DETECTION OF AN ANOMALOUS CLUSTER IN A NETWORK[1]

BY ERY ARIAS-CASTRO, EMMANUEL J. CANDÈS AND ARNAUD DURAND

*University of California, San Diego, Stanford University
and Université Paris-Sud 11*

We consider the problem of detecting whether or not, in a given sensor network, there is a cluster of sensors which exhibit an "unusual behavior." Formally, suppose we are given a set of nodes and attach a random variable to each node. We observe a realization of this process and want to decide between the following two hypotheses: under the null, the variables are i.i.d. standard normal; under the alternative, there is a cluster of variables that are i.i.d. normal with positive mean and unit variance, while the rest are i.i.d. standard normal. We also address surveillance settings where each sensor in the network collects information over time. The resulting model is similar, now with a time series attached to each node. We again observe the process over time and want to decide between the null, where all the variables are i.i.d. standard normal, and the alternative, where there is an emerging cluster of i.i.d. normal variables with positive mean and unit variance. The growth models used to represent the emerging cluster are quite general and, in particular, include cellular automata used in modeling epidemics. In both settings, we consider classes of clusters that are quite general, for which we obtain a lower bound on their respective minimax detection rate and show that some form of scan statistic, by far the most popular method in practice, achieves that same rate to within a logarithmic factor. Our results are not limited to the normal location model, but generalize to any one-parameter exponential family when the anomalous clusters are large enough.

1. Introduction. We discuss the problem of detecting whether or not, in a given network, there is a cluster of nodes which exhibit an "unusual behavior." Suppose that we are given a set of nodes with a random variable attached to each node. We observe a realization of this process and would like to tell whether all the variables at the nodes have the same behavior, in the sense that they are all sampled from a common distribution, or whether there is a cluster of nodes at which the variables have a different distribution.

1.1. *A wide array of applications.* The task of detection in networks is critical for an increasing number of applications, for example, in surveillance and environment monitoring. We describe a few of these applications below.

Detection in sensor networks. The advent of sensor networks [3, 20, 74] has multiplied the amount of data and the variety of applications where the task of detection is central. Surveillance and environment monitoring are prime areas of application for sensor networks. Take, for example, the transport of hazardous materials. Currently, some major traffic bottlenecks (e.g., airports, subways and borders) use portal monitoring systems [27, 28]. Sensor networks offer a more flexible, decentralized alternative and are considered for the detection of radioactive, biological or chemical materials [15, 19, 35]. Sensor networks are also extensively used in other target tracking settings [11, 50].

Detection in digital signals and images. A digital camera may be seen as a sensor network, with CCD or CMOS pixel sensors. As imaging systems have been available for quite some time, the literature on detection in images is quite extensive, spanning several decades, particularly in satellite imagery [18, 29, 58, 65], computer vision [68, 75] and medical imaging [14, 39, 51, 53].

Disease outbreak detection. The presence of a biological or chemical material in a given geographical region may also be detected indirectly through its impact on human health. In this context, early detection of the disease outbreak is crucial in order to minimize the severity of the epidemic. For that purpose, some specific information networks are used, with surveillance systems now incorporating data from hospital emergency visits, ambulance dispatch calls and pharmacy sales of over-the-counter drugs [34, 60, 70].

Virus detection in a computer network. Diseases affect computers as well, in the form of viruses and worms spreading from host to host in a computer network [66]. Affected machines may exhibit slightly anomalous behavior (e.g., a loss of performance or violations of specific rules) which may be hard to detect on an individual machine.

Detection from field measurements. In [54], the water quality in a network of streams in Pennsylvania is assessed by field biologists performing a variety of analyses at various locations along the streams; the objective is to determine whether there are regions of low biological integrity based on the collected data, and to identify these regions. Other field measurements include census data and surveys involving geographical location.

Detection is, of course, closely related to estimation (i.e., the localization or extraction of the anomalous cluster of nodes), but different. This distinction is rarely made clear, however. Indeed, reliable detection is possible at lower signal-to-noise ratios than reliable estimation and it may be important to detect the presence of signals from noisy data without being able to estimate them. For example, one could imagine developing a surveillance system performing detection at relatively low energy/bandwidth costs, yet efficient at low signal-to-noise ratios, and

then switching to estimation mode whenever the presence of a signal is detected. Another example would be a low cost preliminary survey involving fewer field measurements, with findings subsequently confirmed by a larger, more expensive survey.

1.2. *Mathematical framework.*

1.2.1. *Purely spatial model.* We loosely model a network with a set of m nodes, denoted by $\mathbb{V}_m$. In our examples, we will either assume that $\mathbb{V}_m$ is embedded in a Euclidean space or we will equip $\mathbb{V}_m$ with a graph structure. Our analysis is in the setting of large networks, that is, $m \to \infty$. To each node $v \in \mathbb{V}_m$, we attach a random variable X_v. The nodes represent the sources of information (e.g., sensors) and the variables represent the data they collect. In some settings, the data collected by each unit is multidimensional, in which case X_v is a random vector. Our discussion readily generalizes to that setting.

The random variables are assumed to be independent. For concreteness, we consider a normal location model, popular in signal and image processing, to model the noise. Our analysis, however, generalizes to any exponential family under some conditions on the sizes of the anomalous clusters, such as Bernoulli models which arise in sensor arrays where each sensor collects one bit (i.e., makes a binary decision) or Poisson models which come up with count data, for instance, arising in infectious disease surveillance systems [43]. The extension to exponential families is detailed in Section 4.1.

The situation where no signal is present, that is, "business as usual," is modeled as

$$\mathrm{H}_0^m : X_v \sim \mathcal{N}(0, 1) \qquad \forall v \in \mathbb{V}_m.$$

Let K be a cluster, which we define for now as a subset of nodes, that is, $K \subset \mathbb{V}_m$. In fact, we will be interested in classes of clusters that are either derived from a geometric shape, when $\mathbb{V}_m$ is embedded in Euclidean space, or connected components, when $\mathbb{V}_m$ has a graph structure. The situation where the nodes in K behave anomalously is modeled as

$$\mathrm{H}_{1,K}^m : X_v \sim \mathcal{N}(\mu_K, 1) \qquad \forall v \in K; \qquad X_v \sim \mathcal{N}(0, 1) \qquad \forall v \notin K,$$

where $\mu_K > 0$. We choose to decompose μ_K as $\mu_K = |K|^{-1/2} \Lambda_K$, where $|K|$ denotes the number of nodes in K and Λ_K is the signal strength. Indeed, with this normalization, for any cluster K,

$$(1.1) \qquad \min_T \mathbb{P}(T = 1 | \mathrm{H}_0^m) + \mathbb{P}(T = 0 | \mathrm{H}_{1,K}^m) = 2\mathbb{P}(\mathcal{N}(0, 1) > \Lambda_K / 2),$$

where the minimum is over all tests for H_0^m versus $\mathrm{H}_{1,K}^m$ and the lower bound is achieved by the likelihood ratio (Neyman–Pearson) test. We define

$$\overline{\Lambda}_m = \max_{K \in \mathcal{K}_m} \Lambda_K, \qquad \underline{\Lambda}_m = \min_{K \in \mathcal{K}_m} \Lambda_K.$$
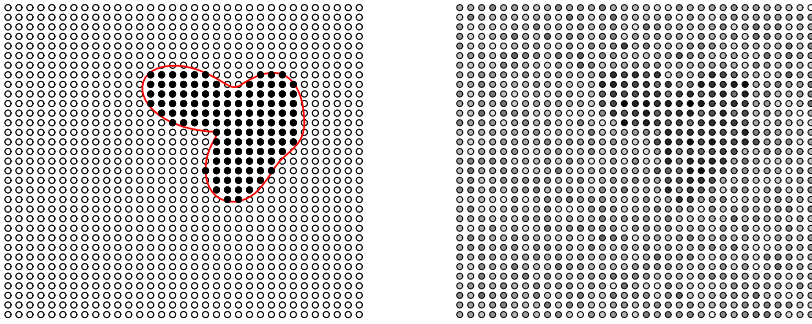
FIG. 1. *Left*: *a thick cluster is defined as the nodes within a closed curve, which is a mild deformation of a circle. Right*: *corresponding noisy data.*

Figures 1–4 illustrate the setting for various types of clusters.

Let $\mathcal{K}_m$ be a class of clusters within $\mathbb{V}_m$ and define

$$H_1^m = \bigcup_{K \in \mathcal{K}_m} H_{1,K}^m.$$

We are interested in testing H_0^m versus H_1^m. In other words, under the alternative, the cluster of anomalous nodes is only known to belong to $\mathcal{K}_m$. We adopt a minimax point of view. For a test T, we define its worst-case risk as

$$\gamma_{\mathcal{K}_m}(T) = \mathbb{P}(T = 1 | H_0^m) + \max_{K \in \mathcal{K}_m} \mathbb{P}(T = 0 | H_{1,K}^m).$$

The minimax risk for H_0^m versus H_1^m is defined as

$$\gamma_{\mathcal{K}_m} = \inf_T \gamma_{\mathcal{K}_m}(T).$$

We say that H_0^m and H_1^m are asymptotically inseparable (in the minimax sense) if

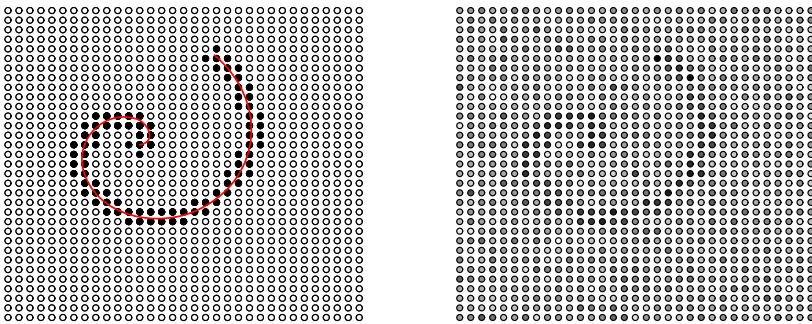$$\lim_{m \to \infty} \gamma_{\mathcal{K}_m} = 1,$$

