

# False Discovery Rates for Random Fields

M. PERONE PACIFICO, C. GENOVESE, I. VERDINELLI, L. WASSERMAN <sup>1</sup>

*Carnegie Mellon University and Università di Roma “La Sapienza”.*

February 25, 2003

## ABSTRACT

This paper extends False Discovery Rates to random fields, where there are uncountably many hypothesis tests. This provides a method for finding local regions in the field where there is a significant signal while controlling either the proportion of area or the number of clusters in which false rejections occur. We develop confidence envelopes for the proportion of false discoveries as a function of the rejection threshold. This yields algorithms for constructing a confidence superset for the locations of the true nulls. From this we derive rejection thresholds that control the mean and quantiles of the proportion of false discoveries. We apply this method to scan statistics and functional neuroimaging.

*Keywords:* false discovery rates, multiple hypothesis test, random fields.

## 1. INTRODUCTION

In this paper, we extend the False Discovery Rate (FDR), introduced by Benjamini and Hochberg (1995), to random fields, where there are uncountably many hypothesis tests. Our method produces a threshold that controls the fraction of errors, expressed as a ratio of areas within the field’s domain. As in the standard multiple testing problem, controlling the fraction of errors is an alternative to the traditional approach of strong control of the familywise error rate. For the familywise approach in the context of random fields see, for example, Cao and Worsley (1999), Siegmund and Worsley (1995) and Worsley (1994, 1995). Another approach to using FDR in the case of spatial signals is given in Shen, Huang, and Cressie (2002).

Consider a set  $S$  and a random field  $X = \{X(s) : s \in S\}$  on  $S$  with mean function  $\mu(s) = \mathbb{E}X(s)$ . We use the realized value of  $X$  to test the collection of one-sided hypotheses

$$H_{0,s} : \mu(s) = 0 \text{ versus } H_{1,s} : \mu(s) > 0. \quad (1)$$

---

<sup>1</sup>Perone Pacifico’s research supported by *Progetto Giovani Ricercatori “La Sapienza”* and *Cofinanziamento MIUR*. Genovese’s research supported by NSF Grant SES 9866147, ITR 0121671. Verdinelli’s research supported by *Progetto di Ateneo “La Sapienza”* and *Cofinanziamento MIUR*. Wasserman’s research supported by NIH Grant R01-CA54852-07, NIH grant number MH57881, NSF Grant DMS-98-03433, ITR 0121671 and NSF Grant DMS-0104016.

Let  $S_0 = \{s \in S : \mu(s) = 0\}$  denote the unknown set on which the null hypothesis is true.

Let  $\lambda$  denote Lebesgue measure on  $S$ , or counting measure if  $S$  is discrete. Following Genovese and Wasserman (2001, 2003), we define the False Discovery Proportion (FDP) process  $\Gamma = \{\Gamma(t) : t \in \mathbb{R}\}$  by

$$\Gamma(t) = \frac{\lambda(S_0 \cap \{s \in S : X(s) \geq t\})}{\lambda(\{s \in S : X(s) \geq t\})}. \quad (2)$$

The FDR is then defined as  $\mathbb{E}(\Gamma(t))$  for each threshold  $t$ .

Our general goal is to choose a data-dependent threshold  $T$  that guarantees a bound on some measure of error. We consider (i) controlling FDR when, for a pre-specified  $0 < \alpha < 1$ , we choose  $T$  such that  $\mathbb{E}(\Gamma(T)) \leq \alpha$  (Benjamini and Hochberg, 1995), (ii) controlling quantiles of the FDP distribution when, for pre-specified  $0 < \alpha < 1$  and  $0 < c < 1$ , we choose  $T$  so that  $\mathbb{P}(\Gamma(T) \leq c) \geq 1 - \alpha$ . Genovese and Wasserman (2001, 2003) call such a  $T$  a *confidence threshold*. (They also consider a version in which  $c$  is itself data dependent).

One approach for constructing such thresholds is to find first a confidence envelope for the FDP process  $\Gamma(t)$  defined in (2). For this, we use the strategy in Genovese and Wasserman (2001, 2003) based on inverting a collection of goodness of fit tests. See also Genovese and Wasserman (2003) where this idea is explored more fully in the standard multiple testing problem.

Our strategy for finding a confidence envelope for  $\Gamma$  is as follows:

1. For every  $A \subset S$ , test at level  $\alpha$ , the hypothesis,

$$H_0 : A \subset S_0 \quad \text{versus} \quad H_1 : A \not\subset S_0, \quad (3)$$

using the test statistic

$$X(A) = \sup_{s \in A} X(s). \quad (4)$$

2. Let  $\mathcal{C}$  denote all subsets  $A$  not rejected in the previous step. Our construction ensures that  $\mathcal{C}$  is closed under unions.

3. Define  $U = \bigcup_{A \in \mathcal{C}} A$ . The set  $U$  is a  $(1 - \alpha)$  *confidence superset* for  $S_0$ , meaning that

$$\mathbb{P}(U \supset S_0) \geq 1 - \alpha. \quad (5)$$

4. Define

$$\bar{\Gamma}(t) = \sup_{A \in \mathcal{C}} \frac{\lambda(A \cap \{s \in S : X(s) \geq t\})}{\lambda(\{s \in S : X(s) \geq t\})} = \frac{\lambda(U \cap \{s \in S : X(s) \geq t\})}{\lambda(\{s \in S : X(s) \geq t\})}, \quad (6)$$

where the second equality follows because  $\mathcal{C}$  is closed under unions. The curve  $\bar{\Gamma}$  is a *confidence envelope*, meaning that

$$\mathbb{P}(\Gamma(t) \leq \bar{\Gamma}(t), \text{ for all } t) \geq 1 - \alpha. \quad (7)$$

Given the confidence envelope  $\bar{\Gamma}$ , define  $T_c = \inf\{t : \bar{\Gamma}(t) \leq c\}$ . Then  $\mathbb{P}(\Gamma(T_c) > c) \leq \alpha$ . In Genovese and Wasserman (2001),  $T_c$  is called a  $(1 - \alpha)$  confidence threshold with ceiling  $c$ , or a  $(c, \alpha)$  threshold. An alternative is the minimum rate  $(1 - \alpha)$  confidence threshold, defined in Genovese and Wasserman (2003) as  $T_m = \arg \min_t \bar{\Gamma}(t)$ , which satisfies  $\mathbb{P}(\Gamma(T_m) > C) < \alpha$  with  $C = \bar{\Gamma}(T_m)$ . To control FDR, let  $c \in (0, \alpha)$  and take  $T$  to be a  $1 - \beta$  confidence threshold where  $\beta = (\alpha - c)/(1 - c)$ . We show in Lemma 6 that  $\mathbb{E}\Gamma(T) \leq \alpha$ .

Our methods have strong connections to recent research related to controlling FDR in standard multiple testing problems. Among the many contributions we mention the following. Benjamini and Yekutieli (2001) extended the Benjamini-Hochberg method to a class of dependent tests. Efron, Tibshirani, Storey and Tusher (2001) developed an empirical Bayes approach to multiple testing and made interesting connections with FDR. Storey (2001, 2002) connected the FDR concept with a certain Bayesian quantity and proposed a new FDR method which has higher power than the original Benjamini-Hochberg method. Finner and Roters (2002) discuss the behavior of the expected number of type I errors. Sarkar (2002) considers a general class of stepwise multiple testing methods. Storey, Taylor and Siegmund (2003) study FDR procedures in some generality.

## 2. FINDING CONFIDENCE SUPERSETS FOR $S_0$ .

In this section, we describe a method for finding the set  $U$ , which is a confidence superset for  $S_0$  as defined in (5).

The tail area for the test in (3) is defined for  $A \subset S_0$  by

$$p(z, A) = \mathbb{P}(X(A) \geq z). \quad (8)$$

The following monotonicity properties of  $p(z, A)$  will be useful in what follows:

$$z_1 < z_2 \quad \text{implies} \quad p(z_1, A) \geq p(z_2, A), \text{ for all } A \quad (9)$$

$$A_1 \subset A_2 \quad \text{implies} \quad p(z, A_1) \leq p(z, A_2), \text{ for all } z. \quad (10)$$

These are immediate consequences of definitions (4) and (8).

Let  $x(A)$  denote the realized value of the test statistic  $X(A)$  in equation (4), thus the p-value is  $p(x(A), A) = \mathbb{P}(X(A) \geq x(A))$ . We define the class  $\mathcal{C}$  of all sets such that the null hypothesis is retained by

$$\mathcal{C} = \{A \subset S : p(x(A), A) \geq \alpha\} \quad (11)$$

and  $U = \bigcup_{A \in \mathcal{C}} A$  as above.

**Lemma 1** *The class  $\mathcal{C}$  is closed with respect to the union.*

**Lemma 2** *For each  $A \subset S_0$ ,  $\mathbb{P}(\mathcal{C} \ni A) \geq 1 - \alpha$  and hence  $\mathbb{P}(U \supset S_0) \geq 1 - \alpha$ .*

Proofs are provided in Section 10.

It appears from the definition above that to carry out our strategy requires considering every subset of  $S$ . However, in the next two subsections, we propose a procedure that reduces the number of sets  $A$  for which p-values must be computed to a level that is feasible in practice.

## 2.1 Finite multiple testing problems.

Because our strategy for random fields will involve solving a sequence of finite problems, we first consider a coarser set of hypotheses based on a finite partition of  $S$ . The results in this section also solve the problem of finding a confidence superset when  $S$  is finite.

Let  $S_1, \dots, S_N$  be a partition of  $S$  and let  $X(S_1), \dots, X(S_N)$  denote the test statistic values for those  $N$  elements. We make no assumptions about the dependence among these statistics, but we assume that the null distribution of  $\sup_{j \in J} X(S_j)$  can be computed for any collection of indexes  $J \subset \{1, \dots, N\}$ .

The following strategy requires at most  $N$  steps even though it is equivalent to testing all  $2^N$  unions of elements of the partition.

