ELSEVIER

Journal of Statistical Planning and Inference 59 (1997) 45–60

journal of statistical planning and inference

Semiparametric nonhomogeneity analysis[1]

Carey E. Priebe[a,*], David J. Marchette[b], George W. Rogers[b]

[a] *Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, USA*

[b] *Naval Surface Warfare Center, Code B10, Dahlgren, VA 22448, USA*

Received 25 July 1995; revised 3 April 1996

Abstract

Let $\xi(x,\omega)$ be a 'piecewise stationary' random field, defined as an embedding of stationary random fields $\xi^i(x,\omega)$ via the polytomous field $m(x,\omega)$. The domain of definition is partitioned into disjoint regions R^i. Denote the marginals for each $\xi^i(x,\omega)$ by $\alpha^i(\xi)$ so that $\xi(x,\omega) \sim \alpha^i(\xi)$ for $x \in R^i$. Define homogeneity as the situation in which all the α^i are identical versus nonhomogeneity in which there exist at least two regions with differing marginals. To perform a test of these hypotheses without assuming parametric structure for the α^i or choosing a specific type of nonhomogeneity in the alternative requires estimates $\hat{\alpha}^i$ for each region. However, the competing requirements of estimation without restrictive assumptions versus small-area investigation to determine the unknown locations of potential nonhomogeneities lead to an impasse which cannot easily be overcome and has led to a dichotomy of approaches — parametric versus nonparametric. This paper develops a borrowed strength methodology which can be used to improve upon the local estimates which are obtainable by either fully nonparametric methods or by simple parametric procedures. The approach involves estimating the marginals as a generalized mixture model, and the improvement derives from using all the observed data, borrowing strength from potentially dissimilar regions, to impose constraints on the local estimation problems.

AMS 1991 classification: Primary 62M40; secondary 62G10

Keywords: Borrowed strength; Scan process; Mixture model; Profile likelihood; Random field

1. Introduction and summary

In many situations one wishes to perform an analysis of the homogeneity of a random field, often as a precursor to more advanced analysis. For instance, a conclusion of nonhomogeneity may imply a requirement for further analysis, particularly of the suggested regions of nonhomogeneity. A finding of nonhomogeneity may warrant more

* Corresponding author. E-mail: priebe@kronecker.mts.jhu.edu.

[1] This work is partially supported by Office of Naval Research Grant N00014-95-1-0777, Office of Naval Research Grant R&T 4424314, and the Naval Surface Warfare Center Independent Research Program. The authors are grateful to an associate editor and an anonymous referee for many useful suggestions, and to Edward J. Wegman for helpful discussions and support.

PII S0378-3758(96)00095-X

involved change point or change curve analysis (see Carlstein et al. 1994; or the Proc. Applied Change Point Conference, 1994). Uniformity of background conditions is relevant in applications as diverse as astronomy, ecology, epidemiology, etc. (Cressie, 1993). In image analysis testing for homogeneity is often the first step: for PET scan analysis of brain functions homogeneity is the 'no-change' condition and regions of honhomogeneity are of interest for their functionality implications (O'Sullivan, 1995; Worsley, 1995); in mammographic analysis homogeneity may imply the 'uniformly healthy tissue' case while regions of nonhomogeneity warrant closer inspection (Miller and Astley, 1992); a finding of homogeneity in minefield detection implies 'no minefield' while nonhomogeneity again requires further analysis (Smith, 1991; Muise and Smith, 1992; Hayat and Gubner, 1994; Basawa, 1993).

This paper develops a semiparametric scan analysis approach for testing for nonhomogeneity which will serve as a preprocessing step in image analysis and pattern recognition tasks.

1.1. The random field

Let $\xi(x,\omega)$: $R^0 \times \Omega \to \Xi$ be a random field with domain of definition $R^0 \subset \mathbf{R}^n$. Given a polytomous field $m(x,\omega)$ taking on the values $1,\ldots,r$ and r strictly stationary and ergodic fields $\xi^i(x,\omega)$, each with the same domain, we construct ξ as an embedding. Following Carlstein and Lele (1994), let $\xi(x,\omega) = \sum_{i=1}^{r} \xi^i I_{\{m_x=i\}}$. The field $\xi(x,\omega)$ is termed *piecewise strictly stationary*. Here and hereafter m_x denotes the observed value of the random field $m(x,\omega)$ at location x and I_S is the indicator function for the set S.

We will consider the case in which the domain in question is a subset of the integer lattice $\mathbf{Z}^n$ in $\mathbf{R}^n$, $R^0 \subset \mathbf{Z}^n$. The number of *regions* in R^0, sets, not necessarily made up of contiguous lattice sites, consisting only of random variables from a single field ξ^i, is r. Thus the domain is partitioned into a finite number of disjoint regions; $R^0 = \bigcup R^i$ $(i = 1,\ldots,r)$. When the embedding field $m(x,\omega)$ is modelled as random the regions R^i are random sets. Asymptotic considerations involve letting R^0 (the domain of m and the ξ^i) grow. This can be physically realized by obtaining multiple images for which the embedding field $m(x,\omega)$ is identical.

By construction the random variables associated with each region are identically distributed and have the same dependence structure, *class conditional identically distributedness*. Thus $\xi(x,\omega) \sim \alpha^i(\xi)$ for $x \in R^i$ for probability density functions (or, more generally, for distribution functions) α^i.

For instance, in image processing we may consider R^0 to be an $M_1 \times M_2$ lattice of pixel locations and let the value of the field observations $\xi_x \in \Xi = \Re$ represent pixel intensity as in, e.g., German (1990).

1.2. The test for nonhomogeneity

In the simplest case, the goal is to test homogeneity, in the sense of multiple comparisons.

H_0: Homogeneity $(\alpha^i = \alpha^j \ \forall i,j)$

versus

$$H_1: \text{ Nonhomogeneity } (\exists i,j \text{ such that } \alpha^i \neq \alpha^j). \qquad (1)$$

That is, is the statistical structure of the random field the same throughout, or does it vary locally? Note that in the identifiable situation, for which the distributions of the summand fields ξ^i are different from one another, the null hypothesis can be interpreted as the case where the (unobservable) embedding field m_x is identically i for some $i \in \{1,\ldots,r\}$.

