Electronic Journal of Statistics
Vol. 10 (2016) 2641–2684
ISSN: 1935-7524
DOI: 10.1214/16-EJS1188

Exact asymptotics for the scan statistic and fast alternatives

James Sharpnack*

Department of Statistics
University of California
Davis, CA 95616, USA
e-mail: jsharpna@ucdavis.edu

and

Ery Arias-Castro*

Department of Mathematics
University of California
La Jolla, CA 92093, USA
e-mail: eariasca@ucsd.edu

Abstract: We consider the problem of detecting a rectangle of activation in a grid of sensors in d-dimensions with noisy measurements. This has applications to massive surveillance projects and anomaly detection in large datasets in which one detects anomalously high measurements over rectangular regions, or more generally, blobs. Recently, the asymptotic distribution of a multiscale scan statistic was established in [18] under the null hypothesis, using non-constant boundary crossing probabilities for locally-stationary Gaussian random fields derived in [8]. Using a similar approach, we derive the exact asymptotic level and power of four variants of the scan statistic: an oracle scan that knows the dimensions of the activation rectangle; the multiscale scan statistic just mentioned; the adaptive variant; and an ϵ-net approximation to the latter, in the spirit of [3]. This approximate scan runs in time near-linear in the size of the grid and achieves the same asymptotic level and power as the adaptive scan, and has a polylogarithmic time parallel implementation. We complement our theory with some numerical experiments, and make some practical recommendations.

MSC 2010 subject classifications: 62G10, 62M40, 60G32.
Keywords and phrases: Sensor networks, image processing, multiscale detection, scan statistic, suprema of Gaussian random fields.

Received December 2015.

1. Introduction

Detecting anomalies in networks is important in a number of areas, such as sensor arrays [6, 9], digital images (incl. satellite, medical, etc.) [7, 29, 23, 16, 24], syndromic surveillance systems [31, 15, 33], and many more. The scan statistic

*Work partially supported by NSF grant DMS-1223137.

[19] is by far the most popular approach, and is given different names in engineering, such as the method of matched filters or deformable templates [23]. It was perhaps first introduced for finding patterns in point clouds [25, 12] and is now applied to any setting where the goal is to detect a "localized" anomaly. In statistics, it corresponds to the generalized likelihood ratio test after a particular model is assumed, and as such is even more widely applicable, being the most omnibus approach to hypothesis testing.

In its simplest version, the scan is done over a window of fixed size. In practice, however, the size of the anomaly may not be known, in which case it makes sense to scan over windows of different sizes. This was considered in a number of papers. In particular, some approximations to the distribution of the resulting scan statistic is provided in [26, 28] in dimension 1, and in [13, 35, 14] in dimension 2. First order theoretical performance bounds are established in a small number of papers, such as [34, 10] in dimension 2 and [3, 2] in arbitrary dimension. More refined results establishing weak convergence are even fewer. [17] considers the scan over rectangles in a grid of independent random variables with negative expectation, while [5] study the scan over intervals of given length in a Bernoulli sequence. Both works rely heavily on the Chen-Stein Poisson approximation. Still in the context of the one-dimensional lattice, but now with standard normal random variables, [32] provide a weak convergence for the normalized scan over all intervals. Concretely, suppose that $y(1), \dots, y(n)$ are iid standard normal, and define

$$Z_n = \max_{1 \le i_1 \le i_2 \le n} \frac{1}{\sqrt{i_2 - i_1 + 1}} \sum_{i=i_1}^{i_2} y(i).$$

Then [32] show that, for all $\tau \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\big(Z_n \ge u_n(\tau)\big) = 1 - e^{-e^{-\tau}}, \quad u_n(\tau) := \sqrt{2\log n} + \frac{\frac{1}{2}\log(2\log n) + \kappa + \tau}{\sqrt{2\log n}},$$

for some numeric constant κ. This was recently extended to higher dimensions, for scans over hypercubes and hyperrectangles, by [18]. Formally, define $[n] = \{1, \dots, n\}$ and assume that $(y(\mathbf{i}) : \mathbf{i} \in [n]^d)$ are iid standard normal. A (discrete) hyperrectangle is of the form $[a_1, b_1] \times \cdots \times [a_d, b_d] \subset [n]^d$. Let $\mathcal{R}$ denote the class of all discrete hyperrectangles of $[n]^d$ and define the scan over $\mathcal{R}$ as

$$Z_n = \max_{R \in \mathcal{R}} \frac{1}{\sqrt{|R|}} \sum_{\mathbf{i} \in R} y(\mathbf{i}), \tag{1.1}$$

where $|R|$ denotes the number of nodes in R, equal to $\prod_j (b_j - a_j + 1)$ when $R = \times_j [a_j, b_j]$. [18] shows that, for all $\tau \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\big(Z_n \ge u_n(\tau)\big) = 1 - e^{-e^{-\tau}} \tag{1.2}$$

for the threshold function,

$$u_n(\tau) := \sqrt{2d\log n} + \frac{(d - \frac{1}{2})\log(2d\log n) + \kappa + \tau}{\sqrt{2d\log n}}, \tag{1.3}$$

for some constant κ depending only on the dimension d. These results allow, in theory, to control the (asymptotic) level of the test based on the scan statistic if, indeed, the data is iid standard normal when no anomaly is present and an anomaly comes in the form of a rectangle with elevated mean. This is what we assume throughout the paper.

> *Contribution 1.* We establish a weak convergence result when an anomaly is present which, in theory, allows for precise (asymptotic) power calculations.

Besides the scan statistic (1.1), we study other variants. One of them, already considered in [2, 34], is based on a finer normalization for the scans at different scales. The tests at different scales are combined with a Bonferroni correction. Similar multiple testing approaches are also proposed in [13, 35]. In detail, define the class of rectangles with shape $\mathbf{h} \in [n]^d$ as

$$\mathcal{R}(\mathbf{h}) = \left\{ \times_{j=1}^{d} [v_j, v_j + h_j] : v_j \in [n - h_j], \forall j \in [d] \right\}, \tag{1.4}$$

and let $Z_{n,\mathbf{h}}$ denote the scan over $\mathcal{R}(\mathbf{h})$, defined as in (1.1) but with $\mathcal{R}(\mathbf{h})$ in place of $\mathcal{R}$. We consider the test that rejects if there is $\mathbf{h}$ such that $Z_{n,\mathbf{h}} \geq u_{n,\mathbf{h}}(\tau)$, for some explicit critical values $u_{n,\mathbf{h}}(\tau)$ defined later. We refer to this procedure as the (scale or shape) adaptive scan. We note that in the first order analyses found in [2, 34], $u_{n,\mathbf{h}}(\tau)$ only depends on $\prod_{j \in [d]} h_j$, which is not quite true in our situation.

> *Contribution 2.* We establish weak convergence results for the adaptive scan, both when an anomaly is absent and when it is present.

Both the scan and the adaptive scan are computationally intensive. With proper implementation, they can be computed in $O(n^{2d})$ basic operations, which may nevertheless be prohibitive for scans over large networks. For example, a typical 2D digital image is of size $n \times n$, where n is in the order of 10^3, resulting a computational complexity on of order 10^{12} basic operations. Aware of that, [3, 2] and [34] propose to approximate the scan statistic by scanning over a subset of rectangles that is sufficiently dense in $\mathcal{R}$. For a given metric δ over $\mathcal{R}$, we say that $\mathcal{R}_\epsilon$ is an ϵ-covering if, for all $R \in \mathcal{R}$, there is $R' \in \mathcal{R}_\epsilon$ such that $\delta(R, R') \leq \epsilon$. Both [3, 2] and [34] construct different ϵ-coverings which can be scanned in roughly $O(n^d)$ basic operations, and show that, when $\epsilon = \epsilon_n \to 0$ sufficiently slowly, scanning over an ϵ-covering yields the same first-order asymptotic performance.

> *Contribution 3.* We establish weak convergence results for the adaptive scan over a given ϵ-covering, both when an anomaly is absent and when it is present. We also construct a new ϵ-covering and design an efficient way to scan over it using on the order of $O(n^d)$ basic operations when ϵ is not too small, and a poly-logarithmic parallel implementation.

As a benchmark we consider an oracle which knows the shape $\mathbf{h}^\star$ of the anomalous rectangle (but is ignorant of its location) and therefore only scans over rectangles with the same shape, meaning, over $\mathcal{R}(\mathbf{h}^\star)$.

Contribution 4. We establish weak convergence results for the oracle scan, both when an anomaly is absent and when it is present.

We note that our method of proof is largely borrowed from [18], whose approach is based on extensive work of [8] on the extrema of Gaussian random fields, and related topics, and the Chen-Stein Poisson approximation [4].

We complement our theoretical findings with some numerical experiments that we performed to compare these various methods, meaning, the oracle scan, the scan, the adaptive scan, and the adaptive scan over an ϵ-covering.

The rest of the paper is organized as follows. In Section 2, we set the framework and state the theoretical results announced above, and in Section 3, we present the result of our numerical experiments. We briefly discuss some extensions and open problems in Section 4, while the technical proofs are gathered in Section 5.

Before we continue, we pause to introduce some notation. We already used the notation $[n] = \{1, \dots, n\}$ for any positive integer n. Cartesian products of sets are denoted with the $\times$ operator and for a set A and integer $k \geq 1$, $A^k = A \times \cdots \times A$, k times. All vectors are bolded and scalars are not. Some special vectors are $\mathbf{0} = (0, \dots, 0)$, $\mathbf{1} = (1, \dots, 1)$, and the jth canonical basis vector $\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$ with the 1 in the jth component. A vector $\mathbf{a}$ in dimension d will have components denoted $a_1, \dots, a_d$. The rectangle with endpoints $\mathbf{a}, \mathbf{b}$ will be denoted $[\mathbf{a}, \mathbf{b}] = \times_{i=1}^d [a_i, b_i]$. The symbol $\circ$ indicates the component-wise product for vectors and matrices, division between vectors denoted $\mathbf{a}/\mathbf{b}$ is component-wise. The Lebesgue measure in $\mathbb{R}^d$ will be denoted by λ. For a discrete set R, $|R|$ denotes its cardinality. For a vector $\mathbf{a}$, $\|\mathbf{a}\|$ and $\|\mathbf{a}\|_1$ denote its Euclidean and ℓ_1 norms, respectively. For a set A, $I\{A\}$ (sometimes $\mathbb{1}_A$) will denote the indicator of A. We use the Bachmann-Landau notation to compare infinite sequences. For example, if $\{a_n\}_{n=1}^\infty, \{b_n\}_{n=1}^\infty$ are such that $a_n/b_n \to 0$ then $a_n = o(b_n)$ and $b_n = \omega(a_n)$. For stochastic sequences, if the convergence is in probability then this is denoted by a subscript as in $a_n = o_{\mathbb{P}}(b_n)$.

2. Model, methodology, and theory

We assume that we are given one snapshot of measurements from a sensor array in d-dimensional space. This array is arranged by placing one sensor at each grid point in $[n]^d$. An important example is that of digital images, from CCD or CMOS cameras, or other modalities such as MRI. It also encompasses video (incl. fMRI), by letting one dimension represent time (in some unit), although the time dimension is often treated in a special way.

We denote the measurement at sensor $\mathbf{i} \in [n]^d$ by $y(\mathbf{i})$ and model this as a signal vector with additive white Gaussian noise,

$$y(\mathbf{i}) = x(\mathbf{i}) + \xi(\mathbf{i}), \quad \mathbf{i} \in [n]^d, \tag{2.1}$$

where x is the signal and ξ is white standard normal noise, or in vector notation,

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\xi},$$

where $\mathbf{y}, \mathbf{x} \in \mathbb{R}^{n^d}$ and $\boldsymbol{\xi}$ is a standard normal vector in d dimensions. We address the problem of deciding whether the signal x is nonzero, formalized as the following hypothesis testing problem:

$$\begin{aligned} &H_0 : \mathbf{x} = \mathbf{0}, \\ &H_1 : \mathbf{x} \in \mathcal{X}_\mu, \end{aligned} \tag{2.2}$$

for some parameter μ, that will be interpreted as the signal size, and some class $\mathcal{X}_\mu \subset \mathbb{R}^{n^d}$ parametrized by μ and with the property $\mathbf{0} \notin \mathcal{X}_\mu$. While H_0 represents 'business as usual', H_1 would indicate that there is some anomalous activity, here modeled by x.

We address the situation where the signal has substantial 'energy' over a rectangle of unknown shape. Given a signal $\mathbf{y} = (y(\mathbf{i}) : \mathbf{i} \in [n]^d)$, define its Z-score over a subset $R \subset [n]^d$ as

$$y[R] = \frac{1}{\sqrt{|R|}} \sum_{\mathbf{i} \in R} y(\mathbf{i}). \tag{2.3}$$

Recalling the class $\mathcal{R}(\mathbf{h})$ of rectangles of with shape $\mathbf{h}$ defined in (1.4), let $\mathcal{X}_\mu(\mathbf{h})$ denote the following set of signals:

$$\mathcal{X}_\mu(\mathbf{h}) = \Big\{ \mathbf{x} \in \mathbb{R}^{n^d} : \exists R \in \mathcal{R}(\mathbf{h}) \text{ s.t. } x[R] \geq \mu \;\&\; x(\mathbf{i}) = 0, \forall \mathbf{i} \in [n]^d \backslash R \Big\}. \tag{2.4}$$

Rectangles are useful in practice because of their ease of interpretation and implementation, and also because they are building blocks for more complicated shapes. They are also more amenable to a sharp asymptotic analysis, which is the focus in this paper. See the discussion in Section 4.

Let $\mathbf{h}^\star$ denote the shape of rectangle of activation, defined as the shape $\mathbf{h}$ such that $\operatorname{supp}(\mathbf{x}) \in \mathcal{R}(\mathbf{h})$. For the sake of clarity and ease of analysis, we assume that we are given integers $1 \leq \underline{h} \leq \overline{h} \leq n/e$ (where $e = \exp(1)$) such that $\mathbf{h}^\star \in [\underline{h}, \overline{h}]^d$. Redefine $\mathcal{R}$ as

$$\mathcal{R} = \bigcup_{\mathbf{h} \in [\underline{h}, \overline{h}]^d} \mathcal{R}(\mathbf{h}). \tag{2.5}$$

We know that, under the alternative, the signal is elevated over a rectangle in $\mathcal{R}$, namely

$$\mathbf{x} \in \mathcal{X}_\mu := \bigcup_{\mathbf{h} \in [\underline{h}, \overline{h}]^d} \mathcal{X}_\mu(\mathbf{h}).$$

Our analysis is asymptotic with respect to the grid size diverging to infinity, $n \to \infty$. While the grid dimension d remains fixed, μ, $\overline{h}$, and $\underline{h}$ are allowed to depend on n. In fact, throughout this paper we assume that

$$\underline{h} = \underline{h}_n \text{ satisfies } \underline{h}/\log n \to \infty, \text{ as } n \to \infty, \tag{2.6}$$

to avoid special cases and complications that arise when including very small rectangles in the scan.

As mentioned in the Introduction, we will use the oracle scan as a benchmark. Instead of (2.2), the oracle, which knows the shape $\mathbf{h}^\star$, is faced with the simpler alternative:

$$H_1^\star : \mathbf{x} \in \mathcal{X}_\mu(\mathbf{h}^\star). \tag{2.7}$$

We take the asymptotic Neyman-Pearson approach in which we control the asymptotic probability of type I error (aka false rejection). Consider a test $T(\mathbf{y})$ which evaluates to 1 if it rejects H_0 and 0 otherwise. Throughout, we assume that a level $\alpha \in (0, 1)$ is given and we control the tests at the exact asymptotic level α, which means that

$$\lim_{n\to\infty} \mathbb{P}_0\{T(\mathbf{y}) = 1\} = \alpha,$$

where $\mathbb{P}_0$ indicates the distribution of $\mathbf{y}$ under H_0. The left-hand side is called the asymptotic size of the test T. For all of the test statistics that we will study, we provide a threshold that gives us such a type I error control. Once the size of the test is under control, we examine the power of the test against each alternative.

2.1. The Oracle scan

When tasked with finding a rectangle of activation in a d-dimensional lattice, the problem is made easier if one knows the precise shape of the active rectangle. Having access to an oracle that provides the shape of the anomalous region simplifies the alternative down to (2.7). In this situation, one would naturally restrict the scan to rectangles with shape $\mathbf{h}^\star$. We called this procedure the *oracle scan* in the Introduction. Given a critical value u, the oracle scan test is defined as

$$T_o(\mathbf{y}) = I\big\{y[R] > u \text{ for some } R \in \mathcal{R}(\mathbf{h}^\star)\big\}. \tag{2.8}$$

2.1.1. Asymptotic theory

Define the following critical value

$$u_n(\tau) = v_n + \frac{(2d-1)\log(v_n) + \kappa + \tau}{v_n}, \tag{2.9}$$

where

$$v_n = \sqrt{2 \textstyle\sum_j \log(n/h_j^\star)}, \quad \kappa = -\log(\sqrt{2\pi}). \tag{2.10}$$

Given a level $\alpha \in (0, 1)$, we choose

$$\tau = \tau_\alpha = -\log(-\log(1-\alpha)). \tag{2.11}$$

Theorem 1. *Suppose that* $\min_i h_i^\star = \omega(\log n)$. *The oracle scan test* (2.8) *with critical value* (2.9) *and* τ *chosen as in* (2.11), *has the following asymptotic size*

$$\lim_{n\to\infty} \mathbb{P}_0\{T_o(\mathbf{y}) = 1\} = 1 - e^{-e^{-\tau}} = \alpha.$$

Let $\bar{\Phi}$ denote the survival function of the standard normal distribution.

Theorem 2. *Suppose that* $\min_j h_j^\star = \omega(\log n)$. *The oracle scan test* (2.8) *with critical value* (2.9) *(with* (2.11)*) has the following asymptotic power*

$$\lim_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_\mu(\mathbf{h}^\star)} \mathbb{P}_{\mathbf{x}}\{T_o(\mathbf{y}) = 1\} = \begin{cases} 1, & \mu - v_n \to \infty, \\ \alpha + (1-\alpha)\bar{\Phi}(c), & \mu - v_n \to c, \ \textit{for } c \in \mathbb{R}, \\ \alpha, & \textit{otherwise}, \end{cases}$$

where v_n *is defined in* (2.10).

2.1.2. Computational complexity

While a naive implementation runs in $O(n^d \prod_j h_j^\star)$ time, the oracle scan can be computed in $O(n^d \log n)$ time using the Fast Fourier Transform (FFT), which is generally faster when the $h_j^\star$'s are not too small. Specifically, let $b_{\mathbf{h}}$ be the boxcar function with shape $\mathbf{h}$, namely

$$b_{\mathbf{h}}(\mathbf{i}) = \prod_{j=1}^d I\{i_j \le h_j\},$$

and let $*$ denote the convolution operator, so that, for $f : [n]^d \mapsto \mathbb{R}$,

$$(f * b_{\mathbf{h}})(\mathbf{t}) = \sum_{\mathbf{i}\in[n]^d} f(\mathbf{t}+\mathbf{i}) b_{\mathbf{h}}(\mathbf{i}) = \sum_{\mathbf{i}\in[\mathbf{h}]} f(\mathbf{i}+\mathbf{t}).$$

Thus, computing the convolution $y * b_{\mathbf{h}}$ amounts to computing $(y[R] : R \in \mathcal{R}(\mathbf{h}))$, and using the FFT, this convolution can be computed in $O(n^d \log n)$ time. And the oracle scan test is based on the maximum of $y * b_{\mathbf{h}^\star}$.