1. Compute all realized values of the test statistics  $x(S_j)$ , for  $j = 1, \dots, N$ .
2. Sort the values of the single test statistics in decreasing order and denote them as  $x_{(1)} \geq \dots \geq x_{(N)}$ . Let  $S_{(j)}$  be the partition element corresponding to the  $j$ -th ordered test statistic  $x_{(j)}$ .
3. For  $k = 1, \dots, N$  do the following:
  - (a) Set  $V_k = \bigcup_{j=k}^N S_{(j)}$ .
  - (b) Compute  $p(x_{(k)}, V_k)$ .
  - (c) If  $p(x_{(k)}, V_k) \geq \alpha$ : STOP and set  $V^* = V_k$ .
  - (d) If  $p(x_{(k)}, V_k) < \alpha$ : increase  $k$  by 1 and GOTO 3a.

The rationale for Steps 3c and 3d is that we are looking for the largest set  $A$  (union of sets in the partition) with  $p(x(A), A) \geq \alpha$ . Thus, we stop as soon as  $V_k$  satisfies this condition (Step 3c). Otherwise, if  $p(x_{(k)}, V_k) < \alpha$ , then for each  $A$  such that  $S_{(k)} \subset A \subset V_k$  we have  $x(A) = x_{(k)}$  and, from the monotonicity of  $p(x_{(k)}, \cdot)$ ,

$$p(x(A), A) = p(x_{(k)}, A) \leq p(x_{(k)}, V_k) < \alpha.$$

Hence all sets containing  $S_{(k)}$  do not belong to  $\mathcal{C}$ . As a consequence, we proceed in Step 3d by removing  $S_{(k)}$  from the set that we test next.

**Lemma 3** *The output of the above algorithm, the set  $V^*$ , equals the confidence superset  $U$ .*

**Remark 1** Under the assumption that the test statistics  $X(S_j)$  are independent and identically distributed, sorting the observed values  $x(S_j)$  is equivalent to sorting the  $p$ -values, of the corresponding single test problems, in increasing order:

$$p(x_{(1)}, S_{(1)}) \leq p(x_{(2)}, S_{(2)}) \leq \cdots \leq p(x_{(N)}, S_{(N)}).$$

Since the test statistics are IID, then  $p(x_{(k)}, V_k) = 1 - [1 - p(x_{(k)}, S_{(k)})]^{N+1-K}$ . Using the strategy described above, one will retain all the hypotheses  $S_{(k^*)}, \dots, S_{(N)}$  where  $k^*$  is the first index such that  $p(x_{(k)}, V_k) \geq \alpha$ . Thus one rejects all the null hypotheses for which  $p(x_{(k)}, S_{(k)}) < p(x_{(k^*)}, S_{(k^*)})$ , Where

$$k^* = \min \left\{ 1 \leq k \leq N : p(x_{(k)}, S_{(k)}) \geq 1 - (1 - \alpha)^{\frac{1}{N+1-k}} \right\}. \quad (12)$$

Figure 1 compares this procedure with the one proposed by Benjamini and Hochberg (1995) that rejects all null hypotheses for which  $p(x_{(k)}, S_{(k)}) \leq p(x_{(BH)}, S_{(BH)})$ , where:

$$BH = \max \left\{ 1 \leq k \leq N : p(x_{(k)}, S_{(k)}) \leq \alpha \frac{k}{N} \right\}, \quad (13)$$

or take  $BH = 0$  if no such  $k$  exists.

The boundary is more conservative than Benjamini and Hochberg's. This is to be expected, because this strategy searches for the largest set of hypotheses that would be retained, with probability at most  $\alpha$ , while Benjamini and Hochberg's controls the FDR, or the expected value of the false discovery proportion. ■

## 2.2 Testing infinitely many hypotheses.

We now turn to the case where there are infinitely many hypotheses. For each  $n \in \mathbb{N}$ , let  $\mathcal{S}_n = \{S_1, \dots, S_{N_n}\}$  denote a partition of  $S$ . We consider partitions such that  $p$ -values can be computed for each set and for their unions. Let  $U_n$  denote the set in Lemma 3, obtained from applying the strategy of Section 2.1 to  $\mathcal{S}_n$ . The following results show that, if the partitions are chosen properly, the limit of  $U_n$  is a level  $\alpha$  upper confidence set for  $S_0^o$ , the interior of  $S_0$ . We choose the partitions of  $S$  to be nested and degenerating, defined as follows.

**Definition 1** For each  $s \in S$  and each  $n \in \mathbb{N}$ , let  $S_{n,s}$  be the partition element in  $\mathcal{S}_n$  that contains  $s$ . We say that the sequence of partitions  $\mathcal{S}_n$  is *degenerating* if for each  $s \in S$  and for each open neighborhood  $O_s$  of  $s$ , the set  $S_{n,s} \subset O_s$  for sufficiently large  $n$ .

Applying the strategy described in Section 2.1 to each  $\mathcal{S}_n$ , we obtain a sequence of sets  $U_n$  that converge to a confidence superset for  $S_0^o$ , the interior of  $S_0$ . Analogous to the above, for each  $n$  we define

$$\mathcal{C}_n = \left\{ A = \bigcup_{S_j \in \mathcal{S}_n} S_j : p(x(A), A) \geq \alpha \right\}, \quad (14)$$

and take  $U_n = \bigcup_{A \in \mathcal{C}_n} A$ .

**Lemma 4** *The sequences  $\mathcal{C}_n$  and  $U_n$  are both increasing.*

**Theorem 1** *If the sequence of partitions  $\mathcal{S}_n$  is nested and degenerating, then  $\lim_n U_n$  is a  $(1 - \alpha)$  confidence superset for  $S_0^o$*

$$\mathbb{P} \left( \lim_n U_n \supset S_0^o \right) \geq 1 - \alpha.$$

Also,  $\mathbb{P}(U \supset S_0) \geq 1 - \alpha$ , where  $U$  is the closure of  $\lim_n U_n$ .

**Remark 2** The proof of Theorem 1 rests on the monotonicity of  $\mathcal{C}_n$ , as proved in Lemma 4. This makes the procedure more efficient. Once  $U_{n-1}$  is found, one can apply the above algorithm to a partition of  $S \setminus U_{n-1}$ . This gives all elements of  $\mathcal{C}_n \setminus \mathcal{C}_{n-1}$  and  $U_n = U_{n-1} \cup \bigcup_{A \in \mathcal{C}_n \setminus \mathcal{C}_{n-1}} A$ . This is reminiscent of the tree hypothesis testing method in Benjamini and Yekutieli (2003). ■

**Remark 3** This procedure is a continuous analogue of the  $P_{(1)}$  test of Genovese and Wasserman (2003). ■

### 3. EXTRACTING THRESHOLDS

Once the confidence superset  $U$  is available, we can build the confidence envelope  $\bar{\Gamma}$  defined in (6). Here, we summarize briefly how thresholds are extracted from  $\bar{\Gamma}$ .

**Lemma 5** *Let  $\bar{\Gamma}$  be a  $(1 - \alpha)$  confidence band. Let  $T = \inf\{t : \bar{\Gamma}(t) \leq c\}$ . Define  $T = \infty$  if no such  $t$  exists. Then  $T$  is a  $(c, \alpha)$  confidence threshold, i.e.  $\mathbb{P}(\Gamma(T) \leq c) \geq 1 - \alpha$ .*

**Lemma 6** *Let  $c \in (0, \alpha)$  and let  $\beta = (\alpha - c)/(1 - c)$ . Let  $T$  be a  $(c, \beta)$  confidence threshold. Then  $T$  is a level  $\alpha$  FDR threshold, meaning that  $\mathbb{E}(\Gamma(T)) \leq \alpha$ .*

**Remark 4** Lemma 6 defines a family of procedures indexed by  $c \in (0, \alpha)$ . It would be interesting to study the power of the procedure as a function of  $c$  though we do not pursue this here. ■

We can use the confidence superset  $U$  to control the number of false clusters instead of their area; to our knowledge such a method does not presently exist. Decompose the level set  $L_t = \{s \in S : X(s) \geq t\}$  into its connected components  $C_{1t}, \dots, C_{k_t t}$ . We call these the level- $t$  observed clusters. Let  $\epsilon$  be a user-specified tolerance level. Say that a cluster  $C$  is false at tolerance  $\epsilon$  if

$$\frac{\lambda(C \cap S_0)}{\lambda(C)} \geq \epsilon. \quad (15)$$

The false cluster proportion  $\Xi(t)$  is defined to be the number of false clusters above threshold  $t$  divided by the total number of clusters above threshold  $t$ . Define the  $\epsilon$  false cluster envelope  $\bar{\Xi}(t)$  by

$$\bar{\Xi}(t) = \frac{\# \left\{ 1 \leq i \leq k_t : \frac{\lambda(C_{it} \cap U)}{\lambda(C_{it})} \geq \epsilon \right\}}{k_t} \quad (16)$$

where  $U$  is the  $(1 - \alpha)$  confidence superset defined in Section 2.

**Lemma 7** *If  $U$  is a  $(1 - \alpha)$  confidence superset for  $S_0$ , then*

$$\mathbb{P}(\Xi(t) \leq \bar{\Xi}(t) \text{ for all } t) \geq 1 - \alpha. \quad (17)$$

Similarly to Section 3, we can extract thresholds from  $\bar{\Xi}$  that control the proportion of false clusters. An example is given in Section 8.

#### 4. THE FALSE NONDISCOVERY RATE

The false nondiscovery proportion was introduced in Genovese and Wasserman (2001) as the analogue of power in the FDR context. In the current setting, define the False Nondiscovery Proportion (FNP) process by

$$\Lambda(t) = \frac{\lambda(S_1 \cap \{X < t\})}{\lambda(\{X < t\})} \quad (18)$$

where  $S_1$  is the complement of  $S_0$ . The False Nondiscovery Rate (FNR) is  $\mathbb{E}(\Lambda(t))$ . Our goal in this section is to find a confidence band for  $\Lambda(t)$ .

