{4.5}$$

where the final term is the prior density for the two hyperparameters, the obvious choice for which is proportional to $\kappa^{-1}\lambda^{-1}$. However, with this choice, (4.5) is improper because of its behavior near the origin $u = v = 0$, $\kappa = \lambda = 0$. It should be emphasized that this behavior does not stem from any spatial aspects of the formulation but is a common and unpleasant feature of Bayesian hierarchical models in general. Available remedies include banning a neighborhood of $\kappa = \lambda = 0$ in prior(κ, λ) or invoking a proper prior distribution, though here we have chosen prior$(\kappa, \lambda) \propto 1$.

We estimate u, v, κ and λ by approximations to their posterior means,

$$\hat{u} = E(u \mid y), \quad \hat{v} = E(v \mid y), \quad \hat{\kappa} = E(\kappa \mid y), \quad \hat{\lambda} = E(\lambda \mid y),$$

obtained from the Gibbs sampler which, as a by-product, will also produce interval estimates. Any one cycle of the Gibbs sampler requires each of the $2n + 2$ components of (u, v, κ, λ) to be updated by sampling from the salient conditional distribution. However, there is a technical problem in that, although (4.5) is now a proper distribution, it still has a singularity at the origin and this invalidates the Gibbs sampler, because the origin becomes an absorbing state of the Markov chain. We conveniently avoid this problem, with negligible other effects, using the modification

$$\text{prior}(\kappa, \lambda) \propto e^{-\epsilon/2\kappa} e^{-\epsilon/2\lambda}, \qquad \kappa, \lambda > 0, \tag{4.6}$$

where ϵ is a small positive constant having the value 0.01 in our computations. The conditional densities of u_i and v_i are unaltered, while those of κ and λ remain within the family of inverse gamma distributions. For example, u_i has conditional density

$$P_i(u_i \mid u_{-i}, v, \kappa, \lambda, y) \propto \exp\left\{-c_i e^{u_i + v_i} + u_i y_i - \frac{n_i}{2\kappa}(u_i - \bar{u}_i)^2\right\}, \qquad u_i \in \mathcal{R},$$

which can be sampled efficiently by carefully designed rejection methods; and κ has conditional density

$$P(\kappa \mid u, v, \lambda, y) \propto \kappa^{-n/2} \exp\left\{-\frac{1}{2\kappa}\left[\epsilon + \sum_{i \sim j}(u_i - u_j)^2\right]\right\}, \qquad \kappa > 0,$$

which can be sampled using the standard technique for the chi-squared distribution.

In practice, we typically run the Gibbs sampler for an initial period of 1000 cycles and then collect information from a further 10000 cycles of which we store every 10th or 20th for the subsequent construction of approximate interval estimates etc. The posterior means are estimated by the corresponding sample means. However, note that the logarithm of the joint posterior density of u and v, given κ, λ and y, is a strictly concave differentiable function of u and v and therefore possesses a single maximum, a result that holds whenever $\phi(z)$ is a differentiable

convex function of z. Thus, the conditional m.a.p. estimates u^* and v^* of u and v, given $\hat{\kappa}$, $\hat{\lambda}$ and y, provide an appealing alternative to $\hat{u}$ and $\hat{v}$ (cf. Section 2) and can be located by any deterministic hill-climbing method. For convenience, we use the iterated conditional modes (ICM) algorithm in Besag (1983, 1986) to find u^* and v^*. Finally, note that, as a bonus, u^* and v^* satisfy

$$\sum_{i=1}^{n} v_i^* = 0, \qquad \sum_{i=1}^{n} c_i e^{u_i^* + v_i^*} = \sum_{i=1}^{n} y_i,$$

so that the fitted total number of cases matches the observed total.

We illustrate the above methodology on three sets of data. The first two involve the 94 départements of mainland France; labels, assigned in alphabetical order, and contiguities are identified in Fig. 3. The first data set concerns a total of 2588 deaths from thyroid cancer reported among women during the period 1971–1978. The observed mortality rates, relative to the overall mean rate, are shown in Fig. 4. We used the Gibbs sampler to generate successive samples from the posterior distribution (4.5) with the prior (4.6) for κ and λ, and hence estimate the posterior means of u, v, κ and λ; in particular, we found $\hat{\kappa} = 0.129$ and $\hat{\lambda} = 0.011$. The corresponding conditional m.a.p. estimates u^* and v^* were calculated by ICM and these provide the estimate $e^{u^*+v^*}$ of true relative risks shown in Fig. 5. There is very little difference between these values and those obtained either from the means or from the medians of the posterior distribution for $x = u + v$ given by the Gibbs sampler. Residuals, calculated as the ratio of the observed mortality rates to the estimated risks, are displayed in Fig. 6. Figures 7 and 8 provide marginal 10% and 90% points of the posterior distribution of x.

The second example concerns 4340 deaths from multiple myeloma among men during the same period. Here, we found $\hat{\kappa} = 0.009$ and $\hat{\lambda} = 0.009$. Figures 9 through 13 correspond to Figs. 4 through 8.

Briefly, the first analysis supports the existence of the spatial effects that are suggested in the raw data. High and low observed rates shrink somewhat towards the overall rate, as would be expected. The results for the second data set suggest an almost uniform risk among the départements, with the possible exception of zone 56 (Moselle). Of course, any firm conclusions for either data set would require more detailed epidemiological and statistical study, including an examination of the residuals.