FIG. 2. *Left*: *a thin cluster is defined as the nodes within a band around a given curve. Right*: *corresponding noisy data.*
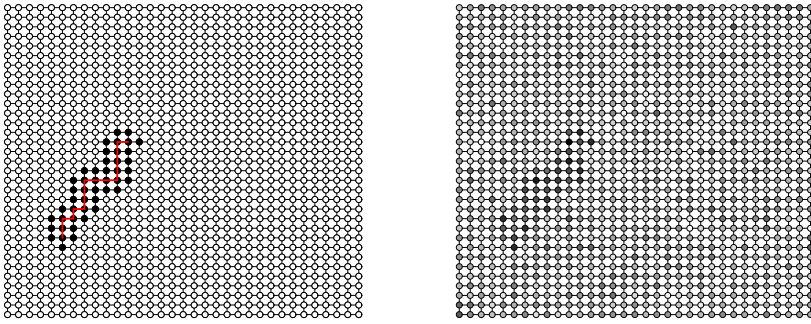
FIG. 3. *Left*: *a band defined around a path. Right*: *corresponding noisy data.*

which is equivalent to saying that, as m becomes large, no test can perform substantially better than random guessing, without even looking at the data. A sequence of tests (T_m) is said to asymptotically separate H_0^m and H_1^m if

$$\lim_{m \to \infty} \gamma_{\mathcal{K}_m}(T_m) = 0,$$

and H_0^m and H_1^m are said to be asymptotically separable if there is such a sequence of tests. For example, in view of (1.1), for any sequence of clusters $K_m \subset \mathbb{V}_m$, H_0^m and H_{1,K_m}^m are asymptotically inseparable if $\Lambda_{K_m} \to 0$ and they are asymptotically separable if $\Lambda_{K_m} \to \infty$. For convenience, we assume that no cluster in the class $\mathcal{K}_m$ is of size comparable to that of the entire network, that is, $\max\{|K| : K \in \mathcal{K}_m\} = o(m)$. This simplifies the statement of our results and detecting such clusters can easily be achieved using the test that rejects for large values of $\sum_{v \in \mathbb{V}_m} X_v$.

The situation we just described is purely spatial and relevant in some applications not involving time. Such situations are common in image processing. In other applications, especially in surveillance, time is an intrinsic part of the setting. In the following section, we modify the model above to incorporate time.
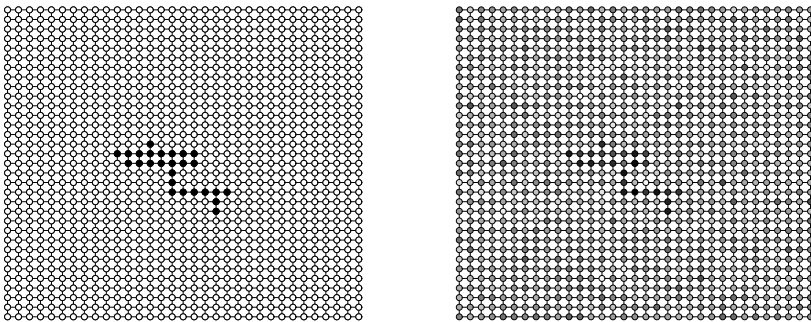

FIG. 4. *Left*: *an arbitrary connected component. Right*: *corresponding noisy data.*

1.2.2. *Spatio-temporal model.* Building on the framework introduced in the previous section, we assume that each X_v is now a (discrete) time series, $(X_v(t), t \in \mathbb{T}_m)$, where $\mathbb{T}_m \subset [0, \infty)$ is finite with $|\mathbb{T}_m| \to \infty$; let $t_m = \max\{t \in \mathbb{T}_m\}$. Let $\mathcal{K}_m$ be a class of cluster sequences of the form $(K_t, t \in \mathbb{T}_m)$ such that $K_t \subset \mathbb{V}_m$ for all $t \in \mathbb{T}_m$. For example, assuming that $\mathbb{V}_m$ is embedded in a Euclidean space, with norm denoted by $\|\cdot\|$, a space–time cylinder (e.g., one used in disease outbreak detection [44]) is a cluster sequence $(K_t, t \in \mathbb{T}_m)$ of the form $K_t = \{v \in \mathbb{V}_m : \|v - x_0\| \le r_0\}$ if $t \ge t_0$, and $K_t = \varnothing$ otherwise, so that t_0 is the origin of the cluster in time and x_0 its center. Note that the radius remains constant here. Another example is that of a space–time cone, of the form $K_t = \{v \in \mathbb{V}_m : \|v - x_0\| \le C(t - t_0)\}$ if $t \ge t_0$, and $K_t = \varnothing$ otherwise, so that (x_0, t_0) is the origin of the cluster in space–time. The random variables $\{X_v(t) : v \in \mathbb{V}_m, t \in \mathbb{T}_m\}$ are assumed to be independent. This spatio-temporal setting is a special case of the purely spatial setting with the set of nodes $\mathbb{V}_m \times \mathbb{T}_m$. Understood as such, we are interested in testing H_0^m versus H_1^m as before.

1.3. *Structured multiple hypothesis testing.* Although the detection problem formulated above seems of great practical relevance, the statistics literature is almost silent on the subject, with the notable exception of the closely related topics of change-point analysis [16] and sequential analysis [64]. Indeed, the former is a special case of the spatial setting with the one-dimensional lattice, while the latter is a special case of the spatio-temporal setting where $\mathbb{V}_m$ has only one node. In our context, these two settings are actually equivalent.

What is further puzzling is that a number of publications addressing the task of detection in sensor networks all assume overly simplistic models. For example, in [4, 49, 52, 55, 69, 73], the values at the sensors are assumed to all have the same distribution under the null and the alternative. That is, either all of the nodes are all right or they are all anomalous—in our notation, $\mathcal{K}_m = \{\mathbb{V}_m\}$. First, this is not a subtle statistical problem since, in such circumstances, it suffices to apply the optimal likelihood ratio test. Second, this assumption does not make sense in all of the applications described above, where the event to be detected is expected to only affect a small fraction of locations in the network.

In stark contrast, in all of the applications described earlier, the set of alternatives is composite. Viewing each node as performing a test of hypotheses, which is common in the literature on sensor networks, our problem falls within the framework of multiple comparisons. Multiple hypothesis testing is a rich and active line of research which is receiving a considerable amount of attention within the statistical community at the moment; see [32] and references therein. The vast majority of the papers assume that the tests are independent of each other, which is clearly not the case here since, in general, the class contains clusters that intersect. This is particularly true in engineering applications, although this assumption is often made [23, 62, 63].

1.4. *The scan statistic.*   We will focus on the test that rejects for large values of the following version of the scan statistic:

$$\text{(1.2)} \qquad \max_{K \in \mathcal{K}_m} \frac{1}{\sqrt{|K|}} \sum_{v \in K} X_v.$$

The chosen normalization is such that each term in the maximization is standard normal under the null and allows us to compare clusters of different sizes. It corresponds to the generalized likelihood ratio test in our context if Λ_K is independent of $K \in \mathcal{K}_m$. The scan statistic was originally proposed in the context of cluster detection in point clouds [30]. This is the method of matched filters which is ubiquitous in problems of detection in a wide variety of fields, sometimes in the form of deformable templates in the engineering literature [38, 51] or their nonparametric equivalent, active contours or snakes [72]. Note that the scan statistic is the prevalent method in disease outbreak detection, with many variations [25, 45–47].

As advocated in [8], we will not use the scan statistic directly in most cases, but rather restrict the scanning to a subset of $\mathcal{K}_m$. More precisely, we will introduce, on subsets of nodes $K, L \subset \mathbb{V}_m$, the metric

$$\text{(1.3)} \qquad \delta(K, L) = \sqrt{2}\left(1 - \frac{|K \cap L|}{\sqrt{|K||L|}}\right)^{1/2}$$

and will restrict the scanning to an ε-net of $\mathcal{K}_m$ with respect to δ, that is, a subset $\{K_j : j \in J\} \subset \mathcal{K}_m$ with the property that for each $K \in \mathcal{K}_m$, there is a $j \in J$ such that $\delta(K, K_j) \leq \varepsilon$. We will elaborate on this approach in the Supplement. When J is minimal, we call the resulting statistic an ε-scan statistic. The approximation precision ε will be chosen appropriately, depending on the situation.

We focus on ε-scan statistics for two reasons. First, their performance is easier to analyze than that of the scan statistic itself; in fact, the main approach to analyzing the scan statistic, the chaining method of Dudley [26, 67], is via a properly chosen ε-scan statistic. Second, some of the classes we consider are rather large and we believe that it would be computationally impractical to scan through all of the clusters in the class; furthermore, our results show that, from an asymptotic standpoint, no substantial improvement would be gained by using the full scan statistic.

We also note that the tuning parameter ε may be dispensed with if we scan over subsets of different sizes in a multiscale fashion and use a scale-dependent threshold.

1.5. *Existing theoretical results.*   The vast majority of the literature assumes that the set of nodes is embedded in some Euclidean space, that is, $\mathbb{V}_m \subset \mathbb{R}^d$. This is the case when the nodes represent spatial locations, such as in most sensor networks. In this context, the cluster class $\mathcal{K}_m$ is often derived from a class of domains $\mathcal{A}$ in $\mathbb{R}^d$, in the following way:

$$\text{(1.4)} \qquad \mathcal{K}_m = \{K = A \cap \mathbb{V}_m : A \in \mathcal{A}\}.$$

Most of the literature assumes that the class $\mathcal{A}$ is parametric, exemplified by deformable templates, for which theoretical results are available, especially in the case of the square lattice [8, 13, 24, 40, 57, 71]. In particular, with a normal location model, the scan statistic performs well, in the sense that it is asymptotically minimax; this is shown in [8] in a slightly different context tailored to image processing applications. We also mention the recent work [33], which considers the detection of multiple clusters (intervals) of various amplitudes in the one-dimensional lattice. As for nonparametric classes of domains, [8] argues that the scan statistic is asymptotically minimax for the case of star-shaped clusters with smooth boundaries.