As in all power calculations, we need to restrict the alternatives to get non-trivial results. For a fixed  $\epsilon > 0$ , consider as alternative hypotheses the class of distributions

$$\mathcal{P}_\epsilon = \left\{ \mathbb{P} \in \mathcal{P} : \inf_{s \in S_1} \mu(s) \geq \epsilon \right\}.$$

We want to find  $\bar{\Lambda}(t)$  such that

$$\mathbb{P}(\Lambda(t) \leq \bar{\Lambda}(t) \text{ for all } t) \geq 1 - \alpha, \quad \text{for all } \mathbb{P} \in \mathcal{P}_\epsilon \quad (19)$$

which is equivalent to

$$\inf_{\mathbb{P} \in \mathcal{P}_\epsilon} \mathbb{P}(\Lambda(t) \leq \bar{\Lambda}(t) \text{ for all } t) \geq 1 - \alpha.$$

We use a strategy similar to that used earlier. For any  $A \subset S$  consider testing

$$H_0 : A \subset S_1 \quad \text{versus} \quad H_1 : A \not\subset S_1. \quad (20)$$

Suppose we have a test for (20) which is level at most  $\alpha$  for every  $\mathbb{P} \in \mathcal{P}_\epsilon$ . Let  $\mathcal{A}$  be the set of non-rejected  $A$ . Then

$$\bar{\Lambda}(t) = \sup_{A \in \mathcal{A}} \frac{\lambda(A \cap \{X < t\})}{\lambda(\{X < t\})}$$

is a confidence band.

**Theorem 2** *Let  $Y = -(X - \epsilon)$ . Let  $\bar{\Gamma}(t)$  denote the  $(1 - \alpha)$  upper envelope obtained by applying the procedure from Section 3 to the field  $Y$ . Let  $\bar{\Lambda} = \bar{\Gamma}$ . Then*

$$\inf_{\mathbb{P} \in \mathcal{P}_\epsilon} \mathbb{P}(\Lambda(t) \leq \bar{\Lambda}(t) \text{ for all } t) \geq 1 - \alpha.$$

## 5. THE GAUSSIAN CASE.

Assume that  $S = [0, 1]^d$  and that, under the null hypothesis, the test statistic  $X(s)$  is a homogeneous Gaussian random field over  $S$  that has almost surely continuous paths, zero mean and covariance structure

$$\mathbb{C}(X(r), X(s)) = \sigma^2 \rho(r - s), \quad (21)$$

where  $\sigma^2 > 0$  is the variance of  $X(s)$ .

For applying the procedure described in Section 2, we need to define a nested and degenerating sequence  $\mathcal{S}_n$  of partitions. For each  $n \in \mathbb{N}$ , we partition the set  $S$  into  $2^{dn}$  hypercubes, each with edge length  $2^{-n}$ :

$$\mathcal{S}_n = \left\{ \left\{ (s_1, \dots, s_d) \in S : \frac{a_j - 1}{2^n} \leq s_j < \frac{a_j}{2^n}, \quad j = 1, \dots, d \right\}, \quad a \in \{1, \dots, 2^n\}^d \right\}. \quad (22)$$

The partitions in (22) are nested and degenerating.

We turn now to computing the  $p$ -value functions  $p(z, A)$ , defined in (8), for each  $z \in \mathbb{R}$  and for each set  $A$ , a finite union of hypercubes (22). Note first that, since the paths

of  $X$  are continuous almost surely, the supremum of  $X$ , over any set, coincides with the supremum over the closure of the same set.

A common method for approximating  $p(z, A)$  is to use the topological properties of the level sets of the random field, especially their expected Euler characteristic (Piterbarg 1996, Theorem 5.1; Adler 1981, 1990, 2000; and Worsley 1994, 1995). We will consider, instead, an approximation based on Theorem 7.1 in Piterbarg (1996). The reasons for using this alternative approximation are explained in Section 6. Briefly, this choice is needed when using the procedure of Section 2, where we tests hypotheses on sets that are union of hypercubes (22) and that might fail to be convex. Our procedure requires monotonicity of the tail area approximation which will fail for the Euler-based approximation. We take the following definition from Piterbarg (1996).

**Definition 2** A Gaussian random field  $X$  over  $A \subset \mathbb{R}^d$  is *locally stationary with quadratic covariance* if for any  $\varepsilon > 0$  there exists  $\delta > 0$  such that for all  $s \in A$  there exists a matrix-valued function  $C_s$  for which the field's covariance structure satisfies

$$\|(1 - \varepsilon)(r - t)\|^2 \leq 1 - \mathbb{C}(X(C_s r), X(C_s t)) \leq \|(1 + \varepsilon)(r - t)\|^2, \quad (23)$$

for all pairs  $r, t \in S$  with  $\|s - r\| \leq \delta$ ,  $\|s - t\| \leq \delta$ .

If  $X$  is a homogeneous Gaussian field, the normalized field  $X/\sigma$  is locally stationary with quadratic covariance if, for some matrix  $C$

$$\rho(Cs) = 1 - \|s\|^2 + o(\|s\|^2) \quad \text{as } s \rightarrow 0 \quad (24)$$

or, equivalently,  $\rho(s) = 1 - s^T B s + o(\|s\|^2)$ , where  $B = C^{-2}$ .

**Theorem 3** [THEOREM 7.1 FROM PITERBARG (1996)] *Let  $A \subset \mathbb{R}^d$  be a closed set whose boundary has zero Lebesgue measure and let  $X$  be a Gaussian field over  $A$ , with zero mean. Assume also that  $X$  is locally stationary with quadratic covariance  $\mathbb{C}(X(r), X(s))$  and that the matrix-valued function  $C_s$  is continuous in  $s$  and non-degenerate everywhere on  $A$ . Then if  $\mathbb{C}(X(r), X(s)) < 1$  for all  $r \neq s \in A$ ,*

$$\mathbb{P} \left( \sup_{s \in A} X(s) \geq z \right) = \left( \int_A |\det C_s|^{-1} ds \right) \pi^{-\frac{d}{2}} z^d (1 - \Phi(z))(1 + o(z)) \quad (25)$$

where  $\Phi$  is the cdf of a standard normal.

Finite unions of closed hypercubes are closed sets with zero-measure boundary, thus if (24) is satisfied, the following approximation for the  $p$ -value function is obtained from Theorem 3:

$$p(z, A) = \mathbb{P} \left( \sup_{s \in A} \frac{X(s)}{\sigma} \geq \frac{z}{\sigma} \right) \simeq \frac{\pi^{-\frac{d}{2}}}{|\det C|} \lambda(A) \left( \frac{z}{\sigma} \right)^d \left[ 1 - \Phi \left( \frac{z}{\sigma} \right) \right]. \quad (26)$$

In our applications, we deal with the homogeneous correlation function

$$\rho(s) = \frac{a \cdot \exp(-s^T B s) + c}{a + c} \quad (27)$$

where  $s^T$  denotes the transpose of  $s$ . The correlation structure (27) satisfies (24) for any symmetric positive definite matrix  $B$ , and any nonzero constant  $a$ . In fact, if  $C = (\frac{a}{a+c}B)^{-\frac{1}{2}}$  then  $(Cs)^T B(Cs) = \frac{a+c}{a} s^T s = \frac{a+c}{a} \|s\|^2$  and

$$\rho(Cs) = \frac{a \cdot \exp(-\frac{a+c}{a} \|s\|^2) + c}{a + c} = 1 - \|s\|^2 + o(\|s\|^2). \quad (28)$$

Approximation (26) applied to a field with correlation (27) gives

$$p(z, A) \simeq \pi^{-\frac{d}{2}} \left( \frac{a}{a+c} \right)^{\frac{d}{2}} \sqrt{\det B} \lambda(A) \left( \frac{z}{\sigma} \right)^d (1 - \Phi(z)). \quad (29)$$

**Remark 5** Approximation (26) is not a decreasing function of  $z$  over the whole real line. In fact  $z^d(1 - \Phi(z))$  is increasing for  $z \in (0, \bar{z})$  and decreasing for  $z \geq \bar{z}$ , and the approximation to  $p(z, A)$  in (26) behaves in the same way. But, we are considering level  $\alpha$  tests, and are interested in large values of  $z$  in the right tail of the distribution of the test statistic. For a level  $\alpha$  test, there might be 0, 1 or 2 real roots of the equation  $p(z, A) = \alpha$ , for  $z > 0$ . In most practical cases, at least one such root exists. Let  $z_\alpha$  denote the largest root  $z_\alpha = \max\{z > 0 : p(z, A) = \alpha\}$ . The testing rule will reject the hypothesis if the observed test statistics is greater than  $z_\alpha$  and will retain it otherwise. ■

**Remark 6** Another inconsistency of approximation (26) is that its right-hand-side depends on  $A$  through  $\lambda(A)$ , that satisfies  $\lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2) - \lambda(A_1 \cap A_2)$ . As a consequence we have that:

$$\mathbb{P} \left( \sup_{s \in A_1 \cup A_2} X(s) \geq z \right) = \mathbb{P} \left( \sup_{s \in A_1} X(s) \geq z \right) + \mathbb{P} \left( \sup_{s \in A_2} X(s) \geq z \right) - \mathbb{P} \left( \sup_{s \in A_1 \cap A_2} X(s) \geq z \right),$$

while the true distribution must satisfy:

$$\begin{aligned} \mathbb{P} \left( \sup_{s \in A_1 \cup A_2} X(s) \geq z \right) &= \mathbb{P} \left( \sup_{s \in A_1} X(s) \geq z \right) + \mathbb{P} \left( \sup_{s \in A_2} X(s) \geq z \right) \\ &\quad - \mathbb{P} \left( \sup_{s \in A_1} X(s) \geq z, \sup_{s \in A_2} X(s) \geq z \right) \end{aligned}$$

so that (26) yields the incorrect equality

$$\mathbb{P} \left( \sup_{s \in A_1} X(s) \geq z, \sup_{s \in A_2} X(s) \geq z \right) = \mathbb{P} \left( \sup_{s \in A_1 \cap A_2} X(s) \geq z \right).$$

Obviously, this same inconsistency exists when dealing with the expected Euler characteristic. This is due to the use of a function (either Lebesgue measure  $\lambda$  or Euler characteristic) that is additive with respect to  $A$  while  $p(z, \cdot)$  is not. Note that Piterbarg (1996, proof of Theorem 7.1) shows that, as the sets become smaller and smaller, as it is the case here, the probability of pairwise intersections gets negligibly small and thus this inconsistency is irrelevant.  $\blacksquare$

## 6. NON-MONOTONICITY OF EULER-BASED P-VALUE APPROXIMATIONS

In this section, we explain why the p-value approximation based on the Euler characteristic of the level sets can fail to be monotone. We require monotonicity in our method since we need to test all subsets in an efficient way.