There are several aspects of this type of analysis that require thorough investigation. Here, we merely describe two simple simulation exercises we carried out on the data for thyroid cancer. In the first, we took the estimated values $u_i^* + v_i^*$ to represent the true log relative risks x_i and used these and the corresponding populations at risk to generate an independent Poisson observation y_i for each zone i. We then carried out an analysis of the new y_i's to obtain point and interval estimates for the known x_i's and for the hyperparameters of our model. This enabled some rough assessments to be made. The posterior means for κ and λ were 0.082 and 0.010; the former is rather low in comparison to the notional 0.129, though this value lay well within the 90% interval for κ, since the posterior distribution is rather diffuse. We use the term "notional" because of course the generated y_i's

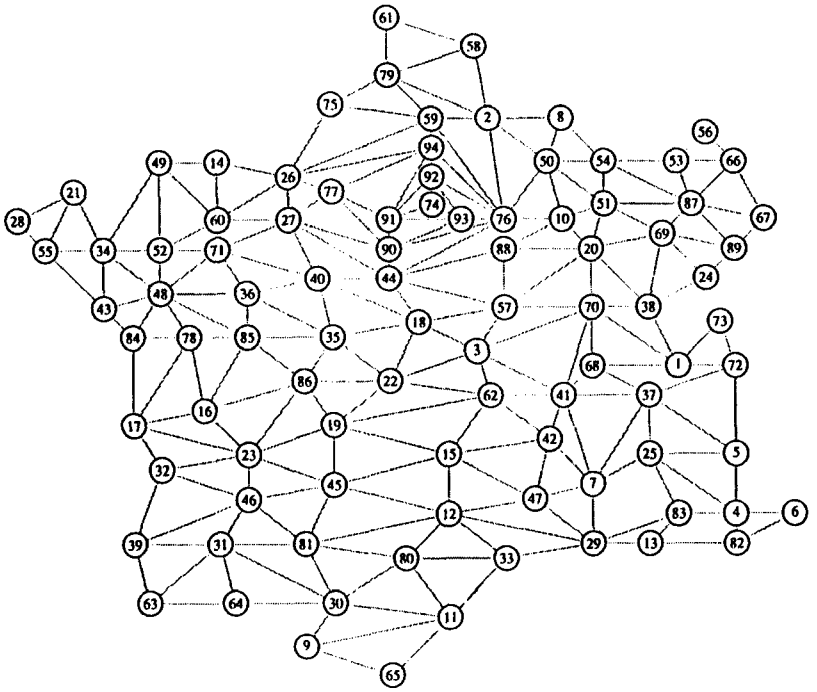

Fig. 3. Labels and contiguities for the 94 départements of France.

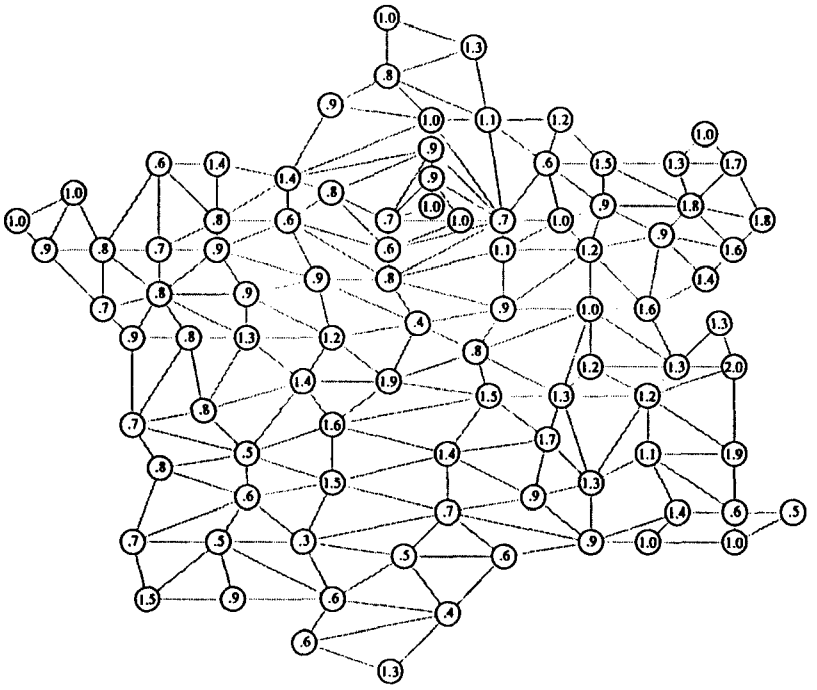

Fig. 4. Observed mortality from thyroid cancer, relative to the overall mean rate.