2.2. The multiscale scan

Perfect knowledge of the shape of the true rectangle of activation is rare. A simple solution to this problem is to scan over all rectangles in the class $\mathcal{R}$ and report the largest observed Z-score. Formally, given a critical value u, the *multiscale scan test* is

$$T_m(\mathbf{y}) = I\big\{y[R] > u \text{ for some } R \in \mathcal{R}\big\}. \tag{2.12}$$

This is the test based on the scan statistic as defined in (1.1), except that $\mathcal{R}$ is now defined as in (2.5).

2.2.1. Asymptotic theory

Define the following critical value

$$u_n(\tau) = v_n + \frac{(4d-1)\log(v_n) + \kappa + \tau}{v_n}, \tag{2.13}$$

where

$$v_n = \sqrt{2d\log(n/\underline{h})}, \quad \kappa = -\log(4^d\sqrt{2\pi}). \tag{2.14}$$

[18, Th 1.2] establishes the asymptotic size of the multiscale scan test when $\underline{h} = 1$ and $\overline{h} = n$. We do the same, when $\underline{h} = \omega(\log n)$ and $\overline{h} \le n/e$. We note that, because of that, the critical value that we use, meaning (2.13), is different from the one that [18] uses, meaning (1.2). The constants denoted by κ in both places are in fact different, and the $(4d-1)$ factor in (2.13) is a $(2d-1)$ factor in (1.2).

Theorem 3. *[18] Suppose that $\underline{h} = \omega(\log n)$. The multiscale scan test (2.12) with critical value (2.13) (with (2.11)) has the following asymptotic size*

$$\lim_{n\to\infty} \mathbb{P}_0\{T_m(\mathbf{y}) = 1\} = 1 - e^{-e^{-\tau}} = \alpha.$$

Theorem 4. *Suppose that $\min_i h_i^\star = \omega(\log n)$. The multiscale scan test (2.12) with critical value (2.13) (with (2.11)) has the following asymptotic power*

$$\lim_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_\mu(\mathbf{h}^\star)} \mathbb{P}_{\mathbf{x}}\{T_m(\mathbf{y}) = 1\} = \begin{cases} 1, & \mu - v_n \to \infty, \\ \alpha + (1-\alpha)\bar{\Phi}(c), & \mu - v_n \to c, \text{ for } c \in \mathbb{R}, \\ \alpha, & \text{otherwise}, \end{cases}$$

where v_n is defined in (2.14).

Compared with the oracle scan test (see Theorem 2), the multiscale scan test (at the same level) has strictly less asymptotic power in general. For example, suppose that $\underline{h} \asymp n^a$ and $h_j^\star \asymp n^b$ for all j, for some fixed $0 < a < b < 1$. In that case, to have power tending to one, the oracle scan requires $\mu - \sqrt{1-b}\sqrt{2d\log n} \to \infty$, while the multiscale scan requires $\mu - \sqrt{1-a}\sqrt{2d\log n} \to \infty$.

2.2.2. *Computational complexity*

Using the FFT, the multiscale scan statistic can be computed in $O\big((n^{2d}/\underline{h}^d)\log n\big)$ time, since each shape can be scanned in $O\big(n^d \log n\big)$ as we saw in Section 2.1.2, and there are $O(n^d/\underline{h}^d)$ shapes in total in $\mathcal{R}$.

2.3. The adaptive multiscale scan

While the multiscale scan uses the same threshold u_m for all rectangle sizes, it ignores the fact that detecting small rectangles (at the finer scales) is more difficult than detecting large rectangles. The approach advocated in [34, 2] is a refinement of the multiscale scan in that a different threshold is used at each scale (i.e., rectangle size). See also [13, 35]. Formally, given (possibly) shape-dependent critical values $u_{\mathbf{h}}$, the *adaptive multiscale scan test* is

$$T_a(\mathbf{y}) = I\big\{y[R] > u_{\mathbf{h}}, \text{ for some } \mathbf{h} \in [\underline{h}, \overline{h}]^d \text{ and } R \in \mathcal{R}(\mathbf{h})\big\}. \tag{2.15}$$

If in fact u does not depend on $\mathbf{h}$, then this is the multiscale scan test (2.12).

2.3.1. Asymptotic theory

Define the following shape-dependent critical value

$$u_{n,\mathbf{h}}(\tau) = v_{n,\mathbf{h}} + \frac{(4d-1)\log(v_{n,\mathbf{h}}) + \kappa + \tau}{v_{n,\mathbf{h}}}, \tag{2.16}$$

where

$$v_{n,\mathbf{h}} = \sqrt{2\sum_j \log\left[\tfrac{n}{h_j}\left(1+\log\tfrac{h_j}{\underline{h}}\right)^2\right]}, \quad \kappa = -\log(4^d\sqrt{2\pi}). \tag{2.17}$$

Theorem 5. *Suppose that* $\underline{h} = \omega(\log n)$. *The adaptive multiscale scan test* (2.15) *with critical value* (2.16) *(with* (2.11)*) has the following asymptotic size*

$$\lim_{n\to\infty} \mathbb{P}_0\{T_a(\mathbf{y}) = 1\} = 1 - e^{-e^{-\tau}} = \alpha.$$

Theorem 6. *Suppose that* $\min_j h_j^\star = \omega(\log n)$. *The adaptive multiscale scan test* (2.12) *with critical value* (2.13) *(with* (2.11)*) has the following asymptotic power*

$$\lim_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_\mu(\mathbf{h}^\star)} \mathbb{P}_{\mathbf{x}}\{T_a(\mathbf{y}) = 1\} = \begin{cases} 1, & \mu - v_{n,\mathbf{h}^\star} \to \infty, \\ \alpha + (1-\alpha)\bar{\Phi}(c), & \mu - v_{n,\mathbf{h}^\star} \to c, \ \textit{for } c \in \mathbb{R}, \\ \alpha, & \textit{otherwise}, \end{cases}$$

where $v_{n,\mathbf{h}}$ *is defined in* (2.17).

The adaptive multiscale scan test (at the same level) happens to achieve the same asymptotic power as the oracle scan (see Theorem 2) in the important case where $\mathbf{h}^\star$ is not too large. Indeed, suppose for example that $\min_j h_j^\star = O(n^b)$ for some fixed $0 < b < 1$. Letting $v_n^\star$ denote the v_n in (2.10), we obviously have $v_{n,\mathbf{h}^\star} \geq v_n^\star$, and also

$$\begin{aligned} v_{n,\mathbf{h}^\star} &\leq v_n^\star\sqrt{1 + (v_n^\star)^{-2} d\log\log n} \leq v_n^\star\left[1 + \tfrac{1}{2}(v_n^\star)^{-2} d\log\log n\right] \\ &= v_n^\star + O\Big(\frac{\log\log n}{\sqrt{\log n}}\Big) = v_n^\star + o(1), \end{aligned}$$

so that $v_{n,\mathbf{h}^\star} = v_n^\star + o(1)$.

2.3.2. Computational complexity

The computational cost for computing the adaptive multiscale scan is the same as that for computing the multiscale scan, i.e., $O\big((n^{2d}/\underline{h}^d)\log n\big)$ time.

2.4. Approximate adaptive multiscale scan

The computational complexity of the adaptive multiscale scan, which is quadratic in the grid size, may be prohibitive in some situations. We provide now an algorithm that has nearly linear computation time while achieving the same asymptotic power. Inspired by the multiscale approximation developed in [3, 2, 34], we accomplish this by effectively scanning only over a subset of the rectangles that form an ϵ-covering for $\mathcal{R}$. We recall that, given a metric δ over $\mathcal{R}$, $\mathcal{R}_\epsilon \subset \mathcal{R}$ is an ϵ-covering of $\mathcal{R}$ for δ if, for all $R \in \mathcal{R}$, there is $R' \in \mathcal{R}_\epsilon$ such that $\delta(R, R') \leq \epsilon$. Recall the definition of ξ in (2.1). We use the canonical metric for the Gaussian random field $\{\xi[R], R \in \mathcal{R}\}$, which is given by

$$\delta^2(R_0, R_1) = \mathbb{E}(\xi[R_0] - \xi[R_1])^2 = 2\left(1 - \frac{|R_0 \cap R_1|}{\sqrt{|R_0||R_1|}}\right), \quad \forall R_0, R_1 \in \mathcal{R}. \tag{2.18}$$

Given an ϵ-covering $\mathcal{R}_\epsilon$ and (possibly) shape-dependent critical values $u_{\mathbf{h}}$, the *ϵ-adaptive multiscale scan test* is

$$T_\epsilon(\mathbf{y}) = I\big\{y[R] > u_{\mathbf{h}},\ \text{for some } \mathbf{h} \in [\underline{h}, \overline{h}]^d \text{ and } R \in \mathcal{R}(\mathbf{h}) \cap \mathcal{R}_\epsilon\big\}. \tag{2.19}$$

2.4.1. Asymptotic theory

Ideally, we would like to select ϵ small enough (in fact, decreasing with n) that the ϵ-adaptive multiscale scan statistic has asymptotically the same distribution as the (full) adaptive multiscale scan statistic. As it turns out, it is sufficient to select ϵ^{-1} on the order of $\sqrt{\log n}$ for this to occur. We will find that with this choice of ϵ it is possible to construct an algorithm that can perform an ϵ-covering scan in near-linear time.

Consider critical values of the form (2.16) (with (2.11)) and define the following P-value

$$\hat{\alpha}_{n,\mathbf{h}}(z) = \inf\{\alpha \in (0,1) : z \geq u_{n,\mathbf{h}}(\tau_\alpha)\}. \tag{2.20}$$

Then the P-value associated with the adaptive multiscale scan test is

$$\hat{\alpha}_n = \min\Big\{\hat{\alpha}_{n,\mathbf{h}}(y[R]) : \mathbf{h} \in [\underline{h}, \overline{h}]^d, R \in \mathcal{R}(\mathbf{h})\Big\}. \tag{2.21}$$

Analogously, the P-value associated with the ϵ-adaptive multiscale scan test is

$$\hat{\alpha}_{n,\epsilon} = \min\Big\{\hat{\alpha}_{n,\mathbf{h}}(y[R]) : \mathbf{h} \in [\underline{h}, \overline{h}]^d, R \in \mathcal{R}(\mathbf{h}) \cap \mathcal{R}_\epsilon\Big\}. \tag{2.22}$$

Theorem 7. *Consider the P-value for the multiscale scan or the adaptive multiscale scan, and ϵ-covering analog, defined in (2.21)–(2.22) respectively. Assuming $\epsilon\sqrt{\log n} \to 0$, we have*

$$|\hat{\alpha}_{n,\epsilon} - \hat{\alpha}_n| = o_{\mathbb{P}}(1), \quad n \to \infty.$$

This implies that any such ϵ-scan test enjoys the same asymptotic size and power as the corresponding full scan, established in Theorems 3 and 4 for the multiscale scan, and in Theorems 5 and 6 for the adaptive multiscale scan.

Algorithm 1 Implementation of the ϵ-adaptive multiscale scan over the ϵ-covering defined in (2.23). n is assumed to be a power of 2 for convenience. The P-values can be as in (2.20) or completely arbitrary, for example based on a different parametric model.

Require: Field $\mathbf{y}$ over $[n]^d$, integers $1 \le \underline{h} \le \overline{h} \le n$, ϵ such that $\epsilon^2 \underline{h} \ge 8d$, P-value functions $\hat{\alpha}_{\mathbf{h}}$

Initialize $\mathrm{dyad}_{\mathbf{1}}(\mathbf{i}) = y(\mathbf{i}), \forall \mathbf{i} \in [n]^d$
for $\mathbf{a} \in [\log_2 n]^d \backslash \{1\}^d$ **do (Downsampling)**
  $j' \leftarrow \min\{j \in [d] : a_j > 1\}$
  for $\mathbf{t} \in [n/2^{\mathbf{a}}]$ **do**
    $\mathrm{dyad}_{\mathbf{a}}(\mathbf{t}) \leftarrow \mathrm{dyad}_{\mathbf{a}-\mathbf{e}_{j'}}(\mathbf{t} \circ (\mathbf{1} + \mathbf{e}_{j'})) + \mathrm{dyad}_{\mathbf{a}-\mathbf{e}_{j'}}(\mathbf{t} \circ (\mathbf{1} + \mathbf{e}_{j'}) - \mathbf{e}_{j'})$
  end for
end for
$\underline{a} \leftarrow \lfloor \log_2(\epsilon^2 \underline{h}/(4d)) \rfloor$
$\overline{a} \leftarrow \lceil \log_2(\epsilon^2 \overline{h}/(4d)) \rceil$
Initialize $\hat{\alpha} \leftarrow 1$
for $\mathbf{a} \in [\underline{a}, \overline{a}]^d$ **do**
  for $\mathbf{f} \in [\lceil 8d/\epsilon^2 \rceil]^d$ **do (Convolution)**
    $\hat{s} \leftarrow \left(\prod_j f_j 2^{a_j}\right)^{-\frac{1}{2}} \max_{\mathbf{t} \in [n/2^{\mathbf{a}}]} (\mathrm{dyad}_{\mathbf{a}} * b_{\mathbf{f}})(\mathbf{t})$
    $\hat{\alpha} \leftarrow \min\{\hat{\alpha}, \hat{\alpha}_{\mathbf{f} \circ 2^{\mathbf{a}}}(\hat{s})\}$
  end for
end for
Ensure: $\hat{\alpha}$

2.4.2. Implementation and computational complexity

The computational complexity of a scan over an ϵ-covering depends, of course, on how the ϵ-covering is designed. We refer the reader to [3, 2, 34] for some existing examples in the literature. We design here another ϵ-covering which we find easier to scan over in practice. Specifically, assuming that n is a power of 2 for convenience, we consider

$$\mathcal{R}_\epsilon = \bigcup_{\mathbf{a} \in [\log_2 n]^d} \Big\{ [2^{\mathbf{a}} \circ \mathbf{t}, 2^{\mathbf{a}} \circ (\mathbf{t} + \mathbf{f})] : f_j \in [\lceil 8d/\epsilon^2 \rceil], t_j \in [n/2^{a_j}], \forall j \in [d] \Big\}. \tag{2.23}$$

Proposition 8. *Suppose that $\epsilon^2 \underline{h} \to \infty$ as $n \to \infty$. When n is large enough, $\mathcal{R}_\epsilon$ defined in (2.23) is an ϵ-covering of $\mathcal{R}$ for the metric δ defined in (2.18).*

Algorithm 1 gives an efficient implementation of a scan over $\mathcal{R}_\epsilon$. As in [3], we start by summing y over dyadic rectangles, which are defined as rectangles whose side lengths are a power of 2. Formally, let $\mathrm{dyad}_{\mathbf{a}}$ denote the result of summing y over all rectangles of shape $2^{\mathbf{a}}$ with the top-left corner at a multiple of $2^{\mathbf{a}}$, thought of as a field over the grid $[n2^{-\mathbf{a}}]$. Using dynamic programming, computing $\{\mathrm{dyad}_{\mathbf{a}} : \mathbf{a} \in [\log_2 n]^d\}$ can be done in time $O(n^d)$. This 'coarsification' allows us to quickly form spatial approximations to the full spatial scan for a specific shape $\mathbf{h}$. Specifically, for a given dyadic scale given by $\mathbf{a} \in [\log_2 n]^d$ and location and scale given by $t_j \in [n/2^{a_j}]$, $f_j \in [\lceil 8d/\epsilon^2 \rceil]$ for all j, we have

$$y\big[[2^{\mathbf{a}} \circ \mathbf{t}, 2^{\mathbf{a}} \circ (\mathbf{t} + \mathbf{f})]\big] = \mathrm{dyad}_{\mathbf{a}}\big[[\mathbf{t}, \mathbf{t} + \mathbf{f}]\big].$$

We note that the P-values that appear on Line 14 of Algorithm 1 can be defined in any way, and in particular could be based on other model assumptions. Put differently, the sole purpose of Algorithm 1 is to compute the P-value (2.22) for a given set of critical values in (2.20), which can be completely arbitrary.

Proposition 9. *Suppose that $\epsilon^2 \underline{h} \to \infty$ as $n \to \infty$. When n is large enough, Algorithm 1 performs a scan over $\mathcal{R}_\epsilon$ defined in (2.23).*

Proposition 10. *Algorithm 1 requires on the order of $\max\{n^d, \epsilon^{-4d}(n/\underline{h})^d \log n\}$ basic operations (the computational complexity on a CPU). Algorithm 1 requires on the order of $(\log n)^d$ downsampling operations, and $(\log n/\epsilon^2)^d$ convolution operations.*

Remark 11. *It is important to enumerate details about the implementation of Algorithm 1 on different computing architectures.*

1. *Regarding computation on a CPU, if $\underline{h} = n^a$ for some fixed $a \in (0,1)$ and $\epsilon = (\log n)^{-1}$ (which is allowed by Theorem 7), then the computational complexity of ϵ-AdaScan is of order $O(n^d)$, which is precisely linear in the grid size.*
2. *Regarding computation on a GPU, by implementing the methodology by using the NVIDIA CUDA deep neural network library, cuDNN, we have been able to parallelize the most expensive operations, downsampling and convolution, which enables Algorithm 1 to run in poly-logarithmic time. Moreover, the multiscale scan statistic is a particular instantiation of a feedforward deep convolutional neural net. Typically, the size of the shared memory is the bottleneck when implementing multiscale scans on GPUs.*
3. *Regarding distributed implementation, the formation of the* dyad *data structure is simple to parallelize in the map-reduce framework. Furthermore, in the event that $\bar{h}$ is of smaller order than n then an image can be broken into overlapping images of size $C\bar{h} \times \ldots \times C\bar{h}$ where C is a constant, whereby each image can be scanned separately where in the computation of $\hat{\alpha}$ using n as the side length of the original larger image.*

In remote sensing applications, real-time detection requires performing a multiscale scan in under 1 second. On a standard CPU, performing a multiscale scan with Algorithm 1 can scale to images of size on the order of 10 kilo-pixels. On a standard GPU, with roughly 1000 cores and 4 gigabytes of shared memory, real-time detection can be performed at the mega-pixel scale. For terapixel sized images that are prevalent in astronomy and satellite surveilance, using the map-reduce framework, real-time detection is possible as long as the size of the largest scanned rectangles are at the megapixel scale.

3. Numerical experiments

In this section, we discuss some findings from simulation experiments. In each of the following experiments, we will generate observations that conform to our assumptions, namely that the random field y is drawn according to (2.1) and

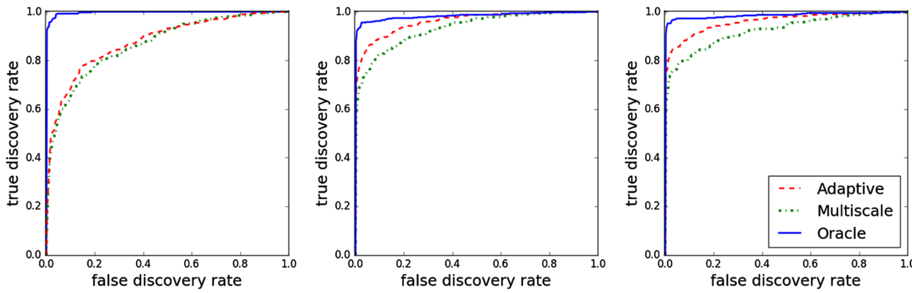

FIG 1. *(ROC curves for varying rectangle size) The percentage of discoveries that are true versus the percentage that are false, obtained by varying the τ parameter for $d = 2$. Constructed with 400 repeats from both H_0 and H_1, with $n = 256$, $\underline{h} = 6$, $\mu = 6$, rectangle of size 34×38 on a CPU (left); $n = 2048$, $\underline{h} = 10$, $\mu = 5$, rectangle of size 148×231 on a GPU; $n = 6656$, $\underline{h} = 10$, $\mu = 7$, rectangle of size 580×326 on 16 distributed GPUs.*

that there is a rectangular activation under H_1 as in (2.2). We will consider three questions.