This scenario can be formulated as a classical multiple comparisons problem (see, e.g., Miller, 1981). Let $\Xi^i = \Xi^i_{n^i} = \{\xi_x\colon\ x \in R^i\} \sim \alpha^i(\xi)$ for $i = 1,\ldots,r$ be the n^i observations in region R^i and perform the test of homogeneity given above. If this test is to be performed without making parametric assumptions on the α^i or choosing a specific type of nonhomogeneity in the alternative it is necessary to develop estimates $\hat{\alpha}^i$ for each i. Large values of a statistic

$$T = \max_{i,j \in \{1,\ldots,r\}} d(\hat{\alpha}^i, \hat{\alpha}^j)$$

for some pseudo-distance $d(\ )$ defined on the space of probability densities will indicate nonhomogeneity. Ghoudi and McDonald (1994) consider the completely nonparametric case.

1.3. The sieve of mixtures

The generalized mixture model assumption (Lindsay, 1995; Lindsay and Lesperance, 1995) which will allow us to utilize a borrowed strength methodology is

$$\alpha^i(\xi) = \int C(\xi;\theta)\,\mathrm{d}F^i(\theta). \qquad (2)$$

The semiparametric estimates $\hat{\alpha}^i$ are constrained to be elements of a sieve of mixture models (Geman and Hwang, 1982; Priebe, 1994). For normal mixtures, used throughout for concreteness, $C(\xi;\theta) = \varphi(\xi;\mu,v)$. Letting $\varepsilon_m, \tau_m, \delta_m, \gamma_m > 0$, we define the elements of the sieve $\{S_m\}$ as

$$\begin{aligned} S_m \equiv \Bigg\{ & \alpha(\xi;\pi,\mu,v) = \textstyle\sum_{t=1}^{m} \pi_t \varphi(\xi;\mu_t,v_t) : \\ & \{\pi_t\} \text{ satisfy } \varepsilon_m \leqslant \pi_t \leqslant 1-\varepsilon_m\ \forall t \text{ and } \textstyle\sum_{t=1}^{m} \pi_t = 1; \\ & \{\mu_t\} \text{ satisfy } -\tau_m \leqslant \mu_t \leqslant \tau_m\ \forall t; \\ & \{v_t\} \text{ satisfy } \delta_m \leqslant v_t \leqslant \gamma_m\ \forall t \Bigg\} \end{aligned}$$

where $\varepsilon_m \to 0$, $\tau_m \to \infty$, $\delta_m \to 0$, and $\gamma_m \to \infty$ as $m \to \infty$. For a given region R^i a sequence $m(n^i)$ is defined and

$$\hat{\alpha}^i_{n^i} \equiv \arg\sup_{\beta \in S_{m(n^i)}} \prod_{\xi \in \Xi^i} \beta(\xi). \qquad (3)$$

Thus $\hat{\alpha}^i$ is a maximum likelihood estimator (a recent review of maximum likelihood algorithms for semiparametric estimation can be found in Bohning (1995)) when the random field observations are independent, and an M-estimator otherwise. In any event, $\hat{\alpha}^i_{n^i}$ is an $m(n^i)$-mixture of normals and can be used in the test for nonhomogeneity. More importantly for our purposes, this semiparametric estimator lends itself to borrowing strength as a fully nonparametric estimator, such as a kernel estimator, cannot.

1.4. Borrowing strength

In image analysis and other random field applications it is often the case that one needs to obtain small-area estimates for some aspect of the local statistical structure of the field. This requirement for local estimates implies, for many real applications, that there is a severe limit on the number of observations available with which to build the estimates. A competing requirement for flexible estimators, ruling out simple parametric models for these estimates, implies that a large number of observations may be necessary in order to obtain sufficient accuracy. This combination of competing demands results in a conundrum for the statistician: expand the extent of the small areas, or restrict the model?

This paper presents a third option, the use of borrowed strength estimators, which is appropriate under certain conditions which can often be assumed in random field analysis. Using observations from regions with different statistical structure in the development of local estimates can often improve estimation accuracy without requiring overly restrictive model assumptions. This paper builds upon a parametric version of borrowed strength presented in Priebe (1996).

The structure imposed by a sieve-of-mixture estimation procedure allows the consideration of semiparametric borrowed strength. The idea of borrowing strength is one of utilizing all the data, the entire random field, to obtain an estimate which is then used to constrain the local estimation problems.

Writing $\Xi^0 = \Xi^0_{n^0} = \{\xi_x : x \in R^0\}$ to be the entire set of n^0 field observations, let

$$\hat{\alpha}^0_{n^0} \equiv \arg\sup_{\beta \in S_{m(n^0)}} \prod_{\xi \in \Xi^0} \beta(\xi). \tag{4}$$

$\hat{\alpha}^0$ is an estimate of the overall field density $\alpha^0 = \sum_{i=1}^{r} (E[n^i]/n^0)\alpha^i$, recalling that n^0 is a fixed constant but the n^i are random variables denoting the number of observations in the random sets R^i derived from the random field $m(x,\omega)$ via the indicator function $I_{\{m_x = i\}}$. The term 'borrowing strength', as used in this paper, indicates the idea of using this overall mixture estimate to constrain the local estimations. $\hat{\alpha}^0$ is an $m(n^0)$-mixture. It will be argued that, under appropriate conditions, local estimates of the α^i constrained to have the same structure as $\hat{\alpha}^0$, i.e., constrained to be $m(n^0)$-mixtures with the same means and variances as $\hat{\alpha}^0$, will be superior to the conventional local estimates $\hat{\alpha}^i$. Thus the borrowed strength estimate is

$$\tilde{\alpha}^i_{n^0,n^i} \equiv \arg\sup_{\beta \in S^{\hat{\alpha}^0}_{m(n^0)}} \prod_{\xi \in \Xi^i} \beta(\xi), \tag{5}$$

where $S^{\hat{\alpha}0}_{m(n^0)} \subset S_{m(n^0)}$ is the restriction of the sieve element S_m to having the means and variances of $\hat{\alpha}^0$. Only the mixing coefficients are free parameters. This idea is explored in more detail in Section 2. $\tilde{\alpha}^i$ is to be compared with the conventional estimate $\hat{\alpha}^i$ defined in Eq. (3).