Consider a homogeneous Gaussian random field with covariance structure over  $S = [0, 1]^2$  as in (34), that is a special case of (27). The approximation of  $p(z, A)$ , based on the expected Euler characteristic, for such a random field is given by:

$$p(z, A) \simeq \left[ \frac{bz}{\pi} \mathcal{A}(A) + \left( \frac{b}{\pi} \right)^{1/2} \frac{\mathcal{P}(A)}{2} \right] \varphi(z) + \mathcal{E}(A)(1 - \Phi(z)), \quad (30)$$

where  $\mathcal{A}(A)$ ,  $\mathcal{P}(A)$ ,  $\mathcal{E}(A)$  denote, respectively, area, perimeter, and Euler characteristic of a set  $A$ , and  $\varphi(\cdot)$  is the pdf of a standard normal. Note that approximation (29) happens to be the first term of (30).

We show that if the sequence of partitions of  $S \subset \mathbb{R}^2$  is chosen as in (22), approximation (30) does not preserve the monotonicity of  $p(z, A)$ . Following the procedure of Section 2, when a set  $V_k$  is rejected, we test the set  $V_{k+1} = V_k \setminus S_{(k)}$  where  $S_{(k)}$  is a square with side length  $l > 0$ .

In the example of Figure 2, the area, perimeter and Euler characteristic of  $V_{k+1}$  and  $V_k$  are in the following relationships:

$$\mathcal{A}(V_{k+1}) = \mathcal{A}(V_k) - l^2 \quad \mathcal{P}(V_{k+1}) = \mathcal{P}(V_k) + 2l \quad \mathcal{E}(V_{k+1}) = \mathcal{E}(V_k).$$

Using approximation (30) we obtain

$$\begin{aligned} p(z, V_{k+1}) &= \left( \frac{bz}{\pi} \mathcal{A}(V_k) - l^2 + \sqrt{\frac{b}{\pi}} \frac{(\mathcal{P}(V_k) + 2l)}{2} \right) \varphi(z) + \mathcal{E}(V_k)(1 - \Phi(z)) \\ &= p(z, V_k) - \left[ \frac{bz}{\pi} l^2 - l \sqrt{\frac{b}{\pi}} \right] \varphi(z). \end{aligned}$$

If  $l < \sqrt{\frac{\pi}{bz^2}}$  the term in square brackets is negative and  $p(z, V_{k+1}) > p(z, V_k)$ , in contradiction with (10).

The above nonmonotonicity is unavoidable for isotropic processes. We show this in the special case of covariance (34): when we remove a square from the last set tested  $V_k$ , for the actual set  $V_{k+1}$  we have

$$\begin{aligned}\mathcal{A}(V_{k+1}) &= \mathcal{A}(V_k) - l^2 \\ \mathcal{P}(V_{k+1}) &= \mathcal{P}(V_k) + 2l\Delta_p \\ \mathcal{E}(V_{k+1}) &= \mathcal{E}(V_k) + \Delta_e\end{aligned}$$

where  $\Delta_p \in \{-2, -1, 0, 1, 2\}$  is related to the number of sides of the square that are in  $V_{k+1}$  and  $\Delta_e$  is the difference in the Euler characteristic. The worst pairs  $(\Delta_p, \Delta_e)$  (worst pairs here means that at each possible  $\Delta_p$  we attach the highest compatible  $\Delta_e$ ) are:

$$(-2, +3) \quad (-1, +2) \quad (0, +1) \quad (+1, 0) \quad (+2, -1). \quad (31)$$

Using approximation (30) we obtain

$$\begin{aligned}p(z, V_{k+1}) &= \left( \frac{bz}{\pi} \mathcal{A}(V_{k+1}) + \sqrt{\frac{b}{\pi}} \frac{\mathcal{P}(V_{k+1})}{2} \right) \varphi(z) + \mathcal{E}(V_{k+1})(1 - \Phi(z)) \\ &= \left( \frac{bz}{\pi} (\mathcal{A}(V_k) - l^2) + \sqrt{\frac{b}{\pi}} \frac{\mathcal{P}(V_k) + 2l\Delta_p}{2} \right) \varphi(z) \\ &\quad + (\mathcal{E}(V_k) + \Delta_e)(1 - \Phi(z)) \\ &= p(z, V_k) + \left( -\frac{bz}{\pi} l^2 + \sqrt{\frac{b}{\pi}} \Delta_p l \right) \varphi(z) + \Delta_e (1 - \Phi(z)).\end{aligned}$$

To keep the correct inequality  $p(z, V_{k+1}) \leq p(z, V_k)$  we need

$$\left( -\frac{bz}{\pi} l^2 + \sqrt{\frac{b}{\pi}} \Delta_p l \right) \varphi(z) + \Delta_e (1 - \Phi(z)) \leq 0.$$

Using inequality (1.1.7) in Adler (2000)

$$(1 - \frac{1}{z^2}) \frac{\varphi(z)}{z} \leq (1 - \Phi(z)) \leq \frac{\varphi(z)}{z},$$

we get the following conditions

$$\begin{aligned}-\frac{bz}{\pi} l^2 + \sqrt{\frac{b}{\pi}} \Delta_p l + \frac{\Delta_e}{z} &\leq 0 && \text{for } \Delta_e \geq 0 \\ -\frac{bz}{\pi} l^2 + \sqrt{\frac{b}{\pi}} \Delta_p l + (1 - \frac{1}{z^2}) \frac{\Delta_e}{z} &\leq 0 && \text{for } \Delta_e < 0.\end{aligned}$$

For  $\Delta_e \geq 0$  the positive solutions are

$$l \geq \sqrt{\frac{\pi}{b} \frac{\Delta_p + \sqrt{\Delta_p^2 + 4\Delta_e}}{2z}};$$

while for  $\Delta_e < 0$  the positive solutions are (if  $\Delta_p > 0$ )

$$0 < l \leq \sqrt{\frac{\pi}{b} \frac{\Delta_p - \sqrt{\Delta_p^2 + 4(1 - \frac{1}{z^2})\Delta_e}}{2z}} \quad l \geq \sqrt{\frac{\pi}{b} \frac{\Delta_p + \sqrt{\Delta_p^2 + 4(1 - \frac{1}{z^2})\Delta_e}}{2z}}.$$

All the pairs in (31) with  $\Delta_e \geq 0$  give

$$l \geq \frac{1}{z} \sqrt{\frac{\pi}{b}}$$

while  $(+2, -1)$  gives

$$0 < l \leq \frac{z-1}{z^2} \sqrt{\frac{\pi}{b}} \quad l \geq \frac{z+1}{z^2} \sqrt{\frac{\pi}{b}}.$$

To prevent non-monotonicity we need to consider the intersection between all the solutions that gives

$$l \geq \frac{z+1}{z^2} \sqrt{\frac{\pi}{b}}. \tag{32}$$

so that the side length of the squares cannot converge to 0, as required by Definition 1.

In contrast, the approximation (25) that we use, does not suffer from this problem. Since  $|\det C_s|^{-1} > 0$  in (25):

$$\begin{aligned} A_1 \subset A_2 \quad \text{implies} \quad & \int_{A_1} |\det C_s|^{-1} ds \leq \int_{A_2} |\det C_s|^{-1} ds \\ & \text{implies} \quad \mathbb{P} \left( \sup_{s \in A_1} X(s) \geq z \right) \leq \mathbb{P} \left( \sup_{s \in A_2} X(s) \geq z \right), \end{aligned} \tag{33}$$

and the approximation of  $p(z, A)$  in (26) is monotone in  $A$ , even if  $A_1$  or  $A_2$  are non-convex.

## 7. SIMULATIONS.

The examples in this section consider four different signal sources in  $S = [0, 1]^2$ , denoted as *Horseshoe*, *Bullets*, *Bubbles* and *Romper Room*, displayed in Figure 3. We add noise from

a Gaussian random field to each of these signals. The noise is generated with zero mean and covariance function

$$\mathbb{C}(X(r), X(s)) = \sigma^2 \exp\{-b||s - r||^2\} \quad (34)$$

using the R package `RandomFields`. We take  $\sigma = 300$  and  $b = 100$ . The signals have different intensities varying from  $2\sigma$  to  $5\sigma$ , and shown from darker to lighter shades. The signal for *Romper room* is set constant to  $5\sigma = 1,500$ . All images are shown at their finest resolution,  $256 \times 256$  pixels.

Figure 4 shows the analysis of a typical realization from the *Bubbles* example and, for comparison, the level set from familywise approach. Table 1 reports the ( $c = 0.1, \alpha = 0.05$ ) threshold  $T$  and the corresponding true values of  $\Gamma(T)$  and  $\Lambda(T)$  for different resolution sizes. Figure 5 shows the true FDP process and its upper envelope  $\bar{\Gamma}$ .

Table 2 summarizes the results of 1000 simulations. The coverage is very close to the nominal value. The true FNP's of the confidence threshold are quite small. In this sense, the method has high “power.” There is also fairly rapid convergence as the resolution size increases. However, the FNP envelope is very conservative. We conjecture that a test statistic other than  $\sup_A X(s)$  might give tighter confidence limits for FNP.

## 8. SCAN CLUSTERING.

A common problem in epidemiology, astronomy, and other fields is to detect unusual clustering of events in a point process. A standard method for detecting the existence of such clusters is the scan statistic, defined as the maximum number of points in a fixed window as the window is shifted across the domain (Glaz, Naus, and Wallenstein 2001). This statistic is used for an omnibus test of the null hypothesis that there is no clustering. If the test rejects, it leaves open the question of where and how much clustering there is. Scan statistics are usually based on rectangular windows whose size is chosen arbitrarily.

In this section, we develop false discovery control for scan statistics. Our approach allows for inferences on where and how much clustering there is. The method can be used for rectangular windows however it just as easily extends to more general kernels.

We begin with an example from astronomy. Figure 6 A shows the smoothed data from an astronomical sky survey with  $N = 135,864$  points. Each point represents one galaxy. The image is a kernel density estimate (described below). Astronomers are interested in identifying clusters of galaxies. Here we formalize this and we adapt our methods from the previous sections for quantifying the error rate.

