Multiscale detection of localized anomalous structure in aggregate disease incidence data

Mary M. Louie[1], Eric D. Kolaczyk[2*†]

[1] *National Center for Health Statistics, 3311 Toledo Road, Room 3215, Hyattsville, MD 20782*
[2] *Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215*

SUMMARY

We present a modeling framework for detection of potentially anomalous structure in aggregate spatial disease incidence data in a manner sensitive to localization at multiple scales and/or positions. The key technical contribution is the re-casting of the components of a multiscale disease mapping methodology, recently introduced by the authors in an earlier paper, into a form appropriate for hypothesis testing. In particular, we describe how hypotheses of spatially clustered variations in disease incidence may be linked in one-to-one correspondence with collections of hypotheses on the values of certain multiscale parameters associated with a user-defined hierarchy of nested partitions of an overall spatial region. A Bayesian hypothesis testing methodology is developed in the context of a standard Poisson measurement model, over the collection of possible multiscale hypotheses. We discuss the specification of hyper parameters and prior distributions on the space of models. The methodology is illustrated on both simulated and real data.

1. Introduction

Two standard tasks in spatial epidemiology are the testing for and mapping of structure in disease incidence data [1]. A large portion of the energy in this field arguably has gone towards the development of methods for the latter of these tasks. However, recently more and more focus is being given to methods of monitoring for detection of potentially anomalous structure in data reporting incidence of a given disease(s). Such methods tend to take the form of spatial or spatio-temporal tests of 'clustering' and similar 'atypical' patterns in disease incidence data.

It has been noted (e.g., [1]) that in spatial epidemiology, as in most other geo-spatial fields of study, the concept of *scale* plays an important role in analysis and inference. Specifically, underlying patterns of disease incidence often have an intrinsic scale(s). As a result, choice of the operable scale of an analysis (e.g., tract size, radius of a scan region, etc.) plays a key role

*Correspondence to: Eric D. Kolaczyk, Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215
†Email: kolaczyk@math.bu.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: BCS0079077, BCS0318209

Contract/grant sponsor: Office of Naval Research; contract/grant number: N00014–99–1–0219

in determining the potential of the analysis to extract the relevant spatial information on the disease. For contexts in which multiple scales are present and at different locations, a so-called 'multiscale' analysis would be desirable, in which the data are analyzed simultaneously across a range of scales and positions within scale. And ideally, such an analysis would be done in a manner that effectively decouples information at each scale and each position within scale, so as to allow for the isolation of each such component of information and its separate assessment. The quintessential example of such techniques are those based on wavelets, in which an object is decomposed with respect to an orthonormal basis of functions that are localized in both scale and position. See [2], for example.

Unfortunately, classical wavelet-based methods are most appropriate for regularly spaced signal and image data and, without modification, are difficult to apply directly to the type of irregularly spaced data typically encountered in geo-spatial applications. In recent years, however, a number of efforts have been directed towards the explicit extension of multiscale principles and structures to these types of applications [3, 4, 5, 6]. With respect to spatial epidemiology, we have recently introduced a framework for multiscale disease mapping that is, to the best of our knowledge, the first such framework [7]. Here, in the present paper, we re-cast our previous framework for the purpose of testing for localized anomalous spatial variations in disease incidence data in a similarly multiscale fashion.

Our proposed framework exploits the fact that, under the standard Poisson measurement model for aggregate count data, hypotheses of spatially localized variation in disease incidence can be usefully linked in one-to-one correspondence with collections of hypotheses on the values of certain multiscale parameters associated with a user-defined hierarchy of nested partitions of an overall spatial region. Hence, the problem of finding localized anomalous spatial variations in the data is replaced by one of finding various patterns within a collection of local hypotheses indexed by position and scale. We develop a Bayesian hypothesis testing machinery to evaluate and compare models within this collection, and show that such comparisons may be done in a computationally efficient manner, given user-specified values for the hyperparameters and the prior probabilities for hypotheses.

The organization of this paper is as follows. In Section 2 we describe necessary background, in the form of multiscale likelihood factorizations of Poisson count models and the correspondence between spatial and multiscale parameterizations. In Section 3, we describe our testing framework and we examine the relative effect of different choices for the hyperparameters and such. The application of our framework is explored in Sections 4 and 5 using simulated and real data, respectively. Some additional discussion may be found in Section 6.

2. Background

2.1. Multiscale Factorization and Reparameterization

Let D be a spatial region of interest. We will represent the occurrences of cases of a given disease by a point process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ that, conditional on an intensity process $\{\mu(\mathbf{s}) : \mathbf{s} \in D\}$, will be taken to be Poisson. We will assume that measurements of the process $Y(\cdot)$ are made available in aggregate form i.e., in the form $Y_B \,|\, \mu_B \sim \text{Poisson}(\mu_B)$, where $\mu_B \equiv \int_B \mu(s)ds$ for subregions $B \subset D$. Models for the process $\mu(\cdot)$ will be specified later, in Section 3.

As in [7], our notion of a multiscale model is associated with a collection $\mathcal{B}^{(J)} =$

$\{\{B_{j,k}\}_{k=1}^{N_j}\}_{j=0}^{J}$ of $J+1$ nested partitions of D. That is, the $B_{j,k}$ represent subregions in D at spatial scales $j = 0, 1, \ldots, J$ and relative positions $k = 1, \ldots, N_j$ within scale, such that

$$\bigcup_{k=1}^{N_j} B_{j,k} \;=\; D \qquad \text{and} \qquad \bigcup_{k' \in ch(k)} B_{j+1,k'} \;=\; B_{j,k}\,. \tag{1}$$

Here, for a given choice of j and k, $ch(k)$ denotes the set of indices k' for which $B_{j+1,k'} \subseteq B_{j,k}$.

The succession of scales $j = J, J-1, \ldots, 0$ corresponds to an assumed hierarchy of aggregations of the original subregions to coarser and coarser levels. In spatial epidemiology, geo-political considerations often drive the choice of subregions chosen. Indeed, such considerations may in fact dictate, or at least limit, such choices, since the conventions adopted by relevant government statistical agencies (e.g., census tracts) tend to be dominant. However, the integration of GIS technologies with spatial epidemiology has led to an increased ability for users to efficiently explore alternative hierarchies (at least at scales coarser than that of the original observations).

By way of a simple, context-free illustration, we will often refer to the hierarchy shown in Figure 1 for reference throughout much of this paper. Note that corresponding to this hierarchy is a tree – the canonical 'quad-tree'. This natural connection with trees has inspired the use of tree-based terminology in the multiscale modeling literature. Specifically, the single region D at the coarsest scale $j = 0$ is said to correspond to the root of the tree; those subregions at the finest scale $j = J$ correspond to leaves. Similarly, a subregion $B_{j,k}$ at scale j that splits into subsubregions $B_{j+1,k'}$, $k' \in ch(k)$, at scale $j+1$, is said to be the parent, and the subsubregions, the children. More generally, if for two subregions, say $B_{j,k}$ and $B_{j',k'}$, the latter can be arrived at from the former solely by moving down the tree in a direction away from the root and towards the leaves, then $B_{j,k}$ is called the ancestor of $B_{j',k'}$, and $B_{j',k'}$ is called a descendent of $B_{j,k}$.

Without loss of generality, we assume that our measurements correspond to those subregions $B_{J,k}$ at the finest scale J, and write these measurements as $Y_{J,k} = Y_{B_{J,k}}$. The random variables $Y_{J,k}$ therefore will be distributed conditionally as independent Poisson variables with means $\mu_{J,k} = \mu_{B_{J,k}}$. Note that for each successive spatial aggregation of subregions $B_{j,k}$, from one scale to another, a similar aggregation of the measurement and mean variables may be defined:

$$Y_{j,k} \;=\; \textstyle\sum_{k' \in ch(k)} Y_{j+1,k'} \qquad \text{and} \qquad \mu_{j,k} \;=\; \textstyle\sum_{k' \in ch(k)} \mu_{j+1,k'} \;\; . \tag{2}$$

Presumably each scale of variables may contain useful information about the process $\mu(\cdot)$, although which scale(s) are most useful for a given purpose typically will be unknown. A simultaneous analysis of the $Y_{j,k}$ at all scales therefore is desirable, but is complicated by the dependence of these variables. A multiscale factorization of the data likelihood can be used to induce a particularly useful decoupling. Let $\mathbf{Y}_j = (Y_{j,1}, \ldots, Y_{j,N_j})^T$ and let $\boldsymbol{Y}_{j+1,ch(k)}$ denote those measurements $Y_{j+1,k'}$ for whom $B_{j+1,k'} \subseteq B_{j,k}$. Define $\boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_{j+1,ch(k)}$ similarly. Then one can write

$$\Pr\left(\boldsymbol{Y}_J \,|\, \boldsymbol{\mu}_J\right) = \Pr\left(Y_{0,1} | \mu_{0,1}\right) \prod_{j=0}^{J-1} \prod_{k=1}^{N_j} \Pr\left(\boldsymbol{Y}_{j+1,ch(k)} \,|\, Y_{j,k}, \boldsymbol{\omega}_{j,k}\right) \;\; , \tag{3}$$

where $Y_{0,1}|\mu_{0,1} \sim \text{Poisson}(\mu_{0,1})$ and $\boldsymbol{Y}_{j+1,ch(k)} \,|\, Y_{j,k}, \boldsymbol{\omega}_{j,k} \sim \text{Multinomial}(Y_{j,k}\,;\, \boldsymbol{\omega}_{j,k})$, with

$$\boldsymbol{\omega}_{j,k} = \mu_{j,k}^{-1} \times \boldsymbol{\mu}_{j+1,ch(k)} \;\; . \tag{4}$$

The multiscale representation above is analogous to that of a classical orthonormal wavelet basis expansion of a function and, as such, separates out the information in the original measurements at the finest scale i.e., in $\boldsymbol{Y}_J$, into conditionally statistically independent components as those measurements are aggregated across each successively coarser scale. Note too that each component $\boldsymbol{Y}_{j+1,ch(k)}|Y_{j,k}$ is accompanied by its own unique parameter $\boldsymbol{\omega}_{j,k}$. The end result is an alternative representation of the original spatially indexed data and parameters in a form that simultaneously separates out the information in both that is local to the scale and position of each $B_{j,k}$. The reader is referred to [3] for a derivation of the above formulas, and to [8] for background on and statistical properties of multiscale likelihood factorizations and statistical methods based thereupon.

2.2. Comparison of Spatial versus Multiscale Parameterizations

The factorization in (3) will be central to the testing framework we introduce in Section 3. Our approach is to characterize hypotheses of interest in the space of the multiscale parameters (4). Accordingly, it is necessary to first explore more closely the relationship between spatial and multiscale parameterizations in the context of disease incidence data.