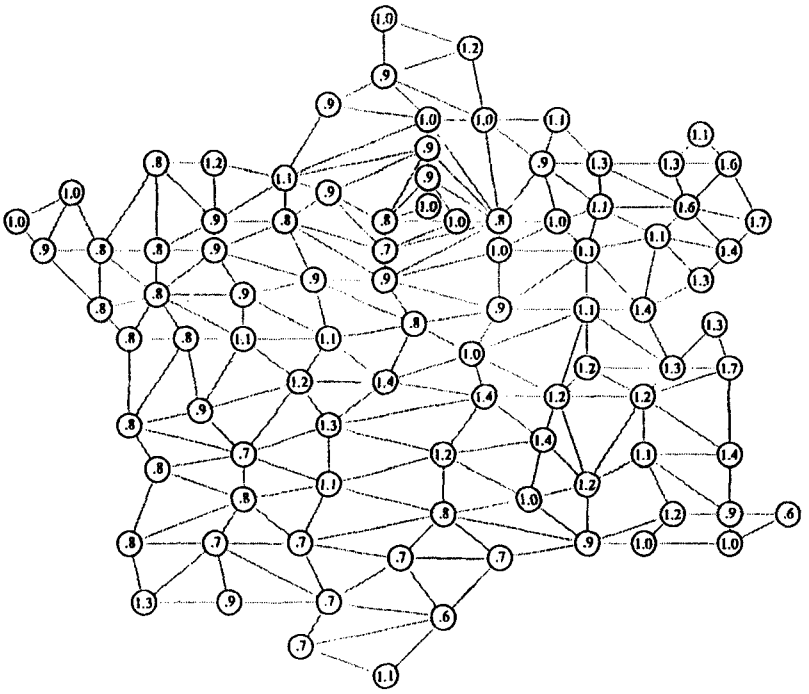

Fig. 5. Estimated relative risks for thyroid cancer, using ICM with fixed $\hat{\kappa}$ and $\hat{\lambda}$.

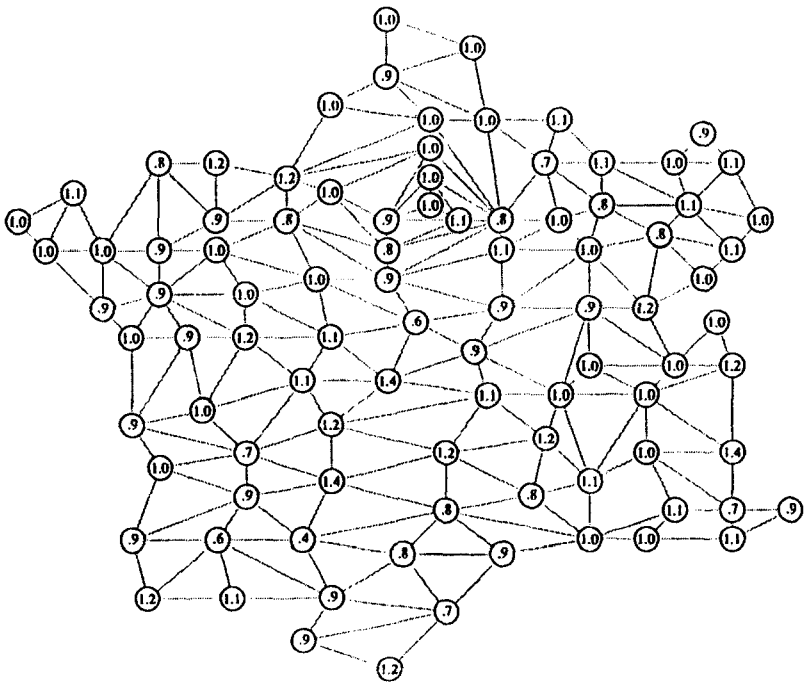

Fig. 6. Residuals for thyroid cancer, calculated as the ratio of the observed mortality rates to the estimated risks.

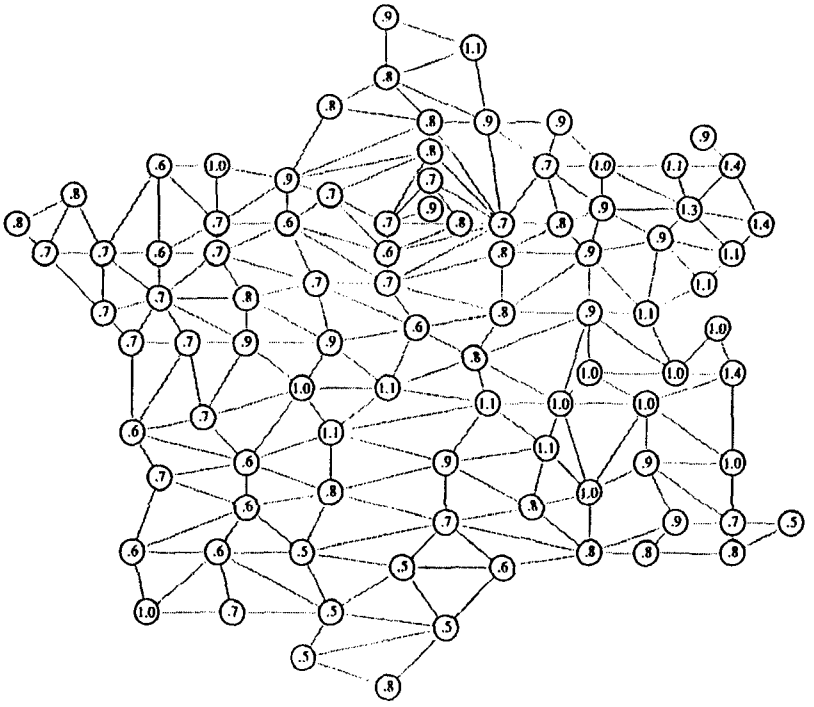

Fig. 7. Marginal 10% points for the posterior distribution of relative risk of thyroid cancer, estimated from the Gibbs sampler.