1. For finite n, does the adaptive test $T_a(\mathbf{y})$ have appreciably superior power compared to the multiscale test $T_m(\mathbf{y})$?
2. For finite n, do the theoretically-derived thresholds (2.13) and (2.16) control the level of the tests $T_m(\mathbf{y})$ and $T_a(\mathbf{y})$ as desired?
3. What is the trade-off between computation time and statistical power as we vary ϵ in the adaptive ϵ-scan (Algorithm 1)?

In all our experiments below, we consider the case of a discrete image ($d = 2$) and the signal under the alternative is proportional to the indicator function of a rectangle, i.e., $x(\mathbf{i}) = \mu/\sqrt{|R^\star|}$ for $\mathbf{i} \in R^\star$ and 0 otherwise, for some rectangle $R^\star$. For the multiscale and adaptive scans, we set $\underline{h} = 6$.

The first experiment will address the effect that adaptive threshold selection has on the statistical power. The experiments are on a 256×256 image ($n = 256$) computed on a 2.80 Ghz CPU, 2048×2048 image on an NVIDIA GPU with 1,536 cores and 4 GB of shared memory, and 6656×6656 image on 16 such distributed GPUs. Each method is run with $\epsilon < 0.9$ using Algorithm 1. We simulate 400 times from both the null H_0 and each instance of the alternative H_1. Under H_1, we set $\mu = 6, 5, 7$ respectively, and consider rectangle sizes — 34×38, 148×231, and 580×326 respectively — with the location of the activation rectangle being chosen uniformly at random. For each method, we simulate the false discovery rate and the true discovery rate — the fraction of the 400 simulations drawn from H_0 that were rejected and the fraction from H_1 that were rejected, respectively — and plot them as the parameter τ varies, producing a receiver operator characteristic (ROC) curve. Our findings (Figure 1) indicate that the adaptive scan test achieves significantly better performance in large images when the hidden rectangle size is large. These advantages only manifest themselves when images are at the megapixel scale or greater, which is the size

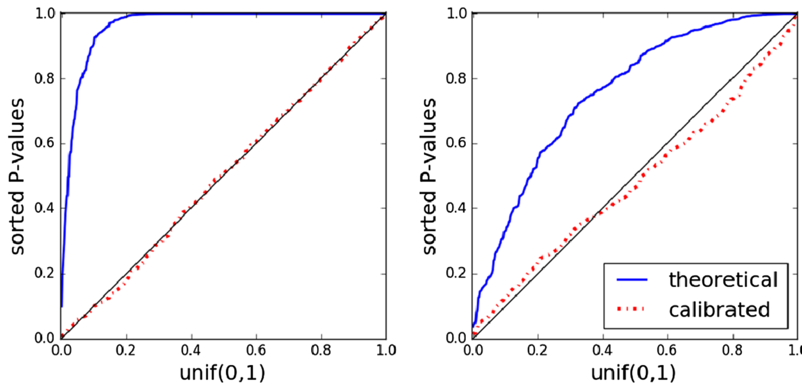

Fig 2. *(P-value qq-plot) The ordered P-values of* 400 *simulations plotted against the quantiles of the uniform*$(0,1)$ *distribution for the P-value in Theorem 5 and the aforementioned calibrated version. The images size is increasing:* 2048×2048 *(left) and* 6656×6656 *(right) and* $\underline{h} = 10$ *in both.*

of most modern cameras. For images of size 256×256 there is no significant advantage to making the statistic adaptive.

We also provide quantile-quantile plots of the P-value statistics against the uniform distribution on $(0,1)$. The motivation for this is to assess if the P-values computed based on the thresholds (2.13) and (2.16) are accurate. Our asymptotic theory in Theorem 5 predicts that this is the case in the large sample limit $n, \underline{h} \to \infty$. We see that the P-values tend to be over-estimated (Figure 2), so that they produce more conservative tests, with this effect diminishing as n increases.

In [1], importance sampling has been proposed as a way to circumvent this issue for scan statistics. It is not clear how to implement importance sampling in the context of adaptive threshold selection. Theorem 5 guarantees that

$$\max_{\mathbf{h} \in [\underline{h}, \overline{h}]^d} \Big(v_{n,\mathbf{h}} \big(\max_{R \in \mathcal{R}(\mathbf{h})} y[R] - v_{n,\mathbf{h}} \big) - (4d-1) \log \big(v_{n,\mathbf{h}} \big) \Big) \tag{3.1}$$

has an asymptotic Gumbel distribution with location κ and shape 1 which has a mean of approximately $\kappa + 0.5772$. We have found that the statistic (3.1) is nearly Gumbel distributed but the location parameter κ differs from the asymptotic theoretical value for most values of $\underline{h}$ and ϵ. We estimate κ via Monte Carlo and the method of moments, specifically we simulate many times from H_0 and average the statistic (3.1) over the simulations and subtract 0.5772. After this calibration, this gives a close fit to the Gumbel distribution (Figure 2). It is important to notice that the specific form of $v_{n,\mathbf{h}}$ was determined in Theorem 5, and is needed to compute (3.1).

The final set of experiments are intended to demonstrate the impact of reducing ϵ on the performance of Algorithm 1 (computed on a CPU). We derive ROC curves for Algorithm 1 by applying it to 480 simulations with two different image sizes: 256×256 and 512×512 pixels (Figure 3) with parameters $\underline{h} = 6, 12$

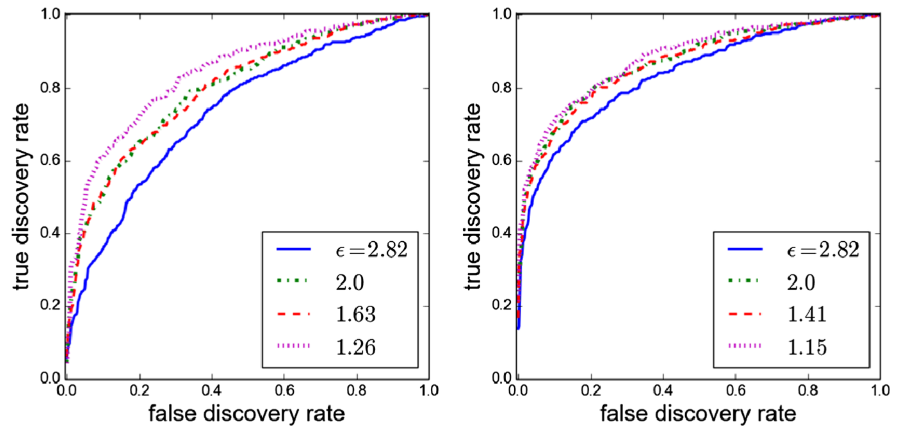

Fig 3. *(ROC curve for the ϵ-adaptive scan) The ROC curve for Algorithm 1 as ϵ decreases for a 256×256 image (left) and a 512×512 image (right) with $\underline{h} = 6, 12$, $\mu = 4, 5$, and a 61×47 and 82×35 active rectangle, respectively. Each setting is repeated 480 times.*

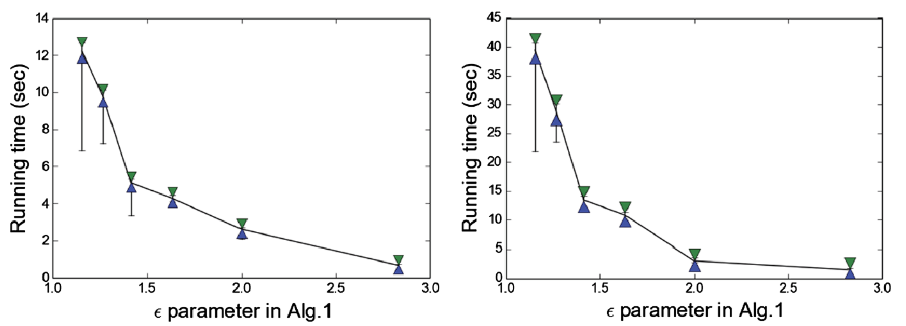

Fig 4. *(Running time) Same setting as in Figure 3. Here we plot the running time in seconds on a 2.80 Ghz virtual CPU as a function of ϵ. The line indicates the average time, the triangles are the 5 and 95 percentiles, and the error bars extend from the minimum to the maximum of the 480 simulations.*

and $\mu = 4, 5$, and active rectangle of size 61×47 and 82×35, respectively. We selected the values for ϵ by making $8d/\epsilon^2$ equal to the integers $1, \ldots, 6$ and selecting from these 4 representative curves. We interpret the results to mean that as ϵ decreases, the performance quickly converges to the optimal ROC plot. We also considered the running time as ϵ changes (Figure 4). In this simple experiment we find that, while it is advantageous to have an ϵ small to increase the power, the improvements in power are generally outweighed by the additional computational burden.

4. Discussion

We briefly discuss some generalizations and refinements of our work here.

More general signals. In this paper we work in a context where the signal has substantial energy over some rectangle of unknown shape and location. This motivates the scans over classes of rectangles that we define and study in detail. Although one could scan over more complicated shapes, to increase power, as done for example in [2, 3, 11, 20], the implementation of such scans is generally very complicated and often ad hoc search methods are implemented. For example, the software SaTScan (Kulldorff and Information Management Services, Inc) offers two options (for spatial data): scanning over circles or discs. In contrast, scanning over rectangles (and other simple shapes such as circles) can be done efficiently and the mathematical analysis can be carried to the exact asymptotics, as we show here — see also [3, 34]. Moreover, rectangles are building blocks for more complicated shapes and are representative of 'thick' or 'blob-like' shapes — see [2].

Signals of arbitrary sign. For concreteness and ease of exposition, we consider signals that are 'mostly' positive over a rectangle. This is clear from the class of signals that we consider, defined in (2.4). This would not be the most appropriate definition when one is expecting signals of arbitrary sign, for example, when the signal $\mathbf{x}$ is such that $x(\mathbf{i})$ are IID normal with zero mean and variance τ, over some rectangle R. In that case, assuming R is asymptotically large, we have $x[R] \sim \mathcal{N}(0, \tau)$, and is negative with probability $1/2$. For a signal $\mathbf{x} = (x(\mathbf{i}) : \mathbf{i} \in [n]^d)$ and $R \subset [n]^d$, define $x_2[R] = \sum_{\mathbf{i} \in R} x(\mathbf{i})^2$. Instead of the class of signals defined in (2.4), consider

$$\mathcal{X}_\tau^2(\mathbf{h}) = \Big\{ \mathbf{x} \in \mathbb{R}^{n^d} : \min_{R \in \mathcal{R}(\mathbf{h})} x_2[R] \geq |R| + \tau\sqrt{2|R|} \Big\}. \tag{4.1}$$

In that case, the most natural scans are based on the chi-squared scores $y_2[R], R \in \mathcal{R}$. Presumably, a similar analysis can be carried in this setting, in particular since the results of [8] — upon which Kabluchko's arguments (and therefore ours too) are founded — apply to approximately Gaussian fields, which is the case of $y_2[R], R \in \mathcal{R}$. This brings us to the following.

Other parametric models. Obtaining similar results for other parametric models may be of interest, for example, in epidemiology where the Poisson distribution is used to model counts, and would replace the Gaussian distribution here. [2] extend their first-order analysis to distributions with finite moment generating function, proving that the bounds obtained under normality still apply as long as $\underline{h} \gg \log n$. It is possible that a similar phenomenon (essentially due to the Central Limit Theorem) applies at a more refined level, and that our results apply to such distributions, again, as long as $\underline{h}$ is sufficiently large.

Dependencies. A more involved extension of our results would be to allow the observations $y(\mathbf{i}), \mathbf{i} \in [n]^d$ to be dependent. The results of [8] apply unchanged to the setting where short-range dependencies are present. So, in principle, an extension of our work in that direction is possible following similar lines. But we did not pursue this here for the sake of concreteness and conciseness of presentation.

5. Proofs

Our method of proof is largely based on [18], which relies on the work of [8] on the extrema of Gaussian random fields and the Chen-Stein Poisson approximation [4].

Signals that are indicators of rectangles. We will focus the remaining of the paper on signals $\mathbf{x}$ that are proportional to the indicator of a rectangle. This is asymptotically the most difficult case for all the scans that we consider. Indeed, we show in this section that the limits in Theorems 2, 4, and 6, hold when the signal is $\mu|R^\star|^{-1/2}\mathbf{1}_{R^\star}$, while for a signal $\mathbf{x}$ such that $x[R^\star] \geq \mu$, these are seen to hold as lower bounds when taking the limit inferior. Together, this shows that the minimax asymptotics stated in Theorems 2, 4, and 6 hold.

We will often leave n implicit, but even then, all the limits are with respect to $n \to \infty$, unless otherwise stated.

5.1. Locally stationary Gaussian random fields

We will begin the proof section with an introduction to some theory for locally-stationary Gaussian random fields (GRFs), particularly their smoothness and extreme value properties. Throughout this work, we approximate the discrete GRF given by $\{\xi[R], R \in \mathcal{R}\}$ with a continuous version. For that, define the continuous analog to $\mathcal{R}$, that is,

$$\bar{\mathcal{R}} = \Big\{[\mathbf{t}, \mathbf{t}+\mathbf{h}] : \mathbf{h} \in [\underline{h}, \overline{h}]^d, \mathbf{t} \in [\mathbf{0}, n\mathbf{1} - \mathbf{h}]\Big\}.$$

Let Ξ be a (canonical) Gaussian white noise on $\mathbb{R}^d$, meaning a random measure on the Borel sets of $\mathbb{R}^d$ such that, for any integer $k \geq 1$ and any Borel sets $R_1, \dots, R_k$, $\Xi(R_1), \dots, \Xi(R_k)$ are jointly Gaussian, with zero mean and $\mathrm{Cov}(\Xi(R_i), \Xi(R_j)) = \lambda(R_i \cap R_j)$ for all $i, j \in [k]$. Consider the GRF on $\bar{\mathcal{R}}$ defined by $\Xi[R] = \Xi(R)/\sqrt{\lambda(R)}$, where $\lambda(R)$ denotes the Lebesgue measure of R when R is a continuous rectangle. This GRF is denoted Ξ henceforth. It has zero-mean and covariance structure

$$\mathrm{Cov}(\Xi[R_0], \Xi[R_1]) = \frac{\lambda(R_0 \cap R_1)}{\sqrt{\lambda(R_0)\lambda(R_1)}}, \quad R_0, R_1 \in \bar{\mathcal{R}}.$$

Consequently, it is invariant with respect to translations and scalings. Following the approach taken by [18], we approximate the discrete GRF ξ with its continuous counterpart Ξ. Therefore, we will be interested in the excursion probabilities of Ξ, which will require an introduction to locally stationary GRFs. For convenience, consider the parametrization of the rectangles $\bar{\mathcal{R}}$ via the one-to-one map $\mathbf{w} = (\mathbf{h}, \mathbf{t}) \mapsto R(\mathbf{w}) := [\mathbf{t}, \mathbf{t}+\mathbf{h}]$ for $\mathbf{w} \in (0, \infty)^{2d}$. We then use the shorthand $\Xi(\mathbf{w}) = \Xi[R(\mathbf{w})]$ for $\mathbf{w} \in (0, \infty)^{2d}$. In this way, Ξ can be thought of as a GRF over $(0, \infty)^{2d}$ with the following covariance structure,

$$\mathrm{Cov}(\Xi(\mathbf{h}, \mathbf{t}), \Xi(\mathbf{g}, \mathbf{s})) = \prod_{j=1}^{d} (h_j g_j)^{-1/2} [(t_j + h_j) \wedge (s_j + g_j) - t_j \vee s_j]_+ \ , \quad (5.1)$$

for pairs $(\mathbf{h},\mathbf{t}),(\mathbf{g},\mathbf{s})\in(0,\infty)^{2d}$, where $x_+ = x\vee 0$. Furthermore, define the set of shapes and location that correspond to rectangles in $\bar{\mathcal{R}}$ as

$$\mathcal{W}=\Big\{(\mathbf{h},\mathbf{t})\in(0,\infty)^{2d}:\mathbf{h}\in[\underline{h},\overline{h}]^d,\mathbf{t}\in[\mathbf{0},n\mathbf{1}-\mathbf{h}]\Big\}.$$

This describes the continuous version of the data under the null distribution H_0. Under the alternative, there is a signal, and the continuous counterpart to the discrete GRF is

$$\Upsilon(\mathbf{w})=m(\mathbf{w})+\Xi(\mathbf{w}),\tag{5.2}$$

where

$$m(\mathbf{w})=\mu\operatorname{Cov}(\Xi(\mathbf{w}),\Xi(\mathbf{w}^\star)),\tag{5.3}$$

$\mathbf{w}^\star=(\mathbf{h}^\star,\mathbf{t}^\star)$ being the scale and location of the active rectangle. (Recall that under the alternative we are considering the signal $\mu|R^\star|^{-1/2}\mathbf{1}_{R^\star}$.) We are now prepared to review some relevant results on boundary crossing probabilities of locally-stationary GRFs.

5.1.1. Boundary crossing probabilities for locally stationary GRFs

In order to analyze the GRF Ξ we must introduce the notion of local stationarity and the tangent process. The definitions below are given in [8, 30], and utilized in [18].

We note that we work in dimension $p=2d$, except when analyzing the oracle scan, in which case $p=d$, because the shape $\mathbf{h}^\star$ is given. Given $K\subset\mathbb{R}^p$ and $\gamma>0$, define

$$[K]_\gamma=\{\mathbf{w}+\mathbf{u}:\mathbf{w}\in K,\|\mathbf{u}\|\le\gamma\}.$$

A function $L:\mathbb{R}_+\to\mathbb{R}_+$ is said to be *slowly varying* if

$$\lim_{x\to 0}\frac{L(\alpha x)}{L(x)}=1,\quad\forall\alpha>0.$$

Let $\mathcal{S}^{p-1}$ denote the unit sphere in $\mathbb{R}^p$. We say that the GRF Ξ is *locally stationary over the set* $\mathcal{W}$, if for $\mathcal{W}$ within the domain of Ξ, there exists $\alpha\in(0,2]$, $\gamma>0$, and a slowly varying function L, such that $[\mathcal{W}]_\gamma\subset(0,\infty)^{2d}$ and for all $\mathbf{w}\in[\mathcal{W}]_\gamma$,

$$\mathbb{E}[\Xi(\mathbf{w})\Xi(\mathbf{w}+\mathbf{u})]=1-(1+g_{\mathbf{w}}(\mathbf{u}))\|\mathbf{u}\|^\alpha L(\|\mathbf{u}\|)r_{\mathbf{w}}(\mathbf{u}/\|\mathbf{u}\|),\tag{5.4}$$

where $r_{\mathbf{w}}:\mathbb{S}^{p-1}\to\mathbb{R}_+$ are continuous functions such that

$$\sup_{\mathbf{v}\in\mathbb{S}^{p-1}}|r_{\mathbf{w}}(\mathbf{v})-r_{\mathbf{w}+\mathbf{u}}(\mathbf{v})|\to 0,\quad\text{ as }\mathbf{u}\to\mathbf{0},$$

and $g_{\mathbf{w}}:\mathbb{R}^p\to\mathbb{R}$ are such that

$$\sup_{\mathbf{w}\in[\mathcal{W}]_\gamma}|g_{\mathbf{w}}(\mathbf{u})|\to 0,\quad\text{ as }\mathbf{u}\to\mathbf{0}.$$

For such processes, the *local structure* is defined as

$$C_{\mathbf{w}}(\mathbf{u}) = \|\mathbf{u}\|^\alpha L(\|\mathbf{u}\|) r_{\mathbf{w}}(\mathbf{u}/\|\mathbf{u}\|),$$

and we say that the local structure is *homogeneous of order* α. Let the *tangent process* at $\mathbf{w} \in \mathcal{W}$ be $\{H_{\mathbf{w}}(\mathbf{u})\}_{\mathbf{u}\in\mathbb{R}^p}$, and defined as the GRF satisfying

$$\mathbb{E} H_{\mathbf{w}}(\mathbf{u}) = -C_{\mathbf{w}}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^p,$$

and

$$\operatorname{Cov}(H_{\mathbf{w}}(\mathbf{u}_0), H_{\mathbf{w}}(\mathbf{u}_1)) = C_{\mathbf{w}}(\mathbf{u}_0) + C_{\mathbf{w}}(\mathbf{u}_1) - C_{\mathbf{w}}(\mathbf{u}_0 - \mathbf{u}_1), \quad \mathbf{u}_0, \mathbf{u}_1 \in \mathbb{R}^p.$$

The *high excursion intensity* is defined as

$$\Lambda(\mathbf{w}) = \lim_{m\to\infty} \frac{1}{m^p} \mathbb{E} \exp\left[\sup\nolimits_{\mathbf{u}\in[0,m]^p} H_{\mathbf{w}}(\mathbf{u})\right]$$

and has been shown to exist within $(0, \infty)$ in [8, Lem 5.2], which in fact states that this convergence is uniform within $\mathbf{w} \in \mathcal{W}$. [18] proves the following result by observing that Ξ has the same local structure as a tensor product of normalized differences of Brownian motions.