1.5. The scan process

Since we do not wish to assume prior knowledge of the location of the local regions of interest R^i it is necessary to introduce a regional structure on R^0 and test for homogeneity under this structure. Let $R^0 = \bigcup \tilde{R}^i$ $(i = 1, \ldots, \tilde{r})$, where the $\tilde{R}^i$ are (possibly overlapping) neighborhoods with number of observations $\tilde{n}^i$ $(\tilde{\Xi}^i = \tilde{\Xi}^i_{\tilde{n}^i} = \{\xi_x : x \in \tilde{R}^i\})$. For instance, in the investigation of spatial scan statistics (Chen and Glaz, 1995; Kulldorff, 1995) one often considers ε-balls, in which case $\tilde{R}^\tau = B(\tau, \varepsilon) = \{x \in R^0 : \|x - \tau\| < \varepsilon\}$ for $\tau \in R^0$ and $\varepsilon > 0$. Using these artificially introduced regions, large values of

$$T^{\mathrm{BS}}_{i,j} = d(\tilde{\alpha}^i, \tilde{\alpha}^j)$$

and

$$T^{\mathrm{CONV}}_{i,j} = d(\hat{\alpha}^i, \hat{\alpha}^j) \tag{6}$$

will be compared for their ability to indicate potential nonhomogeneities. (Note: Priebe et al. (1996) presents an approach for relaxing the requirement that these standard scan windows (ε-balls) be used; the $\tilde{R}^i$ considered there consist of a stochastic partition of R^0.)

The lack of knowledge of the location of potential nonhomogeneous regions necessitates that the regions $\tilde{R}^i$ be chosen to be relatively small as compared to the anticipated size of the true but unknown R^i. This in turn implies that the $\tilde{n}^i$ are small and hence the estimates $\hat{\alpha}^i$ $(i = 1, \ldots, \tilde{r})$ will have relatively large variance, especially when no simple parametric model is assumed. This is the conundrum alluded to earlier. The borrowed strength methodology presented herein can be used to address the simultaneous requirements of small-area estimation and flexible modelling. Since the $\tilde{R}^i$ are necessarily smaller in their spatial extent than the nonhomogeneities anticipated, there likely will be numerous regions $\tilde{R}^i$ completely contained within a given R^k and hence having the same probabilistic structure. This implies that there will be information relevant to the estimation of α^i in at least some of the regions $\tilde{R}^j$ for $j \neq i$ and the assumption that borrowing strength can improve the local estimations is reasonable.

1.6. Relation to the literature

Various applications of the nonhomogeneity analysis addressed herein were given at the outset, and the relationship to efforts in change curve analysis and low-level image analysis was noted.

A particular feature of the approach presented in this paper which deserves special mention is the use of mixture densities to represent the null hypothesis of homogeneity. Examples abound in which mixtures represent heterogeneity of population. See Titterington et al. (1985, Chapter 2) for instance. In such a model each component of the population is assumed to be represented by a single term in the mixture. For example, each possible class of tissue texture in X-ray mammography might be modelled as a single normal. Such a model is often unrealistic. Priebe et al. (1994) indicates that healthy tissue, for example, is poorly modelled as normally distributed. In this work the marginal density for a field observation is considered itself to be modelled as a generalized mixture density, with homogeneity the case in which the same mixture represents the marginals throughout the field. The recent work of O'Sullivan (1993) has also suggested consideration of a mixture density for the marginal density of individual pixels in image processing. The particular application to PET imagery considered therein does not lend itself to improvement via the borrowed strength methodology developed in this paper, as most pixels are considered to have nearly normal marginals. Therefore, the algorithm presented in O'Sullivan (1993) does not utilize borrowed strength. Their work is nonetheless closely related and of obvious interest, as it seems likely that for selected applications the procedure advanced in O'Sullivan (1993) can be improved by incorporating borrowed strength.

2. Borrowed strength

2.1. Borrowed strength methodology

For nonhomogeneity detection we estimate the density $\hat{\alpha}^0 = \int C(\cdot;\theta)\,d\hat{F}^0(\theta)$ of the overall field using Ξ^0. The support of $d\hat{F}^0(\theta)$ is the 'imposed measure', by which we mean the globally estimated parameter values (individual term means and variances) which when fixed constrain the subregion estimation problems. The sieve of mixtures approach implies that $\hat{\alpha}^0 = \hat{\alpha}^0_{m(n^0)}$ will be an $m(n^0)$-mixture and hence, from Eq. (2), $\mathrm{supp}(d\hat{F}^0(\theta)) = \{\hat{\theta}^i_1, \ldots, \hat{\theta}^i_{m(n^0)}\}$. This fixes the mixture in the parameter space $\Theta^{m(n^0)}$ and restricts the possible density estimates for subfields. We then obtain estimates of the densities α^i for subfields $\tilde{R}^i$ using $\tilde{\Xi}^i$ under this imposed measure. Since the parameter space support is now fixed, the only estimation required for the subfield densities is that of $\{d\hat{F}^i(\theta)\}$ for $\theta \in \mathrm{supp}(d\hat{F}^0(\theta))$; that is, $\{\pi^i_1, \ldots, \pi^i_{m(n^0)}\}$. Thus we have the following methodology.

Borrowed strength methodology

1. Introduce a regional structure on R^0; $R^0 = \bigcup \tilde{R}^i (i = 1, \ldots, \tilde{r})$.
2. Estimate dF^0 for the random field R^0 using the n^0 observations Ξ^0.
3. Impose $\mathrm{supp}(d\hat{F}^0(\theta))$ on each subfield $\tilde{R}^i$.
4. Estimate $dF^i(\theta)$ for $\theta \in \mathrm{supp}(d\hat{F}^0(\theta))$ using the $\tilde{n}^i$ observations $\tilde{\Xi}^i$.

5. Large values of

$$T_{i,j}^{\mathrm{BS}} = d(\tilde{\alpha}^i(\xi;\hat{F}^i), \tilde{\alpha}^j(\xi;\hat{F}^j)) \quad (7)$$

indicate potential subregions of nonhomogeneity.

The claim made here is that using the statistic $T_{i,j}^{\mathrm{BS}}$ thus obtained in the test for nonhomogeneity yields an improvement over the analogous conventionally estimated statistic.

Details for the implementation of the above methodology for nonhomogeneity detection in image analysis (the example presented in Section 3) are as follows, where R^0 is a square $\sqrt{n^0} \times \sqrt{n^0}$ lattice:

1. Let $\{\tilde{R}^i\}$ be the collection of $\sqrt{\tilde{n}^i} \times \sqrt{\tilde{n}^i}$ square regions in R^0. These are the standard 'scan windows'. In practice the size of the $\tilde{R}^i$ directly impacts the size of the anomalies that can reliably be detected.