Let  $Y_1, \dots, Y_N$  be points from a point process on  $S = [0, 1] \times [0, 1]$  with intensity function  $\nu$ . Assume that  $\nu(s) = \nu_0$  for all  $s$  in a set  $S_0 \subset S$  called the “field” and that  $\nu(s) > \nu_0$  for  $s \notin S_0$ . Points outside  $S_0$  are said to be in “clusters”. Define the normalized density be

$f(s) = \nu(s) / \int_S \nu(u) du$ . Consider the hypotheses

$$H_{0,s} : s \in S_0 \quad \text{versus} \quad H_{1,s} : s \notin S_0. \quad (35)$$

Define the random field of scan statistics by

$$X(s) = \frac{1}{n} \sum_{j=1}^n K_H(s - Y_j) \quad (36)$$

where  $K_H$  is the two-dimensional kernel which we take here to be normal with fixed bandwidth matrix

$$H = \begin{pmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{pmatrix}.$$

Under weak conditions,

$$\sqrt{N} (X(s) - f(s)) \rightsquigarrow \mathbb{B}$$

where  $\mathbb{B}$  is a mean zero Gaussian process with covariance

$$R(s, t) = \int_S K_H(s - u) K_H(t - u) f(u) du - \int_S K_H(s - u) f(u) du \int_S K_H(t - u) f(u) du.$$

Hence,  $\mathbb{C}(X(s), X(t)) \approx N^{-1} R(s, t)$ .

We assume that  $\int_{S_0^c} \nu(s) / \nu_0 ds$  is small in which case  $f(s) \approx 1$  for  $s \in S_0$ . Moreover, assuming that  $\max(h_1, h_2)$  is small relative to the area of  $S$  and that  $f$  is bounded,

$$\int_S K_H(s - u) K_H(t - u) f(u) du \approx \int_{\mathbb{R}^2} K_H(s - u) K_H(t - u) du$$

and

$$\int_S K_H(s - u) f(u) du \approx 1.$$

Hence,

$$R(s, t) \approx \int_{\mathbb{R}^2} K_H(s - u) K_H(t - u) du - 1 = K_{2H}(s - t) - 1. \quad (37)$$

Therefore, the correlation function is

$$\rho(s) = \frac{R(s, 0)}{R(s, s)} \approx \frac{\frac{1}{2\pi\sqrt{\det(2H)}} \exp\left(-\frac{1}{2}s^T(2H)^{-1}s\right) - 1}{\frac{1}{2\pi\sqrt{\det(2H)}} - 1}.$$

This is of the form (28) and hence approximation (29) applies with

$$a = \frac{1}{2\pi\sqrt{\det(2H)}}, \quad B = \frac{1}{4}H^{-1} \quad \text{and} \quad c = -1.$$

While this approximation to the correlation function appears to be adequate, we have found it better to estimate the field’s variance empirically, while using the approximation for the correlation function.

We computed  $\Xi(t)$  in the example using a kernel with bandwidth chosen by cross-validation and applied our method. Figure 6 C shows a plot of  $\Xi(t)$  versus  $t$ . With  $\epsilon = 0.1$  and  $c = 0.1$ , the minimum confidence threshold  $T$  for which  $\Xi(T) \leq c$  is  $T = 1.233$ . For that value of  $T$ , 7 clusters are declared to be false out of 73. The retained clusters are shown in figure 6 B.

## 9. IMAGING.

The data from a functional Magnetic Resonance Imaging (fMRI) experiment consist of a sequence of three-dimensional images obtained at regular intervals in a Magnetic Resonance scanner while the subject performs a sequence of behavioral tasks. The behavioral tasks are designed to exercise the brain functions under study. Each image is a three-dimensional array of volume elements or *voxels*, usually arranged in two-dimensional *slices*. Concentrated neural activity gives rise to a localized blood-flow (hemodynamic) response that can be detected as small, systematic changes in the time course of measurements for voxels near the site of activity.

A typical fMRI data set has tens of thousands of voxels. A common approach to identifying voxels of interest in fMRI is to (i) perform a hypothesis test at each voxel (after suitable pre-processing of the data such as spatial smoothing) and then (ii) select those voxels for which the corresponding null hypothesis is rejected. This approach has proved effective for a wide range of tests including linear models (Friston *et al.*, 1995), spectral analysis (Müller *et al.*, 2001), permutation tests (Bullmore *et al.*, 1999; Belmonte and Yurgelun-Todd, 2001; Nichols and Holmes, 2002), Kolmogorov-Smirnov (Aguirre, Zarahn and D’Esposito, 1998), and nonlinear regression (Genovese, 2000).

Worsley *et al.* (1996) approached the image of test statistics derived from an fMRI data set as a random field. For example, a t-test computed at each voxel gives a t-field (Cao *et al.*, 1998). See also Worsley *et al.* (2002). The Worsley method approximates the distribution of the Euler characteristic of a level set of the field to compute a threshold for the tests that controls familywise error.

Taking a similar view of fMRI data, we apply our method for False Discovery Rate control to data selected from an fMRI study. We show one slice of the brain for a single subject in an eye movement experiment. Figure 7 shows a field of t-statistics (1083 degrees of freedom) after enough smoothing to make the assumptions of Section 5 apply. Because the t-statistics have 1083 degrees of freedom we can approximate them as Normal for both methods. The example is presented to illustrate our techniques and is not intended to be a definitive fMRI

analysis.

Figure 7 shows the field, the confidence envelope and the rejected voxels at the ( $c = 0.1, \alpha = 0.05$ ) threshold. These rejected voxels correspond to visual areas in the occipetal cortex.

## 10 PROOFS

**Lemma 1** *The class  $\mathcal{C}$  is closed with respect to the union.*

PROOF. Let  $U = \bigcup_{j \in J} A_j$  where  $J$  is an index set, and  $A_j \in \mathcal{C}$  for each  $j \in J$ . From (3),  $x(U) = \sup_{j \in J} x(A_j)$ , and  $p(\cdot, U)$  is decreasing and continuous from the left, hence

$$p(x(U), U) = p(\sup_{j \in J} x(A_j), U) = \inf_{j \in J} p(x(A_j), U).$$

Each  $A_j \subset U$  and  $p(z, \cdot)$  is increasing, thus  $p(x(A_j), U) \geq p(x(A_j), A_j)$ . Therefore, since  $p(x(A_j), A_j) \geq \alpha$

$$p(x(U), U) = \inf_{j \in J} p(x(A_j), U) \geq \inf_{j \in J} p(x(A_j), A_j) \geq \alpha,$$

and  $U \in \mathcal{C}$ . Hence  $\mathcal{C}$  is closed under unions.  $\square$

**Lemma 2** *For each  $A \subset S_0$ ,  $\mathbb{P}(\mathcal{C} \ni A) \geq 1 - \alpha$  and hence  $\mathbb{P}(U \supset S_0) \geq 1 - \alpha$ .*

PROOF. For a level  $\alpha$  test, the probability of rejecting a set  $A \subset S_0$ , is at most  $\alpha$ , and the probability that  $A \in \mathcal{C}$  is greater or equal to  $1 - \alpha$ . This applies also to the whole null set  $S_0$ , thus  $\mathbb{P}(\mathcal{C} \ni S_0) \geq 1 - \alpha$ . Finally, since  $U = \bigcup_{A \in \mathcal{C}} A$ ,  $\mathbb{P}(U \supset S_0) \geq \mathbb{P}(\mathcal{C} \ni S_0) \geq 1 - \alpha$ .  $\square$

**Lemma 3** *The output of the above algorithm, the set  $V^*$ , equals the confidence superset  $U$ .*

PROOF. Since  $p(x(V^*), V^*) \geq \alpha$  (Step 3c), then  $V^* \in \mathcal{C}$  and  $V^* \subset U$ . Conversely, if some  $S_j$  does not belong to  $V^*$ , monotonicity of  $p(x(A), \cdot)$  implies that  $p(x(A), A) < \alpha$  for each set  $A \supset S_j$ . As a consequence no subset of  $\mathcal{C}$  can contain  $S_j$ , thus  $S_j \not\subset U$  and  $U \subset V^*$ .  $\square$

**Lemma 4** *The sequences  $\mathcal{C}_n$  and  $U_n$  are both increasing.*

PROOF. Since the partitions are nested, each  $A \in \mathcal{C}_n$  can be obtained as the union of elements in  $\mathcal{S}_{n+1}$ . Since  $p(x(A), A) \geq \alpha$  still holds,  $A \in \mathcal{C}_{n+1}$ . Hence  $\mathcal{C}_n \subset \mathcal{C}_{n+1}$ , and  $U_n \subset U_{n+1}$ .  $\square$

**Theorem 1** *If the sequence of partitions  $(\mathcal{S}_n)$  is nested and degenerating, then  $\lim_n U_n$  is a  $(1 - \alpha)$  confidence superset for  $S_0^o$*

$$\mathbb{P}\left(\lim_n U_n \supset S_0^o\right) \geq 1 - \alpha.$$

Also,  $\mathbb{P}(U \supset S_0) \geq 1 - \alpha$ , where  $U$  is the closure of  $\lim_n U_n$ .