Lemma 12 ([18])**.** *The GRF Ξ is locally stationary over $\mathcal{W}$ with $\alpha = 1$ and $L(x) = 1$, with local structure given by*

$$C_{(\mathbf{h},\mathbf{t})}(\mathbf{g}, \mathbf{s}) = \frac{1}{2} \sum_{j=1}^{d} \frac{|s_j| + |s_j + g_j|}{h_j},$$

and high excursion intensity given by

$$\Lambda(\mathbf{h}, \mathbf{t}) = 4^{-d} \prod_{j=1}^{d} h_j^{-2}.$$

Define the function

$$\psi(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

Lemma 13 ([8] Th 2.1)**.** *For $K \subset \mathcal{W} \subset \mathbb{R}^p$ fixed, bounded, and Jordan measurable, and a GRF Ξ that is locally stationary over $\mathcal{W}$ with homogeneity α and high-excursion intensity Λ,*

$$\mathbb{P}\big\{\exists \mathbf{w} \in K : \Xi(\mathbf{w}) > c\big\} \sim c^{\frac{2p}{\alpha}} \psi(c) \int_K \Lambda(\mathbf{w}) d\mathbf{w}, \quad \textit{as } c \to \infty.$$

[8] generalized this theorem for calculating the non-constant boundary crossing probability of a locally stationary GRF, which we will use in our analysis of the adaptive scan. In fact, [8, Th 2.8] allows a non-constant boundary, a set that is growing with n, and holds for non-Gaussian random fields. We specialize the theorem to our setting.

Lemma 14 ([8] Th 2.8)**.** *Let Ξ be a GRF that is locally stationary over $\mathcal{W}$ with homogeneity α and high-excursion intensity Λ, and take a fixed bounded and Jordan measurable set K such that $[K]_\gamma \subset \mathcal{W}$ for some $\gamma > 0$. Let $(c_\zeta : \zeta \in (0,1))$ be a family of real-valued functions defined on $\mathcal{W}$ satisfying*

$$\sup_{\mathbf{w}\in[K]_\gamma} c_\zeta(\mathbf{w})^{-2p/\alpha} = o(\zeta), \quad \zeta \to 0, \tag{5.5}$$

for some $\gamma_0 > 0$ fixed and

$$\sup\left\{c_\zeta(\mathbf{w})^2 - c_\zeta(\mathbf{w}')^2 : \mathbf{w}, \mathbf{w}' \in [K]_{2\zeta}, \|\mathbf{w}-\mathbf{w}'\|_\infty \leq \zeta\right\} = o(1), \quad \zeta \to 0. \tag{5.6}$$

Then

$$\mathbb{P}\big\{\exists \mathbf{w} \in K : \Xi(\mathbf{w}) > c_\zeta(\mathbf{w})\big\} \sim \int_K c_\zeta(\mathbf{w})^{\frac{2p}{\alpha}} \psi(c_\zeta(\mathbf{w}))\Lambda(\mathbf{w})\mathrm{d}\mathbf{w}, \quad \zeta \to 0.$$

Lemma 14 differs from the statement in [8, Th 2.8] which includes additional conditions. This is due to the fact that we assume that K is fixed and Ξ is Gaussian. Their conditions (2.16) and (2.18) are precisely (5.5) and (5.6), while the condition (2.14) is trivially true for fixed K. In the proof of [8, Th 2.1] their conditions (A1)–(A5) were shown to hold for locally stationary GRFs, and as a consequence so do (B1)–(B5) since the process is exactly Gaussian and the domain K is fixed.

5.2. Approximating Ξ with an ϵ-covering

In this section we state and prove results on the covering properties of $\mathcal{W}$ and the continuity of Ξ. The metric δ over $\mathcal{R}$ introduced in (2.18) translates into the following metric on $\mathcal{W}$ (we overload the notation)

$$\delta(\mathbf{w}_0, \mathbf{w}_1) = \delta(R(\mathbf{w}_0), R(\mathbf{w}_1)), \quad \forall \mathbf{w}_0, \mathbf{w}_1 \in \mathcal{W}.$$

An ϵ-covering of $\mathcal{W}$ is defined analogously. To be sure, it is a subset $\mathcal{W}_\epsilon \subset \mathcal{W}$ such that, for all $\mathbf{w} \in \mathcal{W}$, there is $\mathbf{w}' \in \mathcal{W}_\epsilon$ such that $\delta(\mathbf{w}, \mathbf{w}') \leq \epsilon$. The ϵ-covering number for the metric space $(\mathcal{W}, \delta)$, denoted $N(\mathcal{W}, \delta, \epsilon)$, is the cardinality of a smallest ϵ-covering of $\mathcal{W}$ for δ, and $\log N(\mathcal{W}, \delta, \epsilon)$ is the ϵ-entropy of $\mathcal{W}$.

Lemma 15. *For any $0 < \epsilon < \sqrt{4d}$,*

$$\log N(\mathcal{W}, \delta, \epsilon) \leq d \log\left(\frac{32d^2 n}{\underline{h}\epsilon^4}\right).$$

Proof. Let $(\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s}) \in \mathcal{W}$. Starting with (2.18) and (5.1), and using the fact that

$$1 - \prod_{j=1}^d (1 - a_j) \leq \sum_{j=1}^d a_j, \quad \text{for any } a_1, \ldots, a_d \in [0,1],$$

which follows from the union bound or a simple recursion, we have

$$\begin{aligned}
\frac{1}{2}\delta^2((\mathbf{h},\mathbf{t}),(\mathbf{g},\mathbf{s})) &= 1-\prod_{j=1}^d (h_j g_j)^{-1/2}[(t_j+h_j)\wedge(s_j+g_j)-t_j\vee s_j]_+ \qquad (5.7)\\
&\le \sum_{j=1}^d \Big(1-\frac{1}{\sqrt{h_j g_j}}[(t_j+h_j)\wedge(s_j+g_j)-t_j\vee s_j]_+\Big)\\
&\le \sum_{j=1}^d \Big(1-\frac{1}{\sqrt{h_j g_j}}[t_j\wedge s_j+h_j\wedge g_j-t_j\vee s_j]_+\Big)\\
&\le \sum_{j=1}^d \frac{1}{\sqrt{h_j g_j}}[\sqrt{h_j g_j}-t_j\wedge s_j-g_j\wedge h_j+t_j\vee s_j]_+\\
&\le \sum_{j=1}^d \frac{1}{\sqrt{h_j g_j}}[h_j\vee g_j-t_j\wedge s_j-g_j\wedge h_j+t_j\vee s_j]_+\\
&\le \sum_{j=1}^d \theta((h_j,t_j),(g_j,s_j)), \qquad (5.8)\\
&\text{where } \theta((h,t),(g,s))=\frac{|h-g|+|t-s|}{\sqrt{hg}}. \qquad (5.9)
\end{aligned}$$

Notice that because

$$\delta((\mathbf{h},\mathbf{t}),(\mathbf{g},\mathbf{s}))\le\sqrt{2\sum_j \theta((h_j,t_j),(g_j,s_j))},$$

and $\mathcal{W}\subset[\underline{h},n]^d\times[0,n]^d$, it suffices to construct an $(\epsilon^2/2d)$-covering for each $[\underline{h},n]\times[0,n]$ with respect to θ. (We define a covering in θ analogously although it is not necessarily a metric.) We divide by $\underline{h}$ everywhere, so that we may focus on $[1,T]\times[0,T]$, where $T=n/\underline{h}$, by the scale invariance of θ. Fix $\alpha\in(0,1)$. We have

$$[1,T]\times[0,T]\subseteq\bigcup_{k=0}^{[\log T/\log(1/\alpha)]}\ \bigcup_{\ell=0}^{[1/(\alpha^k(1-\alpha))]} I_k\times I_{k,\ell},$$

where $I_k=[\alpha^{k+1}T,\alpha^kT]$ and $I_{k,\ell}=[\ell\alpha^k(1-\alpha)T,(\ell+1)\alpha^k(1-\alpha)T]$. Take $h,g\in I_k$ and $t,s\in I_{k,\ell}$ for some k and ℓ in these ranges. Then

$$\theta((h,t),(g,s))\le\frac{(1-\alpha)\alpha^kT+(1-\alpha)\alpha^kT}{\alpha^{k+1}T}=\frac{2(1-\alpha)}{\alpha}.$$

The cardinality of the resulting covering is equal to

$$\sum_{k=0}^{[\log T/\log(1/\alpha)]}([1/(\alpha^k(1-\alpha))]+1)\le\frac{T}{(1-\alpha)^2}+\frac{\log T}{\log(1/\alpha)}\le\frac{2T}{(1-\alpha)^2},$$

using the fact that $\log(1+x) \geq x^2 \log(2)$ for all $x \in [0,1]$ and $t \geq \log(t)/\log(2)$ for all $t \geq 1$. When we choose $\alpha = 4d/(4d+\epsilon^2)$, the tensor product of these coverings, repeated over $j = 1, \ldots, d$, is an ϵ-covering of $\mathcal{W}$, of cardinality

$$\left(\frac{2T}{(1-\alpha)^2}\right)^d = (2T(4d/\epsilon^2)^2)^d,$$

since $1 - \alpha \leq \epsilon^2/(4d)$. □

We use this bound on the entropy and a continuity property of Ξ to bound Ξ. This bound will be crude relative to the asymptotic guarantees which are the focus of this work, but is necessary as a lemma. For an ϵ-covering, $\mathcal{W}_\epsilon \subset \mathcal{W}$, define the *interpolated GRF* Ξ_ϵ over $\mathcal{W}$, with value at $\mathbf{w} \in \mathcal{W}$ given by $\Xi_\epsilon(\mathbf{w}) = \Xi(\mathbf{w}')$, where $\mathbf{w}' = \operatorname{argmin}\{\delta(\mathbf{w}_0, \mathbf{w}) : \mathbf{w}_0 \in \mathcal{W}_\epsilon\}$; if the minimizer is not unique, then choose a minimizer arbitrarily. For a real-valued function f over $\mathcal{W}$, let $\|f\|_\infty = \sup_{\mathbf{w}\in\mathcal{W}} |f(\mathbf{w})|$.

Lemma 16. *Consider the GRF Ξ introduced in Section 5.1. In our context, it has the following properties.*

1. *The supremum of Ξ has the following behavior*

$$\|\Xi\|_\infty = O_{\mathbb{P}}\left(\sqrt{\log(n/\underline{h})}\right).$$

2. *Let $\mathcal{U} \subset \mathcal{W}$ be such that there exists a constant $C > 0$ with the property that $\max_j |t_j - s_j| \leq C$ and $\max_j |\log h_j - \log g_j| \leq C$ for all $(\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s}) \in \mathcal{U}$. Then*

$$\sup_{\mathbf{w}\in\mathcal{U}} \Xi(\mathbf{w}) = O_{\mathbb{P}}(1).$$

3. *Let Ξ_ϵ be an interpolated GRF built on an ϵ-covering of $\mathcal{W}$ where $\epsilon < 1$. Then*

$$\|\Xi - \Xi_\epsilon\|_\infty = O_{\mathbb{P}}\left(\epsilon\sqrt{\log(n/(\underline{h}\epsilon^4))}\right).$$

Proof. We prove each part in turn.

Part 1. Let $T = n/\underline{h}$. Note that $\delta(\mathbf{w}, \mathbf{w}') \in [0, \sqrt{2}]$. Dudley's metric entropy theorem [22, Th 6.1.2], along with Lemma 15, can be applied to show that

$$\begin{aligned}\mathbb{E}(\|\Xi\|_\infty) &\leq 16\sqrt{2}\int_0^{\sqrt{2}} \sqrt{\log N(\mathcal{W}, \delta, \epsilon)}\mathrm{d}\epsilon \\ &\leq 16\sqrt{2}\int_0^{\sqrt{2}} \sqrt{d\log\left(32d^2T/\epsilon^4\right)}\mathrm{d}\epsilon = O\left(\sqrt{\log T}\right).\end{aligned}$$

The result now follows by Markov's inequality.

Part 2. This can be proven by noticing that for any ϵ, the entropy of $\mathcal{U}$ satisfies

$$\log N(\mathcal{U}, \delta, \epsilon) \leq C_0 \log(1/\epsilon)$$

for some constant C_0, using a construction analogous to that used in the proof of Lemma 15. The rest follows as in the proof of Part 1.

Part 3. As before let $T = n/\underline{h}$. By the definition of Ξ_ϵ,

$$\mathbb{E}\left(\|\Xi_\epsilon - \Xi\|_\infty\right) \leq \mathbb{E}\Big(\sup\{\Xi(\mathbf{w}_1) - \Xi(\mathbf{w}_2) : \mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W} : \delta(\mathbf{w}_1, \mathbf{w}_2) \leq \epsilon\}\Big).$$

Applying Dudley's theorem and Lemma 15, we bound the RHS by

$$\begin{aligned} 99 \int_0^\epsilon \sqrt{\log N(\mathcal{W}, \delta, \eta)} \mathrm{d}\eta &\leq 99\epsilon \int_0^1 \sqrt{d \log\left(32 d^2 T/\epsilon^4\right) + 4d \log\left(1/\eta\right)} \mathrm{d}\eta \\ &= O\left(\epsilon \sqrt{\log\left(T/\epsilon^4\right)}\right). \end{aligned}$$

The result follows by Markov's inequality. □

We will now analyze the P-values resulting from our various scan statistics by their Lipschitz property. This will allow us to demonstrate that if ϵ is decreasing quickly enough, the P-value of each test when evaluated over an ϵ-covering is asymptotically indistinguishable from the P-value when evaluated over the entire set $\mathcal{W}$. In the end, we will have proven Theorem 7, but these results will also be useful to prove other results. For convenience, we work with τ instead of α, related by (2.11). For each scan statistic, let $\hat{\tau}$ be the value of τ such that the scan statistic equals its threshold ((2.9), (2.13), or (2.16)). It takes the form

$$\begin{aligned} \hat{\tau} &= \max_{\mathbf{w} \in \mathcal{W}'} a(\mathbf{w})\left(y[R(\mathbf{w})] - a(\mathbf{w})\right) + b(\mathbf{w}) \\ &= \max_{\mathbf{w} \in \mathcal{W}'} a(\mathbf{w})\left(\xi[R(\mathbf{w})] + m(\mathbf{w}) - a(\mathbf{w})\right) + b(\mathbf{w}), \end{aligned} \tag{5.10}$$

where m is defined in (5.3) (with $m \equiv 0$ under H_0), while a, b and $\mathcal{W}' \subset \mathbb{Z}^{2d} \cap \mathcal{W}$ will depend on which scan statistic we are considering. In all cases,

$$\sqrt{2d} \leq \sqrt{2d \log(n/\overline{h})} \leq a(\mathbf{w}) \leq \sqrt{2d \log(n/\underline{h})}, \quad \forall \mathbf{w} \in \mathcal{W}'.$$

We will relate the statistic $\hat{\tau}$ with the random variable

$$\tilde{\tau} = \max_{\mathbf{w} \in \mathcal{W}} a(\mathbf{w})\left(\Xi(\mathbf{w}) + m(\mathbf{w}) - a(\mathbf{w})\right) + b(\mathbf{w}). \tag{5.11}$$

Lemma 17. *Suppose there are constants $L > 0$ and $\epsilon_0 > 0$ such that*

$$\left|a(\mathbf{w}_1) - a(\mathbf{w}_2)\right| \vee \left|b(\mathbf{w}_1) - b(\mathbf{w}_2)\right| \leq L\delta(\mathbf{w}_1, \mathbf{w}_2), \tag{5.12}$$

for all $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$ such that $\delta(\mathbf{w}_1, \mathbf{w}_2) \leq \epsilon_0$. Then $|\hat{\tau} - \tilde{\tau}| = o_{\mathbb{P}}(1)$ if $\mathcal{W}'$ is an ϵ-covering of $\mathcal{W}$ with

$$\epsilon\big(\mu + \sqrt{\log\left(n/(\underline{h}\epsilon^4)\right)}\big)\sqrt{\log(n/\underline{h})} = o(1). \tag{5.13}$$

Proof. Applying the triangle inequality,

$$
\begin{aligned}
&|\hat{\tau} - \tilde{\tau}| \\
&\le \max_{\mathbf{w}\in\mathcal{W}} \min_{\mathbf{w}'\in\mathcal{W}'} |a(\mathbf{w})\left(\Xi(\mathbf{w}) + m(\mathbf{w}) - a(\mathbf{w})\right) - a(\mathbf{w}')\left(\Xi(\mathbf{w}') + m(\mathbf{w}') - a(\mathbf{w}')\right)| \\
&\quad + \max_{\mathbf{w}\in\mathcal{W}} \min_{\mathbf{w}'\in\mathcal{W}'} |b(\mathbf{w}) - b(\mathbf{w}')|.
\end{aligned}
$$

For the second term, we use the fact that b is Lipschitz and that $\mathcal{W}'$ is an ϵ-covering, to get

$$
\max_{\mathbf{w}\in\mathcal{W}} \min_{\mathbf{w}'\in\mathcal{W}'} |b(\mathbf{w}) - b(\mathbf{w}')| \le \max_{\mathbf{w},\mathbf{w}'\in\mathcal{W}:\delta(\mathbf{w},\mathbf{w}')\le\epsilon} |b(\mathbf{w}) - b(\mathbf{w}')| \le L\epsilon.
$$

For the first term, it is bounded by

$$
\begin{aligned}
&\max_{\mathbf{w}\in\mathcal{W}} \left(\Xi(\mathbf{w}) + m(\mathbf{w}) - a(\mathbf{w})\right) \min_{\mathbf{w}'\in\mathcal{W}'} |a(\mathbf{w}) - a(\mathbf{w}')| \\
&\quad + \max_{\mathbf{w}\in\mathcal{W}} a(\mathbf{w}) \min_{\mathbf{w}'\in\mathcal{W}'} \left[|\Xi(\mathbf{w}) - \Xi(\mathbf{w}')| + |m(\mathbf{w}) - m(\mathbf{w}')| + |a(\mathbf{w}) - a(\mathbf{w}')|\right].
\end{aligned}
$$

We have

$$
\min_{\mathbf{w}'\in\mathcal{W}'} |a(\mathbf{w}) - a(\mathbf{w}')| \le L\epsilon, \quad \forall \mathbf{w}\in\mathcal{W},
$$

by the fact that a is Lipschitz and $\mathcal{W}'$ is an ϵ-covering for $\mathcal{W}$; we have

$$
\max_{\mathbf{w}\in\mathcal{W}} \left(\Xi(\mathbf{w}) + m(\mathbf{w}) - a(\mathbf{w})\right) \le \max_{\mathbf{w}\in\mathcal{W}} \Xi(\mathbf{w}) + \mu = O_{\mathbb{P}}\left(\sqrt{\log(n/\underline{h})}\right) + \mu,
$$

by Lemma 16, the fact that $m(\mathbf{w}) \le \mu$ and $a(\mathbf{w}) \ge 0$ for any $\mathbf{w}\in\mathcal{W}$; we have

$$
\max_{\mathbf{w}\in\mathcal{W}} a(\mathbf{w}) \le \sqrt{2d\log(n/\underline{h})},
$$

as well as

$$
\begin{aligned}
\min\nolimits_{\mathbf{w}'\in\mathcal{W}'} &\left[|\Xi(\mathbf{w}) - \Xi(\mathbf{w}')| + |m(\mathbf{w}) - m(\mathbf{w}')| + |a(\mathbf{w}) - a(\mathbf{w}')|\right] \\
&\le O_{\mathbb{P}}\left(\epsilon\sqrt{\log(n/(\underline{h}\epsilon^4)}\right) + \mu\epsilon/2 + L\epsilon,
\end{aligned}
$$

for all $\mathbf{w}\in\mathcal{W}$, by Lemma 16. From this, we conclude. ☐

For the oracle and multiscale scan statistics, a and b are constant in $\mathbf{w}$ and so they are trivially Lipschitz. For the adaptive multiscale scan, we verify below that they indeed satisfy (5.12).