2. Use the adaptive mixtures algorithm (Priebe, 1994) on the entire data set Ξ^0, all the pixel values in the image, to obtain a normal mixture estimate $\hat{\alpha}^0$. The number of terms in this mixture, $m(n^0)$, is stochastic and determined by the particular instantiation of the random image.

3 and 4. Update the mixing coefficients $\{\pi_1^i, \ldots, \pi_{m(n^0)}^i\}$ using the EM algorithm with the means and variances fixed, yielding $\tilde{\alpha}^i$. (For finite mixtures this is a standard profile likelihood estimate; see Cox and Reid (1987) and Priebe et al. (1996).) Only $\tilde{\Xi}^i$, the pixel values from the particular scan window $\tilde{R}^i$, are used here.

5. $T_{i,j}^{\mathrm{BS}} = \| \tilde{\alpha}^i, \tilde{\alpha}^j \|_{L^2}$, the integrated squared error between the respective local estimates.

The conventional local likelihood procedure against which this borrowed strength approach is compared simply replaces steps 2–4 above with

2′. Use the adaptive mixtures algorithm to estimate $\hat{\alpha}^i$ for $\tilde{R}^i$ using $\tilde{\Xi}^i$ and employ the statistic $T_{i,j}^{\mathrm{CONV}} = \|\hat{\alpha}^i, \hat{\alpha}^j\|_{L^2}$.

2.2. A consistency result

Considering the sieve $\{S_m\}$ introduced in Section 1.3, for which large sample properties have been given in Geman and Hwang (1982) and Priebe (1994), we wish to establish consistency of the borrowed strength estimates. We consider sieves of normal mixtures for concreteness, but the analogous results hold in general for sieves of mixtures of continuous exponential family densities as well.

Let $\Xi_n = \{\xi_1, \ldots, \xi_n\}$ and $m = m(n) > 0$ be a fixed integer and consider the likelihood $L_{\Xi_n}(\beta) = \prod_{\xi \in \Xi_n} \beta(\xi)$. Define

$$M_{\Xi_n, m} \equiv \left\{ \alpha(\xi; \underset{\sim}{\pi}, \underset{\sim}{\mu}, \underset{\sim}{\nu}) \in S_m : L_{\Xi_n}(\alpha) = \sup_{\beta \in S_m} L_{\Xi_n}(\beta) \right\}$$

to be the maximum likelihood set in S_m. For $\beta \in S_m$ define $S_m^\beta \subset S_m$ to be the restriction of the sieve element S_m to having the means and variances of β; only the mixing coefficients are free parameters. Letting $\alpha' = \alpha(\xi; \underset{\sim}{\pi}', \underset{\sim}{\mu}', \underset{\sim}{\nu}') \in S_m$ be fixed, define

$$M_{\Xi_n,m}|_{\alpha'} \equiv \left\{ \alpha \in S_m^{\alpha'} : L_{\Xi_n}(\alpha) = \sup_{\beta \in S_m^{\alpha'}} L_{\Xi_n}(\beta) \right\}$$

to be the restricted maximum likelihood set in S_m (restricted to having $\underset{\sim}{\mu} = \underset{\sim}{\mu}'$ and $\underset{\sim}{\nu} = \underset{\sim}{\nu}'$). Define

$$M_{\Xi_n,m}|_{M_{\Xi_{n'},m}} \equiv \bigcup_{\alpha' \in M_{\Xi_{n'},m}} M_{\Xi_n,m}|_{\alpha'}.$$

Finally, for $0 < q \leqslant 1$ define

$$qM_{\Xi_n,m} \equiv \left\{ \alpha \in S_m : L_{\Xi_n}(\alpha) \geqslant q \sup_{\beta \in S_m} L_{\Xi_n}(\beta) \right\}.$$

Given this machinery, we have the consistency result given by

Theorem 1. *Let α^i $(i = 1, \ldots, r)$ be finite mixture models with arbitrary, unknown complexity. Let $n^0 = \sum n^i$ and $n^i \to \infty$ $\forall i$ in fixed proportion. If a sequence $\{m\} = \{m(n^0)\}$ increasing slowly enough with respect to n^i is chosen, then for* class conditionally iid *field observations* $(\xi^i(x,\omega) \underset{\sim}{\text{iid}}\ \alpha^i(\xi))$ *the procedure indicated by* (4) *and* (5) *yields consistent estimators of the α^i,*

$$\lim_{n^i \to \infty} \int_{-\infty}^{\infty} \alpha^i \log(\alpha^i / \tilde{\alpha}^i) \mathrm{d}\xi = 0 \quad \text{a.s..}$$

Proof. See Appendix. □

3. Nonhomogeneity detection examples

3.1. Class conditional independence

We begin our simulation study by considering a scenario in which the random field of interest, f, is an embedding of two random fields f^1 and f^2 with $f^i \underset{\sim}{\text{iid}}\ \alpha^i$. Letting m be a binary $(0,1)$ Markov random field used to model the presence of local nonhomogeneities we have $f = I_{\{m_x=1\}} f^1 + (1 - I_{\{m_x=1\}}) f^2$. Thus f consists of observations which are class conditionally independent and identically distributed.

For this simulation consider the specific example $\alpha^1(\xi) = (\sqrt{18})\chi_9^2(9 + \sqrt{18}\xi)$ and $\alpha^2(\xi) = (\sqrt{18})\chi_9^2(9 - \sqrt{18}\xi)$. Fig. 1 shows the densities α^1 and α^2 (each with zero mean and unit variance) as well as an instantiation of f^1, f^2, m and f.