PROOF. For each  $n$ , let

$$I_n = \bigcup_{S_j \in \mathcal{S}_n: S_j \subset S_0} S_j$$

be the union of all sets in the  $n$ -th partition that belong entirely to  $S_0$ . The sequence of sets  $I_n$  is increasing and it has  $\lim_n I_n = \bigcup_n I_n$ . For each  $n$  we have that

$$1 - \alpha \leq \mathbb{P}(\mathcal{C}_n \ni I_n) \leq \mathbb{P}(U_n \supset I_n)$$

where  $U_n$  is a random sequence of sets, while  $I_n$  is not random. Hence

$$\mathbb{P}\left(\lim_n U_n \supset \lim_n I_n\right) \geq 1 - \alpha.$$

We are only left with showing that  $S_0^o \subset \lim_n I_n$ . For each  $s \in S_0^o$ , a neighborhood of  $s$  is contained in  $S_0$ . From Definition 1, for  $n$  large enough,  $S_{n,s} \subset S_0$ . Thus  $s \in S_{n,s} \subset \lim_n I_n$ . This holds for all  $s \in S_0^o$ , and it implies that  $S_0^o \subset \lim_n I_n$ .  $\square$

**Lemma 6** *Let  $c \in (0, \alpha)$  and let  $\beta = (\alpha - c)/(1 - c)$ . Let  $T$  be a  $(c, \beta)$  confidence threshold. Then  $T$  is a level  $\alpha$  FDR threshold, meaning that  $\mathbb{E}(\Gamma(T)) \leq \alpha$ .*

PROOF. We have that

$$\begin{aligned} \mathbb{E}(\Gamma(T)) &= \int_0^1 \mathbb{P}(\Gamma(T) > s) ds = \int_0^c \mathbb{P}(\Gamma(T) > s) ds + \int_c^1 \mathbb{P}(\Gamma(T) > s) ds \\ &\leq \int_0^c ds + \int_c^1 \mathbb{P}(\Gamma(T) > s) ds \leq \int_0^c ds + \int_c^1 \mathbb{P}(\Gamma(T) > c) ds \\ &\leq c + \int_c^1 \beta ds = c + (1 - c)\beta = \alpha. \end{aligned}$$

$\square$

**Theorem 2** *Let  $Y = -(X - \epsilon)$ . Let  $\bar{\Gamma}(t)$  denote the  $(1 - \alpha)$  upper envelope obtained by applying the procedure from Section 3 to the field  $Y$ . Let  $\bar{\Lambda} = \bar{\Gamma}$ . Then*

$$\inf_{\mathbb{P} \in \mathcal{P}_\epsilon} \mathbb{P}(\Lambda(t) \leq \bar{\Lambda}(t) \text{ for all } t) \geq 1 - \alpha.$$

PROOF. Let  $\nu(s) = E(Y(s))$ . Note that (20) is equivalent to

$$H_0 : \nu(s) \leq 0 \text{ for all } s \in A \quad \text{versus} \quad H_1 : \nu(s) \geq \epsilon \text{ for some } s \in A.$$

Now the max test we have been using in the previous sections, applied to  $Y$  over  $A$ , is a level  $\alpha$  for  $H_0$ . This is because  $\nu(s) = 0$  is the least favorable (highest level) value under  $H_0$ .  $\square$

## 11 DISCUSSION

We have provided methods for estimating the False Discovery Proportion process and for choosing rejection thresholds that control the FDP. There are several open problems under current study that we will report on in a future paper.

First, the detection problem for scan statistics involves two components: choosing a statistic and controlling errors. We have given a new method for the latter, though more work is needed for the former. For example, we chose smoothing bandwidths using cross-validation. A different approach is to choose the bandwidth to minimize the FNP. The careful choice of kernel and smoothing parameter remains largely unaddressed in the scan-statistic literature.

Second, in applications of our methods to Gaussian fields, we require an estimate of the covariance kernel. Studying the effect of replacing the true covariance with an estimate is beyond the scope of this paper, but we can make a general observation. If the estimate of  $\mathbb{C}(X(s), X(t))$  is biased towards the identity, then the p-values will be over-estimated and the procedure is conservative. This suggests shrinking empirical covariance estimates towards the identity.

Third, our results are based on the statistic  $\sup_{s \in A} X(s)$ . We are exploring other statistics and other types of fields besides Gaussian fields. We also are examining the benefits of using different statistics on different subsets.

## REFERENCES

- Adler, R.J. (1981). *The Geometry of Random Fields*, Wiley.
- Adler, R.J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics.
- Adler, R.J. (2000). On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, **10**, 1-74.
- Aguirre, G., Zarahn, E. and D'Esposito, M. (1998). A critique of the use of the KolmogorovSmirnov statistic for the analysis of BOLD fMRI data. *Magnetic Resonance in Medicine*, **39**, 500-505.

- Belmonte, M. and Yurgelun-Todd, D. (2001). Permutation testing made practical for functional magnetic resonance image analysis. *IEEE Transactions on Medical Imaging*, **20**, 243–248.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57**, 289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.
- Benjamini, Y. and Yekutieli, D. (2003). Hierarchical FDR testing of trees of hypotheses. *Technical report*, Department of Statistics and Operations Research, Tel Aviv University.
- Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe Hesketh, S., Taylor, E. and Brammer, M.J. (1999). Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, **18**, 32–42.
- Cao, J. and Worsley, K. (1999). The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, **9**, 1021–1057.
- Efron, B., Tibshirani, R., Storey, J.D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, **96**, 1151–1160.
- Finner, H. and Roters, M. (2002) Multiple hypotheses testing and expected number of type I errors. *The Annals of Statistics*, **30**, 220–238.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D. and Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging - A general linear approach. *Human Brain Mapping*, **2**, 189–210.
- Genovese, C. R. (2000). A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data (with discussion), *Journal of the American Statistical Association*, **95**, 691–703.
- Genovese, C. and Wasserman L. (2003). Exact control of the false discovery process. Manuscript in progress.

- Genovese, C. and Wasserman L. (2002). Operating characteristics and extensions of the FDR procedure. *Journal of the Royal Statistical Society, Series B*, **64**, 499–517.
- Genovese, C. and Wasserman L. (2001). A stochastic process approach to False discovery rates. Technical report, Department of Statistics, Carnegie Mellon University.
- Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*, Springer, New York.
- Müller, K., Lohmann, G., Bosch, V., and von Cramon, D. Y. (2001). On multivariate spectral analysis of fMRI time series. *Neuroimage* 14, 347–356.
- Nichols, T.E. and Holmes, A.P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**, 1–25.
- Piterbargh, V.I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, American Mathematical Society.
- Sarkar, S.K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *The Annals of Statistics*, **30**, 239–257.
- Shen, X. Huang, H. and Cressie N. (2002). Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association*, **97**, 1122–1140.
- Siegmund, D.O. and Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *The Annals of Statistics*, **23**, 608–639.
- Storey J.D. (2001). The positive False Discovery Rate: A Bayesian interpretation and the q-value. *Technical Report*, Department of Statistics, Stanford University.
- Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, **64**, 479–498.
- Storey J.D., Taylor J.E. and Siegmund D. (2003). A unified estimation approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, to appear.
- Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of  $\chi^2$ ,  $F$  and  $t$  fields. *Advances in Applied Probability*, **26**, 13–42.
- 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. *The Annals of Statistics*, **23**, 640–669.

- Worsley, K. J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., and Evans, A.C. (2002). A General Statistical Analysis for fMRI data. *NeuroImage*, **15**, 1–15.
- Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. and Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**, 58–73.

<b>Bubbles</b>	32 <sup>2</sup> squares	64 <sup>2</sup> squares	128 <sup>2</sup> squares	256 <sup>2</sup> squares
$T$	845.5	879.8	923.0	948.4
$FDP(T)$	0.017	0	0	0
$FNP(T)$	0.043	0.045	0.048	0.049

**Table 1: Results from one simulation**

	32 <sup>2</sup> pixels	64 <sup>2</sup> pixels	128 <sup>2</sup> pixels	256 <sup>2</sup> pixels
<b>Horseshoe</b>				
Coverage (FDP envelope)	0.928	0.947	0.949	0.950
Coverage (FDP Threshold)	0.962	0.973	0.978	0.982
Coverage (FNP envelope)	0.998	0.998	0.998	0.998
FNP of Threshold	0.034	0.038	0.040	0.042
<b>Bullets</b>				
Coverage (FDP envelope)	0.946	0.950	0.950	0.950
Coverage (FDP Threshold)	0.969	0.980	0.984	0.987
Coverage (FNP envelope)	0.998	0.998	0.998	0.998
FNP of Threshold	0.052	0.056	0.059	0.061
<b>Bubbles</b>				
Coverage (FDP envelope)	0.929	0.944	0.947	0.947
Coverage (FDP Threshold)	0.984	0.993	0.995	0.995
Coverage (FNP envelope)	0.999	0.999	0.999	0.999
FNP of Threshold	0.062	0.066	0.068	0.070
<b>Romper Room</b>				
Coverage (FDP envelope)	0.947	0.952	0.954	0.955
Coverage (FDP Threshold)	1.000	1.000	1.000	1.000
Coverage (FNP envelope)	0.991	0.999	1.000	1.000
FNP of Threshold	0.488	0.530	0.550	0.561

**Table 2: Coverage results from 1000 simulations**

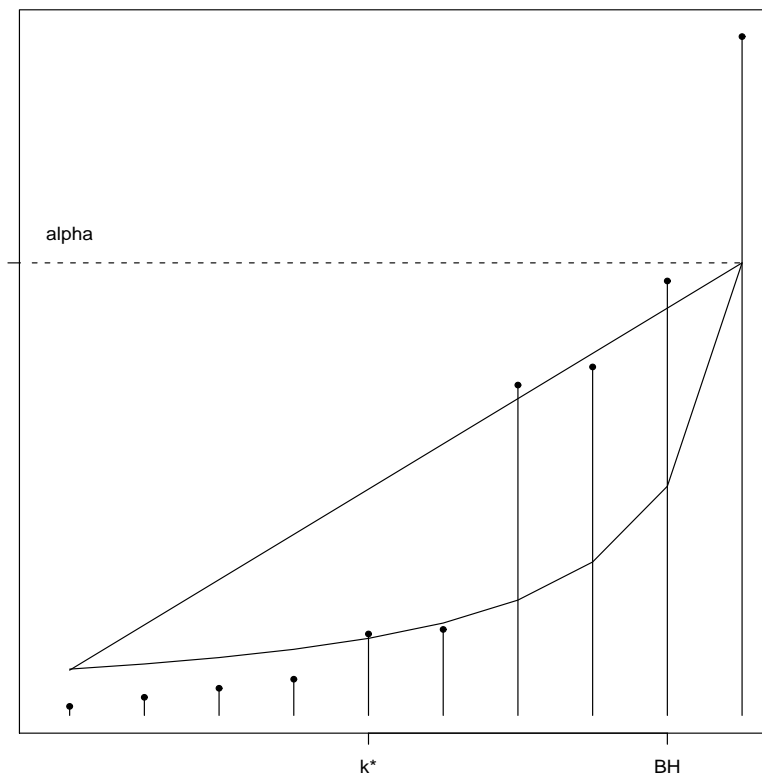


Figure 1: Comparison to BH.