Lemma 18. *For the adaptive multiscale scan, based on (2.17), we have $a((\mathbf{h},\mathbf{t})) = v_{n,\mathbf{h}}$ and $b((\mathbf{h},\mathbf{t})) = -\kappa - (4d-1)\log(v_{n,\mathbf{h}})$, and they satisfy (5.12) for some $L > 0$ and $\epsilon_0 > 0$ depending only on d.*

Proof. Let $f(\mathbf{h}) = v_{n,\mathbf{h}}^2/2$. Since $2f(\mathbf{h}) \ge 1$ and $\log x$ has derivative bounded by 1 over $[1,\infty)$, it is sufficient to show that $(\mathbf{h},\mathbf{t}) \to f(\mathbf{h})$ is Lipschitz with respect to δ. From (5.7), we see that $\delta^2((\mathbf{h},\mathbf{t}),(\mathbf{g},\mathbf{s})) \ge \delta_\ddagger(\mathbf{h},\mathbf{g})$, where $\frac{1}{2}\delta_\ddagger(\mathbf{h},\mathbf{g}) =$

$1 - \prod_j (h_j \wedge g_j)/\sqrt{h_j g_j}$, so that we may work with $\delta_\ddagger$ instead of δ^2. By the fact that log has derivative bounded by 1 on $[1, \infty)$,

$$\begin{aligned} |f(\mathbf{h}) - f(\mathbf{g})| &\le \sum_{j=1}^d |\log h_j - \log g_j| \\ &\quad + 2\sum_{j=1}^d |\log(1 + \log(h_j/\underline{h})) - \log(1 + \log(g_j/\underline{h}))| \\ &\le 3 \sum_{j=1}^d |\log h_j - \log g_j|, \end{aligned}$$

with

$$\sum_{j=1}^d |\log h_j - \log g_j| = 2\sum_{j=1}^d \log \frac{\sqrt{h_j g_j}}{h_j \wedge g_j} = -2 \log\left(1 - \tfrac{1}{2}\delta_\ddagger(\mathbf{h}, \mathbf{g})\right) \le 2\delta_\ddagger(\mathbf{h}, \mathbf{g}),$$

when $\delta_\ddagger(\mathbf{h}, \mathbf{g}) \le 1$. Finally, if $\delta((\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s})) \le \epsilon_0$ as in (5.12), then we have $\delta^2((\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s})) \le \epsilon_0 \delta((\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s}))$. □

The following lemma allows us to approximate a discrete scan with its continuous counterpart.

Lemma 19. *$\mathcal{W}' = \mathcal{W} \cap \mathbb{Z}^{2d}$ is a $(\sqrt{4d/\underline{h}})$-covering for $\mathcal{W}$ with respect to δ.*

Proof. Let $\mathbf{w} = (\mathbf{h}, \mathbf{t}) \in \mathcal{W}$ and define $\mathbf{w}' = (\lfloor h_1 \rfloor, \dots, \lfloor h_d \rfloor, \lfloor t_1 \rfloor, \dots, \lfloor t_d \rfloor)$, which is in $\mathcal{W}'$ by construction. By (5.9), and recalling that $\underline{h}$ is an integer,

$$\delta^2(\mathbf{w}, \mathbf{w}') \le 2 \sum_{j=1}^d \frac{|h_j - \lfloor h_j \rfloor| + |t_j - \lfloor t_j \rfloor|}{\sqrt{h_j \lfloor h_j \rfloor}} \le \frac{4d}{\underline{h}}. \qquad \square$$

5.3. Proofs: main results

The following lemma will allow us to derive the asymptotic threshold from the excursion probabilities that we will derive in the following proofs.

Lemma 20. *Let s and t be constants, let (η_m) be a sequence tending to 0, and define*

$$u_m = \sqrt{2 \log m} + \frac{s \log(\sqrt{2 \log m}) + t + \eta_m}{\sqrt{2 \log m}}. \tag{5.14}$$

Then

$$e^t m\, u_m^s e^{-\frac{1}{2}u_m^2} = 1 + O\Big(\eta_m + \frac{(\log \log m)^2}{\log m}\Big), \quad m \to \infty.$$

Proof. We have

$$u_m^2 = 2\log m + 2t + s\log(2\log m) + O\Big(\eta_m + \frac{(\log\log m)^2}{\log m}\Big),$$
$$\log u_m = \tfrac{1}{2}\log(2\log m) + O\Big(\frac{\log\log m}{\log m}\Big).$$

From this, we get that

$$\log\left(e^t m\, u_m^s e^{-\frac{1}{2}u_m^2}\right) = O\Big(\frac{\log\log m}{\log m}\Big) + O\Big(\eta_m + \frac{(\log\log m)^2}{\log m}\Big)$$
$$= O\Big(\eta_m + \frac{(\log\log m)^2}{\log m}\Big),$$

and the result follows by applying the exponential. □

5.3.1. *Proof of Theorem 1*

Since the shape $\mathbf{h}^\star$ is given, in this section we let $\Xi(\mathbf{t}) = \Xi(\mathbf{h}^\star, \mathbf{h}^\star \circ \mathbf{t})$, indexed only by the spatial parameter $\mathbf{t} \in \mathcal{T} = \times_{j=1}^d [0, T_j]$ where $T_j = n/h_j^\star$. This is after rescaling, where we divided the jth coordinate by $h_j^\star$. Specifically, the reparametrized GRF has zero mean and covariance structure,

$$\mathrm{Cov}(\Xi(\mathbf{t}), \Xi(\mathbf{t}')) = \frac{\lambda(R(\mathbf{h}^\star, \mathbf{t}\circ\mathbf{h}^\star) \cap R(\mathbf{h}^\star, \mathbf{t}'\circ\mathbf{h}^\star))}{\sqrt{\lambda(R(\mathbf{h}^\star, \mathbf{t}\circ\mathbf{h}^\star))\lambda(R(\mathbf{h}^\star, \mathbf{t}'\circ\mathbf{h}^\star))}} = \lambda(R(\mathbf{t}) \cap R(\mathbf{t}')),$$

where $R(\mathbf{t}) := [\mathbf{t}, \mathbf{t}+\mathbf{1}]$. The GRF Ξ restricted to $\mathcal{T}$ is stationary, thus it is locally stationary over $\mathcal{T}$, but in $p = d$ dimensions. Moreover, it has the local structure $C_{\mathbf{t}}(\mathbf{s}) = \|\mathbf{s}\|_1$, by evaluating the local structure in Lemma 12 to the case in which $\mathbf{h} = \mathbf{1}$ and $\mathbf{g} = \mathbf{0}$. Hence, we know that it is homogeneous of order $\alpha = 1$ with $L = 1$ and $r_{\mathbf{t}}(\mathbf{u}/\|\mathbf{u}\|) = \|\mathbf{u}\|_1/\|\mathbf{u}\|$. Due to the restriction to $\mathcal{T}$, the tangent process of $\{\Xi(\mathbf{t})\}_{\mathbf{t}\in\mathcal{T}}$ must also be restricted $\mathcal{T}$. This will alter the high-excursion intensity from that given in Lemma 12, which we derive next.

In order to prove Lemma 12, [18] developed a technique for analyzing the tangent process using sums of independent Brownian motions. We use the same approach. First, note that a version of the tangent process is given by

$$U(\mathbf{s}) = \sum_{j=1}^d \sqrt{2} V_j(s_j), \quad \mathbf{s} = (s_1, \dots, s_d) \in \mathbb{R}_+^d,$$

where V_j are independent versions of the standard Brownian motion with drift $-|s_j|/\sqrt{2}$. (Notice that, when calculating the high excursion intensity Λ, the tangent process is restricted to the positive orthant.) To see that U is indeed a version of the tangent process, notice that, for all $\mathbf{s}, \mathbf{s}' \in \mathbb{R}_+^d$, $\mathbb{E}[U(\mathbf{s})] = -\|\mathbf{s}\|_1$ and

$$\mathrm{Cov}(U(\mathbf{s}), U(\mathbf{s}')) = 2\sum_{j=1}^d \mathrm{Cov}(V_j(s_j), V_j(s_j'))$$
$$= 2\sum_{j=1}^d s_j \wedge s_j' = \|\mathbf{s}\|_1 + \|\mathbf{s}'\|_1 - \|\mathbf{s} - \mathbf{s}'\|_1 \ .$$

Evaluating Λ,

$$\begin{aligned}\Lambda &= \lim_{T\to\infty}\frac{1}{T^d}\mathbb{E}\exp\left(\sup_{\mathbf{s}\in[0,T]^d}U(\mathbf{s})\right)\\ &= \left[\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\exp\left(\sup_{s\in[0,T]}\sqrt{2}V_1(s)\right)\right]^d = H_1^d,\end{aligned}$$

where $H_1 = 1$ is Pickands constant for $\alpha = 1$ [27]. We may now apply Lemma 13, and the high excursion probability becomes

$$\mathbb{P}\Big\{\sup_{\mathbf{t}\in K}\Xi(\mathbf{t}) > u\Big\} \sim \frac{\lambda(K)}{\sqrt{2\pi}}u^{2d-1}e^{-u^2/2}. \tag{5.15}$$

Until further notice, we take u to be the critical value (2.9). Recall that λ is the Lebesgue measure, here in $\mathbb{R}^d$. Define

$$\overline{\mathcal{I}} = \times_{j=1}^d[\lfloor T_j\rfloor], \quad \underline{\mathcal{I}} = \times_{j=1}^d[\lfloor T_j\rfloor - 1].$$

Consider the events $E_{\mathbf{i}} = \left\{\sup_{\mathbf{t}\in R(\mathbf{i})}\Xi(\mathbf{t}) > u\right\}$ for $\mathbf{i}\in\overline{\mathcal{I}}$. Notice that by translational invariance,

$$\forall \mathbf{i}\in\overline{\mathcal{I}}, \quad \mathbb{P}(E_{\mathbf{i}}) = \mathbb{P}(E_{\mathbf{0}}), \tag{5.16}$$

where, applying (5.15),

$$\mathbb{P}(E_{\mathbf{0}}) = \mathbb{P}\Big\{\sup_{\mathbf{t}\in R(\mathbf{0})}\Xi(\mathbf{t}) > u\Big\} \sim \frac{\lambda(R(\mathbf{0}))}{\sqrt{2\pi}}u^{2d-1}e^{-u^2/2} \sim e^{-\tau}\prod_{j=1}^d T_j^{-1}, \tag{5.17}$$

where the second equivalence comes from by applying Lemma 20.

We will now establish a Poisson limit for the above process over the entire set $\mathcal{T}$ based on finite range dependence. Two events, $E_{\mathbf{i}}, E_{\mathbf{i}'}$, are independent if $|i_j - i'_j| > 1$, for some $j\in[d]$. Consider thus the 'blanket' sets $B_{\mathbf{i}} = \{\mathbf{i}' \neq \mathbf{i} : |i_j - i'_j| \leq 1, \forall j\in[d]\}$, and note that $|B_{\mathbf{i}}| \leq 3^d$, for all $\mathbf{i}\in\overline{\mathcal{I}}$. Hence, by (5.16) and (5.17), and the fact that $|\overline{\mathcal{I}}| = O(\prod_j T_j)$, we have

$$A_1 := \sum_{\mathbf{i}\in\overline{\mathcal{I}}}\sum_{\mathbf{i}'\in B_{\mathbf{i}}}\mathbb{P}(E_{\mathbf{i}})\mathbb{P}(E_{\mathbf{i}'}) \leq |\overline{\mathcal{I}}|(3^d)\mathbb{P}(E_{\mathbf{0}})^2 = O(\textstyle\prod_j T_j^{-1}) = o(1).$$

Now take $\mathbf{i}\in\overline{\mathcal{I}}$ and $\mathbf{i}'\in B_{\mathbf{i}}$. We have

$$\mathbb{P}(E_{\mathbf{i}}\cap E_{\mathbf{i}'}) = 2\mathbb{P}(E_{\mathbf{0}}) - \mathbb{P}(E_{\mathbf{i}}\cup E_{\mathbf{i}'}).$$

We have (5.16) and (5.17), and as in (5.17), except that $\lambda(R(\mathbf{i})\cup R(\mathbf{i}')) = 2$ when $\mathbf{i}'\neq\mathbf{i}$, we also have

$$\mathbb{P}(E_{\mathbf{i}}\cup E_{\mathbf{i}'}) = \mathbb{P}\big\{\exists\mathbf{t}\in R(\mathbf{i})\cup R(\mathbf{i}') : \Xi(\mathbf{t}) > u\big\} \sim 2e^{-\tau}\prod_{j=1}^d T_j^{-1} \sim 2\mathbb{P}(E_{\mathbf{0}}).$$

This implies that

$$\mathbb{P}(E_{\mathbf{i}} \cap E_{\mathbf{i}'}) = o\big(\textstyle\prod_j T_j^{-1}\big).$$

This holds uniformly over $\mathbf{i}$ by translation invariance (translating the whole blanket set $B_{\mathbf{i}}$) and also uniformly over $\mathbf{i}'$ in the blanket because there are at most 3^d of these. Hence,

$$A_2 := \sum_{\mathbf{i}\in\overline{\mathcal{I}}} \sum_{\mathbf{i}'\in B_{\mathbf{i}}} \mathbb{P}(E_{\mathbf{i}} \cap E_{\mathbf{i}'}) \le |\overline{\mathcal{I}}|(3^d)\ o\big(\textstyle\prod_j T_j^{-1}\big) = o(1).$$

Finally, by (5.16) and (5.17),

$$M := \sum_{\mathbf{i}\in\overline{\mathcal{I}}} \mathbb{P}(E_{\mathbf{i}}) = |\overline{\mathcal{I}}|\mathbb{P}(E_{\mathbf{0}}) \sim e^{-\tau}|\overline{\mathcal{I}}| \prod_{j=1}^d T_j^{-1} = e^{-\tau}\prod_{j=1}^d \frac{\lceil T_j\rceil}{T_j} \to e^{-\tau}.$$

In our context, the Poisson approximation result stated in [4, Th 1] implies that

$$\Big|\mathbb{P}\big(\cap_{\mathbf{i}\in\overline{\mathcal{I}}}\, E_{\mathbf{i}}^{\mathsf{c}}\big) - e^{-M}\Big| \le A_1 + A_2,$$

from which we derive

$$\mathbb{P}\left(\cap_{\mathbf{i}\in\overline{\mathcal{I}}} E_{\mathbf{i}}^{\mathsf{c}}\right) \to e^{-e^{-\tau}}.$$

In exactly the same way, we can also derive

$$\mathbb{P}\left(\cap_{\mathbf{i}\in\underline{\mathcal{I}}} E_{\mathbf{i}}^{\mathsf{c}}\right) \to e^{-e^{-\tau}}.$$

Because

$$\mathbb{P}\left(\cap_{\mathbf{i}\in\overline{\mathcal{I}}} E_{\mathbf{i}}^{\mathsf{c}}\right) \le \mathbb{P}\{\exists \mathbf{t}\in\mathcal{T} : \Xi(\mathbf{t}) \le u\} \le \mathbb{P}\left(\cap_{\mathbf{i}\in\underline{\mathcal{I}}} E_{\mathbf{i}}^{\mathsf{c}}\right)$$

we conclude that

$$\mathbb{P}\{\exists \mathbf{t}\in\mathcal{T} : \Xi(\mathbf{t}) \le u\} \to e^{-e^{-\tau}}. \tag{5.18}$$

We can then express this result in terms of the behavior of $\tilde\tau$, defined in (5.11),

$$\mathbb{P}\left\{\tilde\tau \le \tau\right\} \to e^{-e^{-\tau}} = 1-\alpha,$$

when (2.11) holds. This being true for all fixed τ, by Lemma 17 with Lemma 19, for $\hat\tau$ defined in (5.10), we have

$$\mathbb{P}\{\hat\tau > \tau\} \sim \mathbb{P}\left\{\tilde\tau > \tau\right\} \to \alpha.$$

We then invert this to get

$$\lim_{n\to\infty} \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}\cap\mathbb{Z}^d} \xi[R(\mathbf{h}^\star, \mathbf{h}^\star\circ\mathbf{t})] > u\Big\} = \alpha,$$

which is what we needed to prove.