A Monte Carlo simulation based upon the above scenario is now described. This simulation is designed so as to be relevant to nonhomogeneity detection. That is, there is to be a significant proportion of the 'background' field f^2 and only a small proportion of the 'anomaly' field f^1. Thus the embedding field m is assumed to be approximately

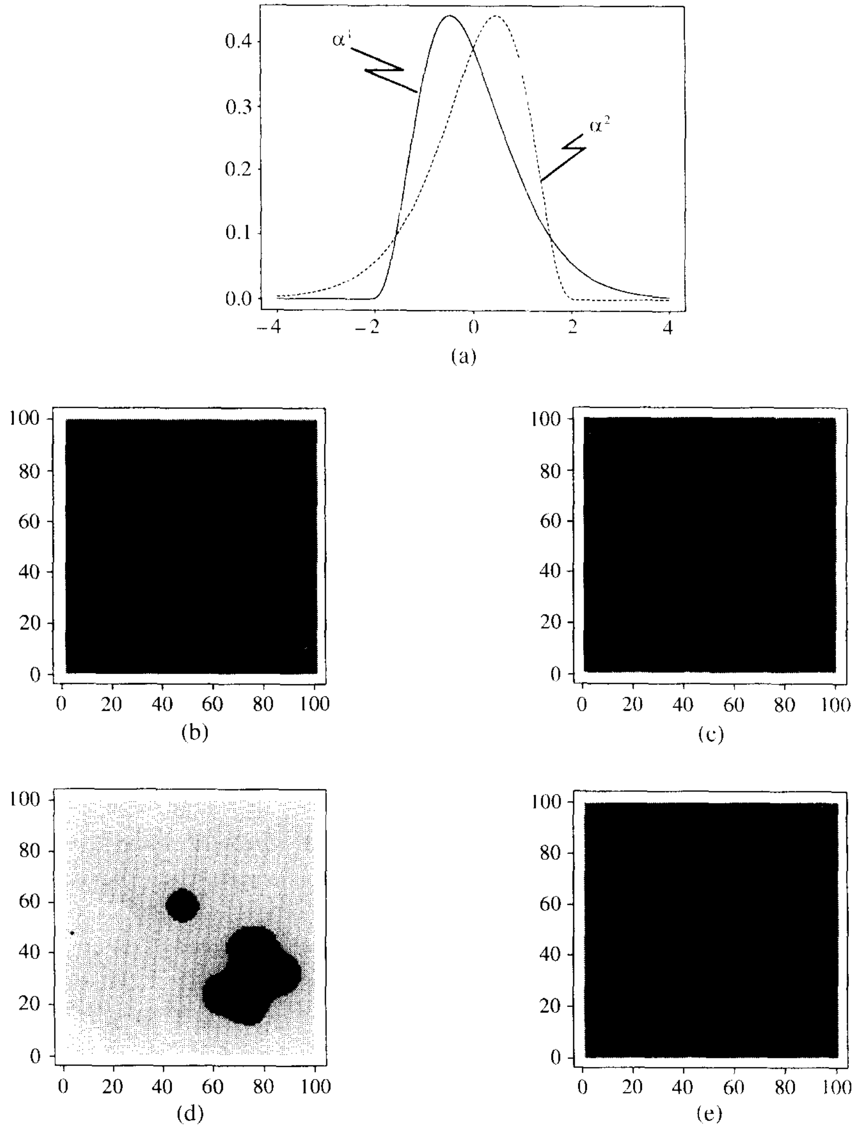


Fig. 1. Scenario for iid simulation and experiment. (a) Probability density functions α^1 and α^2 used in the example. $\alpha^1(\xi) = (\sqrt{18})\chi_9^2(9 + \sqrt{18}\xi)$ and $\alpha^2(\xi) = (\sqrt{18})\chi_9^2(9 - \sqrt{18}\xi)$. (b)–(e) depict f^1, f^2, m and f, random fields representative of the Monte Carlo simulation reported in Tables 1 and 2. These fields are used in the iid nonhomogeneity detection experiment presented in Fig. 2.

90% zeros ($E[n^1] = n^0/10$), and therefore $\alpha^0(\xi) = (1/10)\alpha^1 + (9/10)\alpha^2$. The field m shown in Fig. 1(d) is obtained through a Gibbs sampler and has 8959 zeros out of 10 000 total pixels.

Thus α^1 and α^2 are non-mixture-of-normal densities. The idea, from a nonhomogeneity detection (tumorous tissue detection in digital mammography, for example) standpoint, is that α^1 is the density for the anomalies (tumors), α^2 is for background

Table 1
Number-of-terms results for iid Monte Carlo simulation

	$n^0 = 1000$	$n^0 = 10\,000$	$n^0 = 100\,000$
$m(n^0)$	6.6 ± 1.1	10.1 ± 1.2	16.9 ± 2.5
$m(\tilde{n}^i)$ for $\tilde{R}^i \subset R^1$	2.0 ± 0	4.1 ± 0.9	7.6 ± 1.5
$m(\tilde{n}^i)$ for $\tilde{R}^i \subset R^2$	2.0 ± 0	4.1 ± 0.7	7.2 ± 1.2

Note: Results of Monte Carlo simulation under iid conditions presented in Fig. 1 indicating the performance of borrowed strength and conventional maximum likelihood estimators for scan analysis of nonhomogeneity. The results are based upon 10 Monte Carlo runs with n^0 as shown and $\tilde{n}^i = n^0/100$. The densities α^1 and α^2 are shown in Fig. 1(a). $\alpha^0 = 0.1\alpha^1 + 0.9\alpha^2$. Represented are the performance using the local likelihood estimator $\hat{\alpha}^i$ and the borrowed strength estimator $\tilde{\alpha}^i$.

Quantitative indication of the superiority of the borrowed strength estimator is obtained via the one-sided Wilcoxon test: for $n^0 = 10\,000$ the Wilcoxon test for H_0^2: $\|\tilde{\alpha}^2, \alpha^2\|_{L^2} - \|\hat{\alpha}^2, \alpha^2\|_{L^2} \geqslant 0$ is significant at $p = 0.001$; for $n^0 = 100\,000$ the Wilcoxon test for H_0^1: $\|\tilde{\alpha}^1, \alpha^1\|_{L^2} - \|\hat{\alpha}^1, \alpha^1\|_{L^2} \geqslant 0$ is significant at $p = 0.055$.

(healthy), and under the alternative hypothesis of nonhomogeneity there will be, say, 10% α^1 and 90% α^2 in the field. Recall $\tilde{\alpha}^i$ is the borrowed strength estimator and $\hat{\alpha}^i$ is the conventional local likelihood estimator for $i = 0, 1, 2$. This simulation is designed to support the conjectures (1) $\tilde{\alpha}^0 \to \alpha^0$, and (2) $\tilde{\alpha}^i \to \alpha^i$ faster than $\hat{\alpha}^i \to \alpha^i$ $(i = 1, 2)$, and to give an idea of how this faster convergence leads to superior nonhomogeneity detection.