Again adopting the notation of [7], we introduce a notion of relative risk by re-writing the Poisson intensity parameters $\mu_{J,k}$, at the finest scale J, in the form $\mu_{J,k} = \theta_{J,k} e_{J,k}$. Here $e_{J,k}$ denotes the expected counts in the 'at-risk' population in subregion $B_{J,k}$ and $\theta_{J,k} = \mu_{J,k}/e_{J,k}$ represents the relative risk in $B_{J,k}$. We will assume the $e_{J,k}$ to be fixed and known, deriving perhaps from a separate assessment, user input, or a combination of the two. The assessment may correspond to another data source (e.g., a demographic database) or, in the case of spatio-temporal analysis, it may correspond to values from the preceding time period on a region whose population is being tracked in time. It will also be useful to define an analogous notion of relative risk at coarser scales j, writing $\theta_{j,k} = \mu_{j,k}/e_{j,k}$, where the values $e_{j,k}$ are defined recursively through the expression $e_{j,k} = \mathbf{1}^T \boldsymbol{e}_{j+1,ch(k)}$.

In this paper, we will take the task of detecting potential anomalies in aggregate disease incidence data to mean finding deviations from uniformity in relative risk, with respect to the expected count levels $\{e_{J,k}\}$, and we will aim to do so in a fashion sensitive to localization of deviations in spatial scale and position. The multiscale parameters $\boldsymbol{\omega}_{j,k}$ in (4) capture a natural sense of such localization. For example, note that if $\theta_{J,k} \equiv \theta^*$, for some $\theta^* > 0$, and all (J,k), then $\mu_{j,k} = \theta^* e_{j,k}$ for all (j,k) and, equivalently, $\boldsymbol{\omega}_{j,k} = \mathbf{e}_{j+1,ch(k)}/e_{j,k}$. So, ultimately, detecting localized deviations in the relative risk from a constant θ^* is equivalent to detecting deviations in the $\boldsymbol{\omega}_{j,k}$ from the values $\boldsymbol{e}_{j+1,ch(k)}/e_{j,k}$. We now consider two examples in somewhat more detail.

First, suppose that $\boldsymbol{\omega}_{j,k} = \mathbf{e}_{j+1,ch(k)}/e_{j,k}$ for all but a single index pair (j_0, k_0). It is useful to understand the effect of this single deviation from uniformity at (j_0, k_0) on the various intensity parameters $\mu_{j,k}$. Begin by noting that we will continue to have $\mu_{j,k} = \theta^* e_{j,k}$ for (i) all $j \leq j_0$, and (ii) all (j,k) for which $j > j_0$ but $B_{j,k} \cap B_{j_0,k_0} = \emptyset$. On the other hand, for $B_{j,k} \subset B_{j_0,k_0}$, at least two such $\mu_{j,k}$ will differ from the corresponding $\theta^* e_{j,k}$, since by assumption $\boldsymbol{\omega}_{j_0,k_0} \neq \mathbf{e}_{j_0+1,ch(k_0)}/e_{j_0,k_0}$ but at the same time the constraint $\boldsymbol{\omega}'_{j_0,k_0} \mathbf{1}_{n_{j_0,k_0}} = 1$ must hold. That is, for fixed j, among the set of all k for which $B_{j,k} \subset B_{j_0,k_0}$, the total discrepancy of the corresponding $\mu_{j,k}$ from the values $\theta^* e_{j,k}$ will be zero, since the sum of these $\mu_{j,k}$ must still equal $\theta^* e_{j_0,k_0}$. So the effect of a deviation in a single multiscale parameter is to induce a local 'ripple' in the underlying intensity, with elevation in some subregion(s) and

a corresponding decrease in others.

More realistically, since there is usually particular interest in the case where the relative risk is elevated locally in some manner (i.e., often referred to as a form of 'clustering'), it is useful to consider the effect on the multiscale parameters $\boldsymbol{\omega}_{j,k}$ when there is an elevation within a single subregion. For example, suppose that $\mu_{J,k} = \theta^* e_{J,k}(1+\delta)$, with $\delta \geq 0$ generally and $\delta > 0$ for all $B_{J,k} \subset B_{j_0,k_0}$, for some B_{j_0,k_0}. Then (i) $\boldsymbol{\omega}_{j,k} = \mathbf{e}_{j+1,ch(k)}/e_{j,k}$ for all $j \geq j_0$, and for all (j,k) such that $j < j_0$ and $B_{j,k} \cap B_{j_0,k_0} = \emptyset$, but (ii) $\boldsymbol{\omega}_{j,k} \neq \mathbf{e}_{j+1,ch(k)}/e_{j,k}$ if $B_{j_0,k_0} \subset B_{j,k}$. In particular, in the latter case, that element of $\boldsymbol{\omega}_{j,k}$ corresponding to the unique subregion containing (and possibly equal to) B_{j_0,k_0} will be increased, and at least one of the others will therefore be decreased. Thus, the effect of an elevation in B_{j_0,k_0} is to make the multiscale coefficients differ from their nominal value for *all* ancestor subregions of B_{j_0,k_0}. Put another way, a localized elevation in relative risk at a given scale induces a pattern of deviation across a certain hierarchical subset of the multiscale parameters.

Note that the effect of a reduction in relative risk within a given subregion B_{j_0,k_0} has an analogous effect. Similarly, the effect of compilations of regions of increased and reduced relative risk will result in a compilation of their induced effects on the multiscale coefficients. However, the degree to which these effects may be separated and interpreted will depend on a variety of factors, such as the proximity of the regions affected, the proportion by which the relative risks are changed, etc. This phenomenon is analogous to that which arises in the context of wavelet-based analyses of signals and images, where it is found that such methods are best for the detection and estimation of relatively isolated structure. Examples include changepoints and similarly sharp transitions in signals, and edges in images. Accordingly, it can be expected that our proposed framework similarly will be most appropriate for identifying the locations and scales of isolated disease clusters within a spatial region.

By way of illustration, consider the diagram in Figure 2. The initial square data space and the collection of nested partitions fall under the quad-tree hierarchy illustrated in Figure 1, such that each scale j, for $j = 0, \ldots, 4$, has $2^j \times 2^j$ partitions. The three numbers that appear in each (sub)partition are respectively the value of the Poisson intensity parameter $\mu_{j,k}$, the appropriate element of the multiscale parameter vector $\boldsymbol{\omega}_{j,k}$ (or just $\mu_{0,1}$ at the coarsest scale), and the index of the partition (j = scale, k = position). At the finest scale $J = 4$, in a 2×2 square in the uppermost northwest corner, we have set $\mu_{4,k} = 21$, while for the rest of the spatial region, $\mu_{4,k} = 7$; that is, each of the intensities for the four pixels in the upper corner are three times the intensity of the other pixels. The spatial intensities $\mu_{j,k}$ are printed in bold for those partitions $B_{j,k}$ containing this elevated region.

When we shift our focus to the multiscale parameters, we find that $\boldsymbol{\omega}_{3,k} = 0.25 \times \mathbf{1}_4$ for all $k = 1, \ldots, 4^{J-1}$. Since all the parents at scale $j = J - 1 = 3$ divide up equally among their children at scale $J = 4$, examination of the multiscale parameters between these two scales reveals no evidence of localized elevation. However, upon aggregation from scale $J = 4$ to scale $j = 3$, we now find that the spatial intensity in this uppermost northwest corner is $\mu_{3,1} = 84$ while $\mu_{3,k} = 28$ for $k = 2, \ldots, 4^3$. The corresponding multiscale parameters are now $\boldsymbol{\omega}_{2,1} = \{0.5, 0.167, 0.167, 0.167\}^T$, and $\boldsymbol{\omega}_{2,k} = 0.25 \times \mathbf{1}_4$ for $k = 2, \ldots, 4^2$. So it is between scales $j = 2$ and 3 that the first evidence is seen of the elevation in the uppermost northwest corner. At subsequent coarser scales, we see a similar pattern in the upper northwest corner with $\boldsymbol{\omega}_{1,1} = \{0.333, 0.222, 0.222, 0.222\}^T$ while $\boldsymbol{\omega}_{1,k} = 0.25 \times \mathbf{1}_4$ for $k = 2, \ldots, 4$, and, at the coarsest scale, $\boldsymbol{\omega}_{0,1} = \{0.27, 0.24, 0.24, 0.24\}^T$. Note, however, that the localized elevation, as seen in $\boldsymbol{\omega}_{1,1}$ and $\boldsymbol{\omega}_{0,1}$, is attenuated with decreasing scale due to repeated aggregation of

information from non-elevated regions. A visual summary of the effects just described may be found in Figure 3.

3. Multiscale Detection of Local Deviations in Relative Risk

3.1. A Bayesian Testing Framework

In this section we cast the task of finding local deviations in relative risk within a hypothesis testing framework, and we adopt a Bayesian perspective in deriving our tests. We offer two sets of tests: one general, aimed at detecting arbitrary deviations from uniformity, and the other more specific, aimed particularly at detecting local elevations. In both cases, our approach is to specify null and alternative hypotheses in the spatial domain implicitly, through the explicit specification of collections of simple nulls and alternatives on the values of the $\boldsymbol{\omega}_{j,k}$. In this, we are proceeding in analogy to standard practice in Bayesian wavelet shrinkage methods for nonparametric function estimation which, at their heart, are based on the principle of coefficient by coefficient testing in the space of the wavelet coefficients of observed data. See [9], for example.

Let $\boldsymbol{\gamma} \equiv \{\{\gamma_{j,k}\}_{k=1}^{N_j}\}_{j=0}^{J-1}\}$ be a collection of Bernoulli random variables. Letting $\gamma_{j,k}$ indicate whether (1) or not (0) the parameter vector $\boldsymbol{\omega}_{j,k}$ 'deviates' from $\mathbf{e}_{j+1,ch(k)}/e_{j,k}$, we define the local null and alternative hypotheses $H_{j,k}^{(0)} : \gamma_{j,k} = 0$ and $H_{j,k}^{(1)} : \gamma_{j,k} = 1$. Now let $\mathcal{H}$ be the set of all models $\boldsymbol{\gamma}$ obtainable by combinations of $H_{j,k}^{(0)}$'s and $H_{j,k}^{(1)}$'s, with the obvious restriction that only one of $H_{j,k}^{(0)}$ and $H_{j,k}^{(1)}$ may be included in any given model, for each (j,k). Define $H^{(0)}$ to be the model in which $\boldsymbol{\gamma} \equiv \mathbf{0}$. Detecting deviations from uniformity will be equated with declaring in favor of models $H \in \mathcal{H} \setminus H^{(0)}$ under an appropriate posterior. The precise form of our posterior will follow from the Poisson sampling model defined in Section 2 and the specification of (i) a conditional prior density $p(\{\boldsymbol{\omega}_{j,k}\} \,|\, \boldsymbol{\gamma})$, and (ii) a probability mass function $\Pr(\boldsymbol{\gamma})$. While for completeness we could also specify a prior structure for the global intensity $\mu_{0,1}$, given our focus on testing for the characteristics of the multiscale parameters $\boldsymbol{\omega}_{j,k}$ and the factorization in (3), an explicit definition of this component is unnecessary.