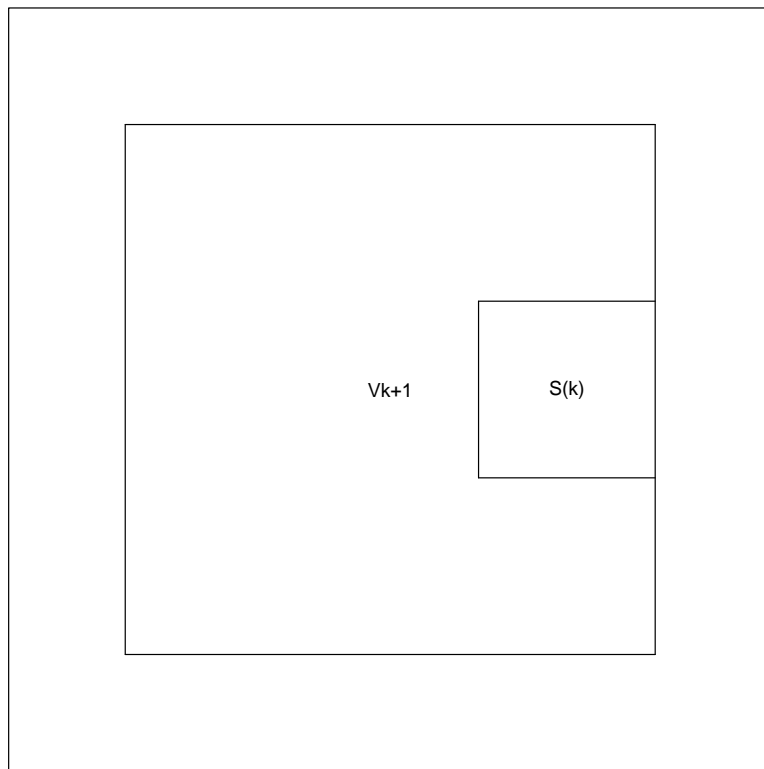


Figure 2: Nonmonotonicity

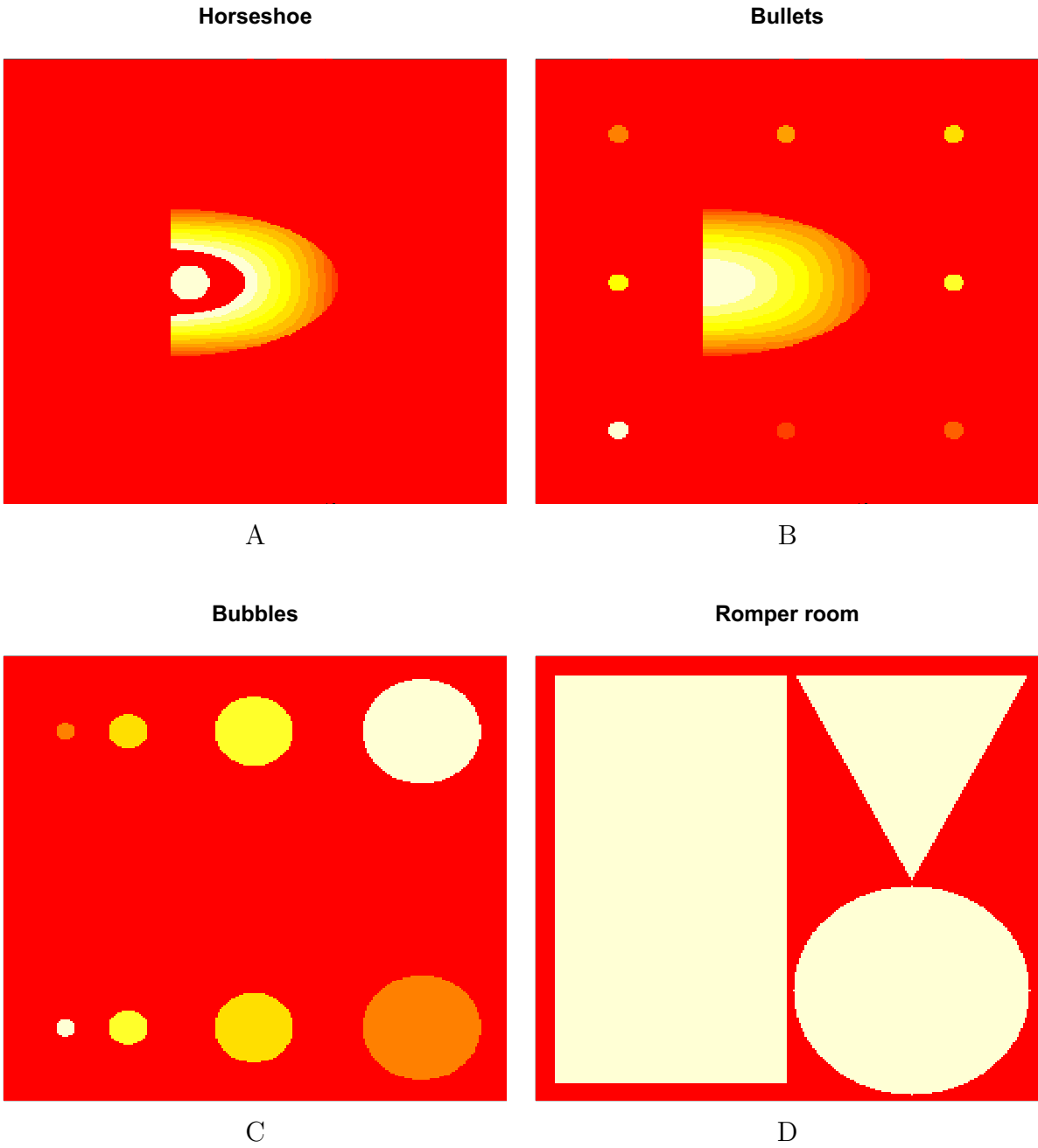


Figure 3: The four Test-signals. A. Horseshoe, B. Bubbles, C. Bullets, D. Romper Room.

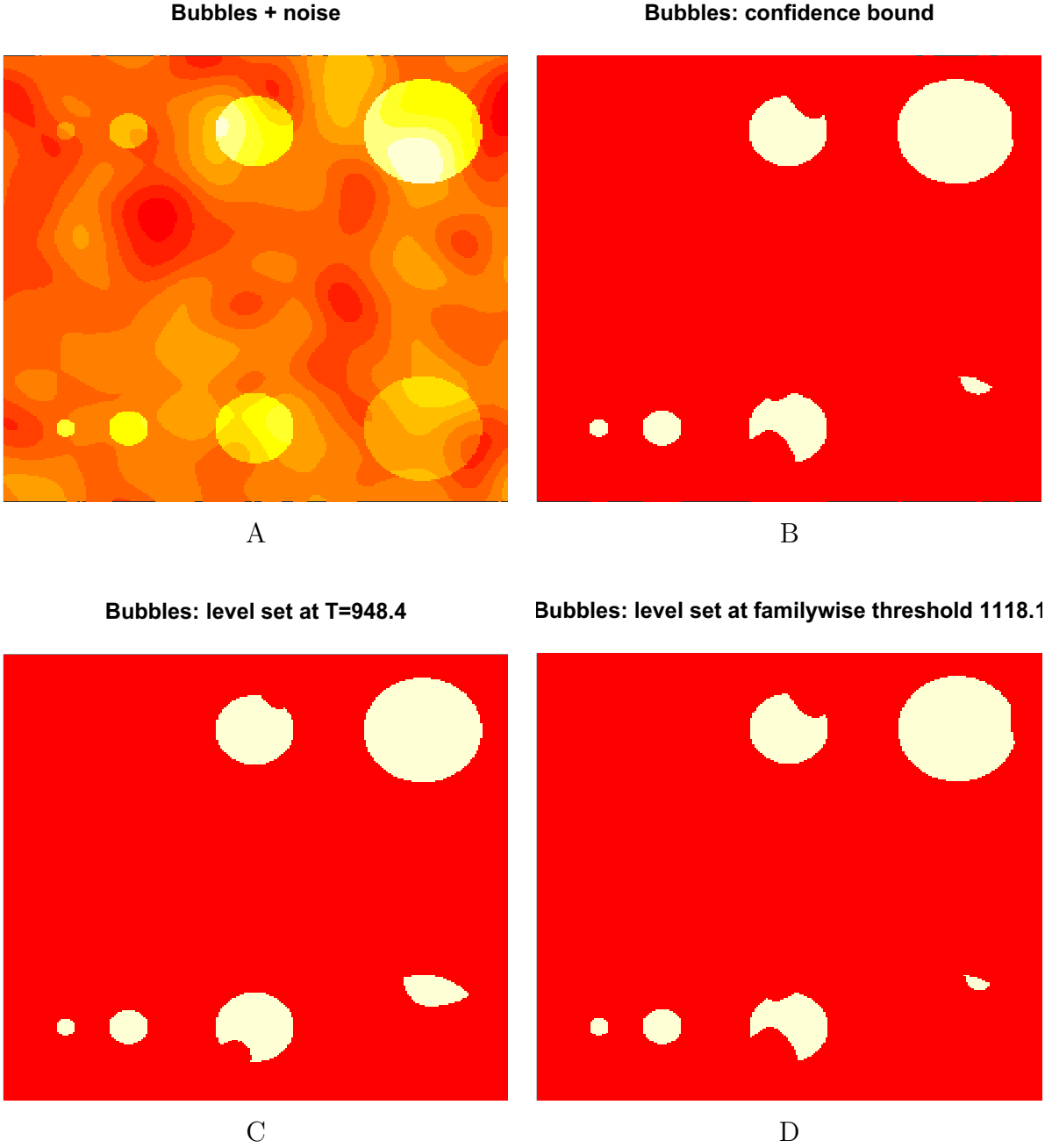


Figure 4: Bubbles example. A. Observed image, B. Confidence set  $U$ , C. Reconstructed image, D. Result from familywise approach.