When $\mathbb{V}_m$ is endowed with a graph structure, [7] considers paths of a certain length. In this setting, the scan statistic is shown to be asymptotically minimax when the graph $\mathbb{V}_m$ is a complete, regular tree and near-minimax for many other types of graphs, such as the d-dimensional lattice for $d \geq 3$. Addario-Berry et al. [1] considers the same general testing problem with a focus on cluster classes defined within the complete graph, such as cliques and spanning trees. Note that part of the material presented here appeared in [5].

1.6. *New theoretical results.* We describe here in an informal way the results we obtain.

In Section 2, we focus on situations where the vertex set $\mathbb{V}_m$ is embedded in a Euclidean space and well spread out in a compact domain. Within this framework, we consider in Section 2.1 a geometric class of clusters obtained as in (1.4) with $\mathcal{A}$ a class of blobs that are mild deformations of the unit ball. The clusters obtained in this way are "thick," in the sense that they are not filamentary. See Figure 1. In particular, this class contains all the common parametric classes obtained from parametric shapes such as hyperrectangles and ellipsoids, as long as the shape is not too narrow. Note that the size, the (exact) shape and the spatial location of the anomalous cluster under the alternative is unknown. In Corollary 1, we show that (under specific conditions) H_0^m and H_1^m are asymptotically inseparable if there is $\eta_m \to 0$ slowly enough such that, for all $K \in \mathcal{K}_m$,

$$\Lambda_K \leq (1 - \eta_m)\sqrt{2\log(m/|K|)};$$

and conversely, we show that a version of the scan statistic asymptotically separates H_0^m and H_1^m if there is $\eta_m \to 0$ slowly enough such that, for all $K \in \mathcal{K}_m$,

$$\Lambda_K \geq (1 + \eta_m)\sqrt{2\log(m/|K|)}.$$

Note that the detection rate is the same as for the class of balls so that, perhaps surprisingly, scanning for the location (and not the shape) is what drives the minimax detection risk.

In Section 2.2, we consider "thin" clusters, obtained as in (1.4) with $\mathcal{A}$ a class of "bands" around smooth curves, surfaces or higher-dimensional submanifolds.

In particular, this class contains hyperrectangles and ellipsoids that are sufficiently thin; see Figure 2. It turns out that, contrary to what happens for thick clusters, scanning for the actual shape impacts the minimax detection risk and is, in fact, the main contributor for some nonparametric classes. The situation is mathematically more challenging, yet we are able to prove the following in Proposition 3. Consider the class of bands of thickness r_m around C^2 curves of bounded curvature. Then (under specific conditions), H_0^m and H_1^m are asymptotically inseparable if

$$\overline{\Lambda}_m r_m^{1/4} (\log m)^{3/2} \to 0.$$

In Theorem 2, we show that, in the same setting, some ε_m-scan statistic asymptotically separates H_0^m and H_1^m if

$$\underline{\Lambda}_m r_m^{1/4} \to \infty.$$

Hence, some form of scan statistics achieves a detection rate within a factor of $(\log m)^{3/2}$ from the minimax rate.

In Section 2.3, we consider the spatio-temporal setting. We first consider cluster sequences that admit a "thick" limit. Cellular automata, which have been used to model epidemics [2], satisfy this condition in some cases. In Proposition 5, we show that scanning over space–time cylinders, as done in disease outbreak detection, achieves the asymptotic minimax risk. We then consider cluster sequences with controlled space–time variations, which may be a relevant model for applications such as target tracking [50]. We consider a fairly general model in Proposition 7.

In Section 3, we assume that $\mathbb{V}_m = \{0, 1, \ldots, m^{1/d} - 1\}^d$, with $m^{1/d}$ an integer, seen as a subgraph of the d-dimensional lattice. We first consider, in Section 3.1, bands around nearest-neighbor paths; see Figure 3. We extend the results obtained in [7] to paths. For example, consider bands of thickness h_m around a path of length ℓ_m, both powers of m. The bounds in Theorem 3 imply that H_0^m and H_1^m are asymptotically inseparable if

$$\overline{\Lambda}_m (\ell_m / h_m)^{-1/2} (\log m)^{3/2} \to 0.$$

Conversely, Proposition 8 states that an ε-scan statistic asymptotically separates H_0^m and H_1^m if

$$\underline{\Lambda}_m (\ell_m / h_m)^{-1/2} \to \infty.$$

Therefore, some form of scan statistic is again within a factor of $(\log m)^{3/2}$ from optimal. In Section 3.2, we consider arbitrary connected components, constraining only the size; see Figure 4. In Proposition 9, we obtain a sharp detection rate for clusters of very small size.

1.7. *Structure of the paper.*   We have just described the contents of Sections 2 and 3. Section 4 is our discussion section. We extend the results obtained for the normal location model to any exponential family in Section 4.1. Other extensions are described in Section 4.2. We state some open problems in Section 4.5. In Section 4.4, we briefly discuss the challenge of computing the scan statistic. The technical arguments are gathered in the Supplement.

1.8. *Notation.*   For two sequences of real numbers (a_m) and (b_m), $a_m \asymp b_m$ means that $a_m = O(b_m)$ and $b_m = O(a_m)$; $a_m \lesssim b_m$ means that $a_m \leq (1+o(1))b_m$. For $a, b \in \mathbb{R}$, we use $a \vee b$ (resp., $a \wedge b$) to denote $\max(a, b)$ [resp., $\min(a, b)$]. For $a \in \mathbb{R}$, let $[a]$ be the integer part of a; $\lfloor a \rfloor = [a]$ if a is not an integer and $[a] - 1$ otherwise; and $\lceil a \rceil = [a] + 1$. For a set A, $|A|$ denotes its cardinality. Define $\log_{\dagger}(x) = \log x$ if $x \geq e$ and $= 1$ otherwise. All the limits in the text are when $m \to \infty$. Throughout the paper, we use C to denote a generic constant, independent of m, whose particular value may change with each appearance. We introduce additional notation in the text.

2. Clusters as geometric shapes in Euclidean space.
We assume that the nodes are embedded in $\Omega_d \subset \mathbb{R}^d$, a compact set with nonempty interior. Let $\|\cdot\|$ denote the corresponding Euclidean norm. For $A \subset \Omega_d$ and $x \in \Omega_d$, let $\text{dist}(x, A) = \inf_{y \in A} \|x - y\|$ and for $r > 0$, define

$$B(A, r) = \{x \in \mathbb{R}^d : \text{dist}(x, A) < r\}.$$

In particular, $B(x, r)$ denotes the (open) Euclidean ball with center x and radius r. On occasion, we will add a subscript d to emphasize that this is a d-dimensional ball.

We consider a sequence $(\mathbb{V}_m)$ of finite subsets of Ω_d, of size $|\mathbb{V}_m| = m$, that are evenly spread out, in the following sense: there is a constant $C \geq 1$, independent of m and a sequence $r_m^* \to 0$ such that

$$(2.1) \qquad C^{-1} m r^d \leq |B(x, r) \cap \mathbb{V}_m| \leq C m r^d \qquad \forall r \in [r_m^*, 1], \forall x \in \Omega_d.$$

In words, the number of nodes in any ball that is not too small is roughly proportional to its volume. For the regular lattice with m nodes in $\Omega_d = [0, 1]^d$, condition (2.1) is satisfied for $r_m^* > \sqrt{d} m^{-1/d}$. This is the smallest possible order of magnitude; indeed, for some constant $C > 0$ and r small enough, there is a set with more than Cr^{-d} disjoint balls with centers in Ω_d, and, by (2.1), they are all nonempty if $r \geq r_m^*$, which forces $r_m^* \geq C m^{-1/d}$. Another example of interest is that of $\mathbb{V}_m$ obtained by sampling m points from the uniform distribution, or any other distribution with a density with respect to the Lebesgue measure on Ω_d, bounded away from zero and infinity; in that case, (2.1) is satisfied with high probability for $r_m^* \geq C (\log(m)/m)^{1/d}$ when C is large enough; for an extensive treatment of this situation, see [56], Chapter 4.

2.1. *Thick clusters.* In this section, we consider clusters as in (1.4), where $\mathcal{A}$ is a class of bi-Lipschitz deformations of the unit d-dimensional ball. This includes the vast majority of all the parametric clusters considered in the literature, such as hyperrectangles and ellipsoids, as long as the shape is not too narrow. Note that a slightly less general situation is briefly mentioned in [8].

We start with a lower bound on the minimax detection rate for discrete balls of a given radius.

PROPOSITION 1.   *Consider $\lambda_m \to 0$ such that $\lambda_m \geq r_m^*$ and let $\mathcal{K}_m$ be the class of all discrete balls of radius λ_m, that is,*

$$\mathcal{K}_m = \{K = B(x, \lambda_m) \cap \mathbb{V}_m : x \in \Omega_d\}.$$

H_0^m *and* H_1^m *are then asymptotically inseparable if*

$$\overline{\Lambda}_m \leq \sqrt{2d \log(1/\lambda_m)} - \eta_m,$$

where $\eta_m \to \infty$.

We now consider a much larger class of clusters and show that, nevertheless, a form of scan statistic achieves that same detection rate. For a function $f : A \subset \mathbb{R}^p \to \mathbb{R}^d$, its Lipschitz constant is defined as

$$\lambda_f = \sup_{x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|}.$$

For $\kappa \geq 1$, let $\mathcal{F}_{d,d}(\kappa)$ be the subclass of bi-Lipschitz functions $f : B(0, 1) \subset \mathbb{R}^d \to \Omega_d$ such that $\lambda_f \lambda_{f^{-1}} \leq \kappa$ or, equivalently,

$$(2.2) \qquad \sup_{x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|} \leq \kappa \inf_{x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|}.$$

For a function $f : A \to \mathbb{R}^d$, define

$$K_f = \text{im}(f) \cap \mathbb{V}_m, \qquad \text{im}(f) := \{f(x) : x \in A\}.$$

Note that λ_f is intimately related to the size of $\text{im}(f)$ and therefore of K_f. Indeed, a simple application of (2.2) implies that, for any $f \in \mathcal{F}_{d,d}(\kappa)$,

$$(2.3) \qquad B(f(0), \lambda_f/\kappa) \subset \text{im}(f) \subset B(f(0), \lambda_f).$$

This implies that sets of the form $\text{im}(f)$, with $f \in \mathcal{F}_{d,d}(\kappa)$, are "thick," in the sense that the smallest ball(s) containing $\text{im}(f)$ and the largest ball(s) included in $\text{im}(f)$ are of comparable sizes.