5.3.2. Proof of Theorem 2

We keep the same notation used in Section 5.3.1. While we worked under the null, we are now working under the alternative. Redefine $\mathbf{t}^\star$ such that $\mathbf{h}^\star \circ \mathbf{t}^\star$ is the true location of the rectangle of activation. Let $\mathcal{T}' = \mathcal{T} \cap \times_j (\mathbb{Z}/h_j^\star)$ and define

$$\mathcal{U}_\eta = \big\{\mathbf{t} \in \mathcal{T} : \lambda(R(\mathbf{t}) \cap R(\mathbf{t}^\star)) \geq 1 - \eta\big\}$$

and

$$\mathcal{U} = \big\{\mathbf{t} \in \mathcal{T} : R(\mathbf{t}) \cap R(\mathbf{t}') \neq \emptyset, \text{ for some } \mathbf{t}' \in \mathcal{U}_\eta\big\}.$$

Recall the definition of Υ in (5.2). In our present context, we can parameterize it by $\mathbf{t} \in \mathcal{T}$, and it satisfies

$$\Upsilon(\mathbf{t}) = \mu\lambda(R(\mathbf{t}) \cap R(\mathbf{t}^\star)) + \Xi(\mathbf{t}).$$

Throughout the following we assume that $\mu - v \to c \in \mathbb{R} \cup \{-\infty, +\infty\}$, where $v = v_n$ is defined in (2.10). Recall the definition of the power and write it as a function of c,

$$\beta(c) = \lim_{n\to\infty} \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}'} y[R((\mathbf{h}^\star, \mathbf{h}^\star \circ \mathbf{t}))] > u\Big\}.$$

Note that $\beta(c)$ is well defined by Slutsky's theorem and is clearly nondecreasing in c. Hence, it suffices to consider the case where $c \in \mathbb{R}$. By Lemma 16, Part 2, and the fact that $u \to \infty$,

$$\mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{U}} \Xi(\mathbf{t}) \geq u\Big\} = o(1).$$

Thus, since

$$\begin{aligned}&\mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}} \Xi(\mathbf{t}) \geq u\Big\} - \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{U}} \Xi(\mathbf{t}) \geq u\Big\} \\ &\qquad \leq \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}} \Xi(\mathbf{t}) \geq u\Big\} \leq \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}} \Xi(\mathbf{t}) \geq u\Big\},\end{aligned}$$

we have

$$\mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}} \Xi(\mathbf{t}) \geq u\Big\} \to \alpha, \tag{5.19}$$

by (5.18). Hence,

$$\begin{aligned}&\mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}} \Upsilon(\mathbf{t}) > u\Big\} \\ &\quad \geq \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}} \Upsilon(\mathbf{t}) > u\Big\} + \mathbb{P}\{\Upsilon(\mathbf{t}^\star) > u\}\, \mathbb{P}\Big\{\sup\nolimits_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}} \Upsilon(\mathbf{t}) \leq u\Big\} \\ &\quad \to \alpha + \bar{\Phi}(c)(1 - \alpha).\end{aligned}$$

Select $\eta \to 0$ such that $\mu\eta \to \infty$. By Lemma 16, Part 2, we know that

$$\sup\nolimits_{\mathbf{t}\in\mathcal{U}_\eta} \Upsilon(\mathbf{t}) - \Upsilon(\mathbf{t}^\star) \leq \sup\nolimits_{\mathbf{t}\in\mathcal{U}_\eta} |\Xi(\mathbf{t}) - \Xi(\mathbf{t}^\star)| = O_{\mathbb{P}}(1).$$

Hence,

$$\mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{U}_\eta}\Upsilon(\mathbf{t}) > u\Big\} \to \bar{\Phi}(c). \tag{5.20}$$

Again by Lemma 16, Part 2,

$$\sup_{\mathbf{t}\in\mathcal{U}\setminus\mathcal{U}_\eta}\Upsilon(\mathbf{t}) \le \mu(1-\eta) + O_{\mathbb{P}}(1).$$

Thus, in probability, $\mu - \sup_{\mathbf{t}\in\mathcal{U}\setminus\mathcal{U}_\eta}\Upsilon(\mathbf{t}) \to \infty$, implying that

$$\mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{U}\setminus\mathcal{U}_\eta}\Upsilon(\mathbf{t}) > u\Big\} \to 0. \tag{5.21}$$

We then have

$$\begin{aligned}\mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{T}}\Upsilon(\mathbf{t}) > u\Big\} &\le \mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}}\Upsilon(\mathbf{t}) > u\Big\} + \mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{U}\setminus\mathcal{U}_\eta}\Upsilon(\mathbf{t}) > u\Big\} \\ &\quad + \mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{U}_\eta}\Upsilon(\mathbf{t}) > u\Big\}\mathbb{P}\Big\{\sup_{\mathbf{t}\in\mathcal{T}\setminus\mathcal{U}}\Upsilon(\mathbf{t}) \le u\Big\} \\ &\to \alpha + \bar{\Phi}(c)(1-\alpha),\end{aligned}$$

where the inequality is by independence of $(\Xi(\mathbf{t}), \mathbf{t}\in\mathcal{U}_\eta)$ and $(\Xi(\mathbf{t}), \mathbf{t}\in\mathcal{T}\setminus\mathcal{U})$, and the convergence is by (5.19), (5.20), and (5.21). We conclude that

$$\beta(c) = \alpha + \bar{\Phi}(c)(1-\alpha),$$

and by Lemma 17 and Lemma 19, we find that this holds for the discrete scan statistic as long as $\underline{h} = \omega(\log n)$, so that (5.13) is satisfied.

5.3.3. Proof of Theorem 3

We now redefine u as the critical value in (2.13). We assume that we are under the null. Applying [18, Th 1.4], with $a \leftarrow 1$ and $n \leftarrow n/\underline{h}$, we get

$$\lim_{n\to\infty}\mathbb{P}\Big\{\sup_{\mathbf{w}\in\mathcal{W}}\Xi(\mathbf{w}) \ge u\Big\} = \alpha.$$

This translates into

$$\lim_{n\to\infty}\mathbb{P}\{\tilde{\tau} > \tau\} = \alpha.$$

Now we may apply Lemma 17 with Lemma 19 to obtain that the statistic $\hat{\tau}$ defined in (5.10) satisfies $|\tilde{\tau} - \hat{\tau}| = o_{\mathbb{P}}(1)$ and, therefore, we also have

$$\lim_{n\to\infty}\mathbb{P}\{\hat{\tau} > \tau\} = \alpha.$$

We then invert this to get

$$\lim_{n\to\infty}\mathbb{P}\Big\{\sup_{\mathbf{w}\in\mathcal{W}\cap\mathbb{Z}^{2d}}\xi[R(\mathbf{w})] \ge u\Big\} = \alpha.$$

5.3.4. Proof of Theorem 4

We assume that we are under the alternative. The arguments are essentially identical to those in Section 5.3.2, except that this time both the scale and location vary. In particular, we work with $\mathcal{W}' = \mathcal{W} \cap \mathbb{Z}^{2d}$, and

$$\mathcal{U}_\eta = \left\{ \mathbf{w} \in \mathcal{W} : \frac{\lambda(R(\mathbf{w}) \cap R(\mathbf{w}^\star))}{\sqrt{\lambda(R(\mathbf{w}))\lambda(R(\mathbf{w}^\star))}} \geq 1 - \eta \right\}$$

$$\mathcal{U} = \left\{ \mathbf{w} \in \mathcal{W} : R(\mathbf{w}) \cap R(\mathbf{w}') \neq \emptyset, \forall \mathbf{w}' \in \mathcal{U}_\eta \right\},$$

where $\mathbf{w}^\star$ denotes the true scale and location of the rectangle of activation. The remaining of the proof is now exactly the same.

5.3.5. Preliminaries

The lemmata stated and proved in this section will be used to prove of Theorems 5 and 6. Until further notice, $u(\mathbf{h})$ (or $u_n(\mathbf{h})$ if we choose not to suppress the dependence on n) denotes the critical value defined in (2.16) while $v(\mathbf{h}) = v_{n,\mathbf{h}}$ denotes the function of $\mathbf{h}$ in (2.17). The parameter τ remains fixed throughout.

The following technical lemma is used throughout this section.

Lemma 21. *There exists $L > 0$ such that for all $\mathbf{w} = (\mathbf{h}, \mathbf{t}), \mathbf{w}' = (\mathbf{h}', \mathbf{t}') \in \mathcal{W}$ such that $\delta(\mathbf{w}, \mathbf{w}') \leq \epsilon_0$ as specified in Lemma 18,*

$$|u(\mathbf{h}) - u(\mathbf{h}')| \leq L\delta(\mathbf{w}, \mathbf{w}').$$

Proof. Recall the notation introduced in (5.10), where we now abbreviate $a(\mathbf{h}) = a((\mathbf{h}, \mathbf{t}))$ and $b(\mathbf{h}) = b((\mathbf{h}, \mathbf{t}))$, and these functions are specified in Lemma 18. We have

$$|u(\mathbf{h}) - u(\mathbf{h}')| \leq |a(\mathbf{h}) - a(\mathbf{h}')| + \tau \Big| \frac{1}{a(\mathbf{h})} - \frac{1}{a(\mathbf{h}')} \Big| + \Big| \frac{b(\mathbf{h})}{a(\mathbf{h})} - \frac{b(\mathbf{h}')}{a(\mathbf{h}')} \Big|.$$

By Lemma 18, $|a(\mathbf{h}) - a(\mathbf{h}')| \leq L\delta(\mathbf{w}', \mathbf{w})$ for some $L > 0$ and $\delta(\mathbf{w}', \mathbf{w}) \leq \epsilon_0$. Working with the second term, we obtain,

$$\Big| \frac{1}{a(\mathbf{h})} - \frac{1}{a(\mathbf{h}')} \Big| = \frac{|a(\mathbf{h}) - a(\mathbf{h}')|}{a(\mathbf{h})a(\mathbf{h}')} \leq L \frac{\delta(\mathbf{w}, \mathbf{w}')}{a(\mathbf{h})a(\mathbf{h}')} \leq L\delta(\mathbf{w}, \mathbf{w}'),$$

using the fact that $a(\mathbf{h}), a(\mathbf{h}') \geq 1$ because $n/\underline{h} \geq n/\overline{h} \geq e$ by assumption. Working with the third term,

$$\begin{aligned} \Big| \frac{b(\mathbf{h})}{a(\mathbf{h})} - \frac{b(\mathbf{h}')}{a(\mathbf{h}')} \Big| &\leq \Big| \frac{b(\mathbf{h})}{a(\mathbf{h})} - \frac{b(\mathbf{h}')}{a(\mathbf{h})} \Big| + \Big| \frac{b(\mathbf{h}')}{a(\mathbf{h})} - \frac{b(\mathbf{h}')}{a(\mathbf{h}')} \Big| \\ &\leq \frac{L\delta(\mathbf{w}, \mathbf{w}')}{a(\mathbf{h})} + \frac{b(\mathbf{h}')}{a(\mathbf{h}')} \Big| \frac{a(\mathbf{h}) - a(\mathbf{h}')}{a(\mathbf{h})} \Big| \leq L(1 + C)\delta(\mathbf{w}, \mathbf{w}'), \end{aligned}$$

because there exists a C such that $b(\mathbf{h})/a(\mathbf{h}) \leq C$ for all allowed $\mathbf{h}$. Combining these we find that $u(\mathbf{h})$ is indeed (locally) Lipschitz with respect to δ, with constant $L' = L(2 + \tau + C)$. ∎

We will introduce some notation for the following lemmata. Let $\mathbf{h} \in [e\underline{h}, \overline{h}]^d$, define $\mathcal{T}(\mathbf{h}) = \times_{j=1}^d h_j[\lceil n/h_j \rceil]$, and let $\mathbf{t} \in \mathcal{T}(\mathbf{h})$. Define the set

$$K_{(\mathbf{h},\mathbf{t})} = \Big\{ (\mathbf{g},\mathbf{s}) \in \mathcal{W} : \mathbf{g}/\mathbf{h} \in [e^{-1}, 1]^d, \mathbf{s} \in [\mathbf{t} - \mathbf{h}, \mathbf{t}] \Big\}$$

and the event

$$E_{(\mathbf{h},\mathbf{t})} = \Big\{ \exists (\mathbf{g},\mathbf{s}) \in K_{(\mathbf{h},\mathbf{t})} : \Xi(\mathbf{g},\mathbf{s}) > u(\mathbf{g}) \Big\}. \tag{5.22}$$

Lemma 22. *Let $\mathbf{k} \in \mathbb{Z}_+^d$ be fixed in n, define $\mathbf{h} = \underline{h} e^{\mathbf{k}}$, and let $\mathbf{t} \in \mathcal{T}(\mathbf{h})$. We have*

$$\mathbb{P}\big(E_{(\mathbf{h},\mathbf{t})}\big) \sim e^{-\tau} n^{-d} \prod_{j=1}^d h_j [k_j^{-1} - (1+k_j)^{-1}].$$

Proof. First, by location invariance

$$\mathbb{P}\Big\{ \exists (\mathbf{g},\mathbf{s}) \in K_{(\mathbf{h},\mathbf{t})} : \Xi(\mathbf{g},\mathbf{s}) > u_n(\mathbf{g}) \Big\} = \mathbb{P}\Big\{ \exists (\mathbf{g},\mathbf{s}) \in K_{(\mathbf{h},\mathbf{h})} : \Xi(\mathbf{g},\mathbf{s}) > u_n(\mathbf{g}) \Big\}.$$

Also, because for $(\mathbf{g}_0, \mathbf{s}_0), (\mathbf{g}_1, \mathbf{s}_1) \in \mathcal{W}$,

$$\frac{\lambda(R(\underline{h}^{-1}\mathbf{g}_0, \underline{h}^{-1}\mathbf{s}_0) \cap R(\underline{h}^{-1}\mathbf{g}_1, \underline{h}^{-1}\mathbf{s}_1))}{\sqrt{\lambda(R(\underline{h}^{-1}\mathbf{g}_0, \underline{h}^{-1}\mathbf{s}_0))\lambda(R(\underline{h}^{-1}\mathbf{g}_1, \mathbf{a}\underline{h}^{-1}\mathbf{s}_1))}} = \frac{\lambda(R(\mathbf{g}_0, \mathbf{s}_0) \cap R(\mathbf{g}_1, \mathbf{s}_1))}{\sqrt{\lambda(R(\mathbf{g}_0, \mathbf{s}_0))\lambda(R(\mathbf{g}_1, \mathbf{s}_1))}},$$

rescaling the set $\mathcal{W}$ by $\underline{h}^{-1}$ does not change the covariance structure of $\Xi[\bar{\mathcal{R}}]$. Thus,

$$\mathbb{P}\Big\{ \exists (\mathbf{g},\mathbf{s}) \in K_{(\mathbf{h},\mathbf{h})} : \Xi(\mathbf{g},\mathbf{s}) \geq u_n(\mathbf{g}) \Big\} = \mathbb{P}\Big\{ \exists (\mathbf{g},\mathbf{s}) \in K_0 : \Xi(\mathbf{g},\mathbf{s}) \geq u_n(\underline{h}\mathbf{g}) \Big\}. \tag{5.23}$$

for

$$K_0 = \underline{h}^{-1} K_{(\mathbf{h},\mathbf{h})} = [e^{\mathbf{k}-\mathbf{1}}, e^{\mathbf{k}}] \times [\mathbf{0}, e^{\mathbf{k}}].$$

Let $\Lambda(\mathbf{h}) = 4^{-d} \prod_{i=1}^d h_i^{-2}$ be the high excursion intensity from Lemma 12. Set $c = 4^d \sqrt{2\pi}$. We now check the conditions of Lemma 14. Here the boundary functions are $u_n(\mathbf{h})$ defined in (2.16), and satisfy (5.5) with $\zeta = (\log n)^{-3/2}$. To see this, first, notice that K_0 is fixed and Jordan measurable. By (5.9), for any fixed $\gamma > 0$ small enough and $\mathbf{w}_0 = (\mathbf{g}_0, \mathbf{s}_0), \mathbf{w}_1 = (\mathbf{g}_1, \mathbf{s}_1) \in [K_0]_\gamma$ such that $\|\mathbf{w}_0 - \mathbf{w}_1\|_\infty \leq \zeta$,

$$\delta^2(\underline{h}\mathbf{w}_0, \underline{h}\mathbf{w}_1) = \delta^2(\mathbf{w}_0, \mathbf{w}_1) \leq 2 \sum_j \frac{|g_{0,j} - g_{1,j}| + |s_{0,j} - s_{1,j}|}{\sqrt{g_{0,j} g_{1,j}}} \leq 4\zeta C_\gamma \sum_j e^{1-k_j}$$

where C_γ is a small constant. Thus by Lemma 21, for ζ small enough,

$$|u_n(\underline{h}\mathbf{g}_0) - u_n(\underline{h}\mathbf{g}_1)| \leq L\delta(\underline{h}\mathbf{w}_0, \underline{h}\mathbf{w}_1) \leq L\sqrt{4\zeta C_\gamma \sum_j e^{1-k_j}}.$$

We also have that

$$\sup_{(\mathbf{g},\mathbf{s})\in[K_0]_{2\zeta}} u_n(\mathbf{g}) = O(\sqrt{\log n}) \tag{5.24}$$

and so

$$|u_n(\underline{h}\mathbf{g}_0)^2 - u_n(\underline{h}\mathbf{g}_1)^2| = O(\sqrt{\zeta \log n}) = o(1) \tag{5.25}$$

uniformly over such $\mathbf{w}_0, \mathbf{w}_1$, which verifies (5.6). Furthermore, recalling that in the notation of Lemma 14, we have $\alpha = 1$ and $p = 2d$, we finally get (5.5)

$$\sup_{(\mathbf{g},\mathbf{s})\in[K_0]_{\zeta}} u_n(\mathbf{g})^{-4d} = O((\log n)^{-2d}) = o(\zeta).$$

Hence, we have established the conditions of Lemma 14. Applying it we have

$$\begin{aligned}
&\mathbb{P}\Big\{\exists(\mathbf{g},\mathbf{s}) \in K_0 : \Xi(\mathbf{g},\mathbf{s}) > u_n(\underline{h}\mathbf{g})\Big\}\\
&\sim \int_{[\mathbf{0},e^{\mathbf{k}}]}\int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} \psi(u_n(\underline{h}\mathbf{g}))\Lambda(\mathbf{g})u_n(\underline{h}\mathbf{g})^{4d}\mathrm{d}\mathbf{g}\mathrm{d}\mathbf{t}\\
&= \frac{e^{\sum_j k_j}}{\sqrt{2\pi}}\int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} \Lambda(\mathbf{g})u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2}\mathrm{d}\mathbf{g}\\
&= \frac{e^{\sum_j k_j}}{4^d\sqrt{2\pi}}\int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} \Big(\prod_j g_j^{-2}\Big)u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2}\mathrm{d}\mathbf{g}.
\end{aligned}$$

We have from Lemma 20 that

$$u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2}\Big/\Big[\frac{c}{e^{\tau}\prod_j(n\log^2(eg_j)/(\underline{h}g_j))}\Big] \to 1, \quad \forall \mathbf{g} \in [e^{\mathbf{k-1}}, e^{\mathbf{k}}],$$

which implies that

$$\frac{n^d}{\underline{h}^d}u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2} \to ce^{-\tau}\prod_j \frac{g_j}{\log^2(eg_j)}, \quad \forall \mathbf{g} \in [e^{\mathbf{k-1}}, e^{\mathbf{k}}].$$

We see by Lemma 20, uniformly over n,

$$\sup_{(\mathbf{g},\mathbf{s})\in K_0} \frac{n^d}{\underline{h}^d}u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2} < +\infty.$$

By dominated convergence,

$$\begin{aligned}
\frac{n^d}{\underline{h}^d}\int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} &\prod_j g_j^{-2}(u_n(\underline{h}\mathbf{g}))^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2}\mathrm{d}\mathbf{g}\\
&\to \int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} ce^{-\tau}\prod_j[g_j\log^2(eg_j)]^{-1}\mathrm{d}\mathbf{g}.
\end{aligned}$$

Thus,

$$\frac{e^{\sum_j k_j}}{4^d\sqrt{2\pi}}\int_{[e^{\mathbf{k-1}},e^{\mathbf{k}}]} \Big(\prod_j g_j^{-2}\Big)u_n(\underline{h}\mathbf{g})^{4d-1}e^{-u_n(\underline{h}\mathbf{g})^2/2}\mathrm{d}\mathbf{g}$$

$$\sim \frac{\underline{h}^d e^{\sum_j k_j}}{n^d e^\tau} \prod_j [k_j^{-1} - (1+k_j)^{-1}].$$