We wish to investigate the relative performance of borrowed strength versus conventional likelihood. Toward this end, consider first the 'conventional' estimation of an unknown density with a mixture of normals. (Such a procedure is quite common; see Chapter 2 of Titterington et al. (1985).) There are two different cases in which this is necessary for our experiment. For conventional likelihood estimation, we obtain $\hat{\alpha}^i$ using the $\tilde{n}^i$ observations available locally. For borrowing strength, we must obtain $\tilde{\alpha}^0$ using n^0 observations. The parameterspace support (means and variances) of this estimate will then be imposed upon the local estimation problems. These estimation tasks are straightforward if the number of terms in the mixture is known or assumed. For our purposes, this is not the case and an automated method for determining an appropriate model complexity is necessary. The adaptive mixtures algorithm (Priebe, 1994) is employed, and $\|\,,\|_{L^2}$, or integrated squared error (ISE), results are presented in Table 2 for a selection of sample sizes: $n^0 \in \{1000, 10\,000, 100\,000\}$; $\tilde{n}^i = n^0/100$. Table 1 presents the actual number of terms used in the estimates. These results are based upon ten Monte Carlo runs and indicate the desired effect. ISE decreases with sample size, even when the number of terms is treated as a random variable and estimated from the data. (For purposes of the nonhomogeneity detection $\|\hat{\alpha}^i, \alpha^0\|_{L^2}$ is reported rather than $\|\hat{\alpha}^i, \alpha^i\|_{L^2}$. The purpose of this deviation from the intuitive is discussed below.)

Table 2
ISE results for iid Monte Carlo simulation

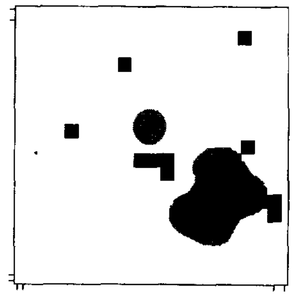
	$n^0 = 1000$	$n^0 = 10\,000$	$n^0 = 100\,000$
$\|\tilde{\alpha}^0, \alpha^0\|_{L^2}$	0.0046 ± 0.0014	0.0008 ± 0.0004	0.0001 ± 0.0001
$\|\hat{\alpha}^2, \alpha^0\|_{L^2}$	0.1110 ± 0.0642	0.0240 ± 0.0163	0.0057 ± 0.0030
$\|\tilde{\alpha}^2, \alpha^0\|_{L^2}$	0.0879 ± 0.0775	0.0053 ± 0.0068	0.0007 ± 0.0008
$\|\hat{\alpha}^1, \alpha^0\|_{L^2}$	0.1423 ± 0.1066	0.0529 ± 0.0202	0.0501 ± 0.0090
$\|\tilde{\alpha}^1, \alpha^0\|_{L^2}$	0.0984 ± 0.0887	0.0359 ± 0.0162	0.0457 ± 0.0079

Note: See footnote to Table 1.

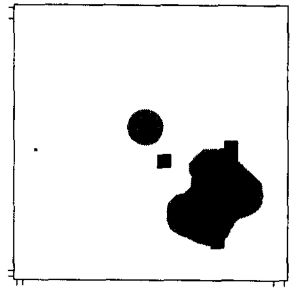
The relevant comparison to be made is $\|\hat{\alpha}^i, \alpha^0\|_{L^2}$ versus $\|\tilde{\alpha}^i, \alpha^0\|_{L^2}$. This comparison, available in Table 2, indicates that the borrowed strength estimates are significantly better than their conventional counterparts. There are two interesting interpretations of this result. First, considering the parameters associated with the terms in the mixture $\tilde{\alpha}^0$ which are more relevant to α^1 than α^2 to be 'nuisance' parameters of a sort in the estimation of α^2, the result indicates that any degradation in performance due to these nuisance parameters is outweighed by improved estimation of the parameters associated with the terms which are relevant to α^2. A second interpretation of these results comes from the viewpoint that one can afford to have more terms in the local estimation problem if the means and variances are kept fixed. Thus a greater complexity, required in the normal mixture estimation of nonnormal densities, is acceptable and superior estimation can be expected as long as this greater complexity is put to good use.

Fig. 2 shows the results of the nonhomogeneity detection methodology applied to the particular realization of the field f shown in Fig. 1(e) using the scan process with and without the borrowed strength. A 10×10 pixel moving window is scanned throughout the region; $\tilde{n}^i = 100$. At each location the density is estimated, using semiparametric borrowed strength maximum likelihood on the 100 observations in the one case and standard maximum likelihood on the 100 observations in the other. Each locality statistic, the estimated marginal density for a given window, is then compared, in terms of ISE, with the overall density which is assumed to be made up mostly of 'background'. Those scan locations which have the largest ISE are considered anomalies, and these locations are shown in Fig. 2. We mark all scan regions with an ISE against $\tilde{\alpha}^0$ among the largest 5%. The superior nonhomogeneity detection afforded by borrowing strength is clear, as is expected from the theory and the Monte Carlo simulation.

As mentioned above we consider the values of $\|\hat{\alpha}^i, \alpha^0\|_{L^2}$ and $\|\tilde{\alpha}^i, \alpha^0\|_{L^2}$ in Table 2 rather than the perhaps more intuitively appealing $\|\hat{\alpha}^i, \alpha^i\|_{L^2}$ and $\|\tilde{\alpha}^i, \alpha^i\|_{L^2}$. The purpose of this becomes clear upon consideration of the description above of the implementation of the detection procedure used to obtain Fig. 2. Given a local density estimate α^i (conventional or borrowed strength) $T = \|\tilde{\alpha}^i, \hat{\alpha}^0\|_{L^2}$ is calculated for each local scan region $\tilde{R}^i$. For the examples presented here we have true values of $\|\alpha^1, \alpha^0\|_{L^2} = 0.0473$ and $\|\alpha^2, \alpha^0\|_{L^2} = 0.0006$. A large value of T indicates that the region $\tilde{R}^i$ is closer to α^1 than α^2 and hence a likely region of nonhomogeneity. This approach is only applicable for the anomaly detection version of nonhomogeneity analysis; i.e., situations in which



(a)



(b)