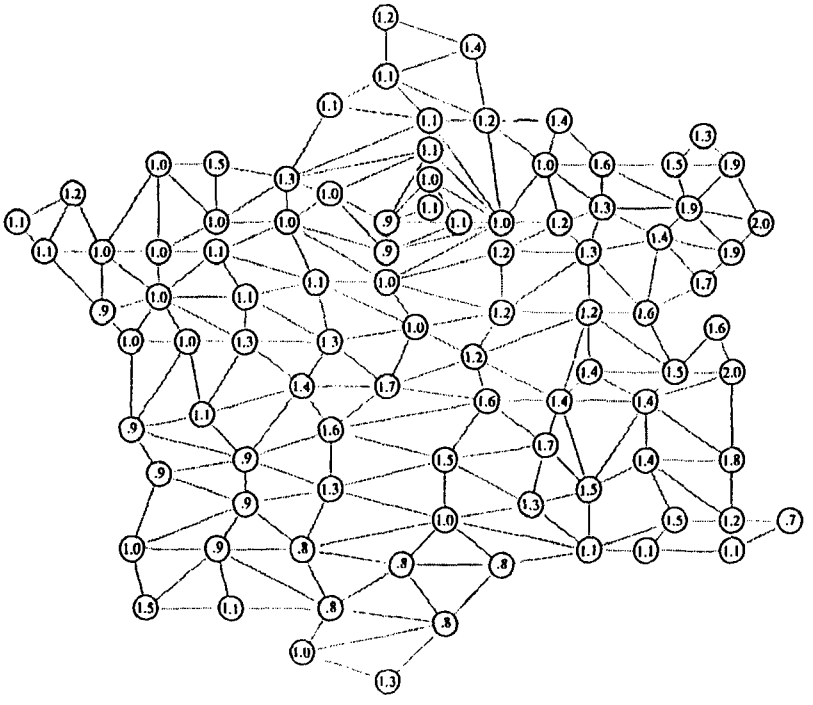

Fig. 8. Marginal 90% points for the posterior distribution of relative risk of thyroid cancer, estimated from the Gibbs sampler.

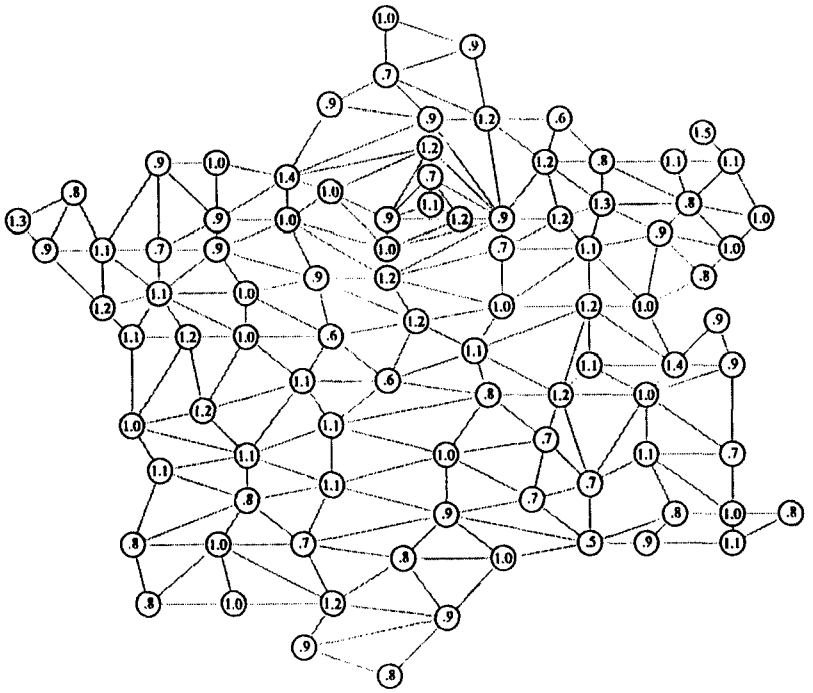

Fig. 9. Observed mortality from multiple myeloma, relative to the overall mean rate.

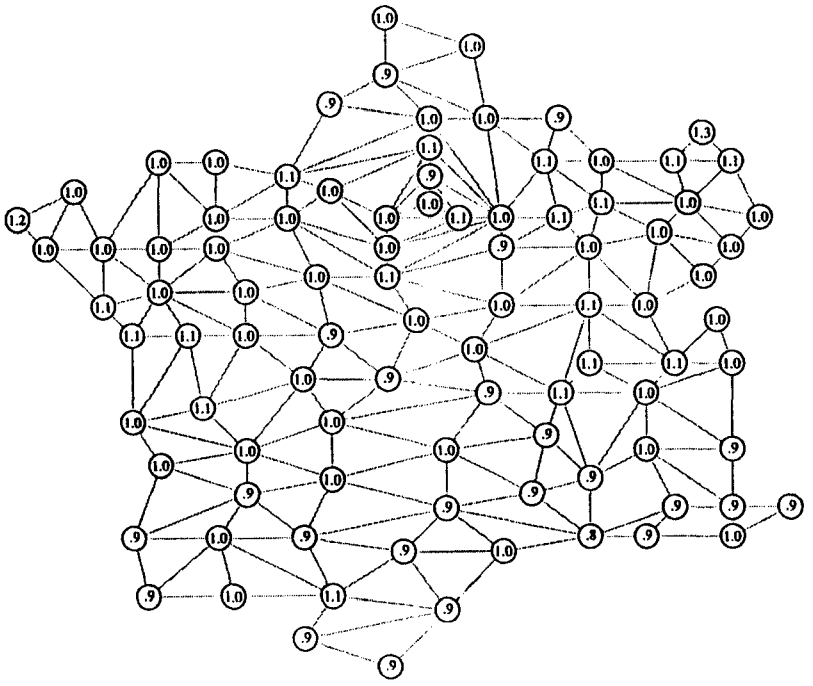

Fig. 10. Estimated relative risks for multiple myeloma, using ICM with fixed $\hat{\kappa}$ and $\hat{\lambda}$