THEOREM 1.   *Consider $\lambda_m \to 0$ such that $\lambda_m \geq r_m^*$ and define*

$$\mathcal{K}_m = \{K_f : f \in \mathcal{F}_{d,d}(\kappa), \lambda_f \geq \lambda_m\}.$$

An ε_m-scan statistic with $\varepsilon_m \to 0$ and $\varepsilon_m(\log(1/\lambda_m))^{1/(2d)} \to \infty$ then asymptotically separates H_0^m and H_1^m if

$$\underline{\Delta}_m \geq \sqrt{2d\log(1/\lambda_m)} + \eta_m,$$

where $\eta_m = \varepsilon_m^2\sqrt{2d\log(1/\lambda_m)}$. Moreover, if $r_m^* \asymp m^{-1/d}$ and

$$\mathcal{K}_m = \{K_f : f \in \mathcal{F}_{d,d}(\kappa)\},$$

then an ε_m-scan statistic with $\varepsilon_m \to 0$ and $\varepsilon_m(\log m)^{1/(2d)} \to \infty$ asymptotically separates H_0^m and H_1^m if

$$\underline{\Delta}_m \geq \sqrt{2\log m} + \eta_m,$$

where $\eta_m = \varepsilon_m^2\sqrt{2\log m}$.

Therefore, on a larger class of mild deformations of the unit ball, some form of scan statistic achieves essentially the same detection rate as for the class of balls stated in Proposition 1.

We note that the lower bound on $\underline{\Delta}_m$ is driven by the smaller clusters in the class and that the performance guarantee is subject to a proper choice of ε_m. A simple fix for both issues is to combine the tests for different cluster sizes with an appropriate correction for multiple testing. We summarize the consequence of Proposition 1 and Theorem 1 with this observation in the following result, inspired by [71].

COROLLARY 1.   *Consider $\lambda_m \to 0$ such that $\lambda_m \geq r_m^*$ and define*

$$\mathcal{K}_m = \{K_f : f \in \mathcal{F}_{d,d}(\kappa), \lambda_m \geq \lambda_f \geq r_m^*\}.$$

H_0^m and H_1^m are then asymptotically inseparable if, for all $K \in \mathcal{K}_m$,

$$\Lambda_K \leq \sqrt{2\log(m/|K|)} - \eta_m,$$

where $\eta_m \to \infty$. Conversely, let T_ℓ be an ε_ℓ-scan statistic for the subclass $\{K_f \in \mathcal{K}_m : 2^{-\ell} \leq \lambda_f < 2^{-\ell+1}\}$ with $\varepsilon_\ell \ell^{1/(2d)} \to \infty$. There is a test based on $\{T_\ell : \ell \geq 0\}$ that asymptotically separates H_0^m and H_1^m if, for all $K \in \mathcal{K}_m$,

$$\Lambda_K \geq \sqrt{2\log(m/|K|)} + \eta_K,$$

where $\eta_K = \varepsilon_{\ell_K}^2\sqrt{2\log(m/|K|)}$ and $\ell_K = \log(m/|K|)$.

The same procedure, that is, combining ε-scan statistics at different (dyadic) scales, may be implemented in any of the settings we consider in this paper to obtain a test that does not depend on a tuning parameter like ε_m and achieves the same optimal rate at every size. This is simply due to the fact that we only need to consider the order of $\log m$ scales and the fast decaying tails of the scan statistics under the null.

Union of thick clusters. In a number of situations, the signal to be detected may be composed of several clusters. Our results extend readily to this case. Let j_m be a positive integer and consider sets of the form $\bigcup_{j=1}^{j_m} K_{f_j}$, where the union is over some $f_j \in \mathcal{F}_{d,d}(\kappa)$ such that, for j, j', $\lambda_{f_j} \le C\lambda_{f_{j'}}$ and

$$\|f_j(0) - f_{j'}(0)\| \le C(\lambda_{f_j} \vee \lambda_{f_{j'}}),$$

so that the sets $\mathrm{im}(f_j)$ and $\mathrm{im}(f_{j'})$ are of comparable sizes and not too far from each other. In that case, Theorem 1 applies unchanged, as long as the number of clusters is not too large, specifically if $j_m = o(\log(1/\lambda_m))^{1/d}$. (This can be improved if the K_{f_j}'s do not overlap too much.) If the proximity constraint is dropped, then the term $\log(1/\lambda_m)$ in Theorem 1 is replaced by $j_m \log(1/\lambda_m)$.

2.2. *Thin clusters.* In this section, we consider clusters that are built from smooth embeddings in Ω_d of the unit p-dimensional ball, where $p < d$. The special case of curves ($p = 1$) is, for example, relevant in road tracking [29] and in modeling blood vessels in medical imaging [36]. As in the previous section, the results we obtain below are valid for (some) unions of such subsets and, in particular, for submanifolds with a wide array of topologies.

For a differentiable function f between two Euclidean spaces, let Df denote its Jacobian matrix. For $\kappa \ge 1$, let $\mathcal{F}_{p,d}(\kappa)$ be the class of twice differentiable, one-to-one functions $f : B(0, 1) \subset \mathbb{R}^p \to \mathrm{im}(f) \subset \Omega_d$ satisfying $\lambda_f \lambda_{f^{-1}} \le \kappa$ and $\lambda_{Df} \le \kappa \lambda_f$. We consider clusters that are tubular regions around the range of functions in $\mathcal{F}_{p,d}(\kappa)$. For a function f with values in $\mathbb{R}^d$ and $r > 0$, define

$$K_{f,r} = B(\mathrm{im}(f), r) \cap \mathbb{V}_m.$$

Again, λ_f is intimately related to the size of $B(\mathrm{im}(f), r)$ and $K_{f,r}$. This relationship is made explicit in the Supplement. We consider classes of clusters of the form $\{K_{f,r} : f \in \mathcal{F}\}$, where $\mathcal{F}$ is a subclass of $\mathcal{F}_{p,d}(\kappa)$.

We start with a result on the performance of the scan statistic. For a class $\mathcal{F}$ of functions with values in $\mathbb{R}^d$ and for $\varepsilon > 0$, let $N_\varepsilon(\mathcal{F})$ denote its ε-covering number for the sup-norm, that is,

$$N_\varepsilon(\mathcal{F}) = \min\Big\{n : \exists f_1, \ldots, f_n \in \mathcal{F}, \text{ s.t. } \max_{f \in \mathcal{F}} \min_j \|f - f_j\|_\infty \le \varepsilon\Big\}.$$

THEOREM 2. *Let C be the constant defined in Lemma B.2 in the Supplement. Consider $\lambda_m, r_m \to 0$ such that $C^{-1}\lambda_m \ge r_m \ge r_m^*$ and let $\mathcal{F}$ be a subclass of $\mathcal{F}_{p,d}(\kappa)$. Define*

$$\mathcal{K}_m = \{K_{f,r} : f \in \mathcal{F}, \lambda_f \ge \lambda_m, C^{-1}\lambda_m \ge r \ge r_m\}.$$

An ε_m-scan statistic with $\varepsilon_m = o(r_m^{1/2})$ then asymptotically separates H_0^m and H_1^m if

$$\underline{\Delta}_m \ge (1 + \varepsilon_m^2)\sqrt{2\log N_{\varepsilon_m^2}(\mathcal{F}) + 2d\log(1/\lambda_m)}.$$

Just as in Theorem 1, if $r_m^* \asymp m^{-1/d}$, we can dispense with the restriction $r_m \geq r_m^*$ and replace the factor $\log(1/\lambda_m^d)$ by $\log m$ in the bound.

For a typical parametric class $\mathcal{F}$, $\log N_\varepsilon(\mathcal{F}) \sim a(\mathcal{F}) \log(1/\varepsilon)$, so the scan statistic (over an appropriate net) is accurate if

$$(2.4) \qquad \underline{\Delta}_m \geq (1 + r_m)\sqrt{2a(\mathcal{F})\log(1/r_m) + 2d\log(1/\lambda_m)}.$$

On the other hand, $\log N_\varepsilon(\mathcal{F}) \asymp (1/\varepsilon)^{a(\mathcal{F})}$ for a typical nonparametric class $\mathcal{F}$ [42], so the scan statistic (over an appropriate net) is accurate if

$$(2.5) \qquad \underline{\Delta}_m r_m^{a(\mathcal{F})/2} \to \infty.$$

Finding a sharp lower bound for the minimax detection rate is more challenging for thin clusters compared to thick clusters. By considering disjoint tubes around p-dimensional hyperrectangles, we obtain a lower bound that matches, in order of magnitude, the rate achieved by the scan statistic when the class $\mathcal{F}$ is parametric, displayed in (2.4).

PROPOSITION 2.   *Consider $\lambda_m, r_m \to 0$ with $\lambda_m \geq r_m \geq r_m^*$. Let $U : \mathbb{R}^p \to \mathbb{R}^d$ be the canonical embedding so that $Ux = (x, 0)$ and let*

$$\mathcal{F} = \{f : B(0,1) \subset \mathbb{R}^p \to \Omega_d, \, f(x) = \lambda_m Ux + b, \text{ where } b \in \mathbb{R}^d\}.$$

Define

$$\mathcal{K}_m = \{K_{f,r_m} : f \in \mathcal{F}\}.$$

H_0^m *and* H_1^m *are then asymptotically inseparable if*

$$\underline{\Delta}_m \leq \sqrt{2(d-p)\log(1/r_m) + 2p\log(1/\lambda_m)} - \eta_m,$$

where $\eta_m \to \infty$.