Fig. 2. Results for iid nonhomogeneity detection experiment.
Results of a nonhomogeneity detection experiment under the iid conditions are presented in Fig. 1 and Tables 1 and 2. The image (field f from Fig. 1(e)) is 100×100 pixels ($n^0 = 10\,000$) and the scan regions are overlapping 10×10 pixel windows ($\bar{n}^i = 100$). The results, those scan regions which have a value for the statistic $T = \|\hat{\alpha}^i, \hat{\alpha}^0\|_{L^2}$ among the largest 5% are overlayed on the Markov random field m depicted in Fig. 1(d) representing the embedded nonhomogeneities. Represented are the performance using (a) the local likelihood estimator $\hat{\alpha}^i$, and (b) the borrowed strength estimator $\tilde{\alpha}^i$.
The superior detection capabilities, fewer false detections in the background region and a higher percentage of correct detections in the anomalous regions, indicate the ability of the borrowed strength estimator to employ increased model complexity and hence obtain superior integrated squares error results, as indicated numerically in Tables 1 and 2.

the alternative hypothesis is small local regions of nonhomogeneity, so that $\alpha^0 \approx \alpha^2$. For such a scenario the practical advantages are significant in terms of both computation time and complexity of analysis, as the alternative involves the nontrivial analysis of the similarity matrix defined by $s_{ij} = \|\tilde{\alpha}^i, \tilde{\alpha}^j\|_{L^2}$.

3.2. Dependence

We now consider a situation similar to that addressed above, with the difference being that the fields are no longer independent. The marginals are g^i id α^i with α^1 and α^2 the same as in the iid case above, and a simple neighborhood dependency is considered. We let $N_x = \{y : \|x - y\| < K\}$ and generate two independent and

Table 3
Number-of-terms results for dependent Monte Carlo simulation

	$n^0 = 0\,000$
$m(n^0)$	9.1 ± 1.3
$m(\tilde{n}^i)$ for $\tilde{R}^i \subset R^1$	4.4 ± 0.7
$m(\tilde{n}^i)$ for $\tilde{R}^i \subset R^2$	4.4 ± 1.1

Note: Results of Monte Carlo simulation under dependency indicating the performance of borrowed strength and conventional maximum lilkelihood estimators for scan analysis of nonhomogeneity. The results are based upon 10 Monte Carlo runs with $n^0 = 90\,000$ and $\tilde{n}^i = 900$. The densities α^1 and α^2 are shown in Fig. 1. $\alpha^0 = 0.1\alpha^1 + 0.9\alpha^2$. Represented are the performance using the local likelihood estimator $\hat{\alpha}^i$, and the borrowed strength estimator $\tilde{\alpha}^i$. Again the Wilcoxon test for H_0^2: $\|\tilde{\alpha}^2, \alpha^2\|_{L^2} - \|\hat{\alpha}^2, \alpha^2\|_{L^2} \geqslant 0$ is significant at $p = 0.001$, quantitatively indicating the superiority of the borrowed strength estimator.

identically distributed fields ε^i. Then $g_x^i = \sum_{y \in N_x} \beta_y^i \varepsilon_y^i$. The marginal density for g_x^i is known, is identical for all x, and is α^i (by construction.) For this example the neighborhoods N_x and the coefficients β_y^i are chosen so that the marginal densities for fields g^1 and g^2 are the densities α^1 and α^2 shown in Fig. 1(a). The field g is constructed as an embedding of g^1 and g^2 using a binary field m as before.

The conventional estimate obtained via (3) and the borrowed strength estimate (4) and (5) are no longer a maximum likelihood estimates. They are however M-estimates and for the simple dependency considered here asymptotic results in the parametric case can be obtained from consideration of blocking (Arcones and Yu, 1994) and decoupling (Doukhan et al., 1995) and are analogous to the empirical subsampling result of Craig (1979).

One would expect results quite similar to those presented in Section 3.1, with an effective sample size effect. That this is indeed the case can be seen by comparing the Monte Carlo simulation result under dependency with the iid results given in Tables 1 and 2. In Tables 3 and 4 we present the results from 10 Monte Carlo replications of the dependent random field g with $n^0 = 90\,000$ and $\tilde{n}^i = 900$. The neighborhood N_x consists of nine observations ($K = 1.5$) and we would therefore expect results analogous to those obtained in the independent case for $n^0 = 10\,000$ and $\tilde{n}^i = 100$. A comparison of Table 2 and Table 4 indicates that this is indeed the case.

4. Conclusions

The motivation for the semiparametric borrowed strength estimation methodology developed in this paper is drawn from consideration of the desire for estimators of

Table 4
ISE results for dependent Monte Carlo simulation

	$n^0 = 0\,000$
$\|\|\tilde{\alpha}^0, \alpha^0\|\|_{L^2}$	0.0001 ± 0.0001
$\|\|\hat{\alpha}^2, \alpha^0\|\|_{L^2}$	0.0064 ± 0.0051
$\|\|\tilde{\alpha}^2, \alpha^0\|\|_{L^2}$	0.0012 ± 0.0017
$\|\|\hat{\alpha}^1, \alpha^0\|\|_{L^2}$	0.0579 ± 0.0235
$\|\|\tilde{\alpha}^1, \alpha^0\|\|_{L^2}$	0.0374 ± 0.0121

Note: See footnote to Table 1.

local characteristics in many random field analyses. In particular, the motivating example of nonhomogeneity detection in low-level image analysis is considered, although the technique is relevant to many small-area random field applications in which local estimates of random field properties are required while stringent parametric assumptions are unwarranted. If it is reasonable to make the generalized mixture model assumption, then the local area estimates may lend themselves to improve estimation under a borrowed strength methodology.

We have shown theoretically that estimating local densities under constraints developed using a larger set of data can yield consistent semiparametric mixture model estimates. Monte Carlo simulations provide an example wherein these borrowed strength estimates outperform conventional likelihood estimates.