3.1.1. Detecting Arbitrary Deviations from Uniform Relative Risk We concentrate first on the case in which general deviations from $H^{(0)}$ are of interest. We model the $\gamma_{j,k}$'s as independent Bernoulli random variables i.e., $\Pr(\boldsymbol{\gamma}) = \prod_{j,k} \alpha_{j,k}^{\gamma_{j,k}} (1-\alpha_{j,k})^{1-\gamma_{j,k}}$, for values $\{\alpha_{j,k}\}$ in $[0,1]$. Then, conditional on $\boldsymbol{\gamma}$, we specify that

$$\boldsymbol{\omega}_{j,k} \,|\, \gamma_{j,k} \;\sim\; \begin{cases} \delta_{\mathbf{e}_{j+1,ch(k)}/e_{j,k}}, & \text{if } \gamma_{j,k} = 0 \\ \text{Dirichlet}(c_j T \mathbf{e}_{j+1,ch(k)}), & \text{if } \gamma_{j,k} = 1, \end{cases} \tag{5}$$

where $\delta_{\mathbf{e}_{j+1,ch(k)}/e_{j,k}}$ indicates a point mass at $\mathbf{e}_{j+1,ch(k)}/e_{j,k}$, T is a constant meant to capture an overall level of relative risk in D, and the c_j are scale-dependent hyperparameters whose effect is to influence the relative variation within each scale. Finally, given $\boldsymbol{\gamma}$ and the $\boldsymbol{\omega}_{j,k}$'s, sampling from the conditional distributions $\boldsymbol{Y}_{j+1,ch(k)} | \boldsymbol{\omega}_{j,k}, Y_{j,k}$, according to the multinomial distributions defined in (3) yields the observations in $\boldsymbol{Y}_J$.

Using standard calculations, it is easy to show that under this model the posterior evidence

in favor of a hypothesis $H \in \mathcal{H}$ i.e., $\Pr(H \mid \boldsymbol{Y}_J) = \Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J)$, for $\boldsymbol{\gamma} \equiv \boldsymbol{\gamma}(H)$, is given by

$$\Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J) = \prod_{j=0}^{J-1} \prod_{k=1}^{N_j} \rho_{j,k}^{\gamma_{j,k}} (1-\rho_{j,k})^{1-\gamma_{j,k}} \quad , \tag{6}$$

where

$$\rho_{j,k} = \frac{O_{j,k}}{1+O_{j,k}} \tag{7}$$

and

$$O_{j,k} = \frac{\Pr(\gamma_{j,k}=1 \mid \boldsymbol{Y}_{j+1,ch(k)}, Y_{j,k})}{\Pr(\gamma_{j,k}=0 \mid \boldsymbol{Y}_{j+1,ch(k)}, Y_{j,k})} = \frac{\alpha_{j,k}}{1-\alpha_{j,k}} \times \frac{\Pr(\boldsymbol{Y}_{j+1,ch(k)}|Y_{j,k}, \gamma_{j,k}=1)}{\Pr(\boldsymbol{Y}_{j+1,ch(k)}|Y_{j,k}, \gamma_{j,k}=0)} \quad , \tag{8}$$

with

$$\Pr(\boldsymbol{Y}_{j+1,ch(k)} \mid Y_{j,k}, \gamma_{j,k}=i) \quad = \\ \int \Pr(\boldsymbol{Y}_{j+1,ch(k)} \mid Y_{j,k}, \boldsymbol{\omega}_{j,k}, \gamma_{j,k}=i) \Pr(\boldsymbol{\omega}_{j,k} \mid \gamma_{j,k}=i)\, d\,\boldsymbol{\omega}_{j,k}, \quad i=0,1, \tag{9}$$

so that

$$\Pr(\boldsymbol{Y}_{j+1,ch(k)} \mid Y_{j,k}, \gamma_{j,k}=1) = \frac{\Gamma\left(c_j\ T\ e_{j,k}\right)}{\prod_{i=1}^{n_{j,k}} \Gamma\left(c_j\ T\ e_{j+1,ch(k,i)}\right)}\ \frac{y_{j,k}!}{\prod_{i=1}^{n_{j,k}} y_{j+1,ch(k,i)}!} \times \\ \frac{\prod_{i=1}^{n_{j,k}} \Gamma\left(c_j\ T\ e_{j+1,ch(k,i)} + y_{j+1,ch(k,i)}\right)}{\Gamma\left(c_j\ T\ e_{j,k} + y_{j,k}\right)}, \tag{10}$$

and

$$\Pr(\boldsymbol{Y}_{j+1,ch(k)} \mid Y_{j,k}, \gamma_{j,k}=0) = \frac{y_{j,k}!}{\prod_{i=1}^{n_{j,k}} y_{j+1,ch(k,i)}!} \prod_{i=1}^{n_{j,k}} \left(\frac{e_{j+1,ch(k,i)}}{e_{j,k}}\right)^{y_{j+1,ch(k,i)}}. \tag{11}$$

Here $ch(k,i)$, in equations (10) and (11) refers to the i-th element of the set $ch(k)$ i.e., the i-th among the children of k.

From the above expressions we see that the posterior is itself a product of independent Bernoulli random variables, with probability of success $\rho_{j,k}$ for the (j,k)-th variable, where the $\rho_{j,k}$ are defined in terms of the posterior odds $O_{j,k}$, and the marginal data likelihoods arising in the latter have the standard, closed-form expressions under a multinomial-Dirichlet model. Therefore, from expression (6) we see that selecting the most likely model in $\mathcal{H}$, which corresponds to selecting the optimal combination of 0's and 1's, reduces to deciding whether $\rho_{j,k}$ or $1-\rho_{j,k}$ is larger, for each (j,k). But each $\rho_{j,k}$ is a monotone increasing function of the posterior odds $O_{j,k}$. So this problem is equivalent to choosing the hypothesis $H_{j,k}^{(i)}$, $i=0,1$, that maximizes $\Pr(H_{j,k}^{(i)}|\boldsymbol{Y}_J)$, for each (j,k).

The end result is that, given values of the $\alpha_{j,k}$ and c_j, comparison across hypotheses in $\mathcal{H}$ may be done in a highly computationally efficient manner under this model, through direct implementation of the formulas above. Of course, one might also wish to incorporate additional structure into our model – modeling the c_j's, for example – in which case an appropriate MCMC sampling procedure may be necessary to explore the posterior model space. In that case, the factorization structure inherent in (6), inherited from that in (3), should still be useful in

simplifying the MCMC sampling. However, as we shall argue below, in Section 3.2, a choice of $c_j \equiv 1$ appears to be quite sensible in the empirical examples we have examined. Therefore, we have chosen not to pursue this particular extension of our model here, but see [7], for example, for a related extension in the context of disease mapping.

3.1.2. Detecting Local Elevation in Relative Risk Next consider the case in which deviations from uniform relative risk are felt likely to occur only in the form of local elevations. That is, analogous to the second example of Section 2.2, we restrict our attention to those models $H \in \mathcal{H} \setminus H^{(0)}$ for which $\theta_{J,k} \geq 1$ for all (J,k), with strict inequality holding for some spatially localized collection(s) of measurement regions $B_{J,k}$. Note that such a collection of $B_{J,k}$ at the finest scale may or may not nest exactly within one of the coarser, user-specified subregions $B_{j,k}$, at some scale j. While the case in which nesting does occur is useful here from a conceptual viewpoint, a lack of nesting is likely to be more typical in practice. We explore the effect of this issue in more detail in the simulations of Section 4, and comment as well in the discussion of Section 6.

Now recall that if the risk is elevated for all $B_{J,k}$ within some single coarser subregion B_{j_0,k_0}, then there are two classes of behaviors among the multiscale parameters $\boldsymbol{\omega}_{j,k}$. Those $\boldsymbol{\omega}_{j,k}$ at scales $j \geq j_0$, and at scales $j < j_0$ for which the corresponding $B_{j,k}$ does not overlap with B_{j_0,k_0}, do not deviate from the nominal values $\mathbf{e}_{j+1,ch(k)}/e_{j,k}$. But for each of those subregions $B_{j,k}$ containing B_{j_0,k_0}, the $\boldsymbol{\omega}_{j,k}$ will deviate from the nominal values.

From a modeling perspective, we wish to reflect these relationships among the $\boldsymbol{\omega}_{j,k}$ in the parameters $\gamma_{j,k}$. In particular, we want to use the $\gamma_{j,k}$ to encode the hierarchical structure among the $\boldsymbol{\omega}_{j,k}$ in the presence of elevation. We accomplish this goal by specifying that (i) if $\gamma_{j,k} = 1$, for a given (j,k), then $\gamma_{j',k'} = 1$ for all (j',k') such that $B_{j,k} \subset B_{j',k'}$, but (ii) if $\gamma_{j,k} = 0$, then $\gamma_{j',k'} = 0$ for all (j',k') such that $B_{j',k'} \subset B_{j,k}$. Put another way, if one pictures a tree associated with the hierarchy of partitions $\{B_{j,k}\}$, such as the quad-tree in Figure 1, and if the $\gamma_{j,k}$ are used to 'decorate' the nodes of this tree, then the above conditions are equivalent to saying that 1's must be traceable in a path from any initial node to the root, while 0's propagate only 0's in moving towards the leaves.

This type of structure occurs in a number of places in the multiscale modeling literature (e.g., [10, 3, 11]). A common probabilistic model used to enforce this structure, while otherwise allowing for stochastic variation, is a scale-recursive Markov model, where

$$\Pr(\gamma_{j,k} = 1 \,|\, \gamma_{j-1,a_{j-1}(k)} = 1) = \beta_{j,k}$$

and

$$\Pr(\gamma_{j,k} = 1 \,|\, \gamma_{j-1,a_{j-1}(k)} = 0) = 0 \ .$$

Here $a_{j-1}(k)$ is the position index k' of that ancestor subregion $B_{j-1,k'}$ that contains $B_{j,k}$, and the $\beta_{j,k}$ are values in $[0,1]$. At scale $j = 0$ the model specifies simply that $\Pr(\gamma_{0,1} = 1) = \beta_{0,1}$, since $B_{0,1}$ has no ancestor.

Adopting this model for a prior distribution on $\boldsymbol{\gamma}$, we specify that

$$\Pr(\boldsymbol{\gamma}) = \prod_{j=0}^{J-1} \prod_{k=1}^{N_j} \Pr(\gamma_{j,k} \,|\, \gamma_{j-1,a_{j-1}(k)}) \ , \tag{12}$$

using the convention that $\Pr(\gamma_{0,1} \,|\, \gamma_{-1,a_{-1}(k)}) \equiv \Pr(\gamma_{0,1})$. Given values for $\boldsymbol{\gamma}$, let the values of

the $\boldsymbol{\omega}_{j,k}$ and of $\boldsymbol{Y}_J$ follow as previously. Then it is clear that the posterior has the form

$$\Pr(\boldsymbol{\gamma} \,|\, \boldsymbol{Y}_J) \propto \prod_{j=0}^{J-1} \prod_{k=1}^{N_j} \Pr(\boldsymbol{Y}_{j+1,ch(k)} | Y_{j,k}, \gamma_{j,k}) \Pr(\gamma_{j,k} | \gamma_{j-1,a_{j-1}(k)}) \ . \tag{13}$$

While the form of the posterior here is not quite as simple as in the setting of Section 3.1.1, it is still in fact possible – given values for the $\beta_{j,k}$ and, if desired, the c_j – to determine which value of $\boldsymbol{\gamma}$ optimizes $\Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J)$ in a computationally efficient manner. The algorithm used for this computation is a dynamic programming algorithm that is a variation on the sort of bottom-up, optimal tree-pruning algorithm that underlies the well-known CART methodology [12]. At scale $j = J-1$, a test of $H^{(0)}_{J-1,k}$ versus $H^{(1)}_{J-1,k}$ is made, for each $k = 1, \ldots, N_{J-1}$, in a manner analogous to the individual Bernoulli tests in Section 3.1.1. Next, at scale $j = J-2$, a comparison is made between (i) the choice $\gamma_{J-2,k} = 1$, with the values of the descendant $\gamma_{J-1,k}$'s set to those optimal choices $\hat{\gamma}_{J-1,k}$ just inferred, and (ii) the choice $\gamma_{J-2,k} = 0$, with all descendant $\gamma_{J-1,k}$'s necessarily zero as well. This process continues, working from fine scales to coarse, comparing local posterior models of increasingly larger size at each step, until at scale $j = 0$ a single optimal vector $\hat{\boldsymbol{\gamma}}$ is returned. The total number of local comparison steps is $\mathcal{O}(N_J)$. See [3], for example, for additional details.