The proof is parallel to that of Proposition 1 and is therefore omitted.

For at least one family of nonparametric curves ($p = 1$), we show that the rate displayed at (2.5) matches the minimax rate, except for a logarithmic factor. For concreteness, we assume that $\Omega_d = [0, 1]^d$. Let $\mathcal{H}(\alpha, \kappa)$ be the Hölder class of functions $g : [0, 1] \to [0, 1]$ satisfying

$$\left|g^{(s)}(x)\right| \leq \kappa \qquad \forall x \in [0, 1], \forall s < \alpha;$$

$$\left|g^{(\lfloor \alpha \rfloor)}(x) - g^{(\lfloor \alpha \rfloor)}(y)\right| \leq \kappa |x - y|^{\alpha - \lfloor \alpha \rfloor} \qquad \forall x, y \in [0, 1].$$

PROPOSITION 3.   *Let $r_m \to 0$ with $r_m \geq r_m^*$. Let $\mathcal{F}$ be the class of functions of the form $f(x) = (x, g_1(x), \ldots, g_{d-1}(x))$, where $g_j \in \mathcal{H}(\alpha, \kappa)$, with $\alpha \geq 2$. Define*

$$\mathcal{K}_m = \{K_{f,r_m} : f \in \mathcal{F}\}.$$

H_0^m *and* H_1^m *are then asymptotically inseparable if*

$$\overline{\Lambda}_m r_m^{1/(2\alpha)} (\log m)^{3/2} \to 0.$$

Thus, for the detection of curves with Hölder regularity, a scan statistic achieves the minimax rate within a poly-logarithmic factor. We prove Proposition 3 by reducing the problem of detecting a band in a graph so that we can use results from Section 3.1. We do not know how to generalize this approach to higher-dimensional surfaces (i.e., $p \geq 2$).

2.3. *The spatio-temporal setting.* In this section, we consider the spatio-temporal setting described in Section 1.2.2. This is a special case of the spatial setting we have considered thus far, with time playing the role of an additional dimension. For their relevance in applications and concreteness of exposition, we focus on two specific models. In Section 2.3.1, we consider cluster sequences with a limit; as we shall see, this assumption is implicit in some popular models for epidemics. In Section 2.3.2, we consider cluster sequences of bounded variations.

In the remainder of this section, we assume, for concreteness, that $\mathbb{T}_m = \{0, 1, \ldots, t_m\}$ with $t_m \to \infty$. Our results apply without any changes if the set of nodes varies with time, that is, with index set of the form $\prod_{t \in \mathbb{T}_m} \mathbb{V}_m^t$, in the case where each $\mathbb{V}_m^t$ satisfies (2.1) with C and r_m^* independent of t.

2.3.1. *Cluster sequences with a limit.* We focus here on cluster sequences obeying $K_{t_m} \neq \varnothing$, that is, the anomalous cluster is present at the last time point. This is a standing assumption in syndromic surveillance systems [44]. To illustrate the difference, consider a typical change-point problem setting, where $\mathbb{V}_m$ contains only one node and, for simplicity, assume that Λ_K is independent of K and that Λ_m denotes this common value. First, let the cluster be any discrete interval (in time), so the signal may not be present at time $t = t_m$. This is a special case of Section 2.1, with time playing the role of a spatial dimension ($d = 1$); we saw in Corollary 1 that the detection threshold is at $\Lambda_m \sim \sqrt{2 \log |\mathbb{T}_m|}$. Now, let the emerging cluster be any discrete interval that includes $t = t_m$. Detecting such an interval is actually much easier since we do not need to search where the interval is located, which is what drives the detection threshold for the thick clusters in Section 2.1—we need only determine its length. Specifically, the scan statistic over the dyadic intervals containing $t = t_m$ asymptotically separates the hypotheses if $\Lambda_m \asymp \sqrt{\log \log |\mathbb{T}_m|}$.

Regarding the actual evolution of the cluster in time, a number of growth models have been suggested, for example, cellular automata [37, 61] and their random equivalent, threshold growth automata [12, 31], which have been used to model epidemics [2]. The latter includes the well-known Richardson model [59]. Under some conditions, these models develop an asymptotic shape (with probability one), a convex polygon in the case of threshold growth automata. Less relevant for modeling epidemics, internal diffusion limited aggregation is another growth model with a limiting shape [48].

The simplest cluster sequences with limiting shape are space–time cylinders, for which we have the equivalent of Proposition 1. (The proof is completely parallel and we omit details.)

PROPOSITION 4. *Consider $\lambda_m \to 0$ with $\lambda_m \geq r_m^*$ and let $\mathcal{K}_m$ be the class of all space–time cylinders of the form $K_t = B(x, \lambda_m) \cap \mathbb{V}_m, \forall t = 0, \ldots, t_m$, where $x \in \Omega_d$. H_0^m and H_1^m are then asymptotically inseparable if*

$$\underline{\Delta}_m \leq \sqrt{2d \log(1/\lambda_m)} - \eta_m,$$

where $\eta_m \to \infty$.

With only one possible shape and known starting point, such a model is rather uninteresting. We now consider a much larger class of cluster sequences with some sort of limit [in the sense of (2.6)] and show that, nevertheless, a form of scan statistic achieves that same detection rate. For a cluster sequence $K = (K_t, t \in \mathbb{T}_m)$, let $t_K = \min\{t : K_t \neq \varnothing\}$, which is the time when K originates. The following is the equivalent of Theorem 1. The metric δ appearing below is defined in (1.3).

PROPOSITION 5. *Consider sequences $\lambda_m \to 0$ with $\lambda_m \geq r_m^*$ and $\log \log t_m = o(\log(1/\lambda_m))$, and a function $v(t)$ with $\lim_{t \to \infty} v(t) = 0$ and $v(t) \leq 1$ for all $t \geq 0$. Let $\mathcal{K}_m$ be a class of cluster sequences such that $t_m - \max\{t_K : K \in \mathcal{K}_m\} \to \infty$ and, for each $K = (K_t, t \in \mathbb{T}_m) \in \mathcal{K}_m$, there exists $f \in \mathcal{F}_{d,d}(\kappa)$ with $\lambda_f \geq \lambda_m$ such that*

$$(2.6) \qquad \delta(K_t, \operatorname{im}(f) \cap \mathbb{V}_m) \leq v(t - t_K) \qquad \forall t \in \mathbb{T}_m.$$

There is then a scan statistic over a family of space–time cylinders that asymptotically separates H_0^m and H_1^m if

$$\underline{\Delta}_m \geq (1 + \xi_m)\sqrt{2d \log(1/\lambda_m)},$$

where $\xi_m \to 0$ slowly enough.

If the starting time is uniformly bounded away from t_m and the convergence to the thick spatial cluster [in the sense of (2.6)] occurs at a uniform speed, then all of the cluster sequences in the class have sufficient time to develop into their "limiting" shapes. The space–time cylinders over which we scan are based on an ε-net for the possible limiting shapes, that is, the class of thick clusters.

Scanning over space–time cylinders (with balls as bases) is advocated in the disease outbreak detection literature [44]. Although seemingly naive, this approach achieves, in our asymptotic setting, the minimax detection rate if the cluster sequences develop into balls and, in general, falls short by a constant factor.

We mention that the equivalent of Corollary 1 holds here as well, in that we can combine the different scans at different space–time scales to obtain a test that does not depend on a tuning parameter (implicit here) and which achieves the same rate for the cluster class defined as above, but with $\lambda_m \geq \lambda_f \geq r_m^*$, which is the class that appears in Corollary 1.

2.3.2. *Cluster sequences of bounded variation.* In target tracking [11, 50], the target is usually assumed to be limited in its movements due to maximum speed and maneuverability. With this example in mind, we consider classes of cluster sequences of bounded variation, meaning that the cluster is limited in the amount it can change in a given period of time. As the rates we obtain in this subsection are the same with or without the condition $K_{t_m} \neq \varnothing$, we do not make that assumption. Let $t_K^+ = \max\{t : K_t \neq \varnothing\}$.

We consider space–time tubes around Hölder space–time curves. For $\alpha \in (0, 1]$ and $\kappa > 0$, let $\mathcal{H}_\infty(\alpha, \kappa)$ be the Hölder class of functions $g : [0, \infty) \to [0, 1]$ satisfying

$$(2.7) \qquad |g(x) - g(y)| \leq \kappa |x - y|^\alpha \qquad \forall x, y \in [0, \infty).$$

The following is the equivalent of Proposition 3.

PROPOSITION 6. *Assume that $\Omega_d = [0, 1]^d$. Consider sequences $r_m \to 0$ with $r_m \geq 2r_m^*$ and ξ_m such that $1 \leq \xi_m \leq t_m$. Let $\mathcal{K}_m$ be the class of all cluster sequences K_g of the form $K_{g,t} = B(g(t/\xi_m), r_m) \cap \mathbb{V}_m$ for all $t = t_K, \ldots, t_K^+$, for some $g = (g_1, \ldots, g_d)$ with $g_j \in \mathcal{H}_\infty(\alpha, \kappa)$. Then, H_0^m and H_1^m are asymptotically inseparable if*

$$\overline{\Lambda}_m (t_m / \lceil \xi_m r_m^{1/\alpha} \rceil)^{-1/2} \log(t_m / \lceil \xi_m r_m^{1/\alpha} \rceil)(\log(\xi_m) + \log\log(t_m))^{1/2} \to 0.$$

Conversely, an ε-scan statistic with $\varepsilon < \sqrt{2}$ asymptotically separates H_0^m and H_1^m if

$$\underline{\Delta}_m ((t_m / \lceil \xi_m r_m^{1/\alpha} \rceil) \log_\dagger(\xi_m^{-1} r_m^{-1/\alpha}) + \log m)^{-1/2} \to \infty.$$

For simplicity, assume that t_m is a power of m. If $\xi_m r_m^{1/\alpha} = O(1)$, then the detection threshold is roughly of order $t_m^{1/2}$, while if $\xi_m r_m^{1/\alpha}$ is large, yet small enough that $t_m/(\xi_m r_m^{1/\alpha})$ is still a power of m, then the detection threshold is roughly of order $(t_m/(\xi_m r_m^{1/\alpha}))^{1/2}$.