We have our result because $h_j = \underline{h} e^{k_j}, j \in [d]$. $\square$

Lemma 23. *Resume the notation of Lemma 22. There exists a constant $C > 0$ not depending on n, $\mathbf{k}$, or $\mathbf{t}$ (but possibly dependent on τ or d) such that*

$$\mathbb{P}\big(E_{(\mathbf{h},\mathbf{t})}\big) \le C \prod_j \frac{h_j}{n k_j^2}.$$

Proof. Let $\mathbf{w} = (\mathbf{h}, \mathbf{t})$. By (5.23),

$$\mathbb{P}(E_\mathbf{w}) = \mathbb{P}\Big\{\exists (\mathbf{g},\mathbf{s}) \in K_0 : \Xi(\mathbf{g},\mathbf{s}) > u(\underline{h}\mathbf{g})\Big\} \le \mathbb{P}\Big\{\exists (\mathbf{g},\mathbf{s}) \in K_0 : \Xi(\mathbf{g},\mathbf{s}) > c_\mathbf{w}\Big\}$$

where

$$c_\mathbf{w} = \min_{(\mathbf{g},\mathbf{s}) \in K_\mathbf{w}} u(\mathbf{g}) = \min_{\mathbf{g} \in [e^{-1}\mathbf{h}, \mathbf{h}]} u(\mathbf{g}).$$

By scale invariance,

$$\begin{aligned} \mathbb{P}\Big\{\exists (\mathbf{g},\mathbf{s}) \in K_0 : \Xi(\mathbf{g},\mathbf{s}) \ge c_\mathbf{w}\Big\} &= \mathbb{P}\Big\{\exists \mathbf{g} \in [e^{-1},1]^d, \mathbf{s} \in [0,1]^d : \Xi(\mathbf{g},\mathbf{s}) \ge c_\mathbf{w}\Big\} \\ &\le C_1 c_\mathbf{w}^{4d} \psi(c_\mathbf{w}) \int_{\mathbf{w}' \in [e^{-1},1]^d \times [0,1]^d} \Lambda(\mathbf{w}')\mathrm{d}\mathbf{w}', \end{aligned}$$

for some constant $C_1 > 0$, by an application of Lemma 13. On the one hand, using the form for Λ given in Lemma 12, we get

$$\int_{\mathbf{w}' \in [e^{-1},1]^d \times [0,1]^d} \Lambda(\mathbf{w}')\mathrm{d}\mathbf{w}' = 4^{-d} \prod_j \int_{e^{-1}}^1 {g_j}^{-2} \mathrm{d}g_j < \infty.$$

On the other hand, by Lemma 20 there is a constant C_2 (not dependent on $\mathbf{k}$, n, or $\mathbf{t}$) such that $\forall (\mathbf{g},\mathbf{s}) \in K_0$,

$$u(\underline{h}\mathbf{g})^{4d} \psi(u(\underline{h}\mathbf{g})) \le e^{-\tau-\kappa} e^{-v(\underline{h}\mathbf{g})^2/2} \Big(1 + C_2 \frac{\log v(\underline{h}\mathbf{g})}{v(\underline{h}\mathbf{g})}\Big).$$

Since $\min_{\mathbf{g} \in [e^{\mathbf{k}-\mathbf{1}}, e^\mathbf{k}]} v(\underline{h}\mathbf{g}) \to \infty$ uniformly over $\mathbf{k} \in \mathbb{Z}_+^d$, we have

$$c_\mathbf{w}^{4d} \psi(c_\mathbf{w}) \le (1+o(1)) e^{-\tau-\kappa} \exp\Big[-\tfrac{1}{2} \min_{\mathbf{g} \in [e^{\mathbf{k}-\mathbf{1}}, e^\mathbf{k}]} v^2(\underline{h}\mathbf{g})\Big] \le C_3 \prod_j \frac{h_j}{n} \log^{-2}(h_j/\underline{h}),$$

where C_3 is some constant, using the fact that $\min_{\mathbf{h}' \in [e^{-1}\mathbf{h}, \mathbf{h}]} v(\mathbf{h}') = v(\mathbf{h})$ for $h_j \ge e\underline{h}, \forall j$. We conclude that there exists a constant C_4 such that, for all such $\mathbf{w}$,

$$\mathbb{P}(E_\mathbf{w}) \le C_4 \prod_j \frac{h_j}{n} \log^{-2}(h_j/\underline{h}). \qquad \square$$

Lemma 24. *For all $A \in \mathbb{Z}_+$, let $\mathcal{U}_A = \{(\mathbf{h},\mathbf{t}) \in \mathcal{W} : h_j \leq \underline{h}e^A, \forall j \in [d]\}$. Then*

$$\lim_{A\to\infty}\lim_{n\to\infty} \mathbb{P}\Big\{\Xi(\mathbf{h},\mathbf{t}) > u_n(\mathbf{h}) \text{ for some } (\mathbf{h},\mathbf{t}) \in \mathcal{U}_A\Big\} = \alpha.$$

Proof. Resume the notation and definitions of Lemma 23. We partition the space $\mathcal{W}$ into blocks in the scale and location parameters. Define

$$\overline{\mathcal{I}} = \Big\{(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \exists \mathbf{k} \in [A]^d, \mathbf{h} = \underline{h}e^{\mathbf{k}}, \mathbf{t} \in \mathcal{T}(\mathbf{h})\Big\}$$

and $\underline{\mathcal{I}} = \overline{\mathcal{I}} \cap \mathcal{U}_A$ whereby $\cup_{\mathbf{w}\in\underline{\mathcal{I}}} K_{\mathbf{w}} \subseteq \mathcal{U}_A \subseteq \cup_{\mathbf{w}\in\overline{\mathcal{I}}} K_{\mathbf{w}}$. Recall that for $\mathbf{k} \in [A]^d$ and $\mathbf{h} = \underline{h}e^{\mathbf{k}}$,

$$\mathbb{P}\big(E_{(\mathbf{h},\mathbf{t})}\big) = \mathbb{P}\big(E_{(\mathbf{h},\mathbf{h})}\big), \quad \forall \mathbf{t} \in \mathcal{T}(\mathbf{h})$$

by translation invariance. By Lemma 22,

$$\mathbb{P}\big(E_{(\mathbf{h},\mathbf{h})}\big) \sim |\mathcal{T}(\mathbf{h})|^{-1} e^{-\tau} \prod_j \big[k_j^{-1} - (1+k_j)^{-1}\big].$$

We partition the set $\mathcal{U}_A$ into the blocks $\{K_{\mathbf{w}} : \mathbf{w} \in \overline{\mathcal{I}}\}$ and then use the Chen-Stein Poisson approximation to derive $\mathbb{P}(\cup_{\mathbf{w}\in\overline{\mathcal{I}}} E_{\mathbf{w}})$. We have

$$\begin{aligned} M := \sum_{\mathbf{w}\in\overline{\mathcal{I}}} \mathbb{P}(E_{\mathbf{w}}) &= \sum_{\mathbf{k}\in[A]^d} \sum_{\mathbf{t}\in\mathcal{T}(\underline{h}e^{\mathbf{k}})} \mathbb{P}\big(E_{(\underline{h}e^{\mathbf{k}},\mathbf{t})}\big) = \sum_{\mathbf{k}\in[A]^d} |\mathcal{T}(\underline{h}e^{\mathbf{k}})|\mathbb{P}\big(E_{(\underline{h}e^{\mathbf{k}},\underline{h}e^{\mathbf{k}})}\big) \\ &\to e^{-\tau} \sum_{\mathbf{k}\in[A]^d} \prod_j \big[k_j^{-1} - (1+k_j)^{-1}\big]. \end{aligned}$$

We then have that

$$\sum_{\mathbf{k}\in[A]^d} \prod_j \big[k_j^{-1} - (1+k_j)^{-1}\big] = \Big(\textstyle\sum_{k=1}^A [k^{-1} - (1+k)^{-1}]\Big)^d = \Big(1 - 1/(1+A)\Big)^d.$$

Thus, we obtain that

$$\lim_{A\to\infty}\lim_{n\to\infty} M \to e^{-\tau}.$$

Two events, $E_{(\mathbf{h},\mathbf{t})}, E_{(\mathbf{g},\mathbf{s})}$, are independent if $|t_j - s_j| > 2(h_j \vee g_j)$, for some $j \in [d]$. Consider then the 'blanket' sets

$$B_{(\mathbf{h},\mathbf{t})} = \Big\{(\mathbf{g},\mathbf{s}) \in \overline{\mathcal{I}} \setminus \{(\mathbf{h},\mathbf{t})\} : \forall j \in [d], |t_j - s_j| \leq 2(h_j \vee g_j)\Big\}.$$

We have

$$\begin{aligned} |B_{(\mathbf{h},\mathbf{t})}| &\leq \sum_{\mathbf{k}\in[A]^d} \Big|\Big\{\mathbf{s} \in \mathcal{T}(\underline{h}e^{\mathbf{k}}) : \exists j \in [d], |t_j - s_j| \leq 2\underline{h}e^A\Big\}\Big| \\ &\leq 4^d \sum_{\mathbf{k}\in[A]^d} \Big\lceil e^{\sum_{j=1}^d (A-k_j)} \Big\rceil. \end{aligned}$$

which is a constant depending only on d and A. Thus,

$$A_1 := \sum_{\mathbf{w}\in\overline{\mathcal{I}}} \sum_{\mathbf{w}'\in B_{\mathbf{w}}} \mathbb{P}(E_{\mathbf{w}})\mathbb{P}(E_{\mathbf{w}'})$$
$$\leq \Big(\max_{\mathbf{w}\in\overline{\mathcal{I}}} |B_{\mathbf{w}}|\mathbb{P}(E_{\mathbf{w}}) \Big) \sum_{\mathbf{w}\in\overline{\mathcal{I}}} \mathbb{P}(E_{\mathbf{w}}) = o\Big(\sum_{\mathbf{w}\in\overline{\mathcal{I}}} \mathbb{P}(E_{\mathbf{w}}) \Big),$$

since $\max_{\mathbf{w}\in\overline{\mathcal{I}}} \mathbb{P}(E_{\mathbf{w}}) = o(1)$ by Lemma 23. Take $\mathbf{w} \in \overline{\mathcal{I}}$ and $\mathbf{w}' = (\mathbf{h}', \mathbf{t}') \in B_{\mathbf{w}}$. We have

$$\mathbb{P}(E_{\mathbf{w}} \cap E_{\mathbf{w}'}) = \mathbb{P}(E_{\mathbf{w}}) + \mathbb{P}(E_{\mathbf{w}'}) - \mathbb{P}(E_{\mathbf{w}} \cup E_{\mathbf{w}'}).$$

By same exact arguments underlying the proof of Lemma 22,

$$\mathbb{P}(E_{\mathbf{w}} \cup E_{\mathbf{w}'}) = \mathbb{P}\Big\{ \exists(\mathbf{h}_0, \mathbf{t}_0) \in K_{\mathbf{w}} \cup K_{\mathbf{w}'} : \Xi(\mathbf{h}_0, \mathbf{t}_0) > u(\mathbf{h}_0) \Big\} \sim \mathbb{P}(E_{\mathbf{w}}) + \mathbb{P}(E_{\mathbf{w}'}).$$

We can also see from Lemma 23 that, uniformly over $\mathbf{w}' \in B_{\mathbf{w}}$,

$$\mathbb{P}(E_{\mathbf{w}'}) = O(\mathbb{P}(E_{\mathbf{w}})).$$

Again by translation invariance and the fact that both $|[A]^d|$ and $|B_{\mathbf{w}}|$ are bounded in n,

$$A_2 := \sum_{\mathbf{k}\in[A]^d} \sum_{\mathbf{t}\in\mathcal{T}(\underline{h}e^{\mathbf{k}})} \sum_{\mathbf{w}'\in B_{(\underline{h}e^{\mathbf{k}},\mathbf{t})}} \mathbb{P}(E_{(\underline{h}e^{\mathbf{k}},\mathbf{t})} \cap E_{\mathbf{w}'})$$
$$= \sum_{\mathbf{k}\in[A]^d} |\mathcal{T}(\underline{h}e^{\mathbf{k}})| \sum_{\mathbf{w}'\in B_{(\underline{h}e^{\mathbf{k}},\underline{h}e^{\mathbf{k}})}} o\big[\mathbb{P}(E_{(\underline{h}e^{\mathbf{k}},\underline{h}e^{\mathbf{k}})}) + \mathbb{P}(E_{\mathbf{w}'})\big]$$
$$= o(M) = o(1).$$

This shows that the events $E_{\mathbf{w}}$ over $\overline{\mathcal{I}}$ have finite-range dependence. Hence, by [4, Th 1] we have that

$$\left| \mathbb{P}\left(\cap_{\mathbf{w}\in\overline{\mathcal{I}}} E^{\mathsf{c}}_{\mathbf{w}} \right) - e^{-M} \right| \leq A_1 + A_2 = o(1).$$

This also holds with $\underline{\mathcal{I}}$ in place of $\overline{\mathcal{I}}$, and with $\lim_{n\to\infty} M$ unaffected. So the proof is complete. □

Lemma 25. *With $\mathcal{U}_A$ defined in Lemma 24, we also have*

$$\lim_{A\to\infty} \lim_{n\to\infty} \mathbb{P}\big\{ \exists(\mathbf{h}, \mathbf{t}) \in \mathcal{W} \setminus \mathcal{U}_A : \Xi(\mathbf{h}, \mathbf{t}) > u(\mathbf{h}) \big\} = 0.$$

Proof. We keep the same notation as in the previous proof. Define the event

$$\mathcal{E}_A = \Big\{ \exists(\mathbf{h}, \mathbf{t}) \in \mathcal{W} \backslash \mathcal{U}_A : \Xi(\mathbf{h}, \mathbf{t}) > u(\mathbf{h}) \Big\}.$$

Note that $\mathcal{E}_A$ depends on n via $u(\mathbf{h})$ in (2.16). By the union bound,

$$\mathbb{P}(\mathcal{E}_A) \leq \sum_{\mathbf{k}\in[\log n]^d \backslash [A]^d} p_{n,\mathbf{k}}, \tag{5.26}$$

where

$$p_{n,\mathbf{k}} := \sum_{\mathbf{t}\in\mathcal{T}(\underline{h}e^{\mathbf{k}})} \mathbb{P}(E_{(\underline{h}e^{\mathbf{k}},\mathbf{t})}) = \left(\prod_j \lceil n/\underline{h}e^{k_j}\rceil\right)\mathbb{P}(E_{(\underline{h}e^{\mathbf{k}},\underline{h}e^{\mathbf{k}})}),$$

by translation invariance. By Lemma 22,

$$\begin{aligned}
&\sum_{\mathbf{k}\in\mathbb{Z}_+^d\setminus[A]^d} \lim_{n\to\infty} p_{n,\mathbf{k}} \\
&= e^{-\tau} \sum_{\mathbf{k}\in\mathbb{Z}_+^d\setminus[A]^d} \prod_{j=1}^d \left[k_j^{-1} - (1+k_j)^{-1}\right] \\
&= e^{-\tau}\Big(\sum_{\mathbf{k}\in\mathbb{Z}_+^d} \prod_j \left[k_j^{-1} - (1+k_j)^{-1}\right] - \sum_{\mathbf{k}\in[A]^d} \prod_j \left[k_j^{-1} - (1+k_j)^{-1}\right]\Big) \\
&= e^{-\tau}\Big(1 - \big(1 - 1/(1+A)\big)^d\Big).
\end{aligned}$$

By Lemma 23,

$$p_{n,\mathbf{k}} \le \Big(\prod_j \lceil n/h_j\rceil\Big) C \prod_j \frac{\underline{h}e^{k_j}}{nk_j^2} \le C_1 \prod_j k_j^{-2}$$

for $C_1 > 0$ not dependent on n or $\mathbf{k}$. Hence, $D_{\mathbf{k}} := C_1 \prod_j k_j^{-2}$ is a dominating sequence that is independent of n and summable over $\mathbb{Z}_+^d$, and satisfies $p_{n,\mathbf{k}} \le D_{\mathbf{k}}$. Thus, we can apply dominated convergence and conclude that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_{n,A}) &\le \lim_{n\to\infty} \sum_{\mathbf{k}\in\mathbb{Z}_+^d\setminus[A]^d} p_{n,\mathbf{k}} \\
&= \sum_{\mathbf{k}\in\mathbb{Z}_+^d\setminus[A]^d} \lim_{n\to\infty} p_{n,\mathbf{k}} = e^{-\tau}\Big(1 - \big(1 - 1/(1+A)\big)^d\Big).
\end{aligned}$$

Since the RHS tends to zero as $A \to \infty$, the proof is complete. ☐

5.3.6. *Proof of Theorem 5*

By Lemma 24 and Lemma 25,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\Big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\Big\} \qquad (5.27)\\
&\quad = \lim_{A\to\infty}\Big[\lim_{n\to\infty}\Big(\mathbb{P}\Big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U}_A : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\Big\} \\
&\qquad + \lim_{n\to\infty} \mathbb{P}\Big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W}\setminus\mathcal{U}_A : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\Big\}\Big] \\
&\quad = \alpha. \qquad (5.28)
\end{aligned}$$

Hence, the random variable $\tilde{\tau}$ defined in (5.11) satisfies

$$\lim_{n\to\infty} \mathbb{P}\{\tilde{\tau} > \tau\} = \lim_{n\to\infty} \mathbb{P}\Big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\Big\} = \alpha.$$

Now, we may apply Lemma 17 with Lemma 19 to obtain that the statistic $\hat{\tau}$ defined in (5.10) satisfies $|\tilde{\tau} - \hat{\tau}| = o_{\mathbb{P}}(1)$ and, therefore,

$$\lim_{n\to\infty} \mathbb{P}\Big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} \cap \mathbb{Z}^{2d} : \xi[R(\mathbf{h},\mathbf{t})] > u(\mathbf{h})\Big\} = \lim_{n\to\infty} \mathbb{P}\{\hat{\tau} > \tau\} = \alpha.$$

5.3.7. Proof of Theorem 6

We resume the notation introduced in Sections 5.3.2 and 5.3.4. The arguments here are very similar, so that we will omit some details. We focus on the case in which $\mu - v(\mathbf{h}^\star) \to c$. By Lemma 16, Part 2, and the fact that $\min_{\mathbf{h}} v(\mathbf{h}) \to \infty$,

$$\begin{aligned}&\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U} : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\ &\quad\le \mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U} : \Xi(\mathbf{h},\mathbf{t}) > v(\mathbf{h}) - o(1)\big\} = o_{\mathbb{P}}(1).\end{aligned}$$

Thus, combining this with (5.28), we have

$$\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W}\backslash\mathcal{U} : \Xi(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \to \alpha.$$

Hence,

$$\begin{aligned}\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} &\ge \mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W}\backslash\mathcal{U} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\ &\quad + \mathbb{P}\big\{\Upsilon(\mathbf{w}^\star) > u(\mathbf{h}^\star)\big\}\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W}\backslash\mathcal{U} : \Upsilon(\mathbf{h},\mathbf{t}) \le u(\mathbf{h})\big\} \\ &\quad \to \alpha + \bar{\Phi}(c)(1-\alpha).\end{aligned}$$

By Lemma 21, there is some $L > 0$ such that for all $\mathbf{w} = (\mathbf{h},\mathbf{t}) \in \mathcal{U}_\eta$, $u(\mathbf{h}^\star) - u(\mathbf{h}) \le L\delta(\mathbf{w}^\star,\mathbf{w})$ for $\delta(\mathbf{w}^\star,\mathbf{w}) \le \epsilon_0$. Select $\eta \to 0$ such that $\mu\eta \to \infty$. For $\mathbf{w} = (\mathbf{h},\mathbf{t}) \in \mathcal{U}_\eta$,

$$\begin{aligned}&(\Upsilon(\mathbf{w}) - u(\mathbf{h})) - (\Upsilon(\mathbf{w}^\star) - u(\mathbf{h}^\star)) \\ &\quad= [\Xi(\mathbf{w}) - \Xi(\mathbf{w}^\star)] + [m(\mathbf{w}) - m(\mathbf{w}^\star)] + [u(\mathbf{h}^\star) - u(\mathbf{h})] \\ &\quad\le |\Xi(\mathbf{w}) - \Xi(\mathbf{w}^\star)| + L\delta(\mathbf{w},\mathbf{w}^\star).\end{aligned}$$

By Lemma 16, Part 2,

$$\sup_{\mathbf{w}\in\mathcal{U}_\eta} |\Xi(\mathbf{w}) - \Xi(\mathbf{w}^\star)| = O_{\mathbb{P}}(1).$$

By this, the fact that if $\eta \to 0$ then $\sup_{\mathbf{w}\in\mathcal{U}_\eta} \delta(\mathbf{w},\mathbf{w}^\star) \to 0$, and that $m(\mathbf{w}^\star) \ge m(\mathbf{w})$,

$$\sup_{\mathbf{w}\in\mathcal{U}_\eta} [\Upsilon(\mathbf{w}) - u(\mathbf{h})] - [\Upsilon(\mathbf{w}^\star) - u(\mathbf{h}^\star)] = O_{\mathbb{P}}(1).$$

Hence,

$$\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U}_\eta : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \to \bar{\Phi}(c).$$

Again by Lemma 16, Part 2,

$$\sup_{\mathbf{w}\in\mathcal{U}\backslash\mathcal{U}_\eta} \Upsilon(\mathbf{w}) \le \mu(1-\eta) + O_{\mathbb{P}}(1).$$

Thus,

$$\mu - \sup_{\mathbf{w} \in \mathcal{U} \setminus \mathcal{U}_\eta} \Upsilon(\mathbf{w}) \to \infty.$$

Hence,

$$\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U} \setminus \mathcal{U}_\eta : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \to 0.$$

The probability of exceedance can be bounded by

$$\begin{aligned}
&\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\
&\quad \le \mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} \setminus \mathcal{U} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\
&\qquad + \mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U} \setminus \mathcal{U}_\eta : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\
&\qquad + \mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{U}_\eta : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\}\mathbb{P}\big\{\nexists(\mathbf{h},\mathbf{t}) \in \mathcal{W}\setminus\mathcal{U} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \\
&\quad \to \alpha + \bar{\Phi}(c)(1-\alpha),
\end{aligned}$$

by independence of $\{\Upsilon(\mathbf{w}) : \mathbf{w} \in \mathcal{U}_\eta\}$ and $\{\Upsilon(\mathbf{w}) : \mathbf{w} \in \mathcal{W} \setminus \mathcal{U}\}$.