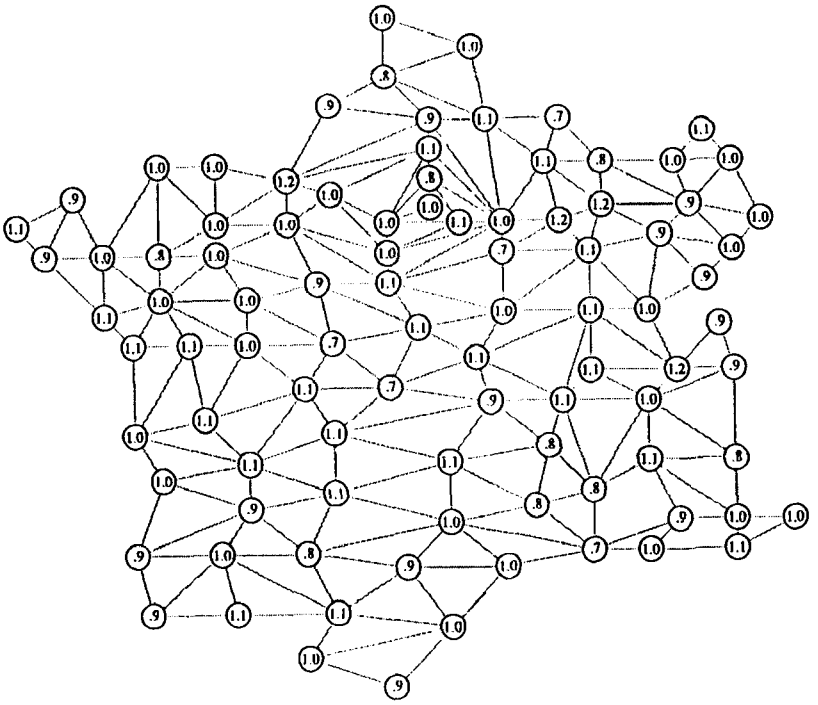

Fig. 11. Residuals for multiple myeloma, calculated as the ratio of the observed mortality rates to the estimated risks.

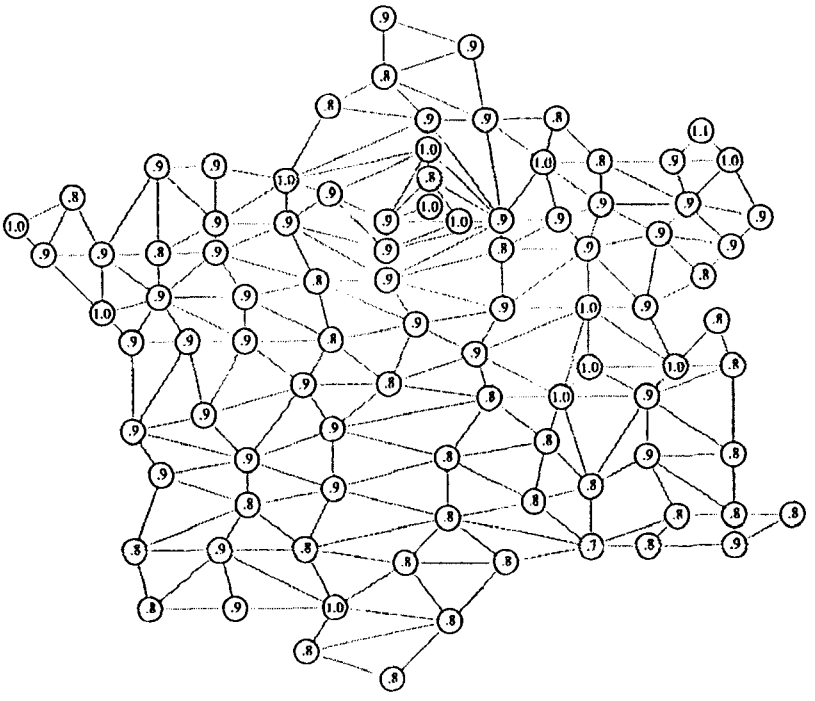

Fig. 12. Marginal 10% points for the posterior distribution of relative risk of multiple myeloma, estimated from the Gibbs sampler.