On a final note, we point out that if one prefers to use Bayes factors [13] for model selection, these quantities are in fact the same in both the model of this section and that of Section 3.1.1. But the space of possible alternative hypotheses in the former case is a restricted subset of that in the latter case. In the numerical work presented in Sections 4 and 5, we have provided results in the form of posterior probabilities.

3.2. Choice of Hyperparameters and Prior Probabilities on Hypotheses

We turn now to the issue of the effect on our tests of the values of the hyperparameters, c_j, and the prior probabilities on hypotheses, $\alpha_{j,k}$ and $\beta_{j,k}$. Because of the scale/position decomposition of the posterior $\Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J)$ in both of our models, and because of the similarity of the local posterior components in this decomposition under both models, we focus primarily on the first model and examine just the behavior of the log-posterior odds ratio $O_{j,k}$ in (8).

The local posterior probability $\Pr(\gamma_{j,k}|\boldsymbol{Y}_J) = \rho_{j,k}$ will be greater than or equal to 0.5 if and only if $\log O_{j,k} \geq 0$. Recall that this log-posterior odds is the sum of two components: the log-prior odds $\log[\alpha_{j,k}/(1-\alpha_{j,k})]$ and the logarithm of the ratio of the marginal data likelihoods. The first takes the form of a logistic function in $\alpha_{j,k}$, and the second is a more complicated function of T, c_j, $\boldsymbol{Y}_{j+1,ch(k)}$, and $\boldsymbol{e}_{j+1,ch(k)}$. We assume that T, a measure of the global relative risk on D can be set relatively easily, given the comparatively quite coarse scale to which it corresponds. For example, in [7] the standardized mortality/morbidity ratio (SMR) at scale $j = 0$ was used successfully as a plug-in estimate of T i.e., $\hat{T} = Y_{0,1}/e_{0,1}$. The effect of the other parameters is the subject of the plots shown in Figures 4 and 5.

In Figure 4 is shown the value of $\log O_{j,k}$ for the scenario underlying the schematic diagram in Figure 2. For each of the scales $j = 3, 2, 1$, and 0, the log posterior odds is plotted as a function of $\alpha_{j,k}$, for k corresponding to a quad $B_{j,k}$ within or containing the region of elevation in the upper left-hand corner. The value of c_j has been set to 1, thus eliminating its role, while the values of $\boldsymbol{Y}_{j+1,ch(k)}$ have been set to their expected values $\boldsymbol{\mu}_{j+1,ch(k)}$. Curves are shown for a number of choices of relative risk parameter $\theta_{J,k}$ in the elevated region (including the

choice $\theta_{j,k} = 1$, corresponding to no elevation). At the highest scale, $j = 3$, the curves show no evidence in favor of the alternative, as should be the case, except under relatively high values of prior probability $\alpha_{j,k}$ for the alternative. At the other three scales, we see various degrees of evidence in favor of the alternative, increasing in the value of $\theta_{j,k}$ and decreasing overall as j decreases. Note that in all of the plots we see that the basic logistic curve is simply shifted vertically by an amount dictated by the log marginal likelihood ratio.

In Figure 5 we again show the value of $\log O_{j,k}$, for the same scenario as in Figure 4, but this time as a function of c_j, with $\alpha_{j,k} = 1/2$ (thus eliminating its role). In the Dirichlet model in (5), as $c_j \to \infty$, the distribution approaches a point mass at $\boldsymbol{e}_{j+1,ch(k)}$, which is equivalent to the prior on $\boldsymbol{\omega}_{j,k}$ under the null model $\gamma_{j,k} = 0$. Hence, all of the curves in Figure 5 approach zero for large values of c_j. For smaller values of c_j, the curves display various behaviors. At the finest scale $j = 3$, the curves all decrease sharply as $c_j \to 0$. Again, this reflects the fact that the relative risk, while elevated, is nonetheless constant in a nested subregion at this scale. At the other scales, where the elevation can now be 'felt', there is generally a value of $c_j < 1$ for which $\log O_{j,k}$ peaks, before decreasing smoothly as c_j continues to decrease. The size of this peak decreases with decreasing $\theta_{j,k}$, within each scale j, and also decreases with decreasing j, for each fixed value of $\theta_{j,k}$. This latter behavior reflects the fact that the effect of a local elevation in relative risk at a certain scale is 'felt' less and less at coarser scales.

Weighing the overall evidence provided by the above plots, we conclude that the value $c_j = 1$ represents a reasonable, conservative choice, yielding a value of $\log O_{j,k} \approx 0$ in the absence of relative risk when $\alpha_{j,k} \approx 1/2$. Further support for this choice may be found in [7], in the context of disease mapping, where use of $c_j = 1$ yields a multiscale estimator of relative risk that is essentially a simple average of estimators in the two extreme cases of $c_j = 0$ and $c_j = \infty$.

Additionally, the plots in Figure 4 suggest that $\alpha_{j,k}$ (and $\beta_{j,k}$, by extension) plays a potentially strong role primarily in the extremes, near zero or one, and a more moderate role otherwise. In the special case of $\alpha_{j,k} = 1/2$ for all (j, k), the role of the $\alpha_{j,k}$ in the test is eliminated. Note that in aggregate this corresponds to assigning equal prior probability among all of the hypotheses in $\mathcal{H}$. This choice is in fact what we have used in the work underlying the numerical results of Sections 4 and 5. Nevertheless, the real potential in our choice of priors on the space of hypotheses is in the manner in which each aggregate hypothesis is broken down into a collection of local hypotheses. In particular, if the user has substantive prior information on the odds of deviations from uniformity in various localities, at various scales, such as might come from pooling reports from administrative centers at different geo-political levels, this information can be incorporated in a natural manner. Alternatively, hypotheses as to the behavior of relative risk as a function of scale might be incorporated through the choice of $\alpha_{j,k} \equiv \alpha_j$, for some set of J parameters $\{\alpha_0, \ldots, \alpha_{J-1}\}$. The effective use of the flexibility allowed by this parameterization of priors on our hypotheses requires further study.

Such study might be aided by the derivation of expressions for spatial covariance under various choices of prior, which should be feasible using calculations similar to those in [14]. It is important to note that, unlike models defined with respect to a spatial indexing, such as the hierarchical Bayes Poisson-lognormal model implemented in the application of Section 5, the multiscale models here have correlation functions that decay in a manner governed by the proximity of spatial regions $B_{j,k}$ on the tree corresponding to the underlying spatial hierarchy. This is in analogy to the temporal or spatial correlation structure induced by wavelet-based models in the signal and image processing literatures. Such tree-based proximities, while not equivalent to actual temporal or spatial distances, have been found to serve as useful proxies.

Detailed spatial covariance expressions for models similar to those proposed here can be found in [14].

4. Simulation

We conducted a simulation study, aimed at illustrating the potential of our multiscale detection framework, as viewed from a handful of simple scenarios. The initial data space D was taken to be a square region, and the nested hierarchy $\mathcal{B}^{(J)}$ was defined through the generic quad-tree structure as illustrated in Figure 1, a structure common in image processing and computer vision. Specifically, D was partitioned into $2^j \times 2^j = 2^{2j}$ identical subregions, for $j = 0, 1, \ldots, 4$, so that each square subregion at scales $j < 4$ consists of four sub-subregions at scale $j + 1$. At the finest scale $J = 4$, there were a total of $16 \times 16 = 256$ subregions, which we refer to as 'pixels', as is standard.

Simulations were conducted under a total of five different scenarios of landscape design. The first landscape has uniform relative risk; that is, each pixel has a relative risk of one. The rest of the landscapes each contain a single area of elevated relative risk. The area in Landscape 2 and Landscape 3, which we will refer to as A2, is comprised of $2 \times 2 = 4$ pixels, each pixel with a relative risk of 3. We consider a larger elevated area in Landscape 4 and Landscape 5, comprised of $4 \times 4 = 16$ pixels, each also with a relative risk of 3. We will refer to this area as A4. Landscape 2 and Landscape 3 differ in the location of A2 within each, as does the location of A4 within Landscape 4 and Landscape 5. The different locations were chosen so as to allow for examination of the effect on our detection framework of different degrees of nesting of the elevated area within the elements of the nested hierarchy.

Figure 6 illustrates all five landscape designs. If we call the pixels at scale $J = 4$ 'children', and subregions at scales $j = 3, 2$, and 1, 'parents', 'grandparents', and 'great-grandparents', respectively, then landscapes 2 through 5 may be characterized as follows. Landscape 2 shows A2 on the uppermost northwest corner, with the four children forming a single parent. The A2 in Landscape 3 is located in the interior and is non-nested, that is, the four children each originate from a different parent. The A4 in Landscape 4 is also located in the uppermost northwest corner. Its 16 children form four parents and a single grandparent. Similar to Landscape 3, the A4 in Landscape 5 is located in the interior and non-nested as well. Its children originate from nine distinct parents, four distinct grandparents and four distinct great-grandparents.

We chose the pixel-level expected counts $e_{J,k}$ to be uniformly equal to 7, which resulted in pixel-level mean parameters $\mu_{J,k} = \theta_{J,k}\, e_{J,k}$ of either $1 \times 7 = 7$ or $3 \times 7 = 21$. The hyperparameters c_j were set to 1 for all $j = 0, \ldots, J-1$, for the reasons described in Section 3.2, and hence were effectively omitted from our study. Under this scenario, we generated 100 independent Poisson samples with intensity $\mu_{J,k}$ for every location k, resulting in 100 data sets of 256 observations for each of the five landscapes on D.