We conclude that

$$\mathbb{P}\big\{\exists(\mathbf{h},\mathbf{t}) \in \mathcal{W} : \Upsilon(\mathbf{h},\mathbf{t}) > u(\mathbf{h})\big\} \to \alpha + \bar{\Phi}(c)(1-\alpha).$$

By Lemma 17 and Lemma 19, we find that this also holds when $\mathcal{W}$ is replaced by $\mathcal{W}'$, as long as $\underline{h} = \omega(\log n)$. And from this we conclude as in Section 5.3.2.

5.3.8. Proof of Theorem 7

For adaptive multiscale scan, Lemma 18 allows us to apply the conclusion of Lemma 17 to the critical value (2.16). For the multiscale scan statistic, Lemma 17 applies to the constant critical value (2.13). Let $\hat{\tau}$ be the result of the scan over the discrete set, $\mathcal{W} \cap \mathbb{Z}^{2d}$, for either the (resp. adaptive) multiscale scan, and let $\hat{\tau}_\epsilon$ be the scan over the ϵ-covering. Then by Lemma 19, $\mathcal{W} \cap \mathbb{Z}^{2d}$ is an ϵ'-covering of $\mathcal{W}$ for $\epsilon' = \sqrt{4d/\underline{h}} = o((\log n)^{-1/2})$. Thus, we may apply Lemma 17, unless $\mu = \omega(\sqrt{\log(n/\underline{h})})$ under H_1, to show that $|\tilde{\tau} - \hat{\tau}| = o_{\mathbb{P}}(1)$. Likewise, when $\epsilon = o((\log n)^{-1/2})$ then $\hat{\tau}_\epsilon$ fulfills the conditions of Lemma 17 unless $\mu = \omega(\sqrt{\log(n/\underline{h})})$ under H_1. But $\mu = \omega(\sqrt{\log(n/\underline{h})})$ implies that $\hat{\tau}, \hat{\tau}_\epsilon, \tilde{\tau} \to \infty$ because then $y[R^\star] = \omega_{\mathbb{P}}(\sqrt{\log(n/\underline{h})})$. In this case, $\hat{\alpha}, \hat{\alpha}_\epsilon \to 0$. When this is not the case then $|\hat{\tau} - \tilde{\tau}| = o_{\mathbb{P}}(1)$ and $|\hat{\tau}_\epsilon - \tilde{\tau}| = o_{\mathbb{P}}(1)$ by Lemma 17, and so $|\hat{\tau} - \hat{\tau}_\epsilon| = o_{\mathbb{P}}(1)$. Because $\hat{\alpha} = 1 - \exp(-\exp(-\hat{\tau}))$ and $\hat{\alpha}_\epsilon = 1 - \exp(-\exp(-\hat{\tau}_\epsilon))$, the result follows by the continuous mapping theorem.

5.3.9. Proof of Proposition 8

We now show that $\mathcal{R}_\epsilon$ is an ϵ-covering of $\mathcal{R}$. Specifically, for each $(\mathbf{h},\mathbf{t}) \in \mathcal{W}$, we construct $(\mathbf{g},\mathbf{s})$ such that $R(\mathbf{g},\mathbf{s}) \in \mathcal{R}_\epsilon$ and $\delta(R(\mathbf{h},\mathbf{t}), R(\mathbf{g},\mathbf{s})) \le \epsilon$. Take $a_j = \lfloor \log_2 \frac{h_j \epsilon^2}{4d} \rfloor \ge \underline{a}$, for each $j \in [d]$. Define

$$g_j = \operatorname{argmin}\{|h - h_j| : h \in 2^{a_j}\mathbb{Z}_+\}, \quad s_j = \operatorname{argmin}\{|s - t_j| : s \in 2^{a_j}\mathbb{Z}_+\}.$$

We know that

$$\frac{4d}{h_j\epsilon^2} = 2^{-\log_2 \frac{h_j\epsilon^2}{4d}} \leq 2^{-a_j} \leq 2^{-\log_2 \frac{h_j\epsilon}{4d}+1} = \frac{8d}{h_j\epsilon^2}.$$

By the construction,

$$2^{-a_j}|g_j - h_j| \leq \frac{1}{2}.$$

Hence, we have that

$$2^{-a_j}g_j \in \left[2^{-a_j}h_j - \frac{1}{2}, 2^{-a_j}h_j + \frac{1}{2}\right] \subseteq \left[\frac{4d}{\epsilon^2} - \frac{1}{2}, \frac{8d}{\epsilon^2} + \frac{1}{2}\right]. \tag{5.29}$$

But because $2^{-a_j}g_j$ is constructed to be in $\mathbb{Z}_+$ then we know that it lies within $[[\lceil 8d/\epsilon^2 \rceil]]$. Therefore, $R(\mathbf{g}, \mathbf{s}) \in \mathcal{R}_\epsilon$. Let $\epsilon^2 < 4d$. It remains to show that $\theta((h_j, t_j), (g_j, s_j)) \leq \epsilon^2/2d$ for all j, so that $\delta(\mathbf{h}, \mathbf{t}), (\mathbf{g}, \mathbf{s})) \leq \epsilon$ by (5.9). We can see that

$$|g_j - h_j|, |s_j - t_j| \leq 2^{a_j - 1}, \text{ and } h_j \in 2^{a_j}\left[\frac{4d}{\epsilon^2}, \frac{8d}{\epsilon^2}\right].$$

Because $4d/\epsilon^2 \geq 1$ then $h_j \geq 2^{a_j}$. Furthermore, $h_j^2 - 2^{a_j-1}h_j$ is an increasing function for $h_j \geq 2^{a_j}$. Hence,

$$g_j h_j = h_j^2 - (h_j - g_j)h_j \geq h_j^2 - 2^{a_j-1}h_j \geq 2^{2a_j}\left(\tfrac{16d^2}{\epsilon^4} - \tfrac{2d}{\epsilon^2}\right).$$

We then have

$$\begin{aligned}\theta((h_j, t_j), (g_j, s_j)) &\leq \frac{|g_j - h_j| + |s_j - t_j|}{\sqrt{g_j h_j}} \leq \left(\frac{16d^2}{\epsilon^4} - \frac{2d}{\epsilon^2}\right)^{-1/2} \\ &\leq \frac{\epsilon^2}{2d}\left(4 - \tfrac{\epsilon^2}{2d}\right)^{-\frac{1}{2}} < \frac{\epsilon^2}{2d},\end{aligned}$$

since $\epsilon^2 < 4d$.

5.3.10. Proof of Proposition 9

First, we establish that, for $\mathbf{a} \in \{\underline{a}, \dots, \overline{a}\}^d$,

$$(\text{dyad}_{\mathbf{a}} * b_{\mathbf{f}})(\mathbf{t}) = (y * b_{2^{\mathbf{a}} \circ \mathbf{f}})(2^{\mathbf{a}} \circ \mathbf{t}), \quad \mathbf{t} \in [n/2^{\mathbf{a}}]. \tag{5.30}$$

An induction on $\|\mathbf{a}\|_1$, based on the recursion in Line 5, gives

$$\text{dyad}_{\mathbf{a}}(\mathbf{t}) = \sum_{\mathbf{i} \in [2^{\mathbf{a}}]} y(2^{\mathbf{a}} \circ \mathbf{t} + \mathbf{i}), \quad \forall \mathbf{t} \in \times_j [n/2^{a_j}].$$

Based on this, we have

$$\begin{aligned}(\text{dyad}_{\mathbf{a}} * b_{\mathbf{f}})(\mathbf{t}) &= \sum_{\mathbf{i} \in [\mathbf{f}]} \text{dyad}_{\mathbf{a}}(\mathbf{i} + \mathbf{t}) = \sum_{\mathbf{i} \in [\mathbf{f}]} \sum_{\mathbf{k} \in [2^{\mathbf{a}}]} y(2^{\mathbf{a}} \circ (\mathbf{i} + \mathbf{t}) + \mathbf{k}) \\ &= \sum_{\mathbf{i} \in [2^{\mathbf{a}}\mathbf{f}]} y(\mathbf{i} + 2^{\mathbf{a}} \circ \mathbf{t}) = (y * b_{2^{\mathbf{a}} \circ \mathbf{f}})(2^{\mathbf{a}} \circ \mathbf{t}).\end{aligned}$$

With (5.30), we can see that the statistic $\hat{s}$ in Algorithm 1 is equivalently expressed as

$$\max_{\mathbf{t}\in[n/2^{\mathbf{a}}]} y\left[[2^{\mathbf{a}}, 2^{\mathbf{a}}(\mathbf{t}+\mathbf{f})]\right],$$

confirming that Algorithm 1 does scan over $\mathcal{R}_\epsilon$.

5.3.11. Proof of Proposition 10

First, the construction of dyad takes $O(n^d)$ operations. Indeed, the computation of $\text{dyad}_{\mathbf{a}}(\mathbf{t})$ over $\mathbf{a} \in [\log_2 n]^d \backslash \{1\}^d$ and $\mathbf{t} \in [n/2^{\mathbf{a}}]$ is done from Line 2 to Line 7 in Algorithm 1, and is easily seen to require on the order of

$$\sum_{\mathbf{a}\in[\log_2 n]^d} \prod_j (n/2^{a_j}) \leq n^d \Big(\sum_{a\geq 1} 2^{-a}\Big)^d = n^d$$

basic operations.

Second, defining $a_+ = \sum_{j=1}^d a_j$, the convolution $\text{dyad}_{\mathbf{a}} * b_{\mathbf{f}}$ takes $O(n^d 2^{-a_+} \log n)$ operations with the FFT, since the convolution happens on a grid of size $\prod_j (n/2^{a_j}) = n^d 2^{-a_+}$. Therefore, the computation on Line 13 requires $O(n^d 2^{-a_+} \log n)$ basic operations. Hence, once dyad is computed, computing $\hat{\alpha}$ requires on the order of

$$\begin{aligned}
&\sum_{\mathbf{a}\in[\underline{a},\overline{a}]^d} d\Big(\textstyle\prod_j |\mathcal{F}_j|\Big)\Big(\frac{n^d}{2^{a_+}} \log n\Big) \\
&\quad = O(\epsilon^{-2d} n^d \log n)\Big(\textstyle\sum_{a\geq\underline{a}} 2^{-a}\Big)^d = O(\epsilon^{-2d} n^d 2^{-d\underline{a}} \log n),
\end{aligned}$$

with $2^{-\underline{a}} = O(1/\epsilon^2 \underline{h})$ since $\epsilon \underline{h} \geq 1$. From this, we conclude.

Acknowledgments

This work was partially supported by a grant from the US National Science Foundation (DMS 1223137). We would like to thank Zakhar Kabluchko for clarifying some technical points appearing in his work. We are also grateful to two anonymous referees for constructive feedback.

References

[1] Amărioarei, A. and C. Preda (2015). Approximation for the distribution of three-dimensional discrete scan statistic. *Methodology and Computing in Applied Probability 17*(3), 565–578. MR3377849

[2] Arias-Castro, E., E. J. Candès, and A. Durand (2011). Detection of an anomalous cluster in a network. *Ann. Statist. 39*(1), 278–304. MR2797847

[3] Arias-Castro, E., D. Donoho, and X. Huo (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory 51*(7), 2402–2425. MR2246369

[4] Arratia, R., L. Goldstein, L. Gordon, et al. (1989). Two moments suffice for poisson approximations: the chen-stein method. *The Annals of Probability 17*(1), 9–25. MR0972770

[5] Boutsikas, M. V. and M. V. Koutras (2006). On the asymptotic distribution of the discrete scan statistic. *J. Appl. Probab. 43*(4), 1137–1154. MR2274642

[6] Brennan, S. M., A. M. Mielke, D. C. Torney, and A. B. Maccabe (2004). Radiation detection with distributed sensor networks. *Computer 37*(8), 57–59.

[7] Caron, Y., P. Makris, and N. Vincent (2002). A method for detecting artificial objects in natural environments. In *Proceedings 16th International Conference on Pattern Recognition*, Volume 1, pp. 600–603. IEEE Comput. Soc.

[8] Chan, H. P. and T. L. Lai (2006). Maxima of asymptotically gaussian random fields and moderate deviation approximations to boundary crossing probabilities of sums of random variables with multidimensional indices. *The Annals of Probability 34*(1), 80–121. MR2206343

[9] Culler, D., D. Estrin, and M. Srivastava (2004). Overview of sensor networks. *IEEE Computer 37*(8), 41–49.

[10] Desolneux, A., L. Moisan, and J.-M. Morel (2003). Maximal meaningful events and applications to image analysis. *Ann. Statist. 31*(6), 1822–1851. MR2036391

[11] Duczmal, L., M. Kulldorff, and L. Huang (2006). Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational & Graphical Statistics 15*(2), 428–442. MR2256152

[12] Glaz, J., J. Naus, and S. Wallenstein (2001). *Scan statistics*. Springer Series in Statistics. New York: Springer-Verlag. MR1869112

[13] Glaz, J. and Z. Zhang (2004). Multiple window discrete scan statistics. *Journal of Applied Statistics 31*(8), 967–980. MR2100425

[14] Haiman, G. and C. Preda (2006). Estimation for the distribution of two-dimensional discrete scan statistics. *Methodology and Computing in Applied Probability 8*(3), 373–382. MR2329304

[15] Heffernan, R., F. Mostashari, D. Das, A. Karpati, M. Kulldorff, and D. Weiss (2004). Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases 10*(5), 858–864.

[16] James, D., B. D. Clymer, and P. Schmalbrock (2001). Texture detection of simulated microcalcification susceptibility effects in magnetic resonance imaging of breasts. *Journal of Magnetic Resonance Imaging 13*(6), 876–881.

[17] Jiang, T. (2002). Maxima of partial sums indexed by geometrical structures. *Ann. Probab. 30*(4), 1854–1892. MR1944008

[18] Kabluchko, Z. (2011). Extremes of the standardized gaussian noise. *Stochastic Processes and their Applications 121*(3), 515–533. MR2763094

[19] Kulldorff, M. (1997). A spatial scan statistic. *Comm. Statist. Theory Methods 26*(6), 1481–1496. MR1456844

[20] Kulldorff, M., L. Huang, L. Pickle, and L. Duczmal (2006). An elliptic spatial scan statistic. *Stat Med 25*(22), 3929–43. MR2297401

[21] Kulldorff, M. and Information Management Services, Inc. Satscan$^{\text{TM}}$ v9.4: Software for the spatial and space-time scan statistics. `http://www.satscan.org/`.

[22] Marcus, M. B. and J. Rosen (2006). *Markov processes, Gaussian processes, and local times.* Number 100. Cambridge University Press. MR2250510

[23] McInerney, T. and D. Terzopoulos (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis 1*(2), 91–108.

[24] Moon, N., E. Bullitt, K. van Leemput, and G. Gerig (2002). Automatic brain and tumor segmentation. In *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part I*, London, UK, pp. 372–379. Springer-Verlag.

[25] Naus, J. I. (1965). The distribution of the size of the maximum cluster of points on a line. *J. Amer. Statist. Assoc. 60*, 532–538. MR0183041

[26] Naus, J. I. and S. Wallenstein (2004). Multiple window and cluster size scan procedures. *Methodology and Computing in Applied Probability 6*(4), 389–400. MR2108559

[27] Pickands, J. (1969). Upcrossing probabilities for stationary gaussian processes. *Transactions of the American Mathematical Society 145*, 51–73. MR0250367

[28] Pozdnyakov, V., J. Glaz, M. Kulldorff, and J. M. Steele (2005). A martingale approach to scan statistics. *Annals of the Institute of Statistical Mathematics 57*(1), 21–37. MR2165605

[29] Pozo, D., F. Olmo, and L. Alados-Arboledas (1997). Fire detection and growth monitoring using a multitemporal technique on AVHRR mid-infrared and thermal channels. *Remote Sensing of Environment 60*(2), 111–120.

[30] Qualls, C. and H. Watanabe (1973). Asymptotic properties of gaussian random fields. *Transactions of the American Mathematical Society 177*, 155–171. MR0322943

[31] Rotz, L. and J. Hughes (2004). Advances in detecting and responding to threats from bioterrorism and emerging infectious disease. *Nature Medicine*, S130–S136.

[32] Siegmund, D. and E. S. Venkatraman (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist. 23*(1), 255–271. MR1331667

[33] Wagner, M., F. Tsui, J. Espino, V. Dato, D. Sittig, R. Caruana, L. Mcginnis, D. Deerfield, M. Druzdzel, and D. Fridsma (2001). The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice 7*(6), 51–59.

[34] Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist. 38*(2), 1010–1033. MR2604703

[35] Wang, X. and J. Glaz (2014). Variable window scan statistics for normal data. *Communications in Statistics-Theory and Methods 43*(10–12), 2489–2504. MR3217827