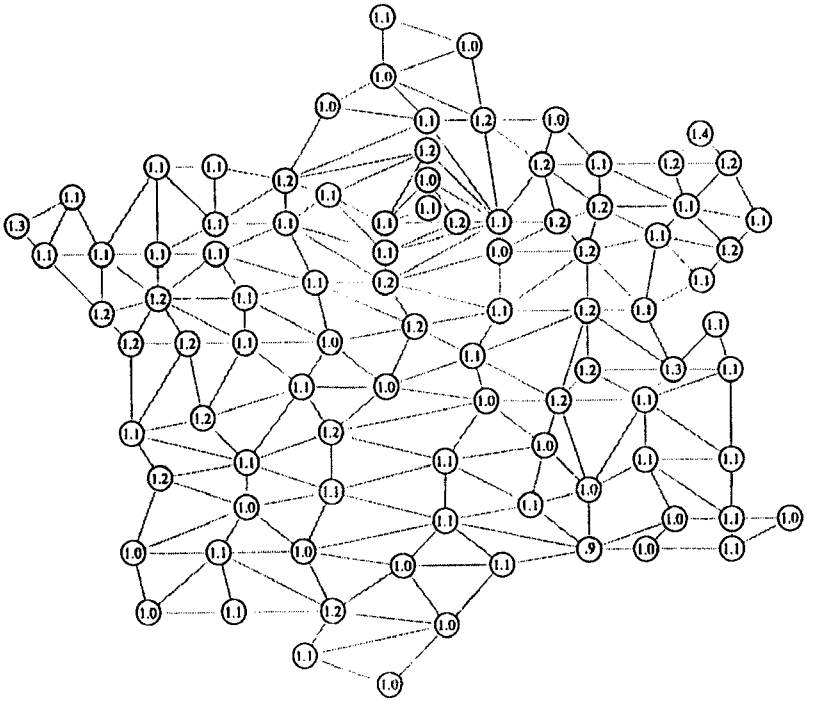

Fig. 13. Marginal 90% points for the posterior distribution of relative risk of multiple myeloma, estimated from the Gibbs sampler.

depend directly only on x and not on the values of u, v, κ and λ. A more important consideration is the accuracy of our eventual estimate x^* of x in comparison to the apparent log relative risks r_i calculated directly from the y_i's. The mean squared error of the r_i's was 5.1, whereas that of the x_i^*'s was only 1.4, so that the Bayesian procedure provides a substantial improvement. Two further simulations suggested that these figures are typical for this example. As regards interval estimation, 81 of the 94 known x_i's lay within their corresponding 80% intervals, and 89 within their 90% intervals. Note that the above method of assessment assumes only the correctness of the Poisson formulation used to generate the y_i's.

In the second part of the simulation exercise, the aim was to investigate the role of κ and λ in the posterior distribution of relative risk; this also has some relevance to the empirical Bayes procedure of Clayton and Kaldor (1987). Here, we ran the Gibbs sampler on the original data for a second time but using fixed values $\kappa = 0.129$ and $\lambda = 0.011$ throughout. On average, such a procedure will produce shorter but erroneous interval estimates for the x_i's, because it does not account for variability in the estimation of the hyperparameters. However, in this example, we found no evidence of a systematic effect on the interval estimates.

The numbers of cases per zone in each of the above examples are sufficiently large that the Poisson likelihood could be replaced by a Gaussian approximation, opening up the possibility of some simplification of the analysis (cf. Clayton and Kaldor (1987)). In complete contrast, we now consider an example in which the

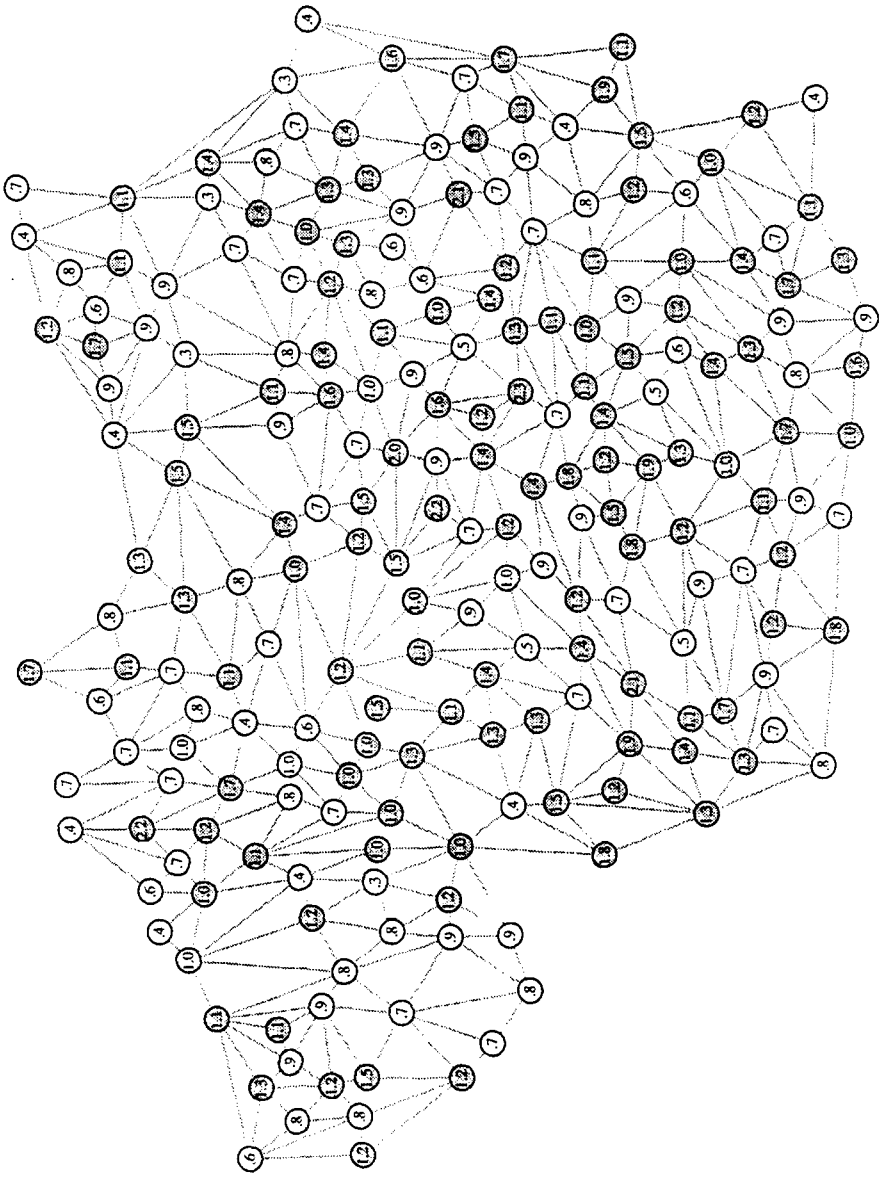

Fig. 14. Electoral wards, contiguities and observed incidence rates, *relative to the* overall mean rate, for all cancers in Greater Manchester, excluding leukemias. Shading identifies wards with above average rates.

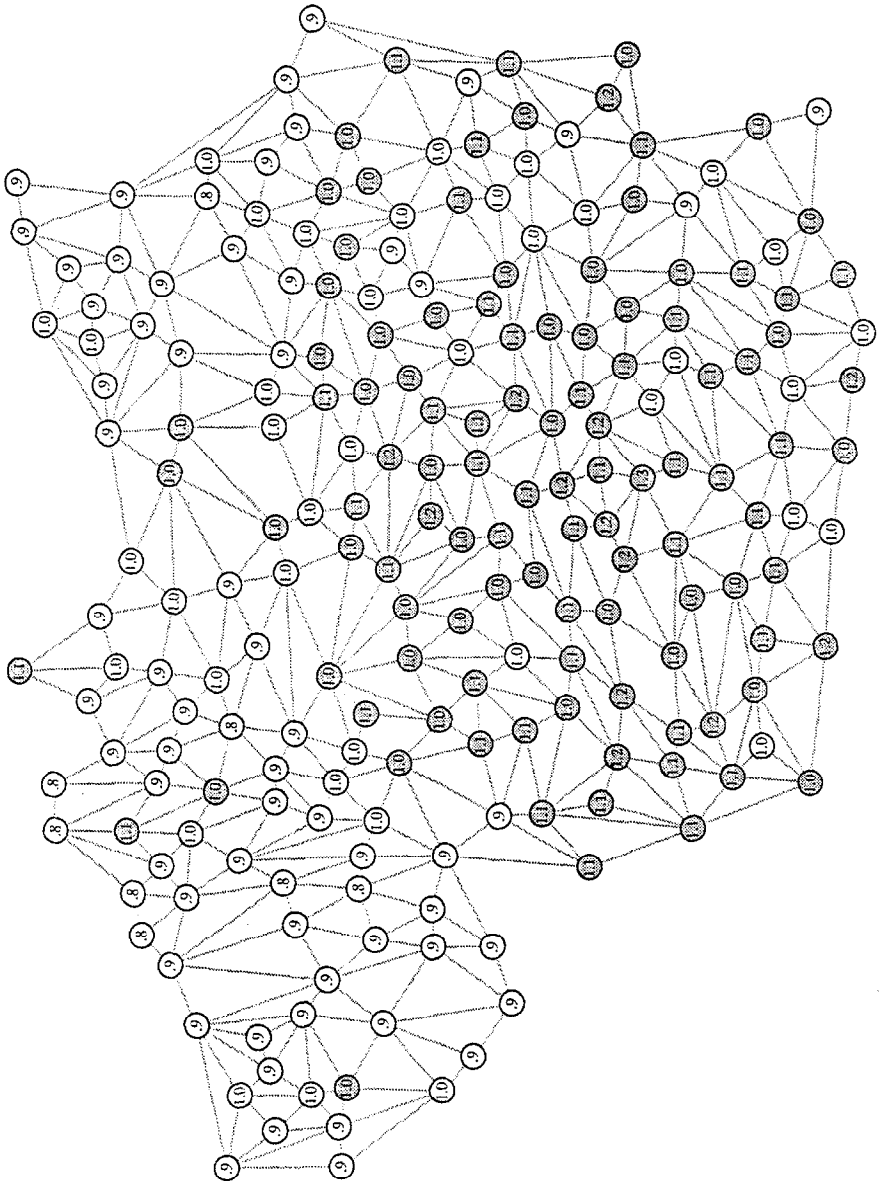

Fig. 15. Estimated relative risks for all cancers, excluding leukemias. Shading identifies wards with above average risks.

numbers of cases in individual zones are very low indeed. The data are part of a larger data set for 1218 wards (zones) in the North of England and refer to the reported incidence of all cancers, excluding leukemias, in the age group 0–24 years, during 1968–1985. The number of cases over the entire region is 4997, ranging from zero to 20 in individual wards; the largest and smallest populations at risk differ by a factor of 134! The region is divided among eight counties, two of which, Greater Manchester, in the North West, and Tyne and Wear, in the North East, form major industrial conurbations. Here we concentrate on the former, for which there are 1904 cases distributed among the 216 wards, with a low of 2 and a high of 20. Figure 14 displays the zones, their contiguity graph and the observed incidence rates, relative to the overall; shaded zones have above average incidence rates. It is difficult to draw any particular conclusions because of the Poisson noise and the different sizes of populations at risk, though here these differ only by a factor of four.

Figure 15 is the result of the same Bayesian analysis as in the first two examples. Now a very clear pattern is evident and has a reasonable explanation: the northwest of Greater Manchester is primarily residential and free from major industrial pollution. The existence of such a pattern is not entirely obvious on *prima facie* grounds, since only the age range 0–24 is under consideration, and, so far as we are aware, has not been pointed out previously. Not surprisingly, all the 90% and indeed almost all the 80% equal-tailed posterior intervals for individual wards include unit relative risk. Had the pattern been hypothesized in advance, it would have been valid to divide the wards into sets, A and A^c, comprising those in the northwest of the county and those elsewhere, and then to construct the posterior distribution of the relative risk,

$$\frac{\sum_{i\in A} c_i e^{x_i}}{\sum_{i\in A} c_i} \div \frac{\sum_{i\in A^c} c_i e^{x_i}}{\sum_{i\in A^c} c_i},$$

from the realizations of the Gibbs sampler. Of course, in this circumstance, a classical significance test on the raw incidence rates for A and A^c would also produce incontestable evidence of a difference; thus, the main point is that the Bayesian analysis has unmasked the pattern. Incidentally, a similar effect was found in Tyne and Wear; also, the estimated risks were almost identical when the prior (4.2) was replaced by (4.4). We end on a somewhat cautionary note, wishing to emphasize that almost any analysis at this scale is likely to be precarious. Nevertheless, we believe that the Bayesian analysis is an improvement on the classical estimation and testing methods that are used currently as one means of allocating health service resources on a ward by ward basis.

Acknowledgements

We are grateful to Dr. Cliff Litton (Department of Mathematics, University of Nottingham, U.K.), who supplied the data in Section 3, and to Dr. Alan Craft, Dr. Louise Parker (Department of Child Health, University of Newcastle upon

Tyne, U.K.) and Dr. Jillian Birch (Manchester Children's Tumour Registry), who provided those in Section 4.

REFERENCES

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion), *J. Roy. Statist. Soc. Ser. B*, **36**, 192–236.

Besag, J. E. (1975). Statistical analysis of non-lattice data, *The Statistician*, **24**, 179–195.

Besag, J. E. (1983). Discussion of paper by P. Switzer, *Bull. Internat. Statist. Inst.*, **50** (Bk. 3), 422–425.

Besag, J. E. (1986). On the statistical analysis of dirty pictures (with Discussion), *J. Roy. Statist. Soc. Ser. B*, **48**, 259–302.

Besag, J. E. (1989). Towards Bayesian image analysis, *Journal of Applied Statistics*, **16**, 395–407.

Besag, J. E. and Mollié, A. (1989). Bayesian mapping of mortality rates, *Bull. Internat. Statist. Inst.*, **53** (Bk. 1), 127–128.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models, *J. Roy. Statist. Soc. Ser. C*, **33**, 38–44.

Buck, C. E., Cavanagh, W. G. and Litton, C. D. (1988). The spatial analysis of soil phosphate data, Tech. Report, Department of Mathematics, University of Nottingham, U.K.

Chow, Y., Grenander, U. and Keenan, D. M. (1988). Hands: a pattern theoretic study of biological shape, Tech. Report, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

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

Geman, D. and Geman, S. (1986). Bayesian image analysis, *Disorderd Systems and Biological Organization* (eds. E. Bienenstock *et al.*), in NATO ASI Series, Vol. F20, Springer, Berlin.

Geman, D., Geman, S., Graffigne, C. and Ping Dong (1990). Boundary detection by constrained optimization, *I.E.E.E. Transactions: Pattern Analysis and Machine Intelligence*, **12**, 609–628.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *I.E.E.E. Transactions: Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geman, S. and Graffigne, C. (1987). Markov random field image models and their applications to computer vision, *Proc. International Congress of Mathematicians* (1986) (ed. A. M. Gleason), 1496–1517, Berkeley, California.

Geman, S. and McClure, D. (1987). Statistical methods for tomographic image reconstruction, *Bull. Internat. Statist. Inst.*, **52** (Bk. 4), 5–21.

Green, P. J. (1990). Penalized likelihood reconstructions from emission tomography data using a modified EM algorithm, *I.E.E.E. Transactions: Medical Imaging*, **9**, 84–93.

Greig, D. M., Porteous, B. T. and Seheult, A. H. (1989). Exact maximum *a posteriori* estimation for binary images, *J. Roy. Statist. Soc. Ser. B*, **51**, 271–279.

Grenander, U. (1983). Tutorial in pattern theory, Tech. Report, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

Kent, J. T. and Mardia, K. V. (1988). Spatial classification using fuzzy membership models, *I.E.E.E. Transactions: Pattern Analysis and Machine Intelligence*, **10**, 659–671.

Künsch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice, *Biometrika*, **74**, 517–524.

Mollié, A. (1990). Représentation géographique des taux de mortalité: modélisation spatiale et méthodes Bayésiennes (unpublished Ph. D. thesis).

Owen, A. (1989). Image segmentation via iterated conditional expectations, Tech. Report, Department of Statistics, Stanford University, California.