Operating under the testing framework of Section 3.1.2, we specified the values of the $\beta_{j,k}$ implicitly by imposing the assumption that all models are equally likely a priori. Values of the posterior $\Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J)$, as a function of candidate models $\boldsymbol{\gamma}$, are then provided by (13). There are a variety of summaries of these posterior probabilities that could be examined to gauge the potential of our framework for detecting and isolating the localized risk elevations in our landscapes. A particularly useful set of quantities for visualization is the set of

local marginal posterior model probabilities $\Pr(\gamma_{j,k} = 1|\boldsymbol{Y}_J)$, in that each such probability summarizes the total posterior evidence for elevated risk in the region of the corresponding partition element $B_{j,k}$. Under our choice of equally likely prior probabilities $\Pr(\boldsymbol{\gamma})$, these marginal posterior probabilities are particularly easy to compute, as they are in fact equal to $\Pr(\gamma_{j,k} = 1 \mid \boldsymbol{Y}_{j+1,ch(k)}, Y_{j,k})$, which is the same as the quantity $\rho_{j,k}$ under the general model of Section 3.1.1. However, we note that higher order marginal distributions (e.g., of $\gamma_{j,k}, \gamma_{j',k'} \mid \boldsymbol{Y}_J$) do not share a similar equivalence under the two testing frameworks.

We calculated these local marginal posterior probabilities, for each pair (j, k), for each of the 100 data sets corresponding to each of the five landscapes. The averages of these quantities, as a function of j and k, across data sets within landscapes, are presented in Figure 7, and serve to indicate that the presence or absence of local elevation in relative risk is clearly 'felt' in our detection framework. For example, consider the case of Landscape 1. Since Landscape 1 has uniform relative risk we would not expect to see across scales any evidence of local elevation in the form of $\Pr(\gamma_{j,k} = 1 \mid \boldsymbol{Y}_J) >> 0.5$. In fact, the mean posterior probabilities across all scales in Figures 7(1a)-7(1d) are between 0.4 and 0.6.

In Landscape 2, even for the area with elevated risk, all the parents at scale $j = 3$ are dividing equally among their children at scale $j = 4$; hence, we should see no evidence of local elevation, which is the case in Figure 7(2a). However, in the upper northwest corner beginning at scale $j = 2$, the grandparent is dividing unequally among the 4 parents at scale $j = 3$; hence, we see a localized elevation with posterior probability above 0.9 in this location and scale, as presented in Figure 7(2b). For the subsequent coarser scales in Figures 7(2c) and 7(2d), we continue to see a localized elevation in the upper northwest corner and a less pronounced elevation in the overall region due to repeated aggregation of the data.

In Figure 7(3a) for Landscape 3, since A2 is non-nested with each child from a different parent, we see immediately from scale 3 to 4 localized elevations pertaining to the four parents in the uppermost northwest corner with posterior probabilities above 0.85. However, since all grandparents at scale $j = 2$ divide equally among the parents at scale $j = 3$, there is no such elevation exhibited in Figure 7(3b). Similar to Landscape 2, there is a localized elevation in Figure 7(3c) and a less pronounced effect in Figure 7(3d).

Landscape 4 is similar to Landscape 2 with the exception that the localized elevation does not arise until the division of the great-grandparents at scale $j = 1$ to grandparents at scale $j = 2$ as in Figure 7(4c); otherwise, the grandparents divide equally among the parents as does parents among their children, as indicated by the nearly uniform posterior probabilities in Figures 7(4b) and 7(4a) respectively.

In Figure 7(5a) for Landscape 5, we note that there is a fairly large area with posterior probabilities above 0.85 surrounding a smaller area with posterior probability below 0.4. This smaller area corresponds to the only parent in A4 which divides equally among its children. In Figure 7(5b), the 2×2 area in the center indicates the highest localized elevation in the southeast region followed by the southwest and northeast regions, with the smallest effect in the northwest. This corresponds directly with the number of descendants exhibiting elevated risk. In other words while the grandparent at scale $j = 2$ in the northwest region has one grandchild at scale $J = 4$ with elevated risk, the grandparent in the southeast region has nine grandchildren with elevated risk while the grandparents in the southwest and northeast each has three grandchildren with elevated risk. Figure 7(5c) indicates a similar trend in posterior probabilities.

Taken together, the results in Figure 7 suggest that effective use may be made of our

framework for detection of localized anomalous structure in aggregate disease incidence data. In the next section we present confirmation of this potential, through the use of our framework within the context of a comprehensive analysis of cancer mortality data.

5. Application to Tuscany Gastric Cancer Mortality Data

We applied our multiscale method to male gastric cancer mortality data obtained from the Tuscan region of Italy during the period 1980-1989 [1, 15]. For our hierarchy of nested partitions we took the set of nested subregions displayed in Figure 8, corresponding to three of the five geo-political statistical units defined under the NUTS system (Nomenclature des Units Territoriales Statistiques) used by the European community. In particular, shown in Figure 8 are the single Tuscan region itself (NUTS Level II), the nine provinces within that region (NUTS Level III), and the 287 municipalities within those provinces (NUTS Level V). Italy itself corresponds to a single subregion at NUTS Level I, while NUTS Level IV is not used in that country. In the notation of this paper, the regional, provincial, and municipality levels correspond to $j = 0, 1$, and 2, respectively. Data were obtained for males over 35 years of age at the municipality level. Italian age specific rates for the same calendar period were used to obtain expected numbers of deaths [15].

Figure 9 shows a map of the local marginal posterior probabilities for this data at the provincial level i.e., $\Pr(\gamma_{1,k} = 1 \mid \boldsymbol{Y}_J)$, for $k = 1, \ldots, 9$. As in Section 4, in computing these probabilities, all models were treated as equally likely a priori and the parameters c_j were set to 1 for all j. The figure indicates varying degrees of evidence for elevated risk throughout the provinces of the Tuscan region. One province, (Massa-Carrara) has fairly strong evidence against elevated relative risk within it. Two other provinces (Pistoia and Siena) have about equal evidence for or against elevation. The rest of the provinces show strong evidence in favor of elevation within. Among these latter provinces, the evidence for the provinces of Lucca, Firenze, Pisa, and Arezzo is very strong, while that for the provinces of Livorno and Grosseto is slightly less so.

The actual values of the probabilities $\Pr(\gamma_{j,k} = 1|\boldsymbol{Y}_J)$, for the 9 provinces and for the overall Tuscan region, are shown in Table I. In addition, the optimal model $\hat{\boldsymbol{\gamma}} = \arg\max \Pr(\boldsymbol{\gamma}|\boldsymbol{Y}_J)$ is given. The model most likely under the posterior indicates the presence of elevation among provinces within the overall Tuscan region, and within all of the individual provinces except Massa-Carrara and Pistoia. These results, which are based on the joint posterior over the full vector γ, essentially confirm the conclusions drawn above based on examination of the local marginal posteriors. The next 11 most likely models also are presented which, together with the optimal model $\hat{\boldsymbol{\gamma}}$, make up just over 90% of the total mass under the posterior. Most of these models confirm the entries in $\hat{\boldsymbol{\gamma}}$, particularly for those provinces with the highest local marginal posterior values. For those provinces with lower values for these probabilities, it can be seen that the evidence for elevation is correspondingly more mixed.

To supplement these results, we present disease maps from a multiscale analysis of this mortality data [7] in Figures 10c and 10d, at the municipal level ($j = 2$) and provincial level ($j = 1$), respectively. Additionally, disease maps summarizing the observed SMR, also at both levels, can be found in Figures 10a and 10b. Both Figures 10a and 10c at the municipal level indicate a potential severity in the northeast border of Tuscany, and to a lesser degree, at the

provincial level as seen in Figures 10b and 10d. Twelve municipalities comprise the northeast border, with the 4 most northern municipalities found in Firenze and the remaining 8 in Arezzo.

Lastly, by way of comparison, we also present the results under a hierarchical Bayes Poisson-lognormal model as implemented in [7]. We note that this is not a one-to-one comparison to our multiscale hypothesis testing framework as the former presents results at each individual and independent scale j, while the latter presents results at multiple scales $j-1$ having taken into account the individual information at scale j. Additionally, such purely spatial (mono-scale) analyses are generally carried out for the data at the finest resolution J, while our method takes into account the data at all scales $j = 0, \ldots, J$. However, since the Poisson log-normal model used here contains components to capture extra-Poisson variation, we can assess whether a general clustering effect exists in the overall region [1, 16].

Specifically, in this model $Y_k \sim Poisson(\mu_k)$, as in our multiscale model, but the log spatial mean is expressed as $log\ \mu_k = log\ e_k + \alpha + u_k + v_k$, where α captures the overall level of the log relative risk, and u_k and v_k are respectively the unstructured heterogeneity and clustering components of extra-Poisson variation. We consider the quantity,

$$\psi = \frac{sd(v)}{sd(v+u)}, \tag{14}$$

where $sd(\cdot)$ is the empirical marginal standard deviation function. The ψ is the proportion of the extra-Poisson variation that is due to clustering. Details of this model can be found in [17]. A discussion on appropriate priors for v_k and u_k can be found in [18]. Using a Bayesian software package BUGS [19] with one sampling chain, a 10000-iteration burn-in period and a 25000-iteration production period, the posterior mean for ψ at the municipal and provincial levels were found to be $\psi_{mun} = .8902$ and $\psi_{prov} = 0.5567$, indicating that the excess variability in the data at each scale, especially at the municipality level, is due to clustering. The corresponding maps summarizing the estimated relative risk for these models are presented in Figures 10e and 10f. These maps also indicate a potential severity in the northeast border of Tuscany.

6. Discussion

Our goal here in this paper has been to lay out the basic framework for a multiscale approach to detecting anomalous localized structure in aggregate disease incidence data. We employ a Bayesian testing machinery to develop our detection mechanisms. The models underlying this machinery are (intentionally) fairly minimal in their complexity and could be, if desired, augmented in various manners. For example, as was pointed out in Section 3.1.1, the hyperparameters c_j might be provided with prior distributions and inferred as part of the overall posterior-based inference procedure. Or, on a grander scale, one might seek to incorporate our models into a spatio-temporal framework for biosurveillance, perhaps through the specification of time-dependent multiscale parameters $\boldsymbol{\omega}_{j,k}$. In addition, it would be interesting to explore the extension of the multiscale modeling perspective to clustering of other types of disease incidence data. For example, case-to-control data, or event data. In the case of the latter, it seems that extensions to point process data of recent work by Jansen and colleagues with multiscale models on two-dimensional irregular, triangulated meshes (e.g., [6]) would be promising.

The technical details of the framework proposed here are analogous to those in our earlier work on disease mapping [7]. Both are grounded upon the two key elements: (i) a multiscale likelihood factorization, and (ii) the placement of prior distributions on the corresponding multiscale parameters. They differ in the underlying inferential task for which they are designed (i.e., estimation and testing, respectively), and as a result they differ in the nature of the priors. In particular, while the priors here are either a point mass or a Dirichlet distribution, depending on the underlying model (i.e., $\boldsymbol{\gamma}$), in the disease mapping work a simple Dirichlet prior is used alone. It would be natural, and indeed straightforward, to combine the two frameworks into a single framework allowing for disease mapping with interpretable testing aspects, by adopting mixtures of point mass and Dirichlet distributions. We have chosen not to do so here purely in the interests of emphasizing the testing aspect of our approach and its connection to detection and clustering.