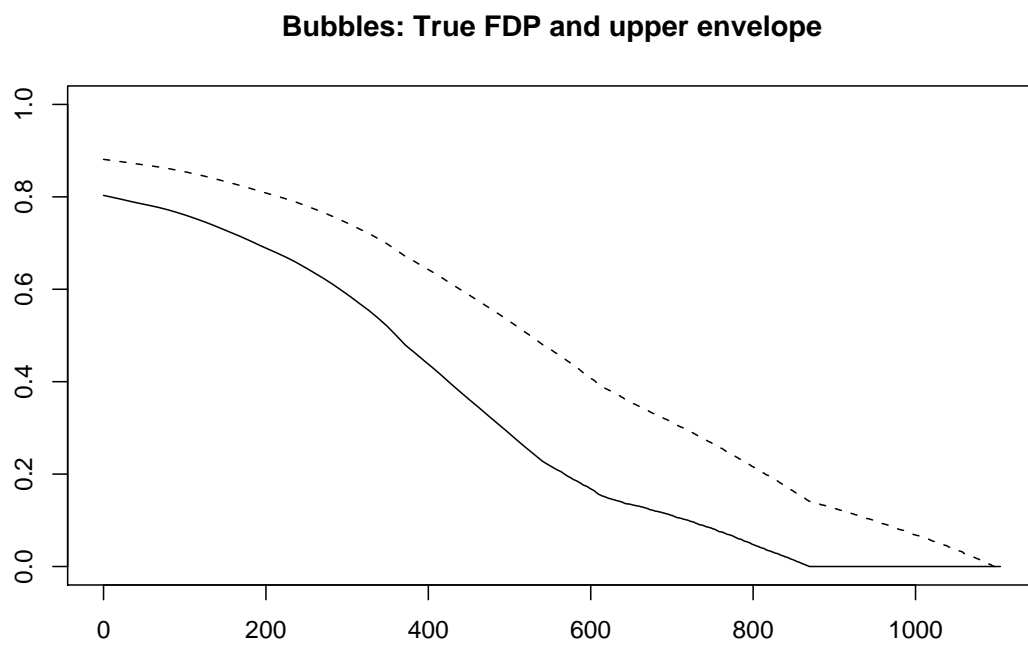


Figure 5: True FDP process (solid) and confidence envelope (dashed).

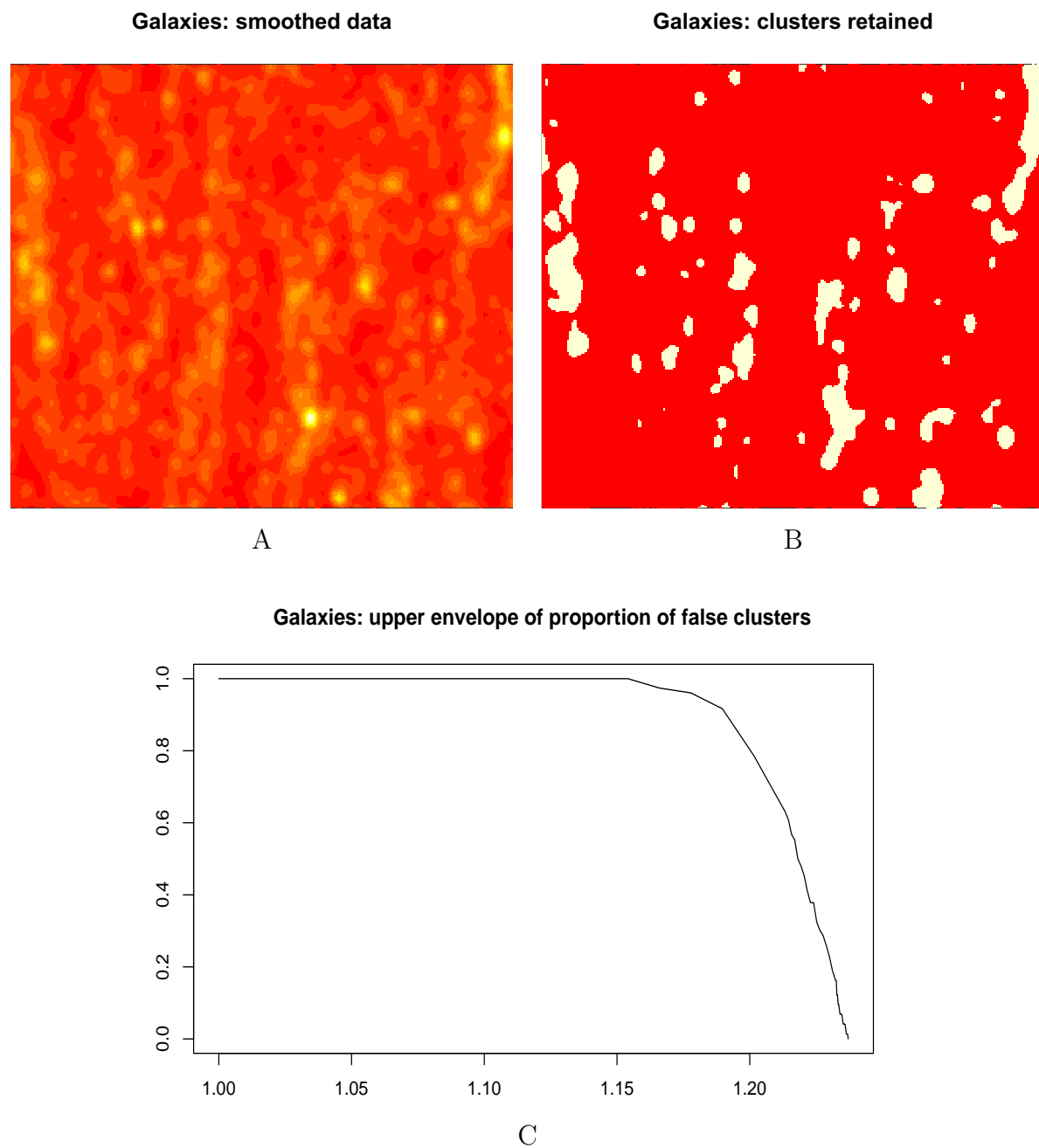


Figure 6: A. Smoothed Galaxy Field. B. Clusters Above Threshold. C. Upper envelope for False Cluster Proportion.

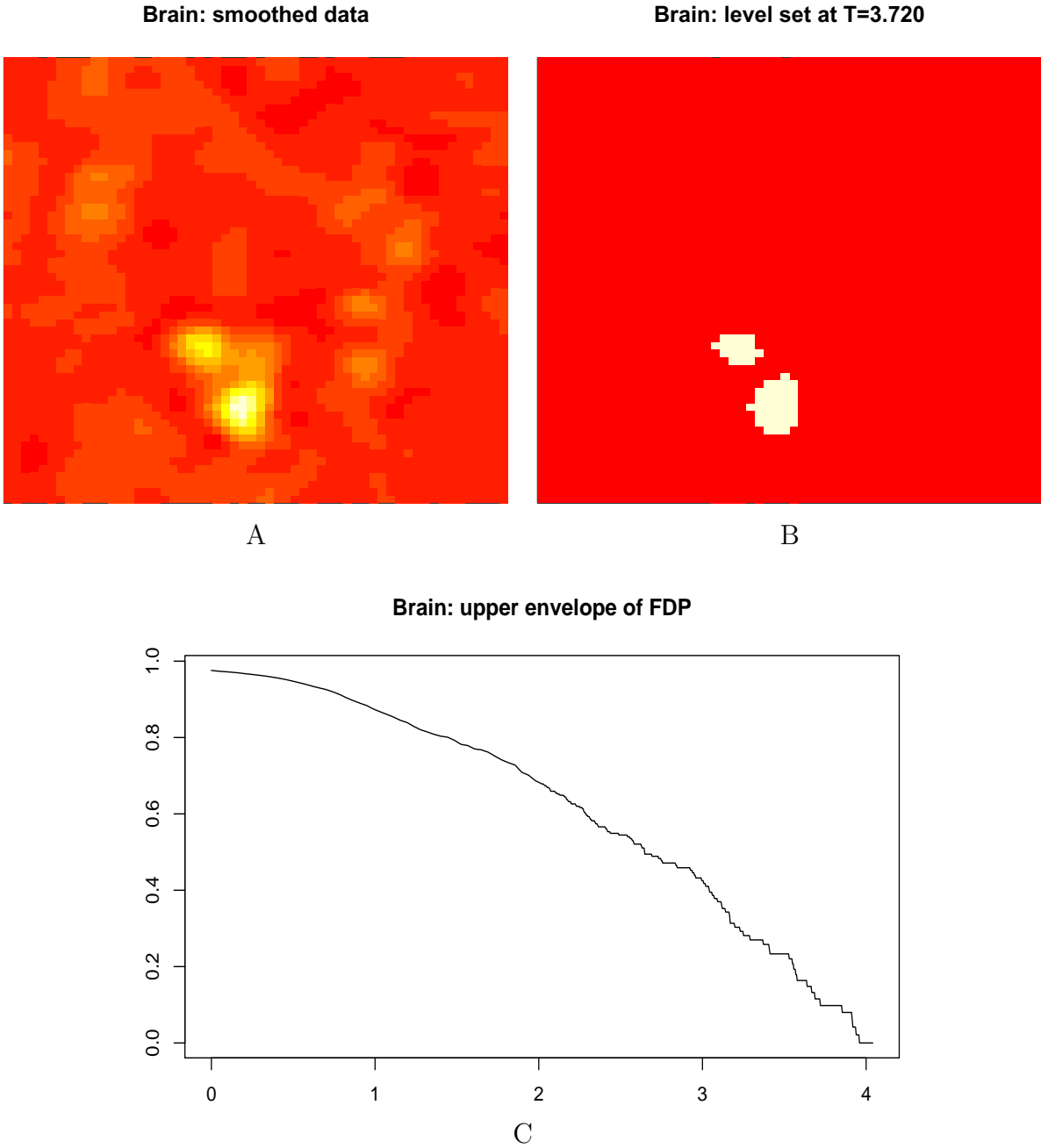


Figure 7: A. Smoothed image of brain. B. Pixels above threshold 3.72. C. FDP envelope for brain image.