A form of scan statistic is actually able to attain the same detection rate when the radius is unknown, but restricted to $r \geq r_m$. In fact, another form of scan statistic achieves a slightly different rate over a much larger class of cluster sequences with bounded variations. Let $\mathcal{S}(r, \kappa)$ be the set of subsets $S \subset \Omega_d$ such that $B(x, r) \subset S \subset B(x, \kappa r)$ for some $x \in \Omega_d$.

PROPOSITION 7. *Consider a sequence ξ_m such that $1 \leq \xi_m \leq t_m$ and a constant $\eta > 0$. Define $\mathcal{K}_m$ as the class of cluster sequences K such that, for each $t = t_K, \ldots, t_K^+$, $K_t = S_t \cap \mathbb{V}_m$, where $S_t \in \mathcal{S}(r_t, \kappa)$ for some $r_t \geq r_m^*$, and, for any $s, t = t_K, \ldots, t_K^+$,*

$$(2.8) \qquad \delta(K_t, K_s) \leq \eta \qquad if \ |t - s| \leq \xi_m.$$

Then, for η small enough, an ε-scan statistic with $\varepsilon < \sqrt{2}$ asymptotically separates H_0^m *and* H_1^m *if*

$$\underline{\Lambda}_m\big((t_m/\xi_m)(\log m) + \log t_m\big)^{-1/2} \to \infty.$$

Consider the condition

(2.9)
$$\delta(K_t, K_s) \le \nu(|t - s|/\xi_m) \qquad \forall s, t \in \{t_K, \ldots, t_K^+\},$$

for a function $\nu : [0, \infty) \to [0, \sqrt{2}]$. Then, (2.8) is satisfied with $\eta = \nu(1)$ and the same ξ_m. The requirement in Proposition 7 is that $\nu(1)$ be small enough. In particular, the cluster sequences considered in Proposition 6 satisfy, for some constant $C > 0$,

$$\delta(K_t, K_s) \le C r_m^{-1/2}\big(r_m^* \vee (|t - s|/\xi_m)^\alpha\big)^{1/2} \qquad \forall s, t \in \{t_K, \ldots, t_K^+\}.$$

This comes from Lemma C.1 in the Supplement and (2.7). Therefore, assuming $\xi_m \ll (r_m^*)^{-1/\alpha}$, (2.9) is satisfied with $\nu(u) = u^{\alpha/2}$ and ξ_m replaced by $\xi_m r_m^{1/\alpha}$. In that case, the detection rates obtained by the scan statistics of Propositions 6 and 7 are of comparable orders of magnitude.

3. Clusters as connected components in a graph. In this section, we model the network with the d-dimensional square lattice; specifically, we assume that $m^{1/d}$ is an integer (for convenience) and consider $\mathbb{V}_m = \{0, 1, \ldots, m^{1/d} - 1\}^d$, seen as a subgraph of the usual d-dimensional lattice. We assume that $d \ge 2$ since the case where $d = 1$ is treated in Section 2.1. We work with the ℓ^1-norm, which corresponds to the shortest-path distance in the graph; let $B(v, h)$ denote the corresponding open ball with center v and radius h so that $B(v, h) = \{v\}$ for $h \in (0, 1]$, and, for a subset of nodes V, let $B(V, h) = \bigcup_{v \in V} B(v, h)$.

3.1. *Paths and bands.* A nearest-neighbor band of length ℓ and width h is of the form $B(V, h)$, where $V = (v_0, \ldots, v_\ell)$ forms a path in $\mathbb{Z}^d$. A band with unit width ($h = 1$) is just a path.

We say that a path $(v_0, \ldots, v_\ell)$ in $\mathbb{Z}^d$ is nondecreasing if, for all $t = 1, \ldots, \ell$, $v_t - v_{t-1}$ has exactly one coordinate equal to 1 and all other coordinates equal to 0. The case of paths was treated in detail in [7]; it corresponds to taking $h_m = 1$ below.

THEOREM 3. *Suppose that $d \ge 2$ and let $\mathcal{K}_m$ be the class of bands of width h_m generated by nondecreasing paths in $\mathbb{V}_m$ of length ℓ_m, starting at the origin, with $m^{1/d} \ge \ell_m \ge h_m$. Then, H_0^m and H_1^m are asymptotically inseparable if*

$$\overline{\Lambda}_m(\ell_m/h_m)^{-1/2}\log_{\dagger}(\ell_m)(\log h_m + \log_{\dagger}\log \ell_m)^{1/2} \to 0 \qquad for\ d = 2,$$

$$\overline{\Lambda}_m(\ell_m/h_m)^{-1/2}(\log_{\dagger} h_m)(\log_{\dagger}\log h_m) \to 0 \qquad for\ d \ge 3.$$

Conversely, an ε-scan statistic with $\varepsilon < \sqrt{2}$ fixed asymptotically separates H_0^m *and* H_1^m *if*

$$\underline{\Delta}_m(\ell_m / h_m)^{-1/2} \to \infty.$$

For the case of nondecreasing paths, a form of the scan statistic achieves the minimax rate in dimension $d \geq 3$, while it falls short by a logarithmic factor in dimension $d = 2$. In the latter setting, Arias-Castro et al. [7] introduces a test that asymptotically separates H_0^m and H_1^m if

$$(3.1) \qquad \underline{\Delta}_m(\ell_m / h_m)^{-1/2} \log_\dagger(\ell_m)^{1/2} \to \infty,$$

coming slightly closer to the minimax rate.

In fact, even when the band has unknown length, width and starting location, and when the path is not restricted to be nondecreasing, a form of scan statistic achieves the same rate, except for a logarithmic factor.

PROPOSITION 8. *Suppose that $d \geq 2$ and let $\mathcal{K}_m$ be the class of all bands of width h and length ℓ, where $\ell_m \geq \ell \geq h \geq h_m$, that are within $\mathbb{V}_m$ and generated by paths that do not self-intersect. An ε-scan statistic, with $\varepsilon < \sqrt{2}$, then asymptotically separates* H_0^m *and* H_1^m *if*

$$\underline{\Delta}_m\big(\ell_m / h_m + \log(m / h_m^d) + \log_\dagger \log \ell_m\big)^{-1/2} \to \infty.$$

3.2. *Arbitrary connected components.* We consider here classes of connected components with a constraint on their sizes. Arbitrary connected components in the square lattice are sometimes called *animals* or *polyominoes* (polycubes in dimension $d \geq 3$), which are well-studied objects in combinatorics, where the goal is to count the number of polyominoes [41]. We mention in passing the results in [22] which provide a law of large numbers for the scan statistic under the null. Otherwise, such objects are fairly new to statistics. Detecting animals is, of course, harder than detecting paths since paths are themselves animals. The result below offers a sharp detection threshold for connected components of sufficiently small size.

PROPOSITION 9. *Let $\mathcal{K}_m$ be the class of animals of size $k_m = o(m)$ within $\mathbb{V}_m$. H_0^m and H_1^m are then asymptotically inseparable if*

$$\overline{\Lambda}_m \leq \sqrt{2 \log m} - \eta_m,$$

where $\eta_m \to \infty$. Conversely, let $\mathcal{K}_m$ be the class of animals of size not exceeding $k_m^+ = o(\log m)$ within $\mathbb{V}_m$. The actual scan statistic then asymptotically separates H_0^m *and* H_1^m *if*

$$\Delta_m \geq \sqrt{2 \log m}.$$

Note that, in general, we can obtain a quick (naive) upper bound on the detection rate for large clusters by considering the simple test that rejects for large values of $\sum_{v \in \mathbb{V}_m} X_v$ (this is the "average test" in [1]). This test asymptotically separates H_0^m and H_1^m if $\underline{\Lambda}_m k_m^{1/2} m^{-1/2} \to \infty$, assuming the clusters in $\mathcal{K}_m$ are of size bounded below by k_m. An open question of theoretical interest is whether, for the class of animals of size $k_m = \sqrt{m}$ in the two-dimensional lattice, there is a test that asymptotically separates H_0^m and H_1^m when $\underline{\Lambda}_m k_m^{-1/2} \to 0$ slowly enough. In dimension three or higher, Theorem 3 implies that this is not possible, even for paths.

4. Discussion.

4.1. *Extension to exponential families.* Although the previous results were stated for the normal location model, they extend to any one-parameter exponential model if the anomalous clusters are large enough. For example, consider a Bernoulli model where the variables are Bernoulli with parameter $1/2$ under the null and with parameter $p_K > 1/2$ when they belong to the anomalous cluster K; or, a Poisson model where the variables are Poisson with mean 1 under the null and $\mu_K > 1$ when they belong to the anomalous cluster K. In general, transforming the variables and/or the parameter if necessary, we may assume that the model is of the form F_θ, with density $f_\theta(x) = \exp(\theta x - \log \varphi(\theta))$ with respect to F_0, where, by definition, $\varphi(\theta) = \mathbb{E}_0[\exp(\theta X)]$, where $\mathbb{E}_0$ denotes the expectation under F_0. We always assume that $\varphi(\theta) < \infty$ for θ in a neighborhood of 0. Let $\sigma^2 = \mathrm{Var}_0(X)$, the variance of $X \sim F_0$. In the Bernoulli model, the correspondence is $\theta = \log(p/(1-p))$ and $\sigma^2 = 1/4$; in the Poisson model, $\theta = \log \lambda$ and $\sigma^2 = 1$. Under the null hypothesis, all of the variables at the nodes have distribution F_0, that is,

$$H_0^m : X_v \sim F_0 \qquad \forall v \in \mathbb{V}_m.$$

Under the alternative, the variables at the nodes belonging to the anomalous cluster $K \in \mathcal{K}_m$ have distribution F_{θ_K} with $\theta_K := \sigma \Lambda_K |K|^{-1/2}$, that is,

$$H_{1,K}^m : X_v \sim F_{\theta_K} \qquad \forall v \in K; \qquad X_v \sim F_0 \qquad \forall v \notin K.$$

As before, the variables are assumed to be independent.