Appendix

Proof of Theorem. From Geman and Hwang (1982) it follows that if a sequence $\{m\} = \{m(n^0)\}$ increasing slowly enough with respect to n^i is chosen such that eventually

$$M_{\Xi^i_{n^i},m}|_{\tilde{\alpha}^0_{n^0}} \subset qM_{\Xi^i_{n^i},m}$$

then

$$\lim_{n^i\to\infty} \alpha^i \in \sup_{M_{\Xi^i_{n^i},m}|_{\tilde{\alpha}^0_{n^0}}} \int_{-\infty}^{\infty} \alpha^i \log(\alpha^i/\alpha')\mathrm{d}\xi = 0 \quad \text{a.s.}$$

as desired. It remains only to show that eventually

$$M_{\Xi^i_{n^i},m}|_{M_{\Xi^0_{n^0},m}} \subset qM_{\Xi^i_{n^i},m} \tag{8}$$

holds for the particular problem at hand, which will imply that the borrowed strength estimator converges to the true unknown density in the Kullback–Leibler distance.

Completion of the proof follows from noting that for any given collection of finite normal mixtures $\{\alpha^i\}_{i=1}^r m(n^0)$ can be chosen to increase slowly enough such that $M_{\Xi^0_{n^0},m}$ converges to the finite-dimensional solution $\alpha^0 = \sum_{i=1}^r (n^i/n^0)\alpha^i$. ('Slowly enough' here is dependent upon the specific $\{\alpha^i\}$ and the rate at which $\varepsilon_m \to 0$, $\tau_m \to \infty$, $\delta_m \to 0$, and $\gamma_m \to \infty$ as $m \to \infty$ for the sieve.) For large enough n^i it follows that a subvector of the vector $\hat{\Psi}^0$ estimating the means and variances of α^0 will be arbitrarily close to Ψ^i, the vector of means and variances for α^i, and (8) will hold. Specifically, for n^i, n^0 large enough, $\varepsilon > 0$ small enough, and $0 < q' < q < 1$,

$$M_{\Xi^i,m}|_{\Psi^0} \subset qM_{\Xi^i,m} \Rightarrow M_{\Xi^i,m}|_{B(\Psi^0,\varepsilon)} \subset q'M_{\Xi^i,m}$$
$$\Rightarrow M_{\Xi^i,m}|_{M_{\Xi^0,m}} \subset q'M_{\Xi^i,m}$$

as desired, where we use the fact that $M_{\Xi^i_{n^i},m}$ converges to a solution which includes the subvector Ψ^i of Ψ^0. □

References

Applied Change Point Conference, University of Maryland Baltimore County/National Security Agency, 17–18 March 1993; Proc. published as a special issue of *J. Appl. Statist. Sci.* **1** (4).

Arcones, M.A. and B. Yu (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* **7**, 47–71.

Basawa, I.V. (1993). Inference for a class of planar point processes with applications to minefield modelling. Preprint.

Bohning, D. (1995). A review of reliable maximum likelihood algorithms for semiparametric mixture models. *J. Statist. Plann. Inference* **47**, 5–28.

Carlstein, E. and S. Lele (1994). Nonparametric change-point estimation for data from an ergodic sequence. *Theory Probab. Appl.* **38**, 726–733.

Carlstein, E., H.-G. Muller and D. Siegmund (Eds.) (1994). *Change point problems.* IMS Lecture Notes-Monograph Series, Vol. 23.

Chen, J. and J. Glaz (1995). Two-dimensional discrete scan statistics. Preprint.

Cox, D.R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser.* B, **49**, 1–39.

Craig, R.G. (1979). Autocorrelation in LANDSAT data. *Proc. 13th Internat. Symp. on Remote Sensing of Environment*, 1517–1524.

Cressie, N.A.C. (1993). *Statistics for Spatial Data.* Wiley, New York.

Doukhan, P., P. Massart and E. Rio (1995). Invariance principles for absolutely regular empirical processes. *Ann. l'Inst. H. Poincare'* **31**, 393–427.

Geman, D. (1990). Random fields and inverse problems in imaging. In: Lecture Notes in Mathematics, Vol. 1427, Springer, Berlin.

Geman, S. and C.-R. Hwang (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10**, 401–414.

Ghoudi, K. and D. McDonald (1994). A nonparametric test for homogeneity: applications to parameter estimation. In: E. Carlstein, H.-G. Muller and D. Siegmund, Eds., *Change Point Problems*, IMS Lecture Notes-Monograph Series, Vol. 23, 149–156.

Hayat, M.M. and J.A. Gubner (1994). Markov-type spatial point processes with an application to minefield modelling. *Proc. 28th Ann. Conf. on Information Sciences and Systems.* Princeton, NJ.

Kulldorff, M. (1995). A spatial scan statistic. Preprint.

Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications.* NSF-CBMS Regional Conf. Series in Probability and Statistics, Vol. 5. IMS Monograph Series, Hayward, CA.

Lindsay, B.G. and M.L. Lesperance (1995). A review of semiparametric mixture models. *J. Statist. Plann. Inference* **47**, 29–39.

Miller, R.G. (1981). *Simultaneous Statistical Inference.* Springer, New York.

Miller, P. and S. Astley (1992). Classification of breast tissue by texture analysis. *Image Vision Comput.* **10**, 277–282.

Muise, R.R. and C.M. Smith (1992). Nonparametric minefield detection and localization. Technical Report CSS-TM-591-91, Naval Coastal Systems Station, Naval Surface Warfare Center.

O'Sullivan, F. (1993). Mixture estimation from multichannel image data. *J. Amer. Statist. Assoc.* **88**, 209–220.

O'Sullivan, F. (1995). A study of least squares and maximum likelihood for image reconstruction in position emission tomography. *Ann. Statist.* **23**, 1267–1300.

Priebe, C.E. (1994). Adaptive mixtures. *J. Amer. Statist. Assoc.* **89**, 796–806.

Priebe, C.E. (1996). Nonhomogeneity analysis using borrowed strength. *J. Amer. Statist. Assoc.* (to appear).

Priebe, C.E., D.J. Marchette and G.W. Rogers (1996). Segmentation of random fields via borrowed strength density estimation. *IEEE Trans. Pattern Anal. Machine Intell.* (to appear).

Priebe, C.E., J.L. Solka, R.A. Lorey, G. Rogers, W. Poston, M. Kallergi, W. Qian, L.P. Clarke and R.A. Clark (1994). The Application of fractal analysis to mammographic tissue classification. *Cancer Lett.* **77**, 183–189.

Smith, C.M. (1991). Two dimensional minefield simulation. Technical Report NCSC-TM-558-91, Naval Coastal Systems Center.

Titterington, D.M., A.F.M. Smith and U.E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

Worsley, K.J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.* **23**, 640–669.