Finally, we point out that many of the same research challenges surrounding the disease mapping framework in [7] hold as well for the testing-based framework introduced here in this paper. These include the modeling of edge effects, the role played by the standardization inherent in the use of SMR-like statistics (i.e., through the role of the $\boldsymbol{e}_{j,k}$'s), and the use of multiple and/or non-nested spatial hierarchies. The role played by the spatial hierarchy is an issue that is particularly important to explore further. As mentioned in Section 4, and reflected through our choice of landscapes in the simulations of that section, it is unlikely in practice that the underlying spatial disease structure in a region will be localized precisely according to the boundaries of subregions specified by a hierarchy. This is essentially the 'spatial misalignment' problem of Mugglin *et al.* [20], and is an aspect of the so-called 'modifiable areal unit problem' (MAUP) in Geography (e.g., [21, 22]).

From the perspective of the multiscale modeling literature, this issue is not unlike the problem of using a fixed, orthonormal wavelet basis in the classical wavelet paradigm. Specifically, due to the rigid constraints placed on wavelet locations by the condition of orthogonality, it is often the case that a sharp feature in a signal or image will not line up precisely with any one, single wavelet. The solution proposed in that literature is the adoption of a much larger, redundant (i.e., non-orthogonal) system of wavelets. Inference then becomes a more complicated task, due to the redundancy of information across the wavelets. We are currently exploring, with colleagues in Geography, the extension of these concepts to the context of spatial epidemiology and our multiscale modeling framework, in which the idea of a redundant basis is replaced by that of a library of different spatial hierarchies.

ACKNOWLEDGEMENTS

This research was supported by NSF Grants BCS0079077 and BCS0318209 and ONR Award N00014-99-1-0219.

REFERENCES

1. Lawson AB. *Statistical Methods in Spatial Epidemiology*. Chichester: Wiley, 2001.
2. Vidakovic, B. *Statistical Modeling by Wavelets*. New York: John Wiley, 1999.
3. Kolaczyk ED, Huang H. Multiscale statistical models for hierarchical spatial aggregation, *Geographical Analysis* 2001; **33**:95-118.

4. Huang H-C, Cressie N, Gabrosek J. Fast, resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational and Graphical Statistics* 2002; **11**: 63-88.
5. Zhu J. and Yue W. A multiresolution tree-structured spatial linear model. *Journal of Computational and Graphical Statistics* 2004; (to appear).
6. Delouille V, Jansen M, von Sachs R. Second generation wavelet methods for denoising irregularly spaced data in two dimensions. *Discussion Paper 0305* 2003, Institut de Statistique, UCL.
7. Louie MM, Kolaczyk ED. A multiscale method for disease mapping in spatial epidemiology. *Statistics in Medicine* 2005; **xx**:yyy-yyy.
8. Kolaczyk ED, Nowak RD. Multiscale likelihood analysis, complexity penalized estimation. *Annals of Statistics* 2004; **32**:500-527.
9. Johnstone, IM, Silverman, BW. Empirical Bayes selection of wavelet thresholds. *Annals of Statistics* 2005; (to appear).
10. Donoho, DL. CART and best-ortho-basis: a connection, *Annals of Statistics* 1997; **25**:1870-1911.
11. Crouse, MS, Nowak, RD, Baraniuk, RG. Wavelet-based statistical signal processing using hidden Markov models, *IEEE Transactions on Signal Processing* 1998; **46**:886-902.
12. Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J. *Classification and Regressions Trees*. Belmont, CA: Wadsworth, 1984.
13. Kass RE, Raftery, AE. Bayes factors. *Journal of the American Statistical Association* 1995; **90**(430):773-795.
14. Louie MM, Kolaczyk ED. On the covariance properties of certain multiscale spatial processes. *Statistics, Probability Letters* 2004; **66**(4):407-416.
15. Dreassi E, Biggeri A. Edge Effect in disease mapping. *Journal of the Italian Statistical Society* 1998; **7**(3):267-283.
16. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. New Jersey: John Wiley, 2004.
17. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC: Boca Raton 2004; 158-167.
18. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* 1995; **14**:2411-2431.
19. Spiegelhalter DJ, Thomas A, Best N, Gilks WR. BUGS: Bayesian inference using Gibbs sampling, Version 0.50. *Technical report* 1995, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.
20. Mugglin AS, Carlin BP, Gelfand AE. Fully model based approaches for spatially misaligned data. *Journal of the American Statistical Association* 2000; **95**:877-887.
21. Quattrochi DA, Goodchild MF, (Eds). *Scale in Remote Sensing and GIS*. Boca Raton, FL: CRC Lewis, 1997.
22. Gotway CA, Young LJ. Combining incompatible spatial data. *Journal of the American Statistical Association* 2002; **97**: 632 - 648.

Table I. Results from the analysis of the Tuscan gastric cancer data. For each of the 9 Tuscan provinces, the local marginal posterior probabilities $\Pr(\gamma_{j,k} = 1|\boldsymbol{Y}_J)$ are given, as well as the 12 most likely models under the full posterior. The Bayes optimal model $\hat{\boldsymbol{\gamma}}$ corresponds to the model with a rank of 1. Size of the provinces, given in parentheses, corresponds to the number of municipalities within each.

		Model Rank											
Province (size)	$Pr(\gamma_{j,k} = 1 \mid \boldsymbol{Y}_J)$	1	2	3	4	5	6	7	8	9	10	11	12
Massa-Carrara (17)	0.212	0	0	0	0	1	1	1	0	1	0	0	0
Lucca (35)	0.994	1	1	1	1	1	1	1	1	1	1	1	1
Pistoia (22)	0.408	0	0	1	1	0	0	1	0	1	0	1	1
Firenze (51)	0.999	1	1	1	1	1	1	1	1	1	1	1	1
Livorno (20)	0.949	1	1	1	1	1	1	1	1	1	1	1	1
Pisa (39)	0.996	1	1	1	1	1	1	1	1	1	1	1	1
Arezzo (39)	0.999	1	1	1	1	1	1	1	1	1	1	1	1
Sienna (36)	0.544	1	0	1	0	1	0	1	1	0	0	1	0
Grosseto (28)	0.849	1	1	1	1	1	1	1	0	1	0	0	0
Tuscany (9)	0.999	1	1	1	1	1	1	1	1	1	1	1	1

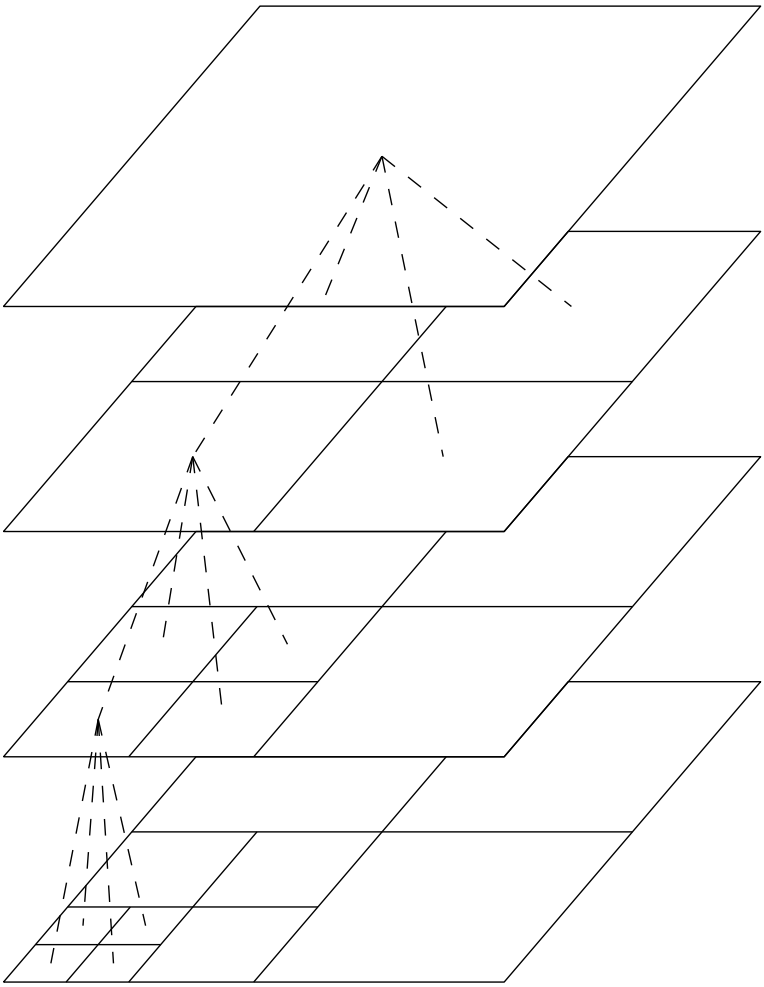

Figure 1. Schematic characterization of a nested spatial hierarchy (a so-called 'quad-tree').

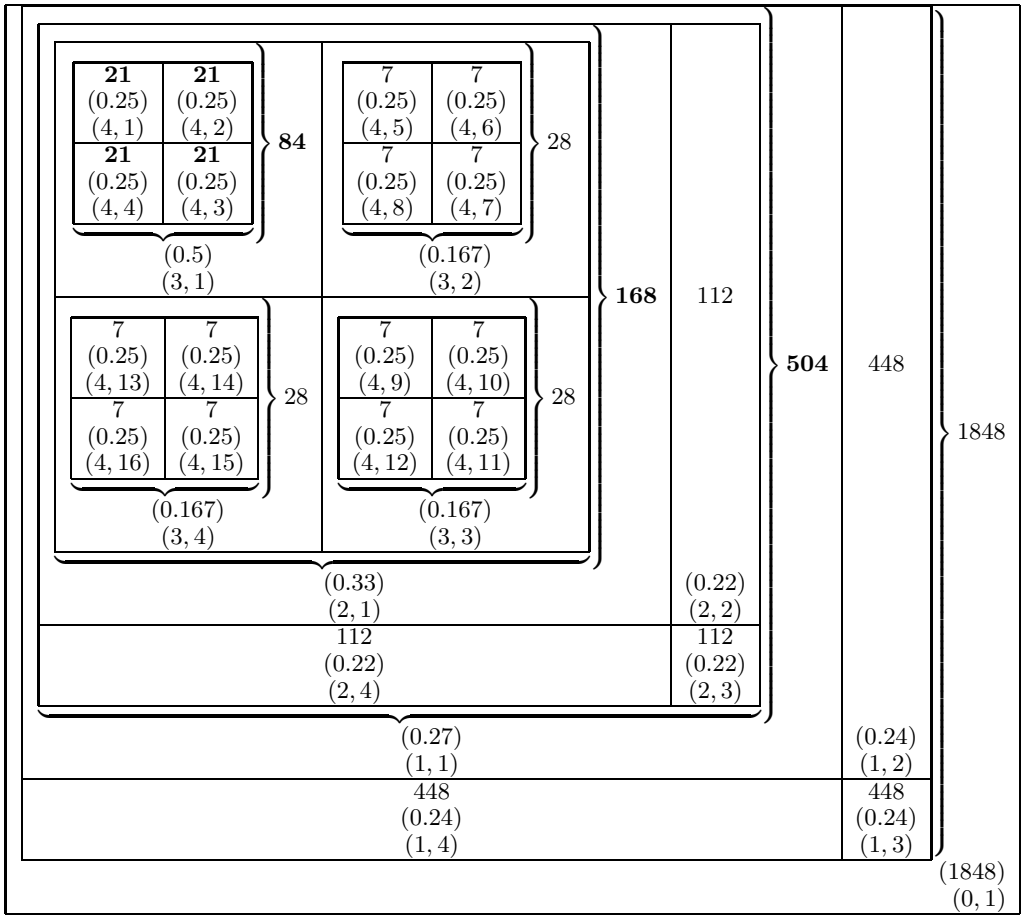

Figure 2. Illustration of the effect of a local elevation in relative risk on the multiscale parameters. The three numbers that appear in each sub-partition are respectively the value of the Poisson intensity parameter $\mu_{j,k}$, the appropriate element of the corresponding multiscale parameter vector $\boldsymbol{\omega}_{j,k}$ (or just $\mu_{0,1}$ at the coarsest scale), and the index of the sub-partition (j = scale, k = position). The spatial intensities are printed in bold for sub-partitions containing elevated areas.

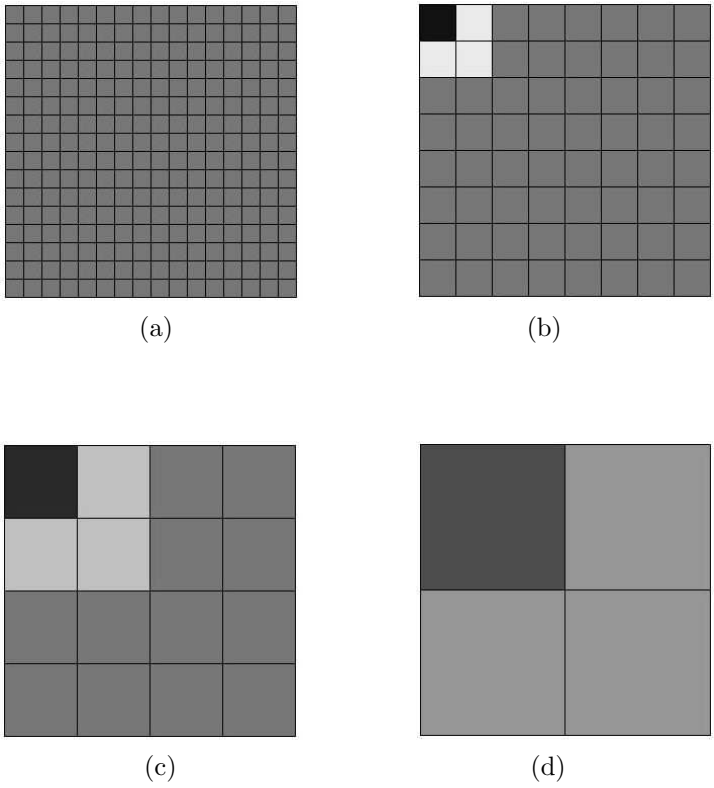

Figure 3. Scale specific images of $\boldsymbol{\omega}_{j,k}$ corresponding to Figure 2

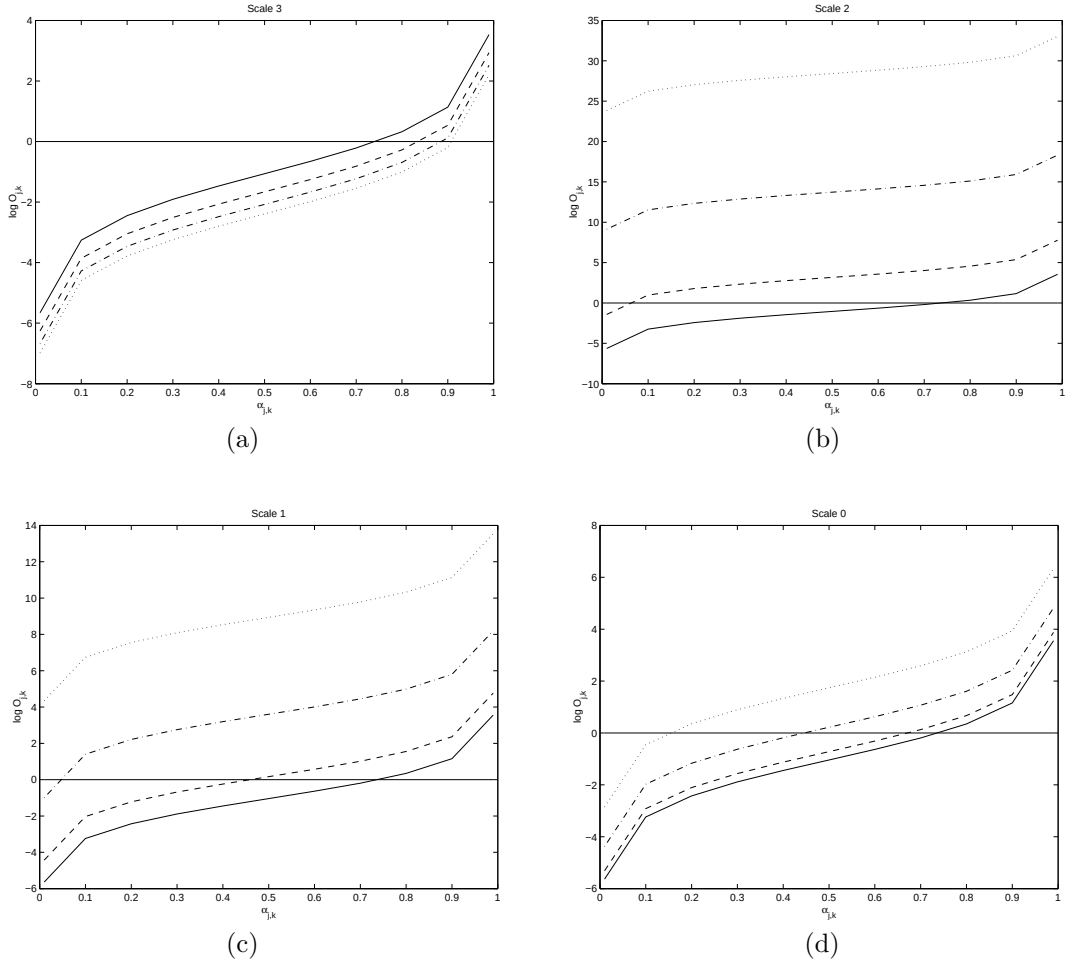

Figure 4. For data $\boldsymbol{Y}_J = \boldsymbol{\mu}_J$ and uniform expected counts $e_{J,k}, \forall(J,k)$, plot of $log\, O_{j,k}$ versus $\alpha_{j,k}$ for $\theta_{J,k} = \{1,2,3,4\}$. For the four plots, $\theta_{J,k} = 1$ is represented by a solid line; $\theta_{J,k} = 2$ by a dashed line; $\theta_{J,k} = 3$ by an alternating dashed and dotted line; and $\theta_{J,k} = 4$ by a dotted line.

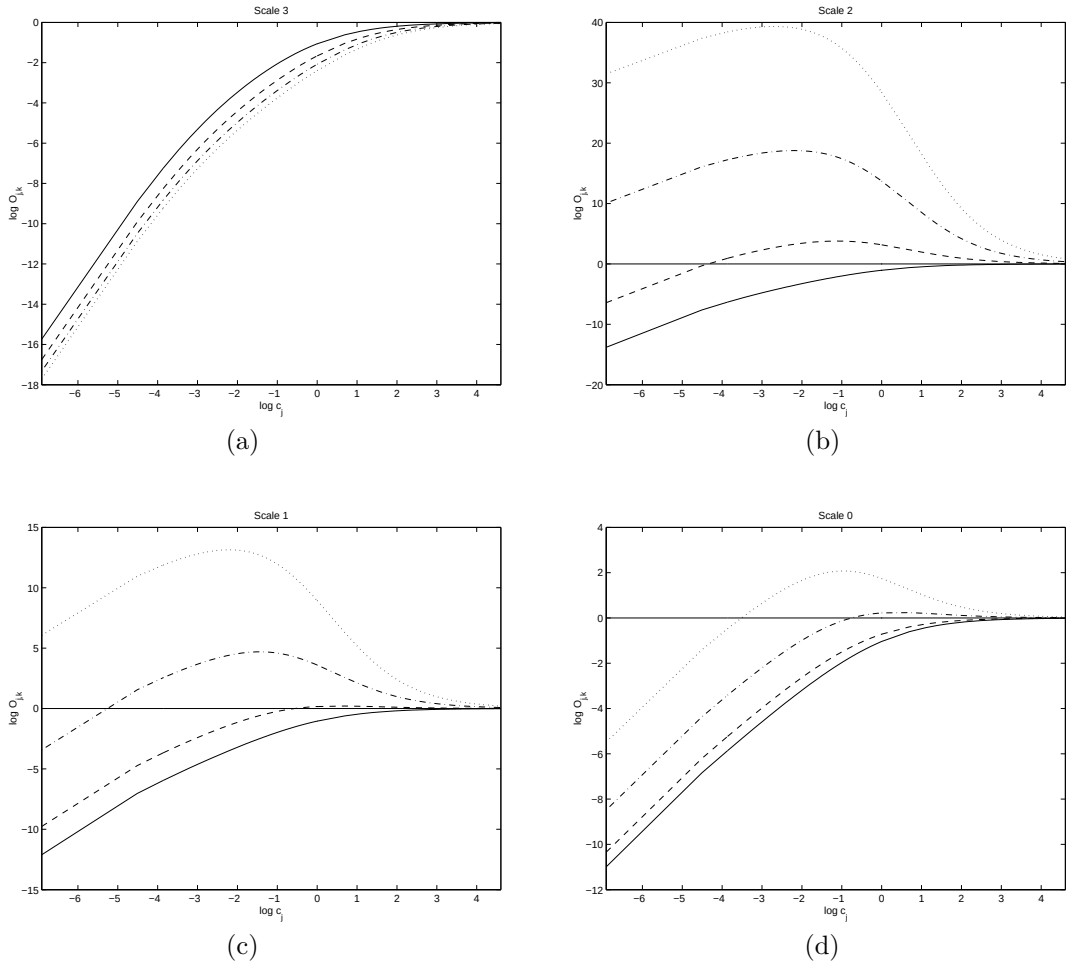

Figure 5. For data $\boldsymbol{Y}_J = \boldsymbol{\mu}_J$ and uniform expected counts $e_{J,k}, \forall(J,k)$, plot of $log\, O_{j,k}$ versus $log\, c_j$ for $\theta_{J,k} = \{1,2,3,4\}$. For the four plots, $\theta_{J,k} = 1$ is represented by a solid line; $\theta_{J,k} = 2$ by a dashed line; $\theta_{J,k} = 3$ by an alternating dashed and dotted line; and $\theta_{J,k} = 4$ by a dotted line.

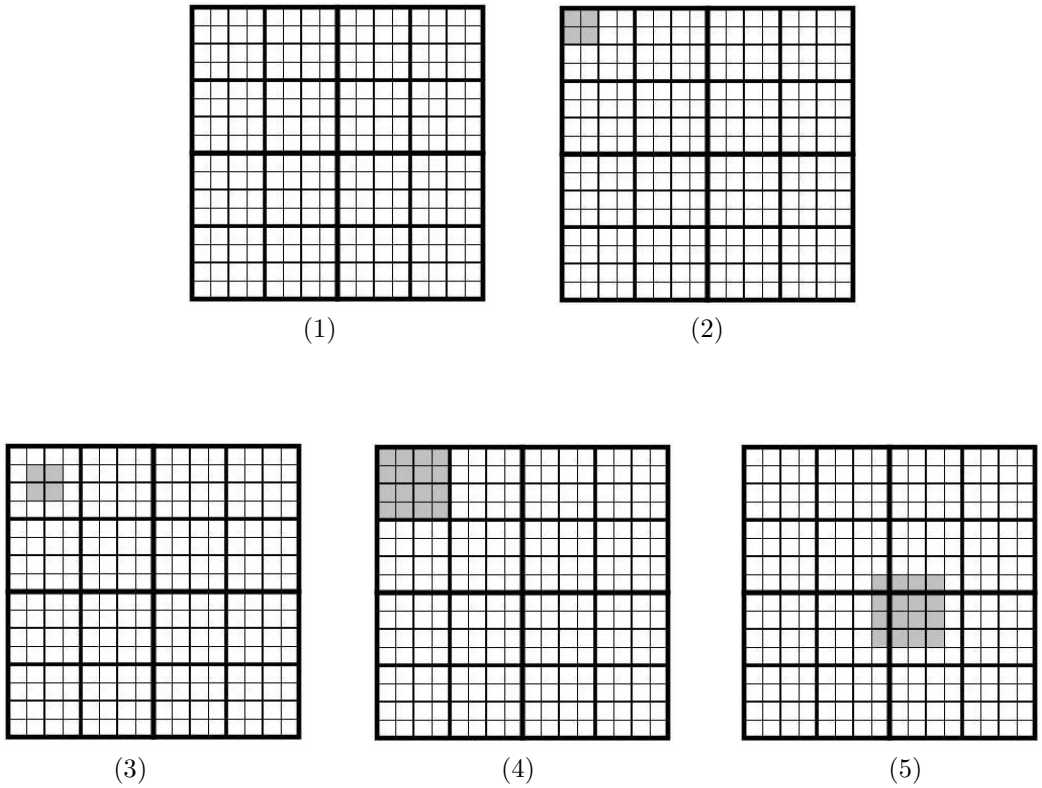

Figure 6. Simulated landscape designs for illustrating performance of the multiscale hypothesis testing framework. Design (1) corresponds to uniform relative risk. Designs (2) and (3) each contain a (shaded) area of elevated risk comprised of 2×2 pixels. Designs (4) and (5) contain a larger area of elevated risk comprised of 4×4 pixels.

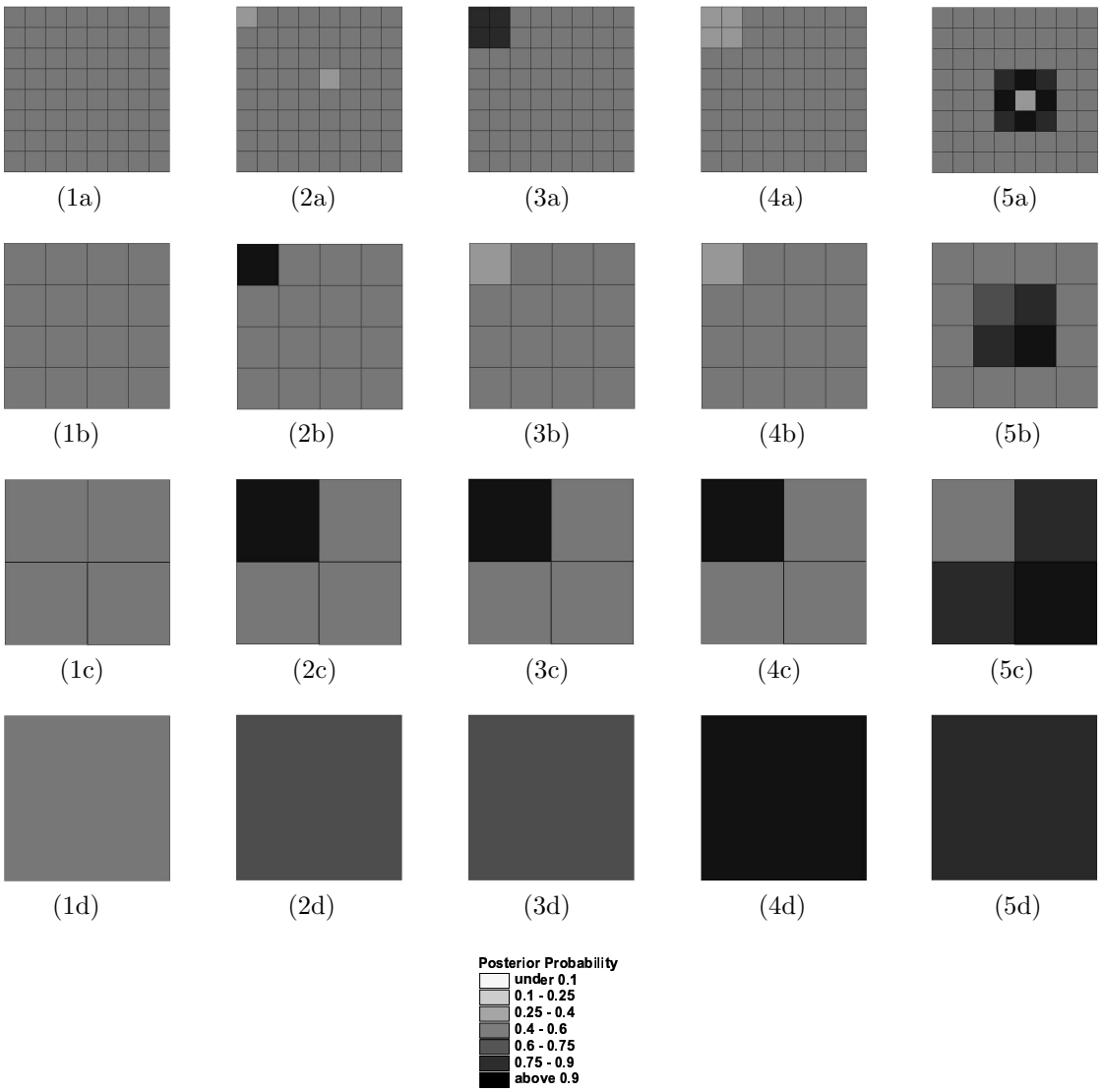

Figure 7. Average of the local marginal posterior probabilities, $\Pr(\gamma_{j,k} = 1 \mid \boldsymbol{Y}_J)$, over 100 samples, corresponding to each of the five landscape designs in Figure 6. Figures #a-#d correspond to Landscape #. Rows, from bottom to top, correspond to results at scales $j = 0$ to 3.

(a)

(b)

(c)

Figure 8. Hierarchy of three nested partitionings of the Tuscan region of Italy: (a) region level, (b) provincial level, and (c) municipality level.

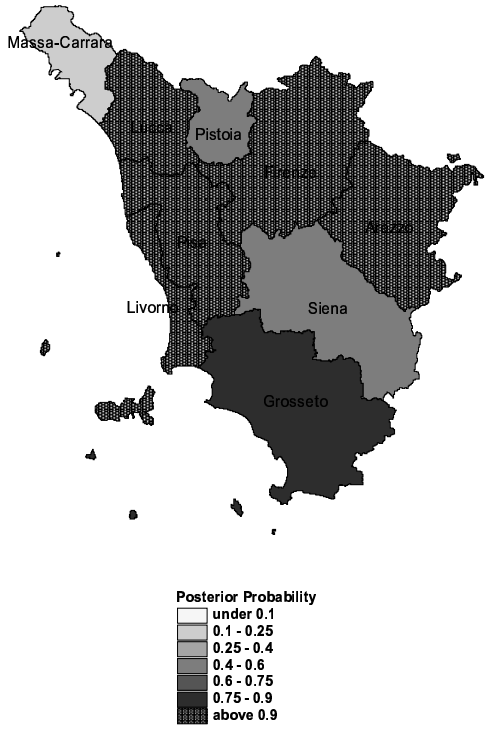

Figure 9. Map, at the provincial level, for posterior probabilities $\Pr(\gamma_{j,k} = 1 \mid \boldsymbol{Y}_J)$, as found in Table I.

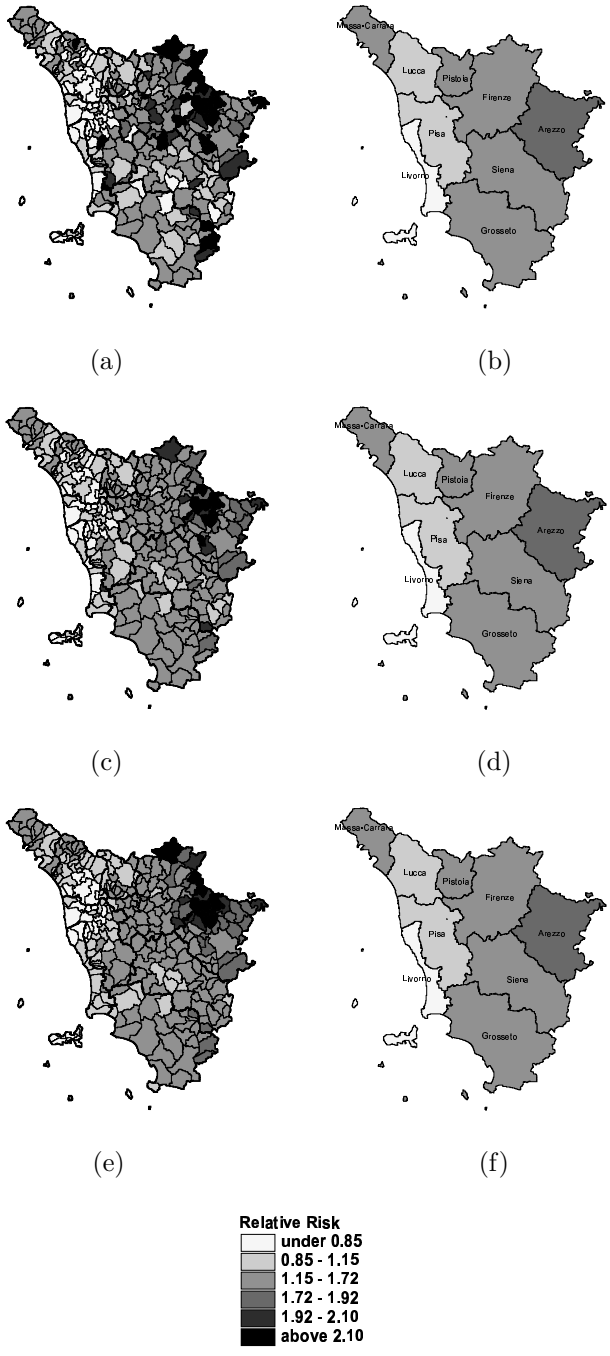

Figure 10. Observed SMR at (a) municipal level and (b) provincial level. Multiscale empirical Bayes estimates of relative risk at (c) municipal level and (d) provincial level. Estimates of relative risk based on Poisson-lognormal model at (e) municipal level and (f) provincial level. Figures (c),(d) and (e) are reproduced from [7].