If the clusters in the class are sufficient large, then the results presented for the normal location family hold unchanged. Intuitively, large enough clusters allow for the sums over them to be approximately normally distributed. Details are provided in the Supplement. For example, we have the following equivalent of Corollary 1 in the context of thick clusters as in Section 2.1. Consider $\lambda_m \geq r_m \geq r_m^*$ with $m r_m^d (\log 1/r_m)^{-3} \to \infty$ (which guarantees that the clusters in the class are large enough) and define the class

$$\mathcal{K}_m = \{K_f : f \in \mathcal{F}_{d,d}(\kappa), \lambda_m \geq \lambda_f \geq r_m\}.$$

In this setting, under the Bernoulli model, the detection threshold is at

$$p_K = \frac{1}{2} + \frac{1}{8|K|^{1/2}} (2\log(m/|K|))^{1/2};$$

under the Poisson model, the detection threshold is at

$$\mu_K = 1 + \frac{1}{|K|^{1/2}} (2\log(m/|K|))^{1/2}.$$

Note that without a lower bound on the minimum size of the anomalous clusters, the general analysis breaks down and the results depend on the specific exponential model. For example, unless $\min\{|K| : K \in \mathcal{K}_m\} \to \infty$ fast enough, detection is impossible in the Bernoulli model, even if the anomalous nodes have value 1 under the alternative.

4.2. *Other extensions.* The array of possible models is as wide as the breadth of real-world applications. We mention a few possible variations below.

Beyond exponential families. Using an exponential family of distributions allows us to obtain sharp detection lower bounds. Otherwise, similar results, although not as sharp, may be obtained for essentially any family of distribution F_θ, where the distance between the null $\theta = 0$ and an alternative θ is in terms of the chi-square distance between F_0 and F_θ; see [7], Section 5.

Different means at the nodes. We could consider a situation where the mean varies over the nodes of the anomalous cluster. This situation is considered in [33] for the case of intervals, and the constant in the detection rate is indeed different. We implicitly considered a worst case scenario where the mean is bounded below over the anomalous cluster and subsequently assumed it was equal to that lower bound everywhere over the anomalous cluster. However, our results hold unchanged if we allow X_v to have any mean above θ_K, for every $v \in K$, K being the anomalous cluster.

Dependencies. Also of interest is the case where the variables are dependent. In the spatial setting, the same paper [33] solves this problem for the case of the one-dimensional lattice, with the correlation between X_v and X_w decaying as a function of distance between v and w. We postulate that the same result holds in higher dimensions. In the spatio-temporal setting, variables could be dependent across time as well, involving a higher degree of sophistication. We plan on pursuing these generalizations in future publications.

Unknown variance or other parameters. We assumed throughout that the variance was known (and equal to 1 after normalization). This is, in fact, a mild as-

sumption, as one can consistently estimate the variance using a robust estimator, say the median absolute deviation (MAD), with the usual $\sqrt{m}$-convergence rate, assuming that the anomalous cluster corresponds to a small part of the entire network. When dealing with one-parameter families such as Bernoulli or Poisson, the issue is to estimate the parameter under the null and a robust version of the maximum likelihood (e.g., trimmed mean for these two examples) can be used for that purpose.

4.3. *Energy, bandwidth and other constraints.* We assume throughout that a central processor has access to all the information measured at the nodes and, based on that, makes a decision as to whether there is an anomalous cluster of nodes in the network or not. This assumption is reasonable in, for example, the context of image processing or syndromic surveillance. However, real-world sensor networks of the wireless type are often constrained by energy and/or bandwidth considerations. A growing body of literature [50] is dedicated to designing efficient (e.g., decentralized) communication protocols for sensor networks under such constraints. As mentioned in Section 1.3, the papers we are aware of consider very simplistic detection settings. In the context of the present paper, it would be interesting to study how the detection rates change when different communication protocols are used.

We also assume that we have infinite computational power. However, all real-world systems operate under finite energy and processing resources. In the same way, it would be interesting to know what detection rates are achievable under such computational constraints.

4.4. *On computing the scan statistic.* In all of the settings we consider in this paper, the scan statistic comes close to achieving the minimax detection rate. Turning to computational issues, however, it is very demanding, even when scanning for simple parametric clusters such as rectangles. For general shapes, Duczmal, Kulldorff and Huang [25] suggests a simulated annealing algorithm, which, from a theoretical point of view, is extremely difficult to analyze. For parametric shapes and blobs, Arias-Castro, Donoho and Huo [8] advocates the use of ε_m-scan statistics based on multiscale nets built out of unions of dyadic hypercubes; similar ideas appear in [71]. Partial results suggest that this approach yields, in theory, a near-optimal algorithm for detecting the more general thick clusters considered in Section 2.1.

For the thin clusters of Section 2.2, or for the bands of Section 3.1, the situation is quite different. Take the latter. After pre-processing the data by performing a moving average with an appropriate radius, it remains to find the maximum over a restricted, yet exponentially large, set of paths. Without further restriction, this problem, known as the "bank robber problem" or "reward budget problem" [21], is NP-hard. Note that DasGupta et al. [21] suggests a polynomial time approximation that deserves further investigation. The case of thin clusters is even harder. In the

context of point clouds, Arias-Castro, Efros and Levi [10] introduces multiscale nets that could be adapted to the setting of a network. It remains to compute the scan statistic over this net, which seems particularly challenging for surfaces of dimension $p \geq 2$, which no longer correspond to paths. In the spatio-temporal setting of Section 2.3, dynamic programming ideas could be used, as done in [9] in the context of point clouds and in [17] in the context of a harmonic analysis decomposition of chirps.

4.5. *Open theoretical problems.* The paper leaves two main theoretical problems unresolved. The first one concerns obtaining sharper bounds for the detection of thin clusters. This is in the context of Section 2.2. For parametric classes, the challenge is to match constants in the rate, while, for nonparametric classes, the challenge is to obtain sharper lower bounds, perhaps closer to what a scan statistic is shown to achieve in Theorem 1. We were only able to do the latter for curves; see Proposition 3.

The second one concerns comparing the detection rates for arbitrary connected components and for paths. At a given size, the thicker the band (relative to its length), the easier it is to detect it; see Theorem 3. It seems, therefore, that the most difficult connected components to detect are paths or unions of paths. But is this true? In other words, are the minimax detection rates for arbitrary connected components and paths of a similar order of magnitude?

SUPPLEMENTARY MATERIAL

Supplement: Technical Arguments (DOI: 10.1214/10-AOS839SUPP; .pdf). In the supplementary file [6], we prove the results stated here. It is divided into three sections. In the first section, we state and prove general lower bounds on the minimax rate and upper bounds on the detection rate achieved by an ε-scan statistic. We do this for the normal location model first and extend these results to a general one-parameter exponential family. In the second section, we gather a number of results on volumes and node counts. In the third and last section, we prove the main results.

REFERENCES

[1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092.

[2] AGUR, S., DIEKMANN, O., HEESTERBEEK, H., CUSHING, J., GYLLENBERG, M., KIMMEL, M., MILNER, F., JAGERS, P. and KOSTOVA, T., EDS. (1999). *Epidemiology, Cellular Automata and Evolution*. Elsevier, Oxford. MR1683061

[3] AKYILDIZ, I., SU, W., SANKARASUBRAMANIAM, Y. and CAYIRCI, E. (2002). A survey on sensor networks. *IEEE Communications Magazine* **40** 102–114.

[4] ALDOSARI, S. and MOURA, J. (2004). Detection in decentralized sensor networks. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2004 (ICASSP'04)* **2** 277–280.

[5] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2009). Detection of an abnormal cluster in a network. In *Proc. 57th Session of the International Statistical Institute, Durban, South Africa.*

[6] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2010). Supplement to "Detection of an anomalous cluster in a network." DOI: 10.1214/10-AOS839SUPP.

[7] ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. MR2435454

[8] ARIAS-CASTRO, E., DONOHO, D. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. MR2246369

[9] ARIAS-CASTRO, E., DONOHO, D. and HUO, X. (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.* **34** 326–349. MR2275244

[10] ARIAS-CASTRO, E., EFROS, B. and LEVI, O. (2010). Networks of polynomial pieces with application to the analysis of point clouds and images. *J. Approx. Theory* **162** 94–130. MR2565828

[11] ARORA, A., DUTTA, P., BAPAT, S., KULATHUMANI, V., ZHANG, H., NAIK, V., MITTAL, V., CAO, H., DEMIRBAS, M., GOUDA, M., CHOI, Y., HERMAN, T., KULKARNI, S., ARUMUGAM, U., NESTERENKO, M., VORA, A. and MIYASHITA, M. (2004). A line in the sand: A wireless sensor network for target detection, classification and tracking. *Comput. Networks* **46** 605–634.

[12] BOHMAN, T. and GRAVNER, J. (1999). Random threshold growth dynamics. *Random Structures Algorithms* **15** 93–111. MR1698409

[13] BOUTSIKAS, M. V. and KOUTRAS, M. V. (2006). On the asymptotic distribution of the discrete scan statistic. *J. Appl. Probab.* **43** 1137–1154. MR2274642

[14] BRAAMS, J., PRUIM, J., FRELING, N., NIKKELS, P., ROODENBURG, J., BOERING, G., VAALBURG, W. and VERMEY, A. (1995). Detection of lymph node metastases of squamous-cell cancer of the head and neck with FDG-PET and MRI. *Journal of Nuclear Medicine* **36** 211.

[15] BRENNAN, S. M., MIELKE, A. M., TORNEY, D. C. and MACCABE, A. B. (2004). Radiation detection with distributed sensor networks. *IEEE Computer* **37** 57–59.

[16] BRODSKY, B. and DARKHOVSKY, B. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications* **243**. Kluwer Academic, Dordrecht. MR1228205

[17] CANDÈS, E. J., CHARLTON, P. R. and HELGASON, H. (2008). Detecting highly oscillatory signals by chirplet path pursuit. *Appl. Comput. Harmon. Anal.* **24** 14–40. MR2379113

[18] CARON, Y., MAKRIS, P. and VINCENT, N. (2002). A method for detecting artificial objects in natural environments. In *Proceedings 16th International Conference on Pattern Recognition* **1** 10600. IEEE Comput. Soc., New York.

[19] CUI, Y., WEI, Q., PARK, H. and LIEBER, C. (2001). Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* **293** 1289–1292.

[20] CULLER, D., ESTRIN, D. and SRIVASTAVA, M. (2004). Overview of sensor networks. *Computer* **37** 41–49.

[21] DASGUPTA, B., HESPANHA, J. P., RIEHL, J. and SONTAG, E. (2006). Honey-pot constrained searching with local sensory information. *Nonlinear Anal.* **65** 1773–1793. MR2252129

[22] DEMBO, A., GANDOLFI, A. and KESTEN, H. (2001). Greedy lattice animals: Negative values and unconstrained maxima. *Ann. Probab.* **29** 205–241. MR1825148

[23] DEMIRBAŞ, K. (1987). Maneuvering target tracking with hypothesis testing. *IEEE Trans. Aerospace Electron. Systems* **23** 757–766. MR0926001

[24] DESOLNEUX, A., MOISAN, L. and MOREL, J.-M. (2003). Maximal meaningful events and applications to image analysis. *Ann. Statist.* **31** 1822–1851. MR2036391

[25] DUCZMAL, L., KULLDORFF, M. and HUANG, L. (2006). Evaluation of spatial scan statistics for irregularly shaped clusters. *J. Comput. Graph. Statist.* **15** 428–442. MR2256152

[26] DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330. MR0220340

[27] FITCH, J., RABER, E. and IMBRO, D. (2003). Technology challenges in responding to biological or chemical attacks in the civilian sector. *Science* **302** 1350–1354.

[28] GEELHOOD, B., ELY, J., HANSEN, R., KOUZES, R., SCHWEPPE, J. and WARNER, R. (2003). Overview of portal monitoring at border crossings. *2003 IEEE Nuclear Science Symposium Conference Record* **1** 513–517.

[29] GEMAN, D. and JEDYNAK, B. (1996). An active testing model for tracking roads in satellite images. *IEEE Trans. Pattern Anal. Mach. Intell.* **18** 1–14.

[30] GLAZ, J., NAUS, J. and WALLENSTEIN, S. (2001). *Scan Statistics*. Springer, New York. MR1869112

[31] GRAVNER, J. and GRIFFEATH, D. (2006). Random growth models with polygonal shapes. *Ann. Probab.* **34** 181–218. MR2206346

[32] HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

[33] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

[34] HEFFERNAN, R., MOSTASHARI, F., DAS, D., KARPATI, A., KULLDORFF, M. and WEISS, D. (2004). Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases* **10** 858–864.

[35] HILLS, R. (2001). Sensing for danger. *Sci. Technol. Rev.* Available at https://www.llnl.gov/str/JulAug01/Hills.html.

[36] HUSBY, O. and RUE, H. (2004). Estimating blood vessel areas in ultrasound images using a deformable template model. *Stat. Model.* **4** 211–226. MR2062101

[37] ILACHINSKI, A. (2001). *Cellular Automata: A Discrete Universe*. World Scientific, River Edge, NJ. MR1849342

[38] JAIN, A., ZHONG, Y. and DUBUISSON-JOLLY, M. (1998). Deformable template models: A review. *Signal Processing* **71** 109–129.

[39] JAMES, D., CLYMER, B. D. and SCHMALBROCK, P. (2001). Texture detection of simulated microcalcification susceptibility effects in magnetic resonance imaging of breasts. *Journal of Magnetic Resonance Imaging* **13** 876–881.

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

[41] KLARNER, D. (1967). Cell growth problems. *Canad. J. Math.* **19** 851–863. MR0214489

[42] KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1961). ε-entropy and ε-capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)* **17** 277–364. MR0124720

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

[44] KULLDORFF, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *J. Roy. Statist. Soc. Ser. A* **164** 61–72. MR1819022

[45] KULLDORFF, M., FANG, Z. and WALSH, S. J. (2003). A tree-based scan statistic for database disease surveillance. *Biometrics* **59** 323–331. MR1987399

[46] KULLDORFF, M., HEFFERNAN, R., HARTMAN, J., ASSUNCAO, R. and MOSTASHARI, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLOS Medicine* **2** 216.

[47] KULLDORFF, M., HUANG, L., PICKLE, L. and DUCZMAL, L. (2006). An elliptic spatial scan statistic. *Stat. Med.* **25** 3929–43. MR2297401

[48] LAWLER, G., BRAMSON, M. and GRIFFEATH, D. (1992). Internal diffusion limited aggregation. *Ann. Probab.* **20** 2117–2140. MR1188055

[49] LEXA, M., ROZELL, C., SINANOVIC, S. and JOHNSON, D. (2004). To cooperate or not to cooperate: Detection strategies in sensor networks. *In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2004 (ICASSP'04)* **3** 841–844.

[50] LI, D., WONG, K., HU, Y. H. and SAYEED, A. (2002). Detection, classification and tracking of targets. *IEEE Signal Processing Magazine* **19** 17–29.

[51] MCINERNEY, T. and TERZOPOULOS, D. (1996). Deformable models in medical image analysis: A survey. *Medical Image Analysis* **1** 91–108.

[52] MEI, Y. (2008). Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks. *IEEE Trans. Inform. Theory* **54** 2072–2089. MR2450850

[53] MOON, N., BULLITT, E., VAN LEEMPUT, K. and GERIG, G. (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* 372–379. Springer, London.

[54] PATIL, G. P., BALBUS, J., BIGING, G., JAJA, J., MYERS, W. L. and TAILLIE, C. (2004). Multiscale advanced raster map analysis system: Definition, design and development. *Environ. Ecol. Stat.* **11** 113–138. MR2086391

[55] PATWARI, N. and HERO, A. (2003). Hierarchical censoring for distributed detection in wireless sensor networks. In *Proceedings of 2003 IEEE International Conference on Acoustics, Speech and Signal Processing 2003 (ICASSP'03)* **4** 848–851.

[56] PENROSE, M. (2003). *Random Geometric Graphs. Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[57] PERONE PACIFICO, M., GENOVESE, C., VERDINELLI, I. and WASSERMAN, L. (2004). False discovery control for random fields. *J. Amer. Statist. Assoc.* **99** 1002–1014. MR2109490

[58] POZO, D., OLMO, F. and ALADOS-ARBOLEDAS, L. (1997). Fire detection and growth monitoring using a multitemporal technique on AVHRR mid-infrared and thermal channels. *Remote Sensing of Environment* **60** 111–120.

[59] RICHARDSON, D. (1973). Random growth in a tessellation. *Proc. Cambridge Philos. Soc.* **74** 515–528. MR0329079

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

[61] SCHIFF, J. L. (2008). *Cellular Automata*. Wiley, Hoboken, NJ. MR2367301

[62] ŞENDUR, L., MAXIM, V., WHITCHER, B. and BULLMORE, E. (2005). Multiple hypothesis mapping of functional MRI data in orthogonal and complex wavelet domains. *IEEE Trans. Signal Process.* **53** 3413–3426. MR2213561

[63] SHEN, X., HUANG, H.-C. and CRESSIE, N. (2002). Nonparametric hypothesis testing for a spatial signal. *J. Amer. Statist. Assoc.* **97** 1122–1140. MR1951265

[64] SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. MR0799155

[65] STRICKLAND, R. and HAHN, H. Wavelet transform methods for object detection and recovery. *IEEE Trans. Image Process.* **6** 724–735.

[66] SZOR, P. (2005). *The Art of Computer Virus Research and Defense*. Addison-Wesley Professional.

[67] TALAGRAND, M. (2005). *The Generic Chaining*. Springer, Berlin. MR2133757

[68] TAN, H. and ZHANG, Y. (2006). An energy minimization process for extracting eye feature based on deformable template. *Lecture Notes in Computer Science* **3852** 663. Springer, Berlin.

[69] THOMOPOULOS, S., VISWANATHAN, R. and BOUGOULIAS, D. (1989). Optimal distributed decision fusion. *IEEE Transactions on Aerospace and Electronic Systems* **25** 761–765.

[70] WAGNER, M., TSUI, F., ESPINO, J., DATO, V., SITTIG, D., CARUANA, R., MCGINNIS, L., DEERFIELD, D., DRUZDZEL, M. and FRIDSMA, D. (2001). The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice* **7** 51–59.

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

[72] XU, C. and PRINCE, J. (1998). Snakes, shapes and gradient vector flow. *IEEE Trans. Image Process.* **7** 359–369. MR1669528

[73] YU, L., YUAN, L., QU, G. and EPHREMIDES, A. (2006). Energy-driven detection scheme with guaranteed accuracy. *Processing of the Fifth International Conference on Information in Sensor Networks 2006 (IPSN'2006)* 284–291.

[74] ZHAO, F. and GUIBAS, L. (2004). *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann, San Francisco.

[75] ZHONG, Y., JAIN, A. and DUBUISSON-JOLLY, M.-P. (2000). Object tracking using deformable templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2** 544–549.

E. ARIAS-CASTRO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093
USA
E-MAIL: eariasca@ucsd.edu

E. J. CANDÈS
DEPARTMENTS OF MATHEMATICS
  AND STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-MAIL: candes@stanford.edu

A. DURAND
LABORATOIRE DE MATHÉMATIQUES, UMR8628
UNIVERSITÉ PARIS-SUD 11
ORSAY, F-91405
FRANCE
E-MAIL: arnaud.durand@math.u-